ICSV13 - Vienna
The Thirteenth International Congress on Sound and Vibration
Vienna, Austria, July 2-6, 2006

DESIGN OF AN INTELLIGENT EXPERT SYSTEM FOR FAILURE DIAGNOSTICS OF ROTATING MACHINERY

Jože Vižintin*[1] and Blaž Suhač[1]

[1]Centre for Tribology and Technical Diagnostics, University of Ljubljana
Bogišičeva 8, 1000 Ljubljana, Slovenia
joze.vizintin@ctd.uni-lj.si

Abstract
This paper proposes a design of a complex intelligent expert system for condition monitoring and failure diagnostics of rotating machinery. The key idea is to combine machine lubrication fluid analysis, temperature analysis and vibration signals analysis into an intelligent hybrid system which is able to perform on-line machine condition monitoring and detection of different faults. We divided the system into three modules: signal acquisition and preprocessing module, condition monitoring module and diagnostics module. A scheme of a such system is proposed in the first part of this paper. In the second part, individual modules are explained in more detail and an overview of signal and information processing methods, which seem to be applicable to the proposed system is presented. The system is being developed in cooperation with a Slovenian steel production plant and a Slovenian power plant. However, we are interested to widen our cooperation to other interested partners worldwide.

INTRODUCTION

Automatic condition monitoring and fault diagnostics of rotating machines are important topics in modern maintenance and are currently extensively researched. They both pose complex multidisciplinary problems, encompassing instrumentation, signal processing, statistics and information processing [1]. Due to complexity, most articles discuss detection of individual faults and apply different methods of signal processing in order to robustly detect them. Condition monitoring and fault recognition is often based on machine temperatures, lubrication fluid properties ([2], [3]) or vibrations ([4], [5], [6]). Even though these methods seem appropriate to detect individual faults, a hybrid system which acquires and analyzes several types of signals and uses several different methods for their analysis should be able to perform a more reliable condition monitoring and detection of multiple faults. This is also demonstrated in [7], where wear particles, contained in lubrication fluid and machine vibration signals were used to detect gearbox faults. This article proposes a design of even more complex system, suitable for application to rotating machines. Issues regarding its structure, functionalities and possible solutions of individual problems are also discussed.

DESIGN OF THE SYSTEM

We propose an intelligent expert system for condition monitoring and fault diagnostics of rotating machinery. A scheme of the proposed system is presented on fig. 1. The proposed

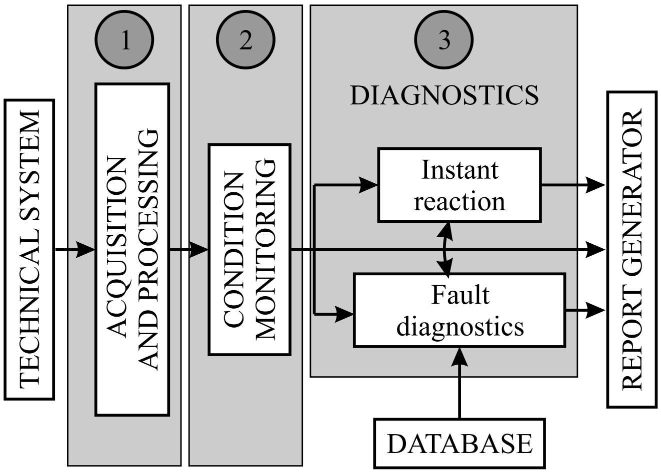

Figure 1: Information processing modules of a failure diagnostic system.

system consists of three modules. The acquisition and preprocessing module (1) performs signal acquisiton and preprocessing. Preprocessed signals are then stored in a database. The condition monitoring module (2) detects whether machine is operating normally or not. For this purpose it extracts features from the preprocessed signals. In case abnormal operation is detected, the diagnostic module (3) activates and determines which fault has occured. The diagnostic module is divided into fault diagnostics and instant reaction submodules. The purpose of the instant reaction submodule is to detect hazardous operation which could seriously damage machine elements. In this case the machine must be stopped immediately. On the other hand, when the machine does not operate normally, but still safely enough, the fault diagnostics submodule activates and determines which fault has caused such operation. Based on this data, the diagnostics module may change machine operation parameters, for example reduce its speed or decrease load in order to prevent more serious damage to occur until the next scheduled maintenance.

