УДК 004.031.6:621.3.07

ALGORITHM FOR INTELLIGENT PREDICTION OF FAILURE MOMENTS IN COMPUTER SYSTEMS

Volodymyr OSYPENKO, Doctor of engineering sciences, professor
Kyiv National University of Technologies and Design
Gregory IVACHNENKO, Professor, Data Scientist in GfK Group,
United Kingdom

Keywords: inductive modeling, computer system, criterion of models selection, prediction, bifurcation points, rare events.

In the projects of the complex technical systems researches it is necessary to decide tasks when some phenomena have a substantial influence on a result, but they take place very rarely and badly predictable. Time intervals between such events are so great that it is possible to consider that they practically do not influence on each other, i.e. it is possible to consider that cross-correlation dependence between them is absent. Such phenomena got the name of «rare» events. The classic example of rare events can be a date (a top) of freezing of the river or water storage in a natural environment. Although freezing are in a middle stripe of Europe take place practically annually, a time intervals between them is so great, that it is possible to consider that each of them not influence on each other. To the category of rare events it is also possible to include natural cataclysms, for example typhoons, hurricanes, earthquakes etc., which can have the substantial negative affect on economy and on the sphere of habitation of mankind in general.

The another example of rare event can be a refuse of some complex technical unit (computer system, computer net, engine, etc.) which characterized by a natural similarity and almost identical operating conditions. The list of rare events it is possible to continue. Therefore undoubtedly, that successful forecasting of rare events is essential for the decision of many tasks in the ecology, economics and in research of reliability while complex technical systems testing

The point of occurrence of a rare phenomenon is often considered the moment (on the forecast interval) when the predictable value passes through a certain "critical" point.

In article [1] the original approach for forecasting rare events has been offered. It uses a technique that is effectively used by experienced specialists in diagnosing and predicting failures of technical devices. Under rare it is suggested to understand an event, coming in the object is looked for which there were not precedents in recent past (not exceeding the most time of delay $\tau_{\max}$).

Statement of the problem of forecasting rare events in technical systems. Under the object of modeling design in the technical systems will understand the few (n) observations of the same type of technical objects, which is observed at the same conditions on a time interval including only one event – failure. It can be when it takes place simultaneously monitoring party of the same type of computers or another technical device from the beginning of exploitation to the moment of their failure.

The task of forecasting the rare event ("bifurcation point" of the process) we will formulate as follows. Let us, that we have the results of monitoring of n the same technical systems behavior on the observation interval $T_{obs} = [t_0, t_k]$, in each of them once took place the event $\xi_i, i = 1, 2, \ldots, n$ – failure, and this moment fixed in a corresponding database. It is required to synthesize a model describing the behavior of this object on the forecast interval $T_f = [t_{k+1}, t_y]$ to predict a new moment of failure in order to eliminate it in advance.

Next, suppose that the whole interval is possible to split up on n intervals: $T_{obs} = \left[t_0, t_{s_1}, \ldots, t_{s_{i-1}}, t_{s_i}, \ldots, t_{s_{n-1}}, t_{s_n}\right]$, where t_{s_i} is a moment of i-th event. Such splitting goes out from supposing that on interval $[t_{s_{i-1}}, t_{s_i}]$ only ones the event ξ_i took place. In addition, every interval $[t_{s_{i-1}}, t_{s_i}]$ is broken on l of more narrow intervals $\Delta t' = [t'_{j-1}, t'_j] = const, j = 1, 2, \ldots l$, and in his knots, the control of parameter tests of the object is produced. Thus, there is the set of n moments of events $\xi_i, i = 1, 2, \ldots, n$ in our task. The forecast of $(n+1)$-th moment of time is the subject of our researches.

Such problem can be described through the regressive equation of model as follow:

$$\begin{aligned} y_f = f\{ & x_{1(0)}, x_{1(-1)}, \ldots, x_{1(-\tau_1)}, \\ & x_{2(0)}, x_{2(-1)}, \ldots, x_{2(-\tau_2)}, \\ & \ldots \\ & x_{m(0)}, x_{m(-1)}, \ldots, x_{m(-\tau_m)}, \theta\}, \end{aligned} \tag{1}$$

where y is an output (forecast) value, $x_i, i = 1, 2, \ldots, m$ arguments, $\tau_1, \ldots, \tau_m$ are the delays of each arguments which took into account, θ is a vector of the estimated parameters.

More laconically, a model (1) can be presented as

$$y_f = f(X, \theta_f). \tag{2}$$

The differences of such an approach from traditional forecasting procedures are: (1) among the arguments of function $f(.)$ the delay arguments of output value y are absent and (2) – output value is the time between the last supervision (control measuring) and beginning of the rare event (bifurcation point of the process).Thus, on the interval $\Delta t' = [t'_{j-1}, t'_j]$ of rare event occurrence $y \le \Delta t'$, and on the interval of «non-occurrence» (precedence) – $y > \Delta t'$.

