Jiří Kléma[1], Ondřej Flek[2], Jan Kout[1], Lenka Nováková[1]

[1]*Department of Cybernetics, CTU Prague, Technická 2*
166 27 Prague, CZECH REPUBLIC
{klema,kout,step}@labe.felk.cvut.cz
[2]*Rockwell Automation Ltd., Research Center Prague,*
Pekařská 695/10a, 155 00 Prague 5, CZECH REPUBLIC
oflek@ra.rockwell.com

This paper addresses the problem of on-line diagnosis of cavitation in centrifugal pumps. The paper introduces an application of the Open Prediction System (OPS) to cavitation diagnosis. The application of OPS results in an algorithmic framework for diagnosis of cavitation in centrifugal pumps. The diagnosis is based on repeated evaluation of a data scan providing full record of input signals which are observed for a fixed short period of time. Experimental verification of the algorithmic framework and the proposed methodology proved that a condition monitoring system built upon them is capable of diagnosing a wide range of cavitation conditions that can occur in a centrifugal pump, including the very early incipient cavitation.

1. INTRODUCTION

Pump is probably the most widespread variety of machinery in the world. Pumping systems, either directly or indirectly, are an essential part of every business throughout the world. As an ultimate illustration, a typical chemical plant utilizes on average one pump per each employee (Hennecke, 2000). It has been estimated that nearly 20% of the energy generated globally is devoted to driving pumping systems (Hart, 2002). Probably the highest percentage of pumps used in industry is accounted to centrifugal pumps. They are relatively simple, inexpensive and generally very reliable pieces of equipment. Nevertheless, the consequences of their unexpected failure include costly machinery repair, extended process downtime, health and safety problems, increased scrap levels, and loss of sales. For this reason, an increasing interest in improved maintenance regimes can be noticed among pump operators.

The traditional machinery maintenance approaches include *reactive* and *preventive* regimes (Day, 1996). The former relies upon allowing the machine to break down before being maintained. In spite of the fact that it has proved to be the most expensive option, it is still widespread among many industries. The preventive maintenance mode, whereby the maintenance is based upon specific time intervals, can lead to savings over the reactive regime, however, it is not an effective use of maintenance resources as work is undertaken regardless of the condition of the

equipment. *Predictive* or *condition-based maintenance* (CBM) is based upon monitoring of condition of the equipment and determining whether corrective action is needed. By detecting the early stages of a fault, maintenance can be scheduled in advance to coincide with planned production stops. The condition-based maintenance approach relies on techniques of machinery diagnostics. Many data acquisition and analysis techniques have been developed for machinery diagnostics. Some of them rely on periodic data collection carried out by means of portable instruments and subsequent batch-mode data interpretation. However, the focus has been on schemes that provide on-line, continuous monitoring and diagnosis of equipment. Improvements in sensor technologies and mass production of a wide variety of sensors have enabled application of on-line machinery diagnostics to a wide range of equipment where such techniques would be thought too expensive just a couple of years ago.

Although the techniques of CBM rely on the ability to detect early stages of a possible failure of equipment, further benefits would certainly emerge with the ability to control the working regime of the equipment in such a way that the potential for occurrence of failure is minimized. In the case of pumps, the most frequent failures are bearing and seal failures (Marscher, 2002). One of the most important root causes of bearing and seal failures is presence of *cavitation* in a pump. The word cavitation refers to formation of vapor bubbles in regions of low pressure within the flow field of a liquid (Brennen, 1994). In the context of turbomachinery, cavitation is generally considered undesirable. Besides of the effect on life of bearings and seals, cavitation causes gradual erosion of internal surfaces of a pump. This may result in an unexpected pump failure with possible disastrous consequences.