Both the diagnostics module and the condition monitoring module generate a report of machine's state. When abnormal operation has been detected, the report generator warns the maintanance crew that a fault had occured and provides additional information about the machine state. It may also suggest a repair process. Otherwise it only notes that machine operates normally. We shall now explain individual modules in more detail.

ACQUISITION AND PREPROCESSING MODULE

The signal acquisition and preprocessing module receives sensory data and transforms them into some more informative state. The system we propose shall acquire temperatures of vital machine elements, lubrication fluid properties and vibration or acoustics signals. These three groups of signals have been found to be a good source of information about the machine state and will be presented later in this section.

An output of the signal processing module is a set of preprocessed signals, e.g. frequency spectra, wavelet transforms, lubrication fluid properties and amount of contained wear particles, temperatures etc. This data is used by other modules in order to extract features, used to describe the state of the machine.

Machine temperatures

Machine elements are designed to perform in a limited temperature interval and are lubricated and cooled accordingly. An increase of temperature either indicates a malfunction of a cooling system or that friction in the contact surface has increased, most likely due to a fault. An increase of temperature changes lubrication fluid viscosity, deforms machine elements, accelerates their fatigue and ultimately leads to overheating and machine element failure. Therefore, temperatures directly indicate whether machine operation is safe or not.

By observing temperature variations it is possible to detect a fault in an early stage of development and also discover its location. Reference [8] states that machine temperatures can be used to estimate maintenance related parameters such as machine service life, service time, running-in time ets. A particularly important parameter is the machine residual service life (RSL), which represents the remaining time of machine operation before service is necessary. A temperature-based methods to estimate RSL are proposed and discussed in [9] and [10]. However, temperature tells us little about the exact root cause of a problem, which can be more precisely determined by the lubrication fluid analysis or the vibration analysis.

Lubrication fluid

The reliability and availability of rotating equipment depends substantially on the lubricant properties. The primary purpose of lubricating oil is to reduce friction and wear, and to take up contaminants by holding them in suspension. It also acts as a cooling medium and provides corrosion protection. Due to ageing, the physical and chemical characteristics of the lubricant are subject to change over its lifetime. When one or more of its properties exceed predescribed limits, the lubricant must be changed in order to avoid equipment failure. However, in order to determine the optimum change interval the actual physical and chemical condition of the oil must to be analyzed.

The most important properties to provide reliable information about lubricant's condition are viscosity, moisture in oil and additive depletion [11], [12]. Viscosity is not only the most important physical property of a lubricant; it can also be effectively used as a measurement of the overall condition of the oil. Many damaging factors can be monitored by the change in viscosity. For example, an increase of the viscosity can indicate: the oxidation of the lubricant, overheating, base stock volatility, coolant contamination, mixing of incompatible oils etc.

Water is a common contaminant in the lubricant and deteriorates its performance. High moisture content increases the risk of corrosion, overheating, equipment malfunction and other problems. Several additive combinations are employed in a lubricant to improve physical-chemical properties (dispersants, viscosity index improvers, antioxidant agents etc.) and modify the chemical and physical properties of a friction surface (friction modifiers, EP and AW additives) [13]. For oil condition monitoring it is important that the decomposition of the additives is realized in an early stage that the oil could be changed in time in case of premature decomposition. Combining these tests provides an invaluable tool for determining lubricant condition. They allow a safe extension of drain intervals and offer an opportunity that the failure of the oil can be detected before other, more serious damages occur.

