Artem AVRAMENKO[1]

PODWYŻSZENIE JEDNORODNOŚCI OCEN DLA KLASYFIKATORA WIELOWARSTWOWEGO

Streszczenie: W pracy zaproponowano metodę optymalizacji struktury systemu kryzysowego monitorowania informacji. Uzyskano redukcję czasu przeprowadzenia optymizacji poprzez zastosowanie klasyfikacji danych wejściowych. Wybrano najlepszy algorytm syntezy modelu dla każdej macierzy danych wejściowych. Efektywność nowej metody została potwierdzona eksperymentalnie. Czas restrukturyzacji modeli zredukowano 3-4 razy. Błędy modelowania nie są znaczące.

Słowa kluczowe: monitorowanie kryzysowe, modelowanie wieloaspektowe, czas restrustrukturyzacji system, klasteryzacja, błąd modelu

INCREASE OF HOMOGENEITY OF VIEW POINTS IN MULTIMODEL CLASSIFIERS

Abstract: Method for optimizing the structure of the crisis monitoring information systems is proposed. Reducing time of optimization has been achieved by classifying the input data arrays. Best algorithm of model synthesis is selected for each input data array. Effectiveness of the new method was experimentally confirmed. Time for restructuring of models was reduced by 3-4 times. Errors of modelling is not significantly worse.

Keywords:crisis monitoring, multilevel modeling, system restructuring time, clasterization, modelling error

1. Setting the problem

Today, the use and development of monitoring systems is a crucial issue as research and practical application, because they can draw conclusions about certain events, while only based on previous data obtained by observation. Thus, such systems successfully used as diagnostic systems in various spheres of life.
The composition of the monitoring system is determined by the following factors: its functional purpose, scope and target set in the list of tasks, functions of information processing assigned for monitoring system and defined by user.

[1] Bogdan Khmelnitsky Cherkassy National University, Department of intelligent decision support systems, email: RedStar929@gmail.com

The main objective of the simulation monitoring systems is to provide information for decision-making. Information obtained by modeling the properties of an object based on monitoring data collected during the measurement numerical characteristics of the object. Decision-making in emergency situations imposes several restrictions on the technologies for information these processes.
Developments of emergencies usually are chain and dynamic process that causes sharp deterioration of an object that is catastrophic for this object and environment. The object usually is a set of areas and economic objects located there. All this means that decisions in these situations should be obtained as soon as possible. Also, poor predictability and dynamism of emergencies generates a large number of parameters for modeling, some of which may appear the first time that means greater possibility of error in pre-derived models.

2. The aim of the article.

The main task is to ensure the reduction of time of re-learning the system while maintaining the quality parameters of the models.
In this paper, method is proposed to solve this problem by improving existing algorithm using classification of AMS to appropriate IDA instead of exhaustive search for ASM.

3. Analysis of recent research and publications

A multi-data conversion technology is implemented as an information system with hierarchical combination of multiparameter models [2, 3]. These models can be implemented using inductive algorithms, neural networks, genetic algorithms and others.
In this system, the choice of best algorithm for synthesis of multiparameter model (AMS) implemented by successive trials and choice of the best by comparing their results [1].
Then the hierarchy constructs from synthesized models. Every level of the hierarchy is needed for solving the local problems of data conversion. In such hierarchical structures, large number of models can be combined, from fifty or more (Fig. 1).

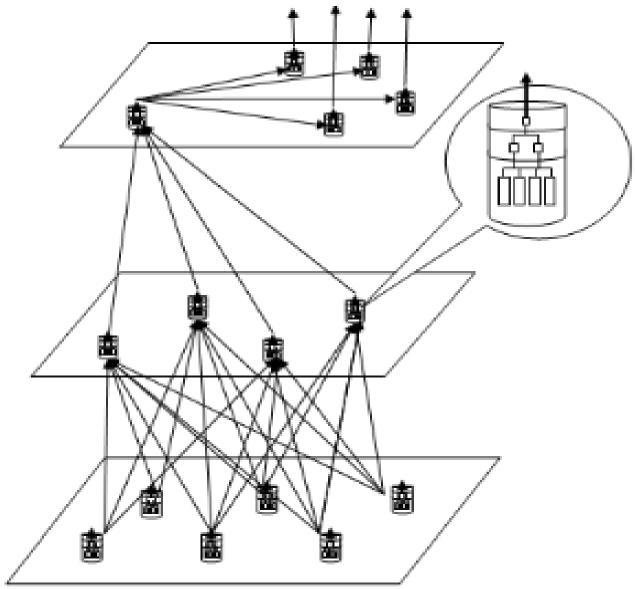

Figure 1. Structure of subsystem for information conversion

4. Pointing out the unsolved aspects of the problem

Since the object of our monitoring is emergencies then in the event when the properties of an IDA change and one of the models stops working, then "damaged" model and all models associated with it must be replaced by re-synthesized models. So, the whole system of fifty or more models need to re-learn.
This process takes a long time, about 40 minutes and more, depending on the number of models in the structure. In conditions of crisis monitoring where the justification for decision takes no more than 2-3 minutes and IDA properties change dynamically, therefore the situation of re-learning the whole system is pretty likely. This means there is a need to reduce the time of restructuring the hierarchy of models.

