THE PROBLEMS OF AUTOMATIC DATA PROCESSING AT THE SYSTEMS OF CONDITION MONITORING OF INDUSTRIAL EQUIPMENT

Alexander Serov,
Machinery Engineering Research Institute, Russian Academy of Science
M.Kharitonievskiy, 4, Moscow, 101990, Russia
alexser16@yahoo.com

Introduction

At each stage of technical development of human society it's possible to pick out the group of priorities, the successful solution of which ones largely determine the future of the technological progress. Currently, these tasks certainly include the automation of the analytical problem-solving activities. These problems are interdisciplinary. The main reason for this is the presence of similar problems to be solved in various branches of modern science and technology.

Objective reasons to introduce the achievements of Technical Cybernetics in solving problems of automation is the ever-increasing degree of specialization of knowledge in every branch of scientific and technical human activities, the continued differentiation of knowledge in our society, leading to the emergence of new scientific and technological areas and some slippage of general level of education relative to the level of scientific and technological progress. At this stage, the period of time in the 10 - 15 years is the interval during which one the many branches of scientific and technological activities take place in its development stages from pure research to the purely technological. The inculcation of automation in such industries usually has two components:

- Building a methodological basis, usually accompanied by a simultaneous construction or adaptation of existing mathematical tools. The use of these tools allows ones to solve research problems in the given industry. Creating of automation allows to perform a set of routine operations, thus speeding up, the research activities.
- Production processes, especially in high-tech industries are characterized by significant demands on the performance accuracy of each of the operations. At these circumstances, the only way to compliance with all requirements is the use of automation.

The field of non-destructive testing (NDT) is not the exception to the above. Advances in technology of nondestructive testing are traditionally used at the solution of control problems of the industrial equipment. Diagnostic tools used at this domain, can be divided into two classes in accordance with the terms of their use [1]:

- Tools used to test the functioning of the equipment working in operational mode.
- Tools used to test the characteristics of the equipment working in non-operational mode (tools used in the manufacture, preventive inspections, repairs, technical testing, preoperational testing, etc.)

Automation of the testing of equipment working in operational mode, is currently one of the most pressing problems. This applies particularly to those industries in which the operation of equipment associated with the risk of man-made disasters. Traditionally, the architectural solution of such control systems identifies two levels of problems to solve. The problems of each of these levels are solved by a separate subsystem. The prescription of the protection subsystem, whose work is carried out in automatic mode, is to prevent an emergency.

The automation of working mode of the protection subsystem is based on the selection of operating parameters of the controlled equipment, a set of values of which ones can characterize the level of criticality of the equipment working mode. Second subsystem, subsystem of diagnostics, is prescribed for the checkup of the current technical condition of the equipment, analysis of changes of technical parameters and prediction of changes in operating characteristics of the object of control. Solving the problem of diagnosing directly related to the tasks of planning and is one of the most important factors determining the efficiency of the equipment operation. In this regard, at present special attention is paid to the process of developing of diagnostic systems.

Stationary systems of automated diagnostics of the equipment: some problems of the development

The basis for the development of methods for the automated diagnosis is a gradation of the technical state of the object of control (see, for example, [2, 3]). This gradation allocates a certain set of levels of technical condition of the equipment. Each of the selected levels of technical condition is associated with a certain set of values of the numerical characteristics that must be determined by measuring the technical parameters of the controlled equipment.

Modern equipment for the control of which one stationary diagnostic systems are developed, usually characterized by a large number of different functional subsystems and high-complexity designs. This leads to the fact that a set of technical parameters used for the testing of the equipment' functioning, is sufficiently large. And these parameters characterize the physical quantities of a fairly broad range. As an example, the work of gas turbine engines is characterized by several tens of parameters. These parameters include oil pressure and temperature in different parts of the engine, temperature and pressure of the gas, the rotational speed of shafts. At this situation, the development of the universal methods of diagnosis is extremely difficult problem: an analysis of objects with different structures, must take into account this difference. In addition, each of the controlled parameters has a certain level of relevance and informativeness in terms of describing the technical state of the object of testing, and this requires consideration in the analysis. The peculiarity of the technical system as an object of diagnosis is the close and often non-linear relationship of electrical, electromechanical and mechanical assemblies and components, which also requires consideration in the process of diagnostics.