Wherever there are rotating equipment and contact between surfaces, particles are being generated due to wear. These particles are removed from the contact surface by the lubrication oil. They carry useful information about tribological conditions of contact surfaces. A sudden increase of the concentration of wear particles in oil alerts the user to a potential problem. Reason that quantitative information can be used to signal a changing wear situation is that the wear particle concentration is in dynamic equilibrium when machine operates normally. Particles have been long recognized as the main cause of failure in hydraulics and high-speed rotational machinery, but they are also a leading indicator of a machine's condition. According to their size, shape and concentration, early warning of impending failure is provided in advance giving time to schedule shutdown and maintenance [14].

Vibrations

Vibrations of the machine are related to forces, acting on machine elements. When machine operates normally and in stationary conditions, it produces stationary vibrations of relatively low intensity. Most faults produce an additional force which increases the vibration level.

Vibrations can be measured as displacements or as accelerations. Acceleration data is most effective if the rotation speed of the machine is large, otherwise, it is more effective to measure displacements [15].

Faults produce characteristic patterns, mixed in the vibration signal. Usually they are not visible directly in its plot versus time. Therefore signal processing is necessary to convert the signal into such form where the presence of a fault is seen more clearly. Since different faults affect different signal characteristic, it may be necessary to use several signal processing methods in order to detect them. Most faults produce either stationary or transient patterns, which can be detected by the frequency analysis or by the wavelet analysis, respectively.

Very common faults of a rotating machine are unbalance, parallel and angular missalignment, eccentric rotor, cocked rotor, deflected shaft etc. They usually occur immediately after service due to incorrect machine reassembly. Further operation of such machine results in a quicker wear of its bearings and gears and increases a probability of their failure. On the other hand, if they are discovered and repaired soon enough, the machine can operate normally until the next scheduled maintenance. A suitable method for detecting such faults soon enough is to analyze the frequency content of a signal. This is done by using the Fourier transform to calculate an amplitude or a power spectrum of a signal, which tells us what are the amplitudes of frequencies which are present in it. The above stated faults increase an amplitude of harmonic frequencies, which is seen in the amplitude or power spectrum plot as peaks. By comparing heights of these peaks it is possible to determine more precisely which fault has occured [16], [17], [18].

Another group of common faults are those which occur due to wear of bearings or gears. These faults usually occur as a consequence of some other fault as described before or simply due to material fatigue. Unlike previously described faults, these faults can not be discovered in the signal's amplitude or power spectrum, because the produced forces are not periodic, but transient. A suitable method for analyzing transients is the wavelet analysis [19], which allows us to split a signal into individual components with chosen frequencies. Such transformation takes into account the moment when certain frequencies have occured and how they evolved, therefore it is more apropriate for non-stationry time series analysis then the Fourier transform. Wavelet transform also allows us to extract transient components from the signal. By the frequency of transient occurence in the signal and its characteristic it is possible to determine more accurately which fault has occured.

The wavelet transform calculates an inner product between the signal function and the wavelet function, which is translated along the time axis and scaled. Wavelet functions are derived from a basic function called the mother wavelet. There are many kinds of mother wavelet functions and a success of the wavelet analysis depends a lot on which one we choose. The Morlet wavelet has proven to be a suitable mother wavelet function in vibration analysis for extraction of transient components. A commonly used method for such extraction is the Donoho's soft thresholding algorithm [20]. An improved version of this algorithm proposed by Lin [21] was succesfully applied for detection bearing and gearbox faults. Recently, Tse et. al. [22] introduced a different approach to select a suitable mother wavelet. They used genetic algorithms to construct a wavelet function which more precisely reveals time and frequency properties of the inspected signal. This method was succesfully used to detect faults of ball bearings. However, it is requires more arithmetic operations comparing to Lin's or Donoho's method, which may be a disadvantage for practical application.

Individual features, calculated by the signal pre-processing module, may require extensive computations, particularly for the vibration signals. Therefore it may be necessary to reduce the number of features being calculated or to perform calculations less frequently. An idea to pull the most out of both options is to contionuously calculate only the most vital features, while a more extensive analysis is performed in a pre-described interval, for example every minute or every hour.

