

# Multi-Stage Intelligent System for Diagnostics of Pumping Equipment for Oil and Gas Industries

V V Bukhtoyarov<sup>1</sup>, V S Tynchenko<sup>1</sup>, E A Petrovsky<sup>1</sup>

<sup>1</sup>Department of Technological Equipment and Machinery of Oil and Gas Industry, Siberian Federal University, 79, Svobodniy pr., Krasnoyarsk, 660041, Russian Federation

E-mail: vbukhtoyarov@sfu-kras.ru

**Abstract.** New approach for classification of the state of technological pumping equipment is presented in the paper. The approach involves the use of pumping equipment parameters operative monitoring data for indirect fault identification. The proposed method is a part of developed integrated approach for decision support in the management of technological equipment of oil and gas fields. The method realizes a multi-stage classification scheme based on an ensemble approach to the intelligent data analysis. The scheme involves the creation of simple classifiers of the first level, which can be implemented on the basis of artificial neural networks or other effective classifying methods. The second level of the scheme is realized by a dynamically tunable aggregators of the first level solutions. The results of an experimental numerical study of the proposed approach and a number of data analysis techniques are presented. The obtained results allow to confirm that it is possible to detect different states of the pumping technological equipment more effectively by the usage of the proposed approach as a part of intelligent data driven diagnostics system.

## 1. Introduction

Modern machine building is being improved together with increasing requirements for quality, reliability and accuracy of technical diagnostics of technological machines and equipment. The science of reliability and durability requires an integrated approach. Requirements to systems for the collection and analysis of diagnostic information are constantly increasing together with introduction of highly efficient technological processes, automated design and production systems, complex systems. Effective means for technological equipment state analysis can prevent the occurrence of serious failures. This contributes to reducing the costs of repair and operation, ensuring the safety of maintenance personnel.

The development of technological data collection tools and oil and gas equipment allows us to conduct in-depth monitoring of the equipment parameters. It makes it possible to switch from repair and maintenance to repair determined by technical state of the equipment. This will reduce the cost of repairs and maintenance. One of the aspects of solving problems of the oil and gas industry machines and equipment technical diagnostics is the further improvement of the methodology for technical state classification. Such a classification involves the determination of mainly hidden operational defects based on information obtained by methods of nondestructive testing and the assignment of the current state of the diagnosed equipment to one of the classes. Each class can then be interpreted by decision

maker for the necessary corrective actions. In this case vibration diagnostics is one of the most informative methods of nondestructive testing for a wide range of types of oil and gas equipment.

An important aspect which affects the effectiveness of vibration diagnostics is the model for detecting and determining equipment defects based on diagnostic information. The analysis of vibration monitoring data manually is practically impossible. Therefore the processing of vibration monitoring data can be effectively performed with the use of automated defect recognition procedures. This forms a scientific problem and the result of its solution leads to the creation of an effective system for the analysis of diagnostic information obtained by vibration diagnostics methods. Analysis of information sources on this issue indicates the possibility of improving the efficiency of diagnostic information processing by using modern means of solving the classification problem. Data mining techniques are significantly effective approaches which demonstrate fairly high efficiency in solving classification problems in many industries.

As for methods of data analysis a number of works suggest using a neural networks for diagnostic information analysis in the oil and gas industry. To reduce the training time of neural networks and improve the quality of the neural network classification a data processing method using parallel neural networks was proposed [1]. Several researchers consider multidivariate adaptive regression splines as the basic models of data analysis [2]. The method makes it possible to obtain models that give a fairly accurate approximation, even in cases where the connections between predictor and dependent variables are nonmonotonic and complex for approximation by parametric models. Automatically generated decision trees and fault tree analysis are also applicable [3].

In this paper we propose a multi-stage approach (system) for diagnostic information analysis. Such a system makes it possible to combine several or all of the above mentioned methods of data analysis within a single classifier. The proposed multi-stage classifier is to be formed using methods of automatic generation of analytical models due to high computational and topological complexity. This will significantly improve the computational efficiency of the approach and achieve a synergistic effect based on the use of heterogeneous classifiers within the framework of a single model.

The main theoretical aspects of the proposed multi-level system for processing diagnostic information are described below. A description of the practical application of the system for processing vibration diagnostics data of centrifugal pumping units which are widely used in the oil and gas industry is also given. The results of the proposed multi-stage classification system application in comparison to usual classification scheme techniques are presented.