This paper is organized as follows. Section 2 gives a brief overview of principal approaches to cavitation diagnosis. It distinguishes two sensor categories that determine whether the diagnosis will be intrusive or non-intrusive. Section 3 introduces a non-intrusive way of cavitation diagnosis based on vibration sensors' data. It describes an experimental setup we have used, defines a structure of measured data and outlines a way in which the phenomenon of cavitation may reflect in the data. A summary of principal questions to be answered by data mining is provided at the end of the section. Section 4 introduces Open Prediction System – the tool used for processing and evaluation of the measured data. The section also theoretically discusses methods relevant to domains with ordinal classifications and possibly dependent samples. Section 5 summarizes reached results and tries to answer questions raised earlier in Section 3. In conclusion, the proposed algorithmic framework for diagnosis of cavitation in centrifugal pumps is recapitulated.

2. PRINCIPAL APPROACHES TO CAVITATION DIAGNOSIS

Quite recently, first commercially available systems have appeared that reflect the trend of integration of pumping system control and condition monitoring (Stavale, 2001). Here, the ability to estimate the presence of cavitation in a pump is based on the knowledge of parameters of the pumping process, especially the pressure in pump suction. Since the phenomenon of cavitation is closely connected with pressure field in the pump, the decision to use such information for estimation of

cavitation is logical. The major disadvantage of the use of pressure information is the fact that pressure sensors are considered intrusive, i.e., come into touch with the pumped fluid. Tapping such sensors into the pipework of the pumping application increases the potential risk of leakage. For that reason, such sensors must be avoided in some applications involving pumping of dangerous fluids. This disqualifies the solutions using such sensors from universal application.

The attention of the industry points towards non-intrusive sensors. Vibration sensors play a dominant role among them. Besides of being a source of information for detection of cavitation, they are capable of providing information that can be used for diagnosis of a wide range of rotating machinery faults (White, 1998). Understanding relationship between pressure pulsation in pumps and mechanical oscillation of solid pump parts creates a background for the use of pump casing vibration as a source of information for cavitation diagnosis. This paper studies possibility of indirect detection of cavitation from mechanical vibrations, the resulting diagnosis scheme must rely exclusively on information from non-intrusive sensors of this type.

3. VIBRATION SENSORS' DATA

3.1 Cavitation Research Setups

The research of cavitation diagnosis methods requires a representative amount of experimental data. The data must cover a wide range of operating conditions of the pump and various degrees of cavitation. Although data collected on a real-world system would be of highest value, there are numerous reasons why this kind of data is normally unreachable: (1) economical and safety reasons, (2) controlling the real-world system deliberately in order to cover a wide range of operating conditions is usually not allowable, (3) fitting the necessary instrumentation to the real-world system may be difficult or even impossible, (4) the cavitation condition present in the pump during experiments is usually not known with a sufficient accuracy as this mostly requires a specially modified pump that allows visual observation of the inside of the pump.

Consequently, in many cases, purpose-built experimental setups are employed to provide the data needed for research. The setup can be divided into four main subsystems: the pump (modified by a transparent material allowing visual observation of cavitation), the motor (energized either directly by mains power line or by variable frequency drive), the flow loop and the data acquisition equipment. Following signals were sensed and recorded: flow rate, pressures in both suction and discharge, temperature of the pumped fluid, shaft rotation frequency. These signals were used to control the experiments and they could not be used during a diagnosis phase. Vibration signals were sensed by accelerometers attached to the casing of the diagnosed pump. For all experiments, two accelerometers were used simultaneously. They were positioned on the casing of the respective experimental pump, in mutually perpendicular directions. One of the sensors was adjusted in the direction of the pump shaft axis (denoted as axial), the other in the direction radial to the pump shaft axis (denoted as radial). A detailed description of the setup employed in presented experiments is given in (Flek, 2002).

A typical experiment, performed with the aim of obtaining data relevant to the phenomenon of cavitation, establishes various levels of cavitation in a system operating under certain conditions. It is important to investigate cavitation in a system working at a (full) range of operating points. The operating point of the system is described by the flow rate Q, the total head rise H and the shaft rotating frequency Ω. It is set by a throttling valve in the discharge pipe. Severity of cavitation in the pump is set either by a throttling valve in the suction pipe (open tank setup) or by modification of pressure above liquid surface in the tank (closed tank setup).