In [3] for solving the same problem the original and effective method of initial informative base forming is also presented. This procedure got the name «floating scale» to indexation of delay arguments. «Floating» indexation means that index «0» appropriated to the control moment of the event has occurred. In this case, we have a situation $y \le (t'_{j-1} - t'_j)$. «Floating» indexation must be used for the creation of the datasheet. Values $x_{i(-\tau_i)}$ must correspond for delays of *i*-th interval, $i = 1, 2, ..., n$.

The procedure for prediction of rare events in the tecknical systems

The feature of such systems is that in complex processes interesting us rare events can take place the very limited times. If to consider the second variant of rare events research, i.e. the similar systems, then the number of rare events in them can be not too much. In both cases, the statistical data of initial supervisions are very limited. Thus, to the algorithms which could be possible to use for modeling and decision of prediction task of the rare event, strict requirements are demanded:

1) algorithms must save operability at limited low times of supervisions (n);
2) algorithms must save operability at high ratio signal/noise to be antijamming;
3) algorithms must have high speed and be able to process large datasheets for modeling optimal results in the form (1).

For today, the techniques of inductive self-organization of complex systems correspond to such strong conditions [2, 3].

The teaching of the model (1) is the task of identification in which from the positions of inductive modeling is exhaustively formulated in [4] as follows. The task of identification consists of forming from observation data $W = (X \vdots y)$ the same set of $\Im$ models having different structures of the kind $\hat{y}_f = f(X, \hat{\theta}_f)$, where θ is a vector of the estimated parameters and selecting of the optimal model under a minimum of the criterion $CR(\cdot)$ [4]:

$$f^* = \arg\min_{f \in \Im} CR(y, f(X, \hat{\theta}_f)), \tag{5}$$

where estimations of parameters $\hat{\theta}_f$ for each $f \in \Im$ are the decision of task

$$\hat{\theta}_f = \arg \min_{f \in R^{s_f}} Q(y, X, \theta_f), \tag{6}$$

where $Q(\cdot) \neq CR(\cdot)$ is a criterion of decision quality in the parametric identification task of private model of complexity s_f generated in the task of structural identification (1).

Most the often applied criterion of the models selection in the indicated algorithms is the criterion of regularity [4]:

$$AR(s) = \left\| y_B - \hat{y}_{Bs} \right\|^2 = \left\| y_B - X_{Bs}\hat{\theta}_{As} \right\|^2 . \qquad (7)$$

This criterion, as well as all criteria in the inductive modeling of the complex systems have properties of external addition [2] which suppose breaking up of set $W = (X \vdots y)$ on two non-overlapping subsets: teaching A (for the evaluation of models parameters) and verification (for the calculation of model errors, $A \cap B = \varnothing$).

Among the often applied criteria of selection $CR(\cdot)$ it is necessary first of all also to name the minimum of deviation criterion and the balance of forecasts criterion. Information about these criteria, conditions and ways of their applying can be found in [2-4].

We will mark in conclusion, that the described approach in one or another way was successfully used both in tasks investigations of complex ecological processes (forecasting of large reservoirs of freezing, forecasting of processes of territories contamination and other ones) and at solving complex technical problems (forecasting the occurrence of anode effect, etc.).

In this article, the original approach for forecasting the so-called rare events that take place in the technical systems is described. Under the term “rare” in the agro-ecological (ecological) systems, it is necessary to understand events that take place during some observed process, and time intervals between them are so great, that it is possible to consider that they practically do not influence each other. Two possible approaches to forming of initial informative base (of data tables) for identification of such processes and phenomena with the possibility of forecasting of rare event beginning moment are described. A multistage procedure of forecasting based on inductive modeling principles as well as the criterion of selection of the best forecasting models is described.

The described approach has a wide field for application in agro-ecological and in technical, medical, biological, and many other applications as well, where it is necessary to have a forecast of not only output value of the process (temperature, for example) but should know the top of the rare event in the investigated process.

Bibliography

1. Ivakhnenko A.G., Osypenko V.V. Prediction of Rare Events on the Basis of the GMDH Algorithm. *Scripta Technica, Inc., U.S.A., SAC*, № 5, vol. 17, 1994. pp. 6-10.

2. Ивахненко А.Г. Индуктивный метод самоорганизации моделей сложных систем. *Киев: «Наук. думка»,* 1982. – 296 с. [In Rus.]

3. Madala, H. R., Ivakhnenko, A.G. Inductive learning algorithms for complex systems modeling. *New York: Boca Raton, CRC Press*, 1994. – 384 с.

4. Степашко В.С. Теоретичні аспекти МГУА як методу індуктивного моделювання. *Управляющие системы и машины,* 2003, № 2. С. 31-44. [In Ukr.]