Method of attractor reconstruction in modeling the integral risk indicator

L M Bogdanova¹, S Ya Nagibin² and O A Rabinovich³

¹ Plekhanov Russian University of Economics, Stremyanny lane, 36, Moscow, 117997, Russian Federation
² Moscow Aviation Institute (National Research University), Volokolamskoe highway, 4, Moscow, 125993, Russian Federation
³ Russian Satellite Communications Company, Baumanskaya Street, 16, Moscow, 105005, Russian Federation

⁴ E-mail: nagininmai@mail.ru

Abstract. This study examines a mathematical model for forecasting the dynamics of changes in the industrial safety risk indicator of an enterprise. The results of the development of a method for reconstructing the attractor of the integral risk indicator based on the Grassberger-Prokacchia algorithm, which is currently the most popular for analyzing time series and allowing automating the process of calculating a parameter that determines the number of points, for making a forecast. There are results of estimating the forecast of the dynamics of the behavior for a strange attractor, obtained on the basis of real data.

1. Introduction

As the economy develops, objects and management tasks become more complex, and this leads to a change in management technologies. Modern control systems must ensure the joint mutually consistent operation of many related heterogeneous technical objects. Industrial Internet of Things (IIoT) [1,2], which is used in industrial enterprises, allows collecting large volumes of structured and weakly structured information from sensors, which allows obtaining, after processing and analysis, an objective assessment of the state of production, industrial safety and business processes.

The classical methods of control theory were created on the assumption that the models of an object describe its state and behavior with the required accuracy. However, in the increase in the amount of collected data, increase in unstructured information and impact of the external environment and deviations from this situation are inevitable. This creates uncertainties in the models for describing the object, complicates its management and leads to the need for a transition to intelligent control, since traditional technologies are not able to provide the required improvement in the quality of control, since they do not take into account all the uncertainties affecting the system.

In addition, modern industrial systems place increased demands on control systems. These requirements include, for example, ensuring automatic tuning to change the parameters of the control object and external disturbances, maintaining operability in conditions of uncertainty, predicting the possible development of events, as well as other requirements, depending on the characteristics of the subject area of production and the wishes of the customer.
In this setting, intelligent control systems using enterprise monitoring, collecting and analyzing data received from IIoT, must monitor, register and compare the obtained parameters of the object with the specified criteria to determine the state of the elements that make up the landscape of the object and about the object as a whole. Thus, intelligent control systems using descriptive analytics should answer the question "What happened" and, with the help of diagnostic analytics, determine the cause of the violation. For especially hazardous production facilities, the disruption of the functioning of which can lead to serious consequences in the form of explosions, death of people and damage to the environment, intelligent control systems must predict the risk of a change in the state of the facility to determine and predict the moment of failure. In this case, the goal of predictive analytics of the control system is to answer the question "What can happen" to prevent the onset of a negative event. In other words, intelligent control systems must predict and assess the dynamics of changes in the metric, which consists of a set of key parameters that determine industrial safety, which are an integral indicator of risk. By risk we mean the probability of an event occurring, which can have a negative impact on the achievement of the set goals and entail the occurrence of damage.

2. Theoretical basis
The main idea of the methods of chaotic dynamics for the analysis of the risk of industrial safety of an industrial enterprise, as a complex dynamic system, is that the basic structure of the behavior of such a system, which contains all information about the system, namely the attractor of a dynamic system, can be restored through measurement only one observable characteristic of this dynamical system [3]. In our case, such a characteristic is the metric of the integral risk indicator of a dynamic system, presented as a time series.

The time series prediction is reduced to the approximation of a function of many variables for a given set of examples using the procedure of immersing the series in the m-dimensional lagged space M[4,5].

For dynamical systems, the following Takens theorem is known [6]: “if a time series is generated by a dynamical system, that is, the value of M is an arbitrary function of the state of such a system, then there is such a depth of immersion m (approximately equal to the effective number of degrees of freedom of the given dynamical system), which provides an unambiguous prediction of the next value of the time series”. Thus, having determined the value of m, it is possible to guarantee an unambiguous dependence of the future value of the series on its m previous values. To restore this unknown function from a set of examples given by the history of a given time series, an artificial neural network is used in this work.