## **2. Multi-stage intelligent classifier for vibration monitoring data analysis**

### *2.1. Ensemble classifier*

It is possible to represent ensemble classifier as a pair  $(C, D)$ , where  $C = (C_1, C_2, \dots, C_n)$  is a set of  $n$  individual classifiers, solution of which are taken into account in the formation of a common ensemble solution;  $D$  is the method of calculating an ensemble solution based on the solutions of individual classifiers (combining strategy). To use an ensemble classifier to solve a specific problem it is necessary to perform the following basic steps:

1. Creation of a set of individual classifiers  $C$  ;
2. Design of a scheme for calculating the ensemble solution  $D$  .

The execution of each stage is a separate problem for the solution of which various methods can be used. The most resource-intensive stage is the first stage. During this stage the structure and parameters of single classifiers are determined. Often such classifiers are initially formed in the form of simple structures. Such simple classifiers are relatively easy to train and they are less likely to retrain because of their simple structure. The choice of simpler classifiers is due to the fact that each of them theoretically solves a simpler problem obtained due to the decomposition of the original problem in an explicit or implicit form.

The second stage is the choice of a strategy for combining solutions. It usually requires less computational resources. Its complexity depends on the chosen combining strategy and the complexity

of the problem being solved. This stage takes considerably less time than the first stage in the case of choosing the most simple combination strategies. One of the main directions for increasing the efficiency of solving problems in the case of using an ensemble approaches is to use more complex combining strategies. It leads to the situation when more and more computing resources are redistributed in favor of the second stage. Accordingly the importance of this stage is growing both in ensuring a higher quality of the solution of the problem, and in the amount of the resource spent on the solution of the problem as a whole. That is why it becomes urgent to improve existing and develop new effective combining strategies. It would ensure intensive use of the computing resource and allow to increase the efficiency of solving problems. An overview of some combination strategies as well as a description of the proposed approach to the formation of a common solution in ensemble classifiers are presented below.

## *2.2. Combining strategy for ensemble classifier*

There are two main types of combining strategies used to solve classification problems using ensembles of classifiers: selection and fusion. Selection approach is based on the assumption that each individual classifier specializes on a specific local subspace of the problem. Each subspace has a separate classifier, as in [4], or, in some cases, a separate classifier can correspond to several local subspaces of the solution of the problem [5]. Fusion strategy assumes simultaneous use of all technologies in the entire space of the problem under consideration.

Concerning the use of an input signal in the formation of a collective solution, methods can be divided into two groups. The first group includes approaches that use static structures to develop a common ensemble solution. Such approaches can be called traditional. The decision-making schemes in such approaches are static and do not depend on the values of the input variables. The second group includes methods that take into account not only the decisions of individual classifiers but also the input variables themselves. Such methods are called “dynamic” because of the direct dependence of the method of forming a general solution on input variables values.

The first group is customary to distinguish two subgroups:

- Approaches using non-adaptive structures. Such schemes include, in particular, the following traditional approaches: averaging, maximum rule, median rule, and voting and the Borda rule.
- Approaches using various techniques for adapting combining strategies schemes. Such approaches are of the greatest interest due to many possible options for their implementation and the possibility of improving the quality of classification without performing the most costly stage of the formation of individual classifiers.

The development of such methods are connected with the search for additional opportunities for the adaptation of ensemble systems to a specific task. Involving some additional computing resource such methods allow increasing the efficiency of ensemble classifiers. It is important that such methods allow intensive use of computing resources. Let's list some known approaches:

- Decision templates method [6].
- Weighted averaging - there are a large number of approaches that differ in the way we calculate the weight coefficients. Different methods also use different procedures to select those classifiers whose solutions will be taken into account when forming a common solution.
- The stacked generalization method - the technology of using collective classifiers, which allows to organize a two-step procedure for the formation of solutions by a group of classifiers with a nonlinear combination of individual solutions. There are also various modifications to this method, for example, a dynamically modified multilevel generalization method [7].

## *2.3. Proposed multi-stage classification approach*

To increase the efficiency of the implementation of the stage of forming solution in classifiers ensembles an evolutionary three-level approach was proposed. It involves decomposition of the problem implicitly at the second level. In general the approach is an extension of the multilevel generalization method by introducing an additional intermediate level in the structure of the process of

forming the solution of the problem. The stages of the proposed three-stage approach of forming an ensemble classifier are described below.

Stage 1. At this stage a pool classifiers are formed. This stage is common to all ensemble approaches. In the general case any available effective method of forming separate classifiers of a selected type can be used. The amount of resources available for use at this stage is determined by the general requirements for the time, the required accuracy and available computing power.