3.2 Data, Preprocessing Phase, Feature Extraction

The vibration of a centrifugal pump casing is governed by numerous excitation forces acting at different frequencies: pump shaft imbalance, misalignment, impeller blade passing pulsation, bearings, cavitation, etc. Many different techniques are used to help the evaluation of vibration signals. The simplest possibilities include evaluation of amplitude information in time-domain signal. However, signals of periodical nature typically require analysis in the frequency domain. The emergence of digital signal processing techniques, especially the Fast Fourier Transform (FFT) algorithm, became the driving force behind the wide spread of frequency-domain analysis techniques.

(Flek, 2002) proposes distinguishing features of following types: (1) power spectral density of frequency band y (denoted as psd$y$), the number of bands depends on frequency resolution of the periodogram, in this paper we use mainly 65 bands of $\Delta f \approx 234$Hz, corresponding to a 128-line FFT with sampling frequency ≈ 30kHz, but other settings were also tested, (2) amplitude of the first 2z harmonic components of shaft rotation frequency, where z is number of impeller blades (denoted as rpm$n$), in pumps with 4 blades 8 features, (3) frequency of rotation of pump shaft (denoted as fr).

Five ordinal cavitation classes can be distinguished: 0 – normal condition (no bubbles), 1 – incipient cavitation (very first bubbles), 2 – tip vortex cavitation (a tiny stream of bubbles), 3 – moderate cavitation (a continuous stream of bubbles), 4 – severe cavitation (severe bubbles, blade cavitation). The classes are assigned on basis of visual observations during the experiments.

3.3 Data Understanding - Visualization

The influence of the phenomenon of cavitation on vibration of pump casing can be better understood with aid of visualization. Having the excitation forces and their effects decomposed into a periodic component (related to the shaft rotation frequency) and a random component (e.g., cavitation), the following visualizations can be carried out.

Figure 1 shows examples of periodograms of pump casing vibration under varying cavitation conditions. Although particular frequencies and the scale of the phenomena necessarily differ among different pumps, the figures demonstrate influence of cavitation on psd$y$ attributes at different operating points. Note that a constructed classifier has to detect cavitation while not knowing the actual operating point.
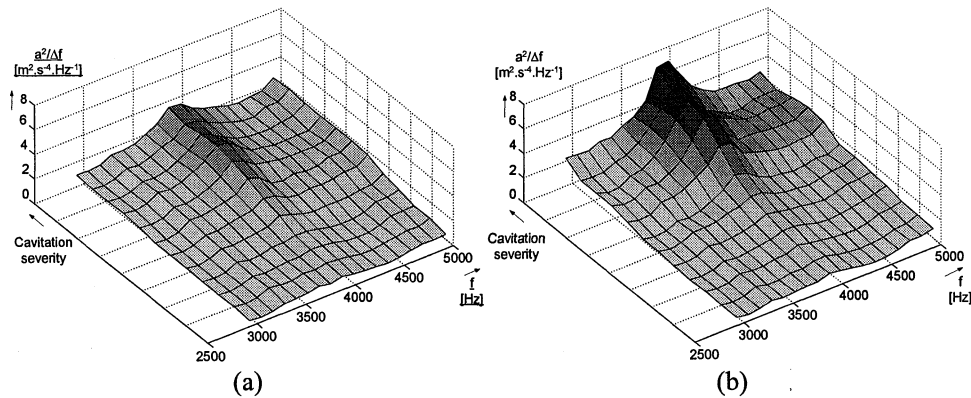
Figure 1 – Comparison of periodograms of pump casing vibration under varying cavitation conditions. Figures (a) and (b) show periodograms for two different operation points (below the flow rate of the best efficiency point and at the best efficiency point).