CONDITION MONITORING MODULE

The condition monitoring module performs on-line periodic characterization whether a machine operates normally or not. If an operation is characterized as normal, the module activates the report generator, which notes that at this moment the machine operates normally. Otherwise it activates the diagnostics module which determines a cause for abnormal operation.

Machine operation characterization is based on a feature vector, derived from the pre-processed signals (provided by the signal acquisition and pre-processing module). Features carry information about some system's property or state in a more compact form. The idea is that when the state of a machine changes, values of certain features will significantly change as well. If we know, which fault affects which features, we can detect a fault by tracking significant feature changes. Some physical properties, such as viscosity or amplitudes of characteristics frequencies may be used as features. But in some cases, it is more convenient to define features based on our observations, without any particular physical meaning. In acoustics, for example, a measure of lowness or highness of some sound is its pitch frequency, which may itself not even exist in the percieved sound. Similarly, we may define features based on our observations of measured signals. A selection of a reliable feature which describes certain abstract property, such as presence of a particular fault, is one of the most challenging tasks in machine condition monitoring and diagnostics.

Since fast performance of the condition monitoring module is important, a limited number of features should be selected in the condition monitoring process. Such features may be temperatures of machine elements, oil viscosity, wear particle concentration, vibration level (RMS) and other features, which can be calculated quickly and are sensitive to as many faults as possible.

Once features to be used for condition monitoring are determined, they are grouped in to a feature vector, which lies in multidimensional feature space. Vectors, characteristic for normal machine operation, point within a subspace in a feature space and our task is to find its boundaries. Since feature vector is a random variable, boundaries could be determined on statistical basis. A common approach is to calculate the prediction interval, which is an interval that contains the next feature vector with a given probability. This interval is determined by a given set of feature vectors, which are known to depict normal machine operation. The idea is that if the machine operates normally, it is very unlikely that a feature vector would lie outside this interval. If it does it is natural to assume that the machine no longer operates normally. Calculation of a prediction interval for a normally distributed random variable is described in [23]. A method for estimation a prediction interval for a random variable with arbitrary probability distribution, generated by a linear process is proposed by Alonso et al. in [24]. For estimation of a non-linear time series prediction interval, a similar method is proposed by Giordano et al. [25]. All these methods assume that the process which generates feature vectors is stationary. In case some of its properties are time-dependant, it may significantly complicate a reliable determination of normal operation's boundaries. In this case it may be more useful to treat a feature vector as a dynamic object and observe its trends, as described in [26].

DIAGNOSTICS MODULE

The diagnostics module is activated when the condition monitoring module detects abnormal operation. In this case it is first necessary to determine whether further machine operation is safe or not. This can be done by observing features like the vibration RMS, motor current and voltage and other variables, which may indicate a direct threat to machine integrity. If such threat is detected, the instant reaction submodule immediately reacts by slowing or shutting down the machine and triggering a red alarm. Otherwise, the failure diagnostics submodule is to determine a root cause of abnormal operation or at least estimate it as well as possible. Since the machine operation has been characterized as abnormal, but safe, only minimal changes of its operating parameters may be necessary.

The failure diagnostics module performs a multiple-class classification, where each class is associated to a region in the feature space where feature vectors of a particular fault point at. Unfortunately, all region’s boundaries may not be known in advance since it is not possible to simulate all kinds of faults on an observed machine. However, when some *a priori* knowledge, either from some other research or previous experience on how specific failures affect the measured signals and features, it is possible to choose additional signal processing methods and features which would at least narrow down the list of possible failures. Nevertheless, this kind of pattern classification is difficult, since it is not easy to determine whether an observed feature vector is merely an unlike member of some known class, or a member of a previously unknown class. It is even a problem to define a quantitative criterion for such distinction. Therefore some robustness should be implemented in the diagnostics module in order to prevent unnecessary alarms due to a single unusual observation. This may slow down the diagnostic process, but since the diagnostics module has already determined that a fault is not critical, it should not be a big disadvantage.

