UDC 004.896

S.V. Golub, S.A. Yemelyanov
(Cherkasy National University named after Bogdan Khmelnitsky, Ukraine)

Intelligent monitoring of cyber attacks

A new approach to monitoring and recognition of cyberattacks is proposed. Stages of cyberattack recognition are selected and thebest algorithm for their classification is selected.

Intelligent monitoring

After the attacks of Petya nobody doubts the relevance of the study of the direction of monitoring cyber attacks. Intelligent monitoring is an information technology providing knowledge of the decision-making process through the organization of continuous observations and processing of their results.

The result of the monitoring is:

1. detecting cyber attacks;
2. classification of cyber attacks;
3. defining measures of protection and compensation of consequences;
4. development of measures to improve the protection of corporate networks.

In order to determine the availability of cyber attacks, it is necessary to solve the problem of pattern recognition. According to the list of signs and their numerical significance, it is necessary to recognize the fact of a cyberattack.

To identify the type of threat solved the problem of classification of cyber attacks.

Classification of cyber attacks

Before beginning the classification, a known list of cyber attacks and numerical values of attributes. This is done expertly. Need to have historical data. As at the identification stage and at the classification stage, it is necessary to construct a decisive rule that can solve each of these local information transformation tasks.

To this end, it is proposed to use the MIST model synthesizer (monitoring intellectual system). The synthesizer contains more than 20 algorithms for model synthesis. They implement inductive methods of synthesis of genetic algorithms of the neural network of their combination in hybrid methods as well as multilayer and multi-level synthesis of models.

The cyberattacks have a special influence on the results of the monitoring of the cyberattacks. The informativeness of the input array and the power of the methods of synthesis of models are informative. This is provided by the qualification of the expert forming the vocabulary of the signs. And the ability of the MIST synthesizer to construct new algorithms for model synthesis adaptively to the properties of an array of input data.

Group method of data handling

Group method of data handling (GMDH) is a family of inductive algorithms for computer-based mathematical modeling of multi-parametric datasets that features fully automatic structural and parametric optimization of models.

GMDH is used in such fields as data mining, knowledge discovery, prediction, complex systems modeling, optimization and pattern recognition. Li et. al. (2017)'s results showed that GMDH neural network performed better than the classical forecasting algorithms such as Single Exponential Smooth, Double Exponential Smooth, ARIMA and back-propagation neural network.

GMDH algorithms are characterized by inductive procedure that performs sorting-out of gradually complicated polynomial models and selecting the best solution by means of the so-called external criterion.

A GMDH model with multiple inputs and one output is a subset of components of the base function:

$$Y(x_1, \ldots, x_n) = a_0 + \sum_{i=1}^{m} a_i f_i$$

where f are elementary functions dependent on different sets of inputs, a are coefficients and m is the number of the base function components.

In order to find the best solution GMDH algorithms consider various component subsets of the base function called partial models. Coefficients of these models are estimated by the least squares method. GMDH algorithms gradually increase the number of partial model components and find a model structure with optimal complexity indicated by the minimum value of an external criterion. This process is called self-organization of models.

The most popular base function used in GMDH is the gradually complicated Kolmogorov-Gabor polynomial:

$$Y(x_1, \ldots, x_n) = a_0 + \sum_{i=1}^{n} a_i x_i + \sum_{i=1}^{n} \sum_{j=i}^{n} a_{ij} x_i x_j + \sum_{i=1}^{n} \sum_{j=i}^{n} \sum_{k=j}^{n} a_{ijk} x_i x_j x_k + \cdots$$

The resulting models are also known as polynomial neural networks.

Combinatorial GMDH

Another important approach to partial models consideration that becomes more and more popular is a brute force combinatorial search that is either limited or full. This approach has some advantages against Polynomial Neural Networks but requires considerable computational power and thus is not effective for objects with more than 30 inputs in case of full search. An important achievement of Combinatorial GMDH is that it fully outperforms linear regression approach if noise level in the input data is greater than zero.

Basic combinatorial algorithm makes the following steps:

- Divides data sample onto parts A and B.
- Generates structures for partial models.
- Estimates coefficients of partial models using Least squares method and sample A.

– Calculates value of external criterion for partial models using sample B.
– Chooses the best model (set of models) indicated by minimal value of the criterion.

In contrast to GMDH-type neural networks Combinatorial algorithm can't be stopped at the certain level of complexity because a point of increase of criterion value can be simply a local minimum, see Fig.1.

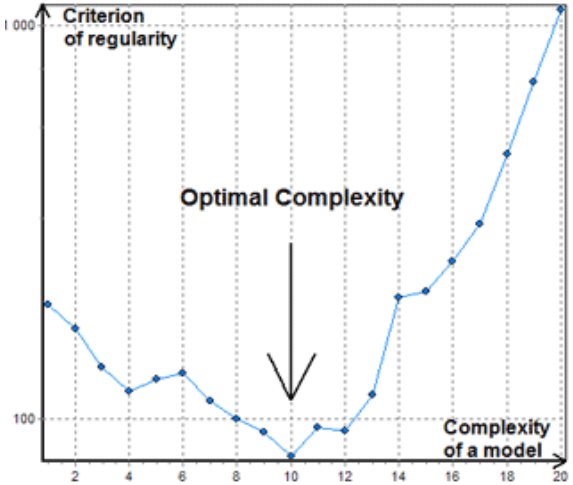

Fig.1. A typical distribution of minimal values of criterion of regularity for Combinatorial GMDH models with different complexity.

References

1. Golub S.V. Multi-level simulation in environmental monitoring technologies. Cherkasy: View. from. ChNU named after Bohdan Khmelnytsky, 2007. - 220 p.
2. Combined algorithm of GMDH with sequential complication of model structures based on recurrent-parallel computing / S.M. Efimenko // Inductive modeling of complex systems: Coll. sciences Ave. - K .: ISTC ITS NAN and Ministry of Education and Science of Ukraine, 2014.
3. Group method of data handling- Modeofaccess: WorldWideWeb:https://en.wikipedia.org/wiki/Group_method_of_data_handling (viewedonSeptember 27, 2018)