Methodology for the Construction of Predictive Analysis Systems as Exemplified by the Mining Equipment in the Conditions of Big Data and Simulation Methods

Rodion Ivchenko[1[0000-0003-4252-4825]] and Andrey Kupin[2[0000-0001-7569-1721]]

[1] Department of Automation Computer Science and Technologies
[2] Department of Computer Systems and Networks
Kryvyi Rih National University
11, Vitaliya Matusevycha str., Kryvyi Rih, Ukraine
kupin.andrew@gmail.com, ivchenko.ra@gmail.com

Abstract. Currently, almost all elements of economic activity somehow exist according to the laws of macroeconomics. This is facilitated by the rapid development of international relations, the acceleration of logistics operations, political, religious, cultural integration and unification at the level of interstate relations and interactions (as a result of the evolution of state-formed market systems). At the same time, every year more and more significant influence on macroeconomics is played by minor global factors that were previously practically not taken into account, such as climate change, growth of the world population, etc. Thus, today, every company seeking to be efficient and profitable needs to focus not only on the laws of the domestic market, but also on global trends when building it's strategies and implementing tactical tasks [1].

All the above prerequisites gave rise to a new concept of production development, called Industry 4.0. Previous scientific and technological revolution led to the automation of individual processes and devices, while Industry 4.0 provides for the end-to-end digitization of all physical assets and their integration into the digital ecosystem along with the assets of partners involved in the value chain.

The creation of the concept Industry 4.0 in the framework of solving the problems of managing modern technological processes and production had several basic prerequisites. One of them is related with the fact that the complication of the so-called material part of production, of course, also leads to the complication of the organizational component. It is becoming more difficult for a modern manager to make the right management decisions. In the progression, the variability of the applied goals, conditions, restrictions, and with them the scale of possible consequences, increases. Another important reason is the fact that in modern management conditions it is necessary not only to obtain statistics and analytics of production, but also to be able to predict using the obtained data. High-performance methods should be applied to isolate the most important and relevant information at the time of the decision, with the possibility of a predictive analysis of possible options for events.

Keywords: Industry 4.0, Big Data, Predictive analysis, Simulation.

1 Object Research

In general, the mathematical model of each stage of the hierarchy can be viewed as a complex object of the mineral dressing technology and presented as a function of variables. Here, three types of actions serve as the input of the object (Fig.1) [2]:

1. Uncontrolled (but monitored) input variables Y = {y1,...yr} constitute a disturbance vector and, as a rule, characterize, as far as concentrating production is concerned, quality indicators of the source material to be processed and those of its intermediate products obtained during the concentrating process;
2. Controlled input variables U = {u1,...,un} constitute a control vector and characterize, as a rule, quantitative indicators (expenditure) of material and energy flows;
3. The uncontrolled factors Z = {z1,...,zk} constitute an interference vector. Basically, this is a disturbance vector, about which the developer of the control system knows very little or nothing at all. Most often that vector is not taken into account at all.

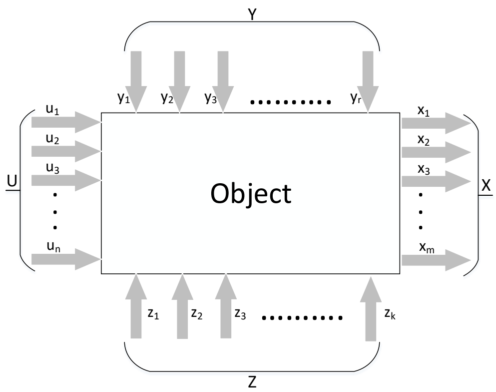

Fig. 1. The structure of a complex control object.

1.1 Monte Carlo Method as a Type of Simulation

The idea of the method is as follows. Instead of describing the process with the help of an analytical apparatus (differential or algebraic equations), a random rally is performed using a specially organized procedure that includes randomness and gives a random result.

The essence of the Monte Carlo method is as follows: it is required to find the value A of a certain quantity under study. To do this, choose a random variable X, the mathematical expectation of which is A, that is, M (X) = A. The probability of events in this case is estimated using the frequency of outcomes in numerous sessions of simulation.

Practically do the following: produce N tests, in which they receive N possible values of X, calculate their arithmetic average and take it as an estimate (approximate value) of the desired number A.

1.2 At the Application Level

In terms of value, this criterion (1) can be expressed in terms of the amount of losses that a combine may incur as a result of a possible failure of an individual equipment segment(s).

$$S_y = (1 - K_1) A \cdot K_{GK} (1 - K_2) Q_K \cdot T \Rightarrow \min \qquad (1)$$

where K_1, K_2 are the coefficients characterizing the share of costs due to a decrease in the quality of products and their quantity, respectively (due to failures or failures) K_{GK} - the coefficient of readiness of operation of the plant; A is the unit cost (UAH / t); Q_K - planned capacity of the plant (tons / hour); T - time of the plant for the year [3].