The described classification process is possible representative when signals of all the faults to be observed are available. In this case it is necessary to find boundary surfaces between different regions. When two regions are linearly separable, the boundary surfaces between them is a hyperplane, which is not difficult to find. But in practice, it is difficult to define such features that produce linearly separable classes. This leads to linearly unseparable classes, whose bounding surfaces are harder to find. However, there always exists a higher dimension feature space where classes are linearly separable. Finally, the third possibility is that regions are overlapping and therefore unseparable. This can be either due to a false distinction between machine states or due to redundancies in the feature space. Redundancies can be reduced by the principle component analysis (PCA) or the Fisher discriminant analysis (FDA) [27]. The PCA is used to decrease the dimension of the feature space by first projecting feature vectors to axes which lie in the direction of greatest variances of features and then selecting only a few axes with greatest variances. Thus the dimension of the feature space can be significantly reduced while obtaining most of variability of samples. On the other hand, the FDA directly deals with the problem of allocation of a feature vector into one of the states. Lately, non-linear versions of the PCA have been developed in order to improve the performance for non-linear signals. In order to decide whether to use linear or non-linear PCA, a measure of non-linearity in data is proposed in [28] and an application of one of the non-linear PCA methods to fault detection in a chemical process is presented in [29]. A succesful application of PCA results in a new feature space with a smaller dimension and more distinct regions, where it is possible to allocate feature vectors more reliably. The most commonly chosen tools for classification are neural networks (NN) [30] and support vector machines (SVM) [31]. Multilayered perceptron or radial basis function ANNs are able to perform multi-class classification, while a SVM can only distinct between two classes at a time. Due to the possibility to increase the feature space dimension by using kernel estimators, both tools are able to separate non-linearly separable classes. Applications of ANNs and SVMs in machine diagnostics show that their classification capabilities are comparable [5].

When the diagnostics module completes its analysis and discovers a failure, it triggers a yellow alarm which warns the machine operator about a failure and provides a diagnostics report which helps in further machine examination by the service crew. Therefore, a final decision on how to react to a yellow alarm is still made by the service crew, the diagnostics module only aids them in making an optimal decision.

FINAL REMARKS

Further development of the proposed condition monitoring and failure diagnostics system shall be conducted by the authors. An idea is to design an experimental system, where a wide range of faults could be simulated and observed by different kinds of sensors. This would provide us with real-life signals which could be used as a benchmark for different signal processing methods. The most promising methods shall also be tested in an industrial environment. The authors would like to use this opportunity to invite all the interested researchers to join us at solving problems that are laid ahead of us.

REFERENCES

[1] B. Rao, *Handbook of condition monitoring*. (Elsevier Advanced Technology, 1996)

[2] T. Kirk, D. Panzera, R. Anamalay, Z. Xu, *Computer image analysis of wear debris for machine condition monitoring and fault diagnosis*, Wear, **181-183**, 717–722 (1995)

[3] J. Vižintin, M. Kambič, I. Lipušček, V. Hudnik, *Application of wear particle analysis to condition monitoring of rotating machinery in iron and steel works*, Journal of the Society of Tribologists and Lubrication Engineers, **51**, 389–393 (1994).

[4] J. K. Spoerre, *Application of the cascade correlation algorithm (CCA) to bearing fault classification problems*, Computers in Industry, **32**, 295–304 (1997).

[5] B. Samanta, K. Al-Balushi, S. Al-Araimi, *Artificial neural networks and support vector machines with genetic algorithm for bearing fault detection*, Engineering Applications of Artificial Intelligence, **16**, 657–665 (2003).

[6] I. Yesilyurt, *Fault detection and location in gears by the smoothed instantaneous power spectrum distribution*, NDT&E International, **36**, 535–542 (2003).