5. Results

To achieve this goal was formed the hypothesis that reducing the time of re-formation of the structure of the monitoring system can be achieved by solving the problem of recognition of best ASM from a predefined list for each IDA. Unlike sequential testing of each available ASM and selecting the best algorithm is proposed to construct a deciding rule to provide for each IDA the most suitable ASM. This rule is, basically, a polynomial model, which is synthesized with one of the ASM used in our system in this case it was GMDH [1]. This model can classify IDA to most suited ASM (Fig. 2).

Formulated hypothesis was tested experimentally. Solving rule was created by using multi GMDH algorithm [2]. Results of monitoring morbidity in Cherkasy region during 2000-2014 years [1] were used as IDA for synthesis of models. The process of model synthesis that contained morbidity dependent on the concentration of pollutants in water, air, food was studied in this work.

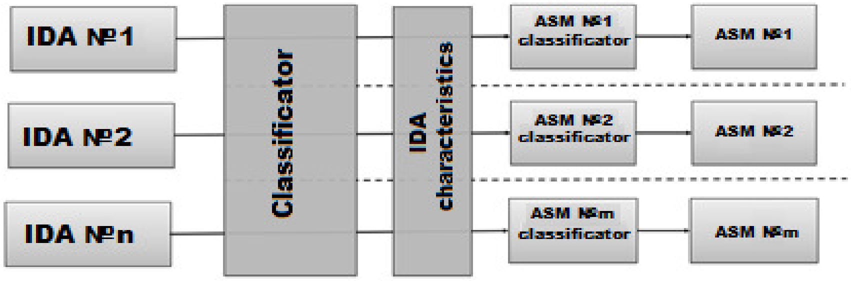

Figure 2. Diagram of proposed method

After analyzing modern methods of matrix and correlation analysis was formulated that the selection of algorithm for model synthesis is possible by pattern recognition technology based on the following table of informative parameters [3]:

– number of observations;
– number of independent variables;
– number of parameters that is strongly combined with the goal function;
– number of parameters not combined with the goal function;
– the average correlation coefficient of independent variables;
– the average correlation coefficient of independent variables and the goal function;
– the average determination coefficient of independent variables;
– the average determination coefficient of independent variables and the goal function;
– determinant of primary description table;
– determinant of the matrix normalized values of the independent variables;
– eigenvalue of the normalized table of primary description;
– eigenvalue of the matrix normalized values of the independent variables;
– maximum singular value of the normalized table of primary description;
– maximum singular value of the matrix normalized values of the independent variables.

Also as a result of research of an offered method the row of defects was found. Namely, problem in providing the classificator ASMs with needed quality data. User first need to test IDA in each ASM and find out which of them is best for each IDA, after that we find characteristics of said IDA and add them as data to synthesize classificators. However each such IDA, that can consist of hundreds thousands rows of data, only forms one row of characteristics. Therefore it requires from user to provide a sufficient amount of IDAs to make sufficient characteristics data to be able to classify them, which requires a lot of effort and not always even possible. To remove this problem the further improvement for method of classification of IDA by clasterization of input data was offered (Fig. 3).

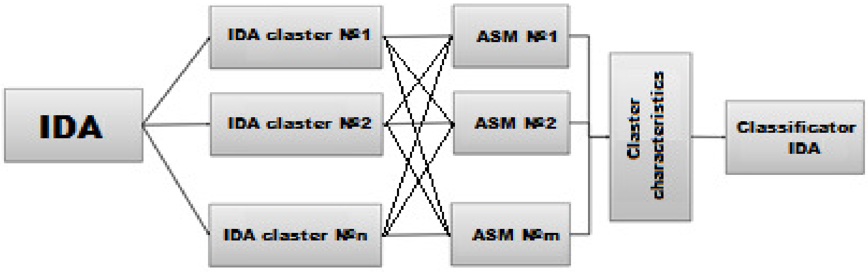

Figure 3. Diagram of proposed clasterizations of IDA

Each claster will have its own best ASM and characteristics. Which means every IDA will provide more than one row of characteristics data for classification. This approach will simplify work on providing the process of classification with data.
A research tool was created based on this technology. It was used for a comparative analysis of the new method and the method of exhaustive search for ASM which is implemented in multi-level monitoring systems [2]. On fig. 4 and 5 shows the results of the analysis.