We have also used Radviz technique of multidimensional visualization implemented in the preprocessing tool SumatraTT (Stepankova et al., 2003). Radviz (Hofman et al., 1997) attaches to each data point fixed springs each of which is also attached at points around a circle. The springs represent dimensions of the data, the spring force for each spring is the value of the data point for that dimension. The data points are displayed at the position where the sum of the spring forces is zero, i.e., points which have one or two coordinate values greater than the others lie closer to those dimensions. The results can be seen in Figure 2.

3.4 Principal Questions To Be Answered

The main goal of the intelligent decision-making system design is to provide a tool allowing reliable and non-intrusive on-line diagnosis of cavitation in centrifugal pumps. Within this process, following principal questions regarding specific task characteristics should be answered:

- What is the optimal placement of the vibration sensors? How many of them one has to use (a minimum number of sensors should be used to save equipment and installation cost)?

- What is the influence of number (and thus resolution) of the power spectral density features? Can we deal with a large number of features having only a limited number of training examples?

- How should we deal with the measured data? Can we increase a number of training instances by generating more examples from a single (longer) signal measured under constant conditions? What is the dependence among signals measured under similar conditions (similar operating points)?

- Class values are ordered. Can we benefit from this ordering?

- How should we evaluate the resulting system and what is an optimal scoring function when developing a model? Shall we use classification accuracy only,

distinguish severity of misclassifications or rely on regression criteria (e.g., mean squared error (MSE))?

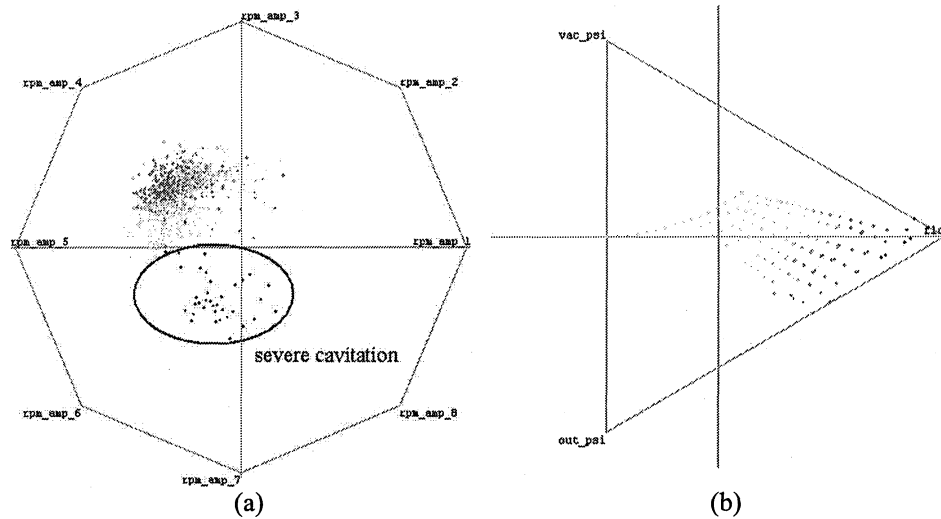(a)   (b)

Figure 2 – Radviz – class distribution in a multidimensional attribute space – the more severe cavitation the darker points. Figure (a) demonstrates ability of rpm*n* features to distinguish severe cavitation. Figure (b) shows strong dependence among cavitation, flow and pressures (these sensors are not available in a real operation).

4. CAVIATION DIAGNOSIS SCHEMES

4.1 Open Prediction System – Experimental Environment

Open Prediction System (OPS) is a prediction tool offering a solution to a wide range of prediction problems. Its main focus is on multivariate time series prediction problems with practical applications bringing cost reduction in utility companies (gas, water, heat, electricity). However, its predictive methodology (Kout et al., 2004) can be understood as fully general and able to solve both regression and classification tasks. The implemented data management structures enable data compression, data filtering, special built-in transformations, problem definition separated from preprocessing of individual models or meta-learning. These features make OPS a suitable tool for processing and evaluation of the pump data.