Stage 2. At the second stage a set of classifiers of the second level is formed independently. Number of classifiers coincides with the number of classes in the problem under consideration. The inputs of the classifiers of this level are the values obtained at the output of the classifiers of the first level. At the same time, for each second-level classifier training is performed according to the following rule: the target value at the output of the classifier for all examples corresponding to the class with the number is equal to 1; for all other examples, the target value of the classifier output is equal to zero.

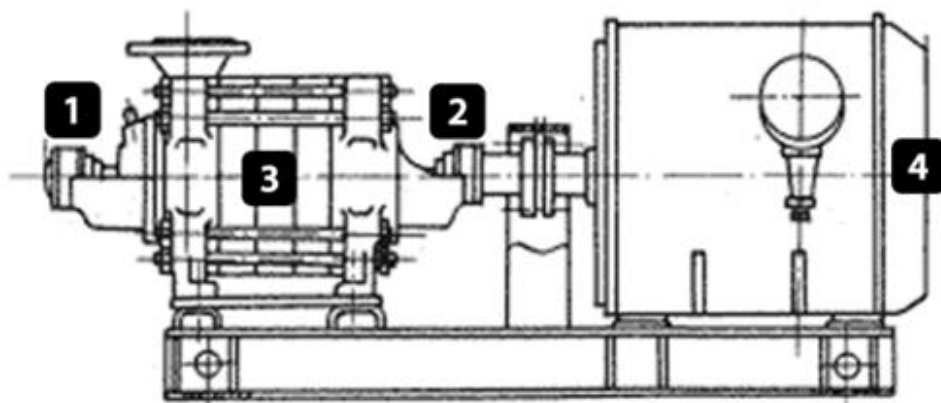
Thus in the second stage the problem decomposition is performed - each classifier of the second level forms a surface in space that cuts objects of one class from objects belonging to any of the other classes. To solve this problem we suggest using a method based on the use of hybrid genetic programming for combining the solutions of individual classifiers [8].

Stage 3. At the third stage the aggregation of the solutions of classifiers of the second level is carried out with the purpose of working out a general solution (class assignment value for the input set). The choice of a scheme for ensemble solution calculation is the subject for further research. Within the framework of the proposed approach we used the following simple and obvious rule: a classifiable object belongs to the class for which the corresponding classifier of the second level produced the maximum value of the output signal.

### 3. Experimental study

#### 3.1. Pumping unit and collection of vibration monitoring data

A monitoring and diagnostic scheme is necessary to obtain raw vibration monitoring data for exploring efficiency of the proposed multi-stage approach. For the processing and analysis, centrifugal pumps will be used for the centrifugal section pump type 60-330. The number of pumps for which data analysis is performed is 87 units. Figure 1 shows the pump and the location of the vibration sensors. For the reading of the vibration sensors, a vibration diagnostic analyzer, ADP-3101, was used. The analyzer has four measuring inputs, which allows simultaneous connection of four vibration transducers.



**Figure 1.** Points for installing vibration sensors on a centrifugal pump model 60-330.

For the analysis such indicators as speed and amplitude of vibration in horizontal, axial and vertical directions for each of four points for installing vibration sensors on a centrifugal pump model 60-330 was used. Each of 89 pumps is assigned a fault class: 1 - problem in the assembly, 2 - bearing failure, 3 - mechanical loosening, 4 - deformation of the shaft supports, 5 - unbalance, 6 normal condition.

### 3.2. Conditions of numerical experiments

To evaluate the effectiveness of the method proposed in Section 2 a number of numerical experiments were carried out. In the course of numerical research for the comparative analysis results were also obtained for other methods of forming ensemble classifiers and approaches using other ("non-ensemble") classification techniques. Results for ensemble approaches were obtained in the program system "IT-Pegas" developed by authors. In the course of the research in addition to the proposed methods results were obtained for the following methods of combining classifier decisions: simple averaging, equal voting, Borda rule and a multilevel generalization method.. A complete list of methods is given in the first column of Table 1.

In the course of the experiments a 5-fold cross-check was used to evaluate the effectiveness. Each time the results were averaged based on the results of designing and solving the problem with the help of five ensemble classifiers of the same type. As an efficiency measure, the average value of the classification reliability estimate was used, which was calculated as the ratio of correctly classified examples to the total number of examples in the exam sample. ANOVA technique was used for the proposed approach to assess the statistical significance of the results [9].