Since it is not known a priori which dimension of the model m should be chosen, special methods are used to estimate its value, for example, method of false nearest neighbors, correlation dimension [7,8], which turn out to be ineffective in the case of the problem of predicting the integral risk indicator due to the specifics of the data. Moreover, these methods are based on exhaustive enumeration: we have to go through different dimension values and build correlation matrices for each in order to determine at which the value of the correlation value will be maximum [9,10]. This approach can be applied in research problems, but is not suitable for industrial purposes that are demanding not only on the accuracy of the forecast, but also on the performance of the software, since the selection of the dimension of the reconstructed attractor becomes part of the whole process of modeling the time series of the integral risk indicator [11-13].

The use of nonlinear dynamics methods reduces the required calculations:

1. full search is replaced by "smart search": calculations are not carried out for all possible cases, but until the condition for stopping calculations is reached;
2. for each iteration of the enumeration, instead of calculating cumbersome correlation matrices, it is required to calculate the correlation integral.

Thus, the task will be reduced to enumerating different values of the dimension, starting with small ones and gradually increasing them until the results are saturated.
3. Methodology

The most acute problem in the management of complex systems is the problem of "big data". The main for such an assessment is a large amount of structured and unstructured information which is collected using the industrial IoT and needs to be processed to determine the industrial safety risk of the enterprise.

From the point of view of mathematics, any dynamical system, whatever it models, describes the motion of a point in phase space.

The most important characteristic of this space is its dimension, or the number of quantities that must be specified to determine the state of the system. Moreover, for complex systems (large enterprises), the dimension can be as many assessed elements (10^7-10^8, for example), which will make the task of analyzing the state of the system in real time unrealizable because the time for its solution may exceed any specified deadlines.

If we assume that a point, moving in phase space, leaves a trail, then a tangle of trajectories in which channels can be distinguished will correspond to dynamic chaos. The channels are the behavior of the system with accuracy suitable for practical use, can be determined by only a few significant variables, while, in the first approximation, all other variables can be neglected. Thus, to determine the channel, it is necessary to know not all the set of parameters of the system, but only a few (several dozen) significant ones. In this regard, an important practical task is to reduce the dimension of the phase space by determining these significant parameters to reduce the requirements for processing computing power.

In this article, for a generalized assessment, it is proposed to use the concept of integral risk of industrial safety, by which we mean a numerical indicator of the quality and efficiency of a process, which determines the state of a metric composed of key indicators of technological production, which to a determining extent affects the risk of industrial safety. For this metric, a confidence interval is calculated, which determines the size of the channel. Thus, to determine the channel, it is necessary to know not all the set of parameters of the system, but only a few (several dozen) significant ones.

Assessment and forecast of the integral risk indicator will allow quantifying the level of industrial safety in real time.

The purpose of this article is to develop a mathematical model of the dynamics of changes in the integral risk indicator to increase the level of industrial safety of especially hazardous industrial facilities, by developing a method for automating the determination of the parameter m for building predictive models in industrial safety monitoring systems.

The integral risk indicator is a dynamically changing value, the mathematical model of which is described by a time series {ui}^N_{i=1} of the variable u(t) [7]:

\[ u_i = u(t_i), i = \overline{1,N}, \]

with regression dependence of current values from retrospective ones. In order to successfully predict the value of the integral risk indicator in umbrella industrial safety monitoring systems, mathematical algorithms are used, for the construction of which it is necessary to determine the minimum number of points m, variables required to simulate the behavior of a dynamic system [14].

3.1. Grassberger-Procaccia algorithm

Proceeding from the fact that the integral risk indicator can be described within the framework of dissipative systems, it is possible to reconstruct the attractor of a given dynamic system, relying on the Takens theorem [6], that is, according to the series {ui}^N_{i=1}, obtained as a result of measurements, construct a new attractor with the same parameters as the original one.

To determine the dimension m of the reconstruction of the attractor, we use an important quantitative characteristic of the attractor, which describes the complexity of the dynamical system under consideration, the correlation dimension \( D_c \). This characteristic determines the dynamic inhomogeneity of the attractor and is calculated based on the calculation of the correlation integral \( C(r,N,m) \):

\[ C(r,N,m) = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \theta(r - \frac{1}{\sqrt{m}} \sum_{k=1}^{m} (u_i - u_j)^2), \]
\[ D_c = \lim_{r \to \infty} \lim_{N \to \infty} \frac{\log C(r, N, m)}{\log r}, \tag{3} \]

where \( r \) is the correlation radius, \( N \) is the number of points, used to determine the correlation dimension, \( \theta(x) \) is the Heaviside function, \( \frac{1}{m} \sum_{k=1}^{m} (u_i - u_j)^2 \) is the distance between images of \( i \)-th and \( j \)-th time series points in the reconstructed space.