4.2 Available Data, Applied Methods

The analysis and results presented in this paper deal with measurements performed by Ondřej Flek on Durco Mark III 1K1.5×1-8 pump in Cleveland, Ohio, USA. In order to answer the questions summarized in subsection 3.4, a number of datasets based on these measurements were generated and processed. Several dimensions that distinguish between the datasets can be identified: (1) the data source is axial,

radial or combination of both sensors, (2) the number of features varies with the selected power spectral density bandwidth – 32 and 65 bands were used, (3) the three-phase asynchronous electric motor can be energized either by direct mains power line or variable frequency drive (VFD). It follows that 3×2×2=12 different datasets were classified and compared.

The OPS offers several classifiers and predictors to be applied. It contains decision tree (DT) and random forest (RF) classifiers as well as neural networks (NN) and support vector machines (SVM). DTs and their visualization give a basic understanding of feature importance and problem complexity, while RF and SVM are techniques well-known to deal with datasets described by a large number of features regarding a number of available instances. Application of NNs gives a chance to reference the results performed in (Flek, 2002) on other two pumps. These four algorithms were applied to the datasets described in the previous paragraph.

The following two subsections describe fundamental approaches to a final task definition considering apparent class ordinality and dependence among measured instances.

4.3 Evaluating Dependent Samples – Dealing with Blocks and Series

The most often used evaluation methods (hold out, N fold cross validation, leave-one-out, bootstrapping) suppose a representative dataset on their input as well as independent instances. At the same time, they suppose that all the instances are drawn with a constant distribution and that the future instances will keep this distribution. These assumptions are violated in many practical domains. A motivational example which tries to find out whether the dataset is representative is as follows. Let us have a company producing pumps. The company produces three different pumps and wishes to equip them with a diagnostic tool that is able to detect their fault states. The fault state model is based on the data measured on pumps, but measurements are expensive. Is it necessary to measure all the pumps? An advisable approach to find it out could be to measure two pumps only. The first model can be trained on the data from the first pump and tested on the data from the second pump, the second model vice versa. When both the models give satisfactory error rates there is a solid chance that the resulting model (based on data from both the pumps) can be valid also for the third pump. Other approach, which first mixes measurements from both the pumps and then splits them randomly between training and testing subsets, is probably not a good indicator of performance on a future data from a different pump.

Under our design of experiments, a representative training data set has to be measured for each type of pump to be diagnosed. Differences between pump types are indispensable, a general cavitation model would be inaccurate. The proposed method relies on a standard pump performance testing procedure carried out by pump manufacturers with virtually every type of pump produced. It has been shown in (Flek, 2002) that such performance testing is capable of producing data covering the whole range of operating conditions.

The dataset can also contain dependent instances. This phenomenon often appears when dealing with time-series measurements, where dependence (expressed in terms of covariance) between ith and $i\pm j$th instance depends only on j and approaches 0 as $j \rightarrow \infty$. (Burman et al., 1994) proposes a modification of cross-

validation called h-block cross-validation. For each instance *i* it is necessary to remove it from the training data along with *h* instances on each side of the *i*th observation. The model is trained on the reduced set and tested on the *i*th instance. This approach reduces to leave-one-out method when h=0. A slightly modified hv-block cross-validation can be found in (Racine, 2000).

Within pump measurements, it is not possible to set and measure an arbitrary number of operating conditions (time and cost reasons, limited number of valve positions, etc.). The measurements usually consist of several series, in which we start in the normal state and gradually stimulate more severe cavitation levels. In order to increase a number of training examples, we generate several instances (6) for every single operating point. A longer scan is subdivided into time slots that are processed separately. These training examples are surely dependent and must be treated in an approach analogical to the block cross-validation described above (we denote it one-block-out, OBO). Dependence among neighboring states in a series can be studied in an approach where a single series makes a single fold in the frame of cross-validation (denoted as one-series-out, OSO).