[7] Z. Peng, N. Kessissoglou, *An integrated approach to fault diagnosis of machinery using wear debris and vibration analysis*, Wear, **255**, 1221–1232 (2003).

[8] L. Burstein, *Relationships for machine thermal diagnostics: theoretical and empirical considerations*, NDT&E International, **38**, 69–75 (2005).

[9] L. Burstein, L. Segal, *Prediction of machine residual service life: method and computing*, NDT&E International, **37**, 517–523 (2004).

[10] L. Burstein, *Theory and computation of the reference dependence and residual service life of a machine*, NDT&E International, **38**, 310–318 (2005).

[11] C. Wincierz, K. Hedrich, *Influence of viscosity index improvers in gear oils on the mechanical efficiency and the operating temperature*, in *International Conference on Gears* (2002).

[12] B. Kržan, J. Vižintin, *Vegetable based oil as a gear lubricant*, in *Intrernational Conference on Gears* (2002).

[13] M. Zundel, *Condition monitoring at the machine element 'oil'*, in *International Conference on Gears* (2005).

[14] B. Kržan, M. Kalin, J. Vižintin, *The lubrication of DLC coated gears with the environmentally adapted ester based oil*, Gear Technology, article in press.

[15] D. Mba, *Applicability of acoustic emissions to monitoring the mechanical integrity of bolted structures in low speed rotating machines: case study*, NDT&E International, **35**, 293–300 (2002).

[16] D. Bently, C. Hatch, *Fundamentals of rotating machinery diagnostics*. (American Society of Mechanical Engineers, 2002).

[17] Y. Chen, R. Du, *Fault features of large rotating machinery and diagnosis using sensor fusion*, Journal of Sound and Vibration, **188(2)**, 227–242 (1995).

[18] J. Vižintin, B. Kržan, *Integrating monitoring system for pitting detection*, in *International Conference on Gears* (2005).

[19] S. Mallat, *A wavelet tour in signal processing*, 2nd edn. (Academic Press, 1999).

[20] D. Donoho, *De-noising by soft-thresholding*, IEEE Transactions on Information Theory, **41**, 613–627 (1995).

[21] J. Lin, *Feature extraction based on morlet wavelet and its application for mechanical fault diagnosis*, Journal of Sound and Vibration, **234(1)**, 135–148 (2000).

[22] P. Tse, W. Yang, H. Tama, *Machine fault diagnosis through an effective exact wavelet analysis*, Journal of Sound and Vibration, **277**, 1005–1024 (2004).

[23] D. Montgomery, G. Runger, *Applied statistics and probability for engineers*, 3rd edn. (Joe Wiley & Sons, 2003).

[24] A. Alonso, D. Peña, J. Romo, *Forecasting time series with sieve bootstrap*, Journal of Statistical Planning and Inference, **100**, 1–11 (2002).

[25] F. Giordano, M. L. Rocca, C. Perna, *Forecasting nonlinear time series with neural network sieve bootstrap*, Computational Statistics & Data Analysis, article in press.

[26] A. Joentgen, L. Mikenina, R. Weber, A. Zeugner, H.-J. Zimmerman, *Automatic fault detection in gearboxes by dynamic fuzzy data analysis*, Fuzzy Sets and Systems, **105**, 123–132 (1999).

[27] W. Haerdle, L. Simar, *Applied multivariate statistical analysis*. (Springer, 2003).

[28] U. Kruger, D. Antory, J. Hahn, G. Irwin, G. McCullough, *Introduction of a nonlinearity measure for principal component models*, Computers and Chemical Engineering, **29**, 2355–2362 (2005).

[29] S. Choi, I. Lee, *Nonlinear dynamic process monitoring based on dynamic kernel pca*, Chemical Engineering Science, **59**, 5897–5908 (2004).

[30] S. Haykin, *Neural networks. A Comprehensive Foundation*. (Prentice Hall 1996).

[31] V. Vapnik, *The nature of statistical learning theory*, 2nd edn. (Springer, 2000).