All these features associated with the development of the automated systems to monitor complex technical objects, led to the selection of certain areas, which have become a priority in creating automation. At the methodological level of this problem the main efforts so far are focused on finding the numerical characteristics, which change would reliably diagnose the changes of technical state of the equipment (see, for example, [4]).

At the architectural level of diagnostic systems' development the preference often is given to multicomponent complex that includes several sub-systems monitoring. The work of each of the subsystems may be based on a separate set of physical principles.

On the one hand, it allows to increase the confidence level of diagnosis of technical condition for the investigated object, but on the other hand, this leads to a number of additional problems. Technical parameters monitored by different subsystems have different sensitivity to the change of the technical state of the investigated object. So it's necessary to take into account these various degrees of sensitivity.

In the case of presence of multiple monitoring subsystems within a single diagnostic system there appears a problem associated with assessing the technical state of the object of control on the basis of a several sets of measurement data. This leads to the need of incorporation of certain level of integration into the architecture of automated diagnostic system, which one is responsible for the joint processing of testing results supplied by autonomous monitoring subsystems. Thus, the problem of developing methods of the processing heterogeneous information arises at the development of modern diagnostic systems.

Methods of data analysis, currently used in the field of technical diagnostics

One of the most promising challenges in the field of diagnostic systems is currently the problem of full automation of all stages of diagnosis. At present, this problem is elusive. One reason for this is the fact that the technical diagnosis still is a domain characterized by a weakly-structured methods of problem solving. Existence of a large number of elements within the object under control, which have nonlinear interaction with each other, as well as the interdisciplinary nature of the phenomena lead to the inability to formulate the ultimate theory of automatic diagnosis at present time.

One major feature of the modern field of technical diagnostics is the need for simultaneous use of the achievements of technical cybernetics, systems analysis and processing of signals in order to achieve success in solving of emerging problems. In the problems of signal processing is now widely used Fourier analysis and the theory of wavelets. In analyzing the experimental data and calculation of quantities characterizing the emerging signs of defects in equipment, typically used statistical treatment. Methods of cluster and discriminant analysis are used in tasks of organizing and classifying of experimental data. In recent years, there are the attempts to use the theory of artificial neural networks for solving the problems of approximation and prediction in the subject area of diagnostics. Full listing of methods that are currently used in solving problems of technical diagnostics, of course, not possible because of the fairly rapid development of this branch of science. However, I would like to focus on technology, the use of which on the one hand, is becoming necessary, and on the other hand can lead to significant progress in the field of automation problems of technical diagnostics.

Data Mining

Since the early 90-ies of the last century began the development of the field of applied science, which is called Data Mining. To date, Data Mining is an open multi-disciplinary field that has emerged and developed on the basis of scientific disciplines such as applied statistics, theory of artificial intelligence, theory of algorithms and other areas of modern scientific knowledge, some methods of which ones can be used in solving problems of system analysis. The emergence of this interdisciplinary field can be a point of view to consider a reaction to the lack of scientific discipline, in the immediate area of interest which would include development of methods for the analysis of complex open dynamic systems. Initially, even at its very origin, the field of Data Mining positioned as a purely applied discipline, combining methods for solving practical problems. Apparently, for this reason so far of Data Mining methods are widely used in fields such as information technology, business analytics, Web Mining, and, unfortunately, still poorly used in the technical disciplines.

One of the methods, which were the basis for the emergence of intellectual automated data processing, as such, is the Group Method of Data Handling (GMDH,) [5], developed by A.G. Ivakhnenko and his colleagues [6, 7]. Areas of possible applications of GMDH is extremely broad. The most promising area of GMDH is currently a technical diagnostics.

Group Method of Data Handling: general information

One of the major theoretical positions of GMDH is the assertion of the fact that with increasing complexity of the problem appears the plurality of variants for its solution, accompanied by a corresponding decrease in the reliability of each of these solutions [6]. This assertion is in some contradiction with the approach of classical mathematics, formulating the theorem on the existence and uniqueness of solutions. Creators of the Group Method of Data Handling first in 60 years of XX century, introduced the concept of self-organization in relation to mathematical objects. The result was the emergence of a number of mathematical disciplines of the modern mathematics.