It should be noted that the exact calculation of the coefficients K_1, K_2 is quite difficult, due to the multi-factorial nature of the latter. Therefore, we investigated the possibilities for the approximate calculation of these quantities, by predicting the probability of equipment failures, as well as their possible consequences using simulation and statistical methods.

For this, as applied to a specific scheme, it is a set of a finite number N of servicing devices. Those any technical device is, in turn, a device for maintenance in the system. The structural model of such a network is presented (see Fig. **1**2). Here, the designations P1, P2, P3, ..., PN correspond to the numbers of instruments for maintenance.

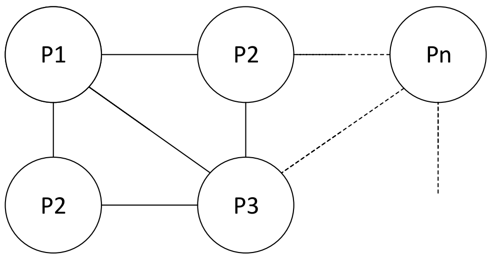

Fig. 2. Structural model example.

It should be noted that absolutely all technical components may not be included in this scheme. It is enough to enter only the most important (key) devices, the failure of which can directly affect the work of the main divisions of the plant.

As the main indicators characterizing the reliability of a device, we will use the time between failures $\bar{t}$ ($\bar{t}$ - expectation of time between adjacent failures of the system being restored) and the average time the device can recover after a failure t_B. These characteristics are usually determined by statistical tests and must be indicated by the manufacturer of the equipment in the passport of the specific equipment.

Thus, for a given structural scheme of the designed IC, we have two vectors in the form:

$$\left.\begin{array}{l} \|\bar{T}\| = [\bar{t}_1, \bar{t}_2, \bar{t}_3, ..., \bar{t}_N] \\ \|T_B\| = [t_{\text{в}1}, t_{\text{в}2}, t_{\text{в}3}, ..., t_{\text{в}N}] \end{array}\right\} \tag{2}$$

where $\bar{t}_1, \bar{t}_2, \bar{t}_3, \ldots, \bar{t}_N$ - the values of MTBF for the corresponding devices in the diagram of Fig.1; $t_{\text{в}1}, t_{\text{в}2}, t_{\text{в}3}, ..., t_{\text{в}N}$ - average recovery time for each of the devices.

Next, a simulation model is built in which the failures of the service devices are reproduced (simulated) and their possible consequences are predicted. The block diagram of the algorithm that implements such a model is shown in Fig. 3.

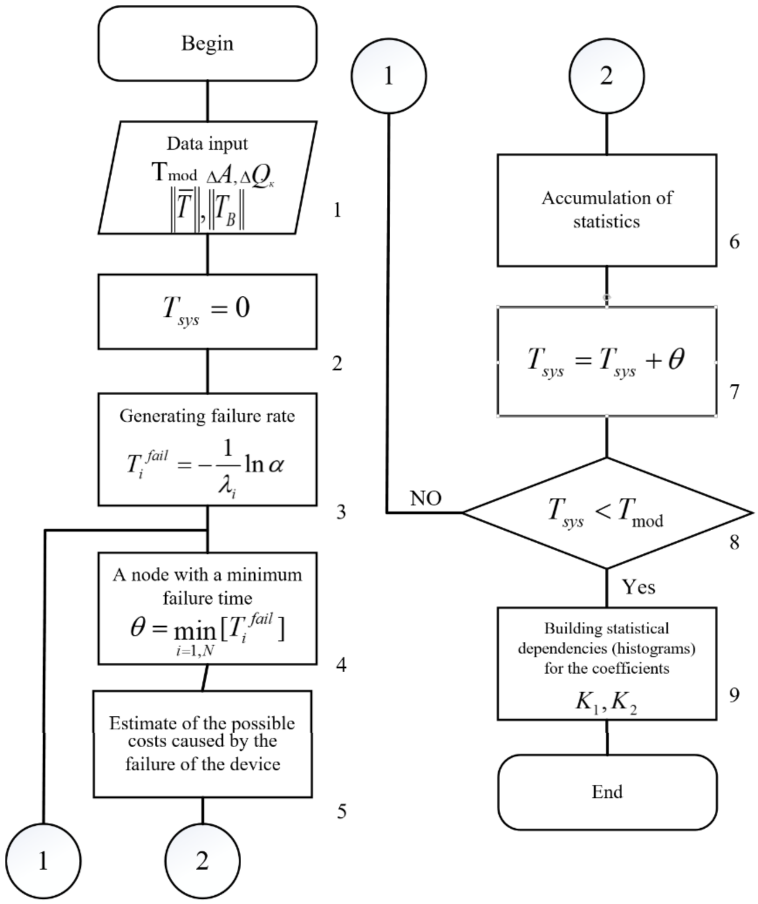

Fig. 3. General algorithm for simulation model of failures.