4.4 Ordinal Classification

Standard classification algorithms assume that the class values are unordered, i.e., do not exhibit any natural order. They treat the class attribute as a nominal quantity – a set of unordered values. Consequently, they cannot make use of the ordering information. (Frank and Hall, 2001) proposed a simple method that enables standard classification algorithms to make use of ordering information in class attributes. In their approach a task is transformed to a series of binary class subtasks that encode the ordering of the ordinal class. The data is first transformed from a k-class ordinal task (class attribute C with values v_1, ..., v_k) to k-1 binary subtasks, where i-th binary attribute represents the test $C > v_i$. This coding is similar to a thermometer code used in neural networks, which encodes into k binary attributes (Smith, 1996). In the next step, k-1 probabilistic models are learnt and used to estimate probabilities of the k original ordinal classes ($P(v_1)=1-P(C > v_1)$, $\forall i=2$, ..., k-1: $P(v_i)=P(C > v_{i-1})(1-P(C > v_i))$, $P(v_k)=1-P(C > v_{k-1})$). The class with maximum probability is assigned to the instance.

Learning to predict ordinal classes can also be rephrased as the regression problem. The class labels defined in subsection 3.2 can be either used immediately or they can be preprocessed prior to learning. We followed the second option, the *grades* transformation (Kramer et al., 2001), which depends also on fractions of values belonging to the individual classes, was carried out. In this paper, we apply the ordinal approach (denoted as ORD) altogether with the standard unordered classification (denoted as 5C) and regression (denoted as REG).

5. REACHED RESULTS

The design of experiments presented in the previous chapter results in 22 different ways of processing of each of the proposed datasets (algorithm x 5C, ORD or REG x OBO or OSO). In order to be able to evaluate and compare all the approaches, the outputs of the models were always transformed into 5 crisp classes first and then the

classification accuracy was calculated. Significance of differences among the individual approaches was evaluated by McNemar's test. There were two different types of classification errors defined. The first type simply involves all misclassifications. Error occurs whenever the desired class does not agree with the generated classification. The second type pays attention to so called hard errors only. These hard errors do not occur whenever the model misclassifies an instance into its neighboring class. The detection rate of each model was also considered (ability to distinguish the normal state form an arbitrary cavitation level).

Table 1 shows a fragment of the final result table. The results proved that the axial placement of the vibration sensor gives a slightly better ability to diagnose cavitation. Moreover, combining the axial and radial data does not help to significantly increase this ability. The framework does not seem to be sensitive to the tested changes of power spectral density bandwidth, we recommend to deal with 65 psd*y* features as better portability to various pumps is assumed. The application of VFD brings another degree of freedom into the system and slightly decreases reliability of fault diagnosis. RFs proved to be the most suitable learning algorithm. Surprisingly, SVMs gave worse results than both RFs and NNs. Utilization of the ordering information significantly increases classification accuracy of DTs and SVMs, but do not help to improve RF classifiers. As for NNs, ORD approach helped to increase overall accuracy, REG approach was better considering hard errors only. OBO cross-validation results in reasonably more optimistic accuracy estimates than OSO. The results were discussed with a domain expert who regards the first sort of estimates as optimistically biased while the second one is biased pessimistically as it leaves out a certain part of the plane where a pump operating point can be set.

Table 1 – An overview of results reached by RFs on the *mains axial* and *radial* datasets with *65* power spectral density bands. The individual cells show error of classification [%]. [#] denotes a difference on 0.05 level of significance between corresponding tests on axial and radial data.

Method\Data	axial			radial		
	all	hard	det	all	hard	det
5C, OBO	16.3	0	7.1	16.7	0.9	8.2
5C, OSO	21.5	1.9	10.6	24.5[#]	6.2[#]	14.7[#]
ORD, OBO	18.2	0.1	8.0	19.2	1.8[#]	9.1
ORD, OSO	21.9	1.9	11.2	25.5[#]	5.2[#]	15.7[#]

5. CONCLUSIONS

The presented approach results in an algorithmic framework for diagnosis of cavitation in centrifugal pumps. The diagnosis is based on repeated evaluation of a data scan based on an axial vibration sensor sampled for a fixed short period of time. Signal is decomposed into periodic and random components, the methodology can deal with a large number of power spectral density features which guarantees its general applicability to various pump types. The proposed diagnosis scheme consists of the following operations: signal sensing, signal pre-processing, feature extraction, classification and presentation of diagnosis. The diagnosis can be presented either as

a crisp classification (5C or REG approach), estimated as a real number (REG approach again) or in a form of a probability vector (ORD approach).