Group Method of Data Handling was established as a basis of software for the specialized computer systems, intended for the direct simulation of complex systems on a small number of experimental data. The basic concept of GMDH is the notion of a mathematical model. This term means a system of regression equations, used for a single forecast of the future course of processes existing in a complex system, or to describe the physical and other laws operating in an investigated object [6]. The method of least squares becomes the numerical basis upon which one the apparatus of the theory of GMDH is grounded.

GMDH has some significant differences from the modern mathematical methods used in data analysis and modeling of complex systems. The first difference, which is methodological in nature, is that the use of GMDH minimizes the so-called "Anthropomorphic" factor. Modern methods of constructing mathematical models are characterized by the fact that a specialist formulating the model determines where is the boundary of the investigated system. It was he who formulated the system of mathematical propositions that will be later used when searching for a solution of the research problem. This fact leads to considerable difficulties in constructing models of dynamical systems. Such phenomena as nonlinear interactions of multiple elements, in particular, characterized by time delay, making it impossible to use traditional approaches in the analysis of systems, whose level of complexity is sufficiently large.

Another key difference of GMDH is the possibility of finding a solution which has an optimal complexity. Modern methods of data analysis in an attempt to solve the problem of forecasting face significant challenges. These problems arises due to fact that it is impossible to obtain an unambiguous assessment of a large number of coefficients of a mathematical model on the basis of using of small number of interpolation nodes, i.e. the small number of experimental values. The result is an actual absence of methods, theoretically justifying the development of the predictive models.

Group Method of Data Handling: short review of numerical method

The general formulation of a mathematical problem using GMDH algorithms is as follows. Let there is a function *f* depending on several arguments x_i:

$$f = f\left(x_1, x_2, \ldots, x_N\right) = f\left(\vec{x}\right) \qquad \textbf{(1)}$$

Let there is several numerical series as an initial data for the analysis:

$$\begin{aligned}
&\{f\} = \{f_1, f_2, ..., f_K\} \\
&\{x_1\} = \{x_{11}, x_{12}, ..., x_{1K}\} \qquad \textbf{(2)} \\
&\{x_2\} = \{x_{21}, x_{22}, ..., x_{2K}\} \\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
&\{x_N\} = \{x_{N1}, x_{N2}, ..., x_{NK}\}
\end{aligned}$$

where: *N* is the number of arguments in dependence of function *f*,
K is the number of elements at each numerical series

The general problem is formulated so that it is necessary to reconstruct the analytical form of function **(1)** on the basis of the existing set of experimental data **(2)**.

In the future when referring to a particular set of values of the arguments we will use the notation: $\vec{x}_i \equiv \left\{x_{1i}, x_{2i}, \ldots, x_{Ni}\right\}$, where $i = 1, 2, \ldots, K$ is the number of the set of arguments values inside of the corresponding numerical series.

In solving this problem the methods of regression analysis are used. In particular, the set of basic functions is introduced:

$$\{\varphi\} = \left\{\varphi_0\left(\vec{x}\right), \varphi_1\left(\vec{x}\right), \varphi_2\left(\vec{x}\right), \ldots, \varphi_M\left(\vec{x}\right)\right\} \qquad \textbf{(3)}$$

The solution of a regression equations **(4)** is performed on the introduced set of functions **(3)**.

$$\vec{f} = D \cdot \vec{\beta} \qquad \textbf{(4)}$$

where: $\vec{\beta}$ is unknown vector of regression coefficients having the size *K*,
D is a matrix of the system of regression equations having the following form in general case:

$$D = \begin{pmatrix} \varphi_1(\vec{x}_1) & \varphi_2(\vec{x}_1) & \dots & \varphi_M(\vec{x}_1) \\ \varphi_1(\vec{x}_2) & \varphi_2(\vec{x}_2) & \dots & \varphi_M(\vec{x}_2) \\ \dots & \dots & \dots & \dots \\ \varphi_1(\vec{x}_K) & \varphi_2(\vec{x}_K) & \dots & \varphi_M(\vec{x}_K) \end{pmatrix} \qquad \textbf{(5)}$$

One illustration of the commonality degree of GMDH approach in the analysis is the conditions imposed on the choice of basis functions φ. Unlike the methods used in a classical functional analysis, GMDH doesn't require that a set of basis functions must be orthonormal.