1.3 The Use of Spreadsheets in Simulation

Imitation using tabular processors is a separate area with its own characteristics. Its proponents argue that using these systems improves the understanding of the processes taking place much better than using specialized software that has a high cost and takes time to study, and also hides the mechanisms used (although such environments are quite widely used because they provide more features and allows simulating complex systems).

One of the advantages of modeling by means of spreadsheets is that the tabular model allows you to visually reflect (see Fig. 4) the behavior of the system under study with the help of charts and graphs.

Using spreadsheets, you can perform simulation modeling of discrete and continuous non-deterministic dynamic systems. The basis for constructing a dynamic model of a system is a description of the system using a system of differential equations or a system of recurrent equations with given initial conditions.

When constructing system-dynamic models using spreadsheets, the time delay is modeled using links to cell ranges corresponding to previous points in model time.

The technology of simulation using spreadsheets includes [4,5]:

- setting discrete points in time;
- description of the state of the system in a certain range of the table;
- setting the rules for the transition of the system to the state corresponding to the next point in model time.

To build non-deterministic models using MS Excel, the functions RANDOMBETWEEN, RANDOMBETWEEN are used.

When creating a dynamic model using a spreadsheet, the modeling horizon is determined by building the required number of ranges (replicating ranges) corresponding to discrete points in time.

If the capabilities of the tabular processor do not allow to cover the entire modeling horizon, then the behavior of the system on a shorter modeling horizon should be investigated to put forward a hypothesis about its behavior throughout the modeling horizon.

A multiple start of a simulation session in a tabular processor is performed by recalculating formulas in all cells of the table, which is started by pressing the F9 key.

	FROM	TO		Device 1	Device 2	Device 3
Months of operation (warranty)	0	120	Detail 1	72	97	67
			Detail 2	112	21	46
			Detail 3	64	16	106
Simulated time (months)	50		Detail 4	45	42	101
Price	150		Detail 5	92	26	44
			Detail 6	79	53	6
			Detail 7	118	104	10
			Detail 8	104	51	64
			Detail 9	23	2	111
			Detail 10	6	51	78
			Detail 11	48	78	21
			Detail 12	3	43	77
		Number of breakdowns		5	6	5
		Finance losses		750	900	750

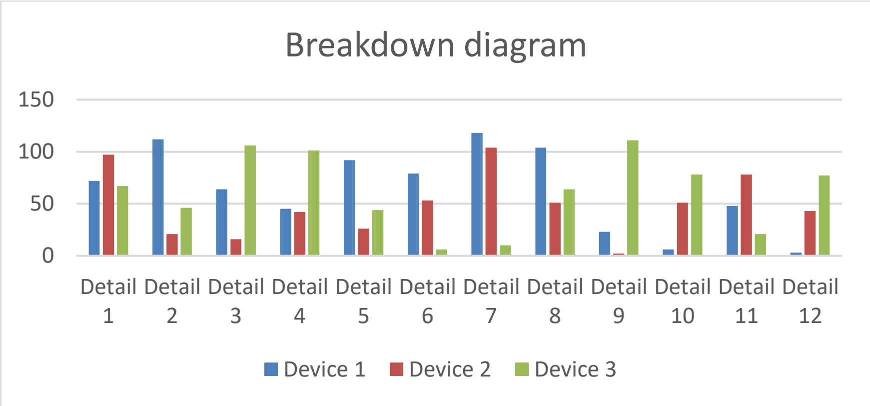

Fig. 4 Simulation in spreadsheet with breakdown diagram.

References

1. Magazine "neftegaz". “Industry 4.0: Key Tasks of Modern Enterprises”. [Electronic resource] – Available: https://neftegaz.ru/science/view/1414-Industriya-4.0-klyuchevye-zadachi-sovremennyh-predpriyatiy-i-resheniya-kompanii-Yokogawa
2. A.M. Mariuta, Yu. G. Kachan, V. A. Bunko: “Automatic Control of Technological Processes of Concentrating Plants”. Moskow «Nedra». p 1983.
3. Molchanov A.A. Modeling and design of complex systems .- K .: Vyshcha school, 1988.- 359p. 5. Denisov A.A., Kolesnikov D.N. Theory of large control systems. - L .: Energoizdat. Leningr. separation, 1982.- 287p
4. “Simulation of dynamic systems”. [Electronic resource] – Available: edu.tltsu.ru/sites/sites_content/site216/html/media67140/lec7_ts.docx
5. A. Kupin, I. Muzyka, “Information Technologies of Processing Big Industrial Data and Decision-Making Methods”, *IEEE “Problems of infocommunications science and technology 2018”*, p. 4, 2018.