Experimental verification of the algorithmic framework and the proposed methodology suggested that a condition monitoring system built upon them is capable of diagnosing a wide range of cavitation conditions that can occur in a centrifugal pump, including the very early incipient cavitation. It can be tuned to individual pump types by means of a standard pump performance testing procedure only. This represents no extra effort since this procedure is a part of common practice exercised throughout the pump manufacturing industry. The future work lies in a verification of the proposed framework on a wider range of pump types including the extension to pump varieties other than centrifugal pumps.

Acknowledgments

6. REFERENCES

1. Brennen, C. E. Cavitation and Bubble Dynamics. Oxford University Press, 1994.
2. Burman, P., Chow, E., Nolan, D. A Cross-Validatory Method For Dependent Data. Biometrika 84, pp. 351-358, 1994.
3. Day, M. J. Condition Monitoring of Fluid Systems – The Complete Approach. In proceedings of the Fifth International Conference on Profitable Condition Monitoring – Fluids and Machinery Performance Monitoring. Harrogate, U.K., BHR Group, pp. 243-256, 1996.
4. Flek, O. Diagnosis of Cavitation in Centrifugal Pumps. PhD. Thesis, Czech Technical University, Department of Cybernetics, 141 p., 2002.
5. Frank, E., Hall, M. A Simple Approach to Ordinal Classification. Proceedings of the European Conference on Machine Learning, Freiburg, Germany. Springer-Verlag, pp. 145-165, 2001.
6. Hart, R. J. Pumps and Their Systems – A Changing Industry. Proceedings of 19th International Pump Users Symposium, Houston, TX, U.S.A., pp. 141-144, 2002.
7. Hennecke, F. W. Reliability of Pumps in Chemical Industry. Proceedings of Pump Users International Forum, Karlsruhe, Germany, 2000.
8. Hoffman, P., Grinstein, G., Marx, K., Grosse, I., Stanley, E. DNA Visual and Analytic Data Mining. IEEE Visualization '97 Proceedings, pp. 437-441, Phoenix, AZ, 1997.
9. Kout, J., Kléma, J., Vejmelka, M. Predictive System for Multivariate Time Series. To appear at European Meetings on Cybernetics and Systems Research (EMCSR), Vienna, 2004.
10. Marscher, W. D. Avoiding Failures in Centrifugal Pumps. Proceedings of 19th International Pump Users Symposium, Houston, TX, U.S.A., pp. 157-175, 2002.
11. Kramer, S., Widmer, G., Pfahringer, B., de Groeve, M. Prediction of Ordinal Classes Using Regression Trees. Fundamenta Informatica, 47(1-2): 1-13, 2001.
12. Open Prediction System, http://ops.certicon.cz.
13. Racine, J. A Consistent Cross-Validatory Method For Dependent Data: hv-Block Cross-Validation. Journal of Econometrics, November 2000.
14. Smith, M. Neural networks for Statistical Modeling. Boston: International Thomson Computer Press, 1996.
15. Stavale, A. E. Smart Pumping Systems: The Time is Now, ITT Industries, Fluid Technology Corporation, Industrial Pumps Group, http://www.gouldspumps.com/download_files/ /Technews/time_is_now.pdf, 2001.
16. Štěpánková, O., Aubrecht, P., Kouba, Z., Mikšovský, P. Preprocessing for Data Mining and Decision Support. In: Data Mining and Decision Support: Integration and Collaboration. Dordrecht : Kluwer Academic Publishers, pp. 107-117, 2003.
17. White, G. D. Introduction to Machine Vibration. Predict DLI, 1998.