Choosing a specific set of functions can be based on a priori information about the physical properties of the considered processes.

If this information is missing or insufficient for the optimal choice of basis functions, this problem can be solved by a serial testing of a number of basis functions with respect to some external criterion. At present polynomial and harmonic sets of basis functions are most popular in the study of smooth and oscillatory dependencies, respectively.

Preparing for the solution of the mathematical problem consists of the following basic steps:

- Definition of a set of basis functions **(3)**
- Determination of an external criteria, which will be implemented at quality analysis for the model
- Partitioning the entire set of experimental values for several subsets, which, in particular, may include teaching, supervisory and examination sample
- Determination of the conditions for the model complicating in finding the optimal model

Within the framework of the concepts introduced in the GMDH, the matrix *D* defines the model analyzed according to **(1)**. Therefore, the solving of the problem of the optimal model' finding is at the periodic change of the form of matrix *D*, the calculation of the corresponding system of regression equations and selection of models based on the given external criterion. So at solving of the numerical problem there is the trying different types of matrix *D*, and the choice of that type which corresponds to the condition of minimum of the external criterion.

One of the main parameters, which are partitioning the modern numerical methods implementing the ideology of GMDH, is a method of models sorting out. There are two main classes of GMDH numerical methods:

- Combinatorial (one-row) methods
- Selection (multi-row) methods

The most general method is the one that implements the combinatorial search. When using a method there is an exhaustive search of all variants of the model imposed on the set of basis functions. The disadvantage of this algorithm is the extremely fast growing number of analyzed variants of the model. Even if the number of an arguments *N* is greater than 10 the amount of computation becomes so large that the use of the method becomes practically impossible.

Selection methods are characterized by periodic restriction of the analyzed set of models. This is achieved by introducing of intermediate stages of assessing the quality of the model and by the corresponding selection of a set of obtained dependencies before the subsequent complication of the model.

Selection methods allow to solve the problem in practice, a search using the retarded arguments, whose number may exceed 10^3. That is impossible for combinatorial methods on the present level of development of computational tools. The introduction of multi-row techniques gave rise to the whole area of research within the GMDH, questions of which one are devoted to search engine optimization solutions.

Common GMDH shortcomings should include low noise immunity of the analysis method, noticed more by the authors of the methodology. Several variants for the algorithmic implementation of GMDH, compensating distortions of the analyzed data or taking into account the features of these distortions are proposed in the monograph [7].

Field of the most active use of modern methods of GMDH is currently at the crossroads of several scientific and technical disciplines.

The ideology of GMDH, in particular, was used at the following set of problems solution such as the construction of the hybrid GMDH-neural networks, development of control algorithms for robotics, development of methods for the parallel processing of a large volumes of data, creation of ways of the information structuring and recovery, development of methods for the constructing of reduced order models, development of methods for the processing of speech information [8]. All of this is the result of a very general provisions of the underlying methodology of GMDH.

Group Method of Data Handling and technical diagnostics

One of the areas in which the use of GMDH ideology most promising is the area of technical diagnostics. Due to high demands for computing resources used by Group Method of Data Handling this method can not yet be applied in solving the problems of an operational nature, which, in particular, include the problems of equipment monitoring. However, a viable alternative to GMDH in solving problems of intelligent data processing for today, in fact, does not exist.

It should be noted that the ideology use of Group Method of Data Handling is possible in several directions, which, in particular, include the following:

- The research work related to the development of new methods of diagnosing
- Immediate implementation of GMDH in the program complexes used in solving problems of technical diagnostics of the equipment

There are several classes of problems, at the solutions of which ones the use of GMDH can lead to significant advancement in the field of a diagnostics as a whole.

Group Method of Data Handling can be used to solve the following classes of problems specific to the field of technical diagnostics:

- Problems of analytical processing of a large volumes of data monitoring
- Problem of forecasting
- Identification problems

The future perspectives of GMDH in addressing these types of problems are wide enough. We will try to consider the examples of some of these applications.