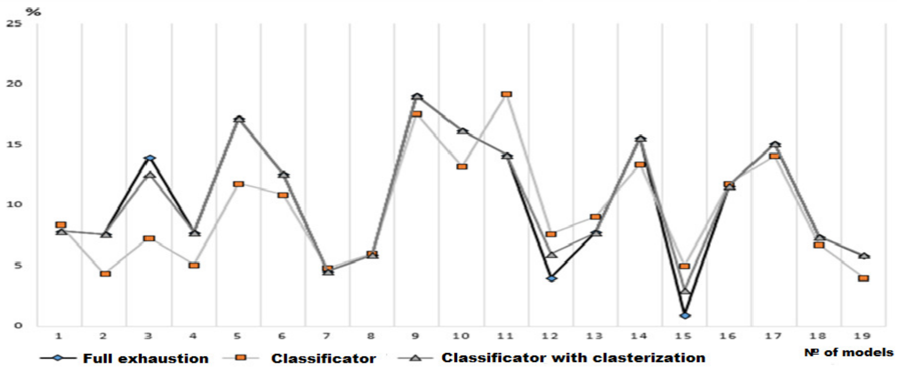

Figure 4. results of comparing modelling errors with classifier.

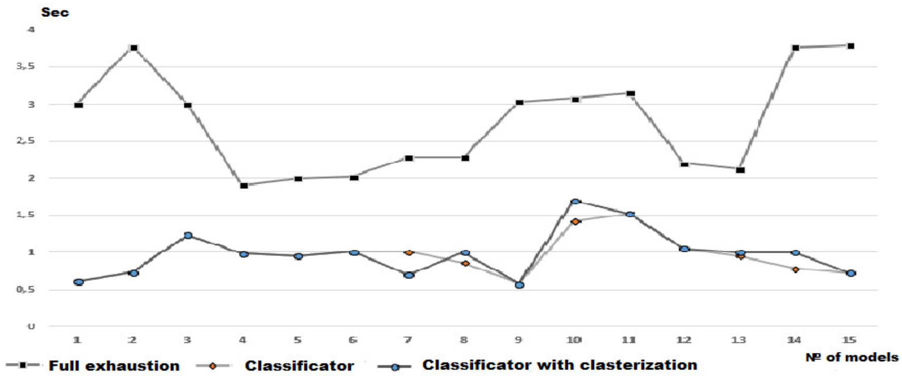

Figure 5. results of comparing time of model synthesis with classifier.

As a result of comparison of methods, synthesis time of model decreased by 5%, with errors of synthesis less then 10% in comparison to standard classification method. At the same time models was synthesized by 3.5 times faster than with full exhaustion method.

Conclusions and suggestions

Growth of modeling error is "payment" for reducing the time of model synthesis. Given the fact that the structure of the information system of multilevel data transformation contains 50 models and more, it is possible to achieve a significant reduction in the time structure by adapting to changes in the properties of IDA. In conditions of crisis monitoring these results give hope for the possibility of monitoring information systems with multi-data conversion technology for support decisions in emergencies localization. Also the proposed clasterization further decreased "payment" in modeling error by 10%.

REFERENCE

1. HOLUB S.V.: Zastosuvannya stratehiyi optymalnosti pry vybori alhorytmiv syntezu modelej u systemax bahatorivnevoho socioekolohichnoho monitorynhu / S.V. Holub, P.O.Kolos // Matematychni mashyny i systemy. 4(2010), 127 - 134.
2. IVAHNENKO, A.G.: Inductive Method of Models Self-organisation for Complex Systems. Kyiv: Naukova Dumka 1982.
3. HOLUB S.V.: Bahatorivneve modelyuvannya v texnolohiyax monitorynhu otochuyuchoho seredovyshha / S.V. Holub. – Cherkasy: Vyd. vid ChNU imeni Bohdana Xmel"nyc"koho, 2007. – 218 s.
4. KOLOS P.O.: Vyznachennya mnozhyny informatyvnyx parametriv tablyci pervynnoho opysu ob’yekta modelyuvannya./ Visnyk Cherkas"koho universytetu, vypusk 173. Cherkasy: Vyd. ChNU, 2009. S. 121-128.