As classifiers of the second level in the proposed three-stage method classifiers obtained by the method of genetic programming were used [10]. The number of classifiers at this step is equivalent to the number of classes in a particular task. The number of generations in the genetic programming method for each classifier is 200, the number of individuals per generation is 100.

### 3.3. Results

Results of numerical experiments are given in table 1.

**Table 1.** Results of numerical experiments.

Classification techniques	Classification Reliability		
	Training Sample	Examination Sample	Test Sample
Ensemble Fuzzy Logic Classifier	0.921	0.821	0.757
Fuzzy Logic Classifier	0.891	0.794	0.725
Bayesian Classifier	0.847	0.679	0.629
Multilayer Perceptron	0.833	0.716	0.693
Boosting	0.760	0.700	0.656
Bagging	0.847	0.684	0.630
Random Subspaces	0.852	0.677	0.632
Neural Network Ensemble with Simple Averaging	0.892	0.805	0.740
Neural Network Ensemble with Voting	0.918	0.815	0.783
Neural Network Ensemble with Borda Rule [11]	0.905	0.831	0.772
Neural Network Ensemble with Stack Generalization	0.925	0.852	0.785
Proposed Approach	0.947	0.857	0.804

In general the results of numerical experiments show that the proposed approach is no less effective than most other methods whose results of the effectiveness evaluation were considered in a

comparative study. The proposed approach shows the best results on the problem of diagnosing the technical condition of the pumping unit.

#### 4. Conclusion

The article outlines the main ideas of ensemble classifiers and describes some particular techniques for combining solutions of individual classifiers used to solve a wide range of problems. A new three-stage ensemble approach for solving classification problems is proposed. Techniques for designing individual classifiers at each stage of the proposed approach are described. The results of a numerical study of the effectiveness of the proposed method and some known competing approaches are presented. The results of experimental study show that the proposed method makes it possible to classify more reliably in comparison with methods using other combination strategies and some known "non-ensemble" approaches. In the future it is planned to test the proposed three-stage scheme on a number of complex problems and use it to solve a wide range of practical problems.

#### 5. References

- [1] Galuev G A, Korovin S Y and Korovin Y S 2015 Management of Technological Processes of Oil and Gas Production Enterprises on the Basis of the Multineuroagent Decision Adoption Support System *Optical Memory and Neural Networks Information Optics* vol 14 no 3 pp 179–184
- [2] Maucec M et al. 2015 Multivariate Analysis and Data Mining of Well-Stimulation Data by Use of Classification-and-Regression Tree with Enhanced Interpretation and Prediction Capabilities *SPE Economics & Management* vol 7 no 2 pp 60–71
- [3] Ruijters E and Stoelinga M 2015 Fault tree analysis: A survey of the state-of-the-art in modeling, analysis and tools *Computer science review* 15, pp 29–62
- [4] Perrone M P and Cooper L N 1993 When networks disagree: ensemble method for neural networks *Artificial Neural Networks for Speech and Vision* pp 126–142
- [5] Plumpton C O and Kuncheva L I 2010 Choosing parameters for Random Subspace ensembles for fMRI classification *Proceedings of Multiple Classifier Systems (MCS'10)* pp 54–63
- [6] Kuncheva L I, Bezdek J C and Duin R P 2001 Decision templates for multiple classifier fusion: an experimental comparison *Pattern recognition* vol 34 no 2 pp 299–314
- [7] Ren Y, Zhang L and Suganthan P N 2016. Ensemble classification and regression-recent developments, applications and future directions *IEEE Computer Intelligence Magazine* vol 11 no 1 pp 41–53
- [8] Bukhtoyarov V and Zhukov V 2014 Ensemble-distributed approach in classification problem solution for intrusion detection systems *Proceedings of International Conference on Intelligent Data Engineering and Automated Learning* pp 255–265
- [9] Rouder J N, Engelhardt C R, McCabe S and Morey R D 2016. Model comparison in ANOVA *Psychonomic bulletin & review* vol 23 no 6 pp 1779–1786
- [10] Pal S K and Wang P P 2017 *Genetic algorithms for pattern recognition*
- [11] Krawczyk B, Minku L L, Gama J, Stefanowski J and Wozniak M 2017 Ensemble learning for data stream analysis: A survey *Information Fusion* vol 37 pp 132–156

#### Acknowledgments

The study has been undertaken as part of the research into the subject MK-1574.2017.8 “Designing the expert system of the analysis and control of reliability, risks and emergencies in support of the operation of petroleum refinery equipment” funded by the Grant Advisory Board for the President of the Russian Federation in a bid to provide governmental support to young Russian scientists.