In problems of analytical data processing Group Method of Data Handling can be used in the analysis of time trend parameters used in the system diagnostics. The sources, for which ones trend characteristics can be calculated may include both directly measured process parameters (which enter into the diagnostic system from the autonomous monitoring sub-systems) and numerical characteristics (used for the analysis of the technical condition of the equipment).

The problem of trend calculation is of particular importance for the technological characteristics that vary non-monotonic way depending on the time. For these characteristics it is important to know not only about the fact of the exceeding of the measured values of a pre-determined range, but also the trend in these variables. From this perspective, the construction of models of the optimal order can be used not only for solving problems of interpolation and also for the extrapolation of experimental data. Using GMDH in the generation of models of the optimal order allows one to build the predictive models. These models can be directly applied in the problem of evaluating the residual life of industrial equipment and planning of repair and restoration operations.

The class of identification problems in the technical diagnosis, which can be solved with the aid of GMDH achievement, is quite wide. Problems of identification of equipment defects typical for the area of technical diagnostics may be included here. Also one can include here the problem of determining the class of technical state of the controlled equipment. Within the computer-aided diagnostic system all these tasks are closely related to the above-mentioned problems of analytical data processing and forecasting. It was to be alone stay on task, which currently has more scientific than practical interest - the problem of complex multi-parameter analysis of monitoring results.

It should be noted that the problem of complex multi-parameter analysis is interdisciplinary in nature. The objectives of this area, having a similar nature, there are at present in various unrelated fields of science and technology: medicine, biology, sociology, economics, and various branches of engineering and technology. Methods developed to solve these problems, may be based on the ideology of GMDH. The fact that there are no any guidelines binding GMDH to a particular discipline of science can make the realization of multi-parameter analysis methods sufficiently general in terms of their subsequent application. Methods of multi-parameter analysis can be directed primarily at identifying of hidden regularities in the process which is investigated. One of the challenges of multi-parameter analysis is to identify functional dependencies in the study of time series, appearing as a result of experimental studies.

There was developed the method, an algorithm of which one refers to a class of multi-row algorithm. It implements the ideology of GMDH in the case of an arbitrary number of arguments and. Studies of this method revealed a number of interesting features. In particular, the criterion of the minimum mean-square error as the external criterion for solving the problem makes it possible to identify a set of solutions close to optimal one. To find a single solution that meets the formulation of the physical problem, it is necessary to use terms that have no statistical background, and describing the investigated systems from a physical point of view. The development of methods applied for multi-parameter analysis of experimental data has the perspective aim of the transition from automated to automatic methods of diagnosis. However, at the early stage of development its results can be used to create decision support systems, which are currently considered to be one of the most promising solutions for the architectural design of control system for the stationary diagnostic systems.

Conclusion

In conclusion, it's necessary to note that the authors of the Group Method of Data Handling considered the limits of its application is much broader than is commonly believed today, on the basis of current data about its use. GMDH has great potential and breadth of use. Alexey Ivakhnenko believed that proposed approach could become the basis for the building a fundamentally new technological basis of human society [6]. In particular, the ideas of GMDH can be successfully used in the construction of artificial intelligence, transforming the whole area of systems analysis, in building of the predictive models for the complex dynamic systems. The ideology of GMDH is now in a good agreement with the steady growth of computational power of the computers and with the advances in technology of parallel computing. This fact makes the task of further development of GMDH particularly attractive.

Literature

1. Kluev V.V., Fursov A.S., Filinov M.V. Approaches to the construction of systems for evaluating the residual life of technical objects // Control and diagnostics. — 2007 — № 3. — p. 18-23.
2. Vibration diagnostics of the emerging defects / [ed. MD Genkin] — M.: Nauka, 1984.
3. Davidov P.S. Technical diagnostics of electronic devices and systems. — M.: Radio i svijaz, 1988.
4. 4-th International Congress on Technical Diagnostics, Olsztyn, Poland, 2008.
5. URL: http://www.gmdh.net
6. Ivakhnenko A.G. Long-term forecasting and management of complex systems: — Kiev: Technika, 1975. — 312 p.
7. Ivakhnenko A.G., Stepashko V.S. Noise Immunity of Modeling: — Kiev: Naukova dumka, 1985. — 216 p.
8. International Workshop on Inductive Modelling, Krynica, Poland, September 15-19, 2009