Grassberger-Proccacia algorithm [14] consists in recovering the attractor by a sequential shift. For sufficiently large \( N \), the value of the correlation integral \( C(r, N, m) \) is a statistical estimate of the amount \( C(r) \) and shows the probability that the original time series contains a pair of points, the distance between which does not exceed \( r \). As follows from Grassberger-Proccacia algorithm [7,14], the dimension of the reconstructed attractor is determined by the slope of the linear section of the dependence graph \( \log C(r) \) from \( \log r \), finding which presents a certain difficulty. The linear section on the logarithmic dependence of the correlation integral on the correlation radius is determined by the condition of the smallness of the second derivative, which is estimated using the least squares method (LSM). Primarily, the first derivative is estimated as a proportional coefficient of LSM of a straight line passing through a given number of points [15]. To estimate the second derivative, we will use the same procedure, only using the estimate of the first derivative as input parameters. For the longest segment that satisfies the smallness of the second derivative, the coefficient of proportionality of LSM of the straight line is found, which determines the correlation dimension \( D_c \) of attractor.

The procedure for finding the dimension of an embedding \( m \) is reduced to the following successive steps:

1. we gradually increase the dimension of space \( m \);
2. for every \( m = 2, 3, \ldots \) we calculate the correlation dimension \( D_c \);
3. we consider the dependence \( D_c(m) \). With an insufficient number of measurements (for small \( m \)), the correlation dimension grows. Starting at some dimension \( \bar{m} \) of attractor, \( D_c \) reaches saturation and stops changing. It is this meaning \( \bar{m} \) is an estimate for the minimum dimension of the attractor and \( D_c(\bar{m}) \) is the estimate of the correlation dimension.

4. Results

Figure 1. Index of integral risk for a weekly period.
Figure 1 shows the integral risk indicator for a weekly period with a frequency of readings equal to 2 minutes from a real industrial facility. In the process of collecting data, due to various kinds of interference, noise, the data is distorted, including the homogeneity of the received data. As we can see from the graph, for specific points in time $t_i$ time series values $u(t_i)$ are atypical, abnormal. In mathematical statistics, such values are called outliers and must be removed from the series with subsequent replacement by an acceptable value by interpolation, since such values do not allow the model to concentrate on the key indicators of the series, distracting to particular atypical cases.

To filter outliers of the integral risk indicator, it is proposed to apply an analytical method based on the interquartile range, which has shown to be effective when considering stochastic processes at an industrial enterprise. To determine the extreme values, it is assumed that the main signal scattering is located between the first and third quartiles, i.e. between the 25th and 75th percentiles (figure 2). It includes the center 50% of the observations in an ordered set, with 25% of the observations below the center point and 25% above:

$$IQR = Q_3 - Q_1,$$  
where $Q_1$ is the lower (first) quartile, $Q_3$ is the upper (third) quartile. We believe that those points with index $i$ are anomalous if they do not meet the condition:

$$(Q_1 - 1.5 \times IQR) < u(t_i) < (Q_3 + 1.5 \times IQR).$$

![Figure 2. Frequency distribution histogram.](image)

For points outside the specified range, the new value was calculated using local cubic spline interpolation, which for each interval $[t_i; t_{i+1}]$ builds a third degree interpolation polynomial:

$$S_3(t) = \frac{(t_{i+1} - t)^2 (t - t_i)}{h^3} y_i + \frac{(t_{i+1} - t)^2 (t - t_{i+1})}{h^3} y_{i+1} + \frac{(t_{i+1} - t)^2 (t - t_i) - (t - t_{i+1})^2 (t - t_i)}{h^3} m_i + \frac{(t - t_i)^2 (t - t_{i+1})}{h^3} m_{i+1},$$

where $t_i \leq t \leq t_{i+1}, i = 1, n - 1, h = \frac{t_{n-1} - t_1}{n}$.  

As it is shown in the figure 3, the new series has no bright outliers.
We calculate the correlation dimension $D_c$ for each $m = 2, 3, \ldots$. The correlation integral for each $m$ is shown in the figure 4.

To determine the number of flows required to build a forecast, we construct the dependence $D_c(m)$ and determine the moment of saturation (transition of the graph to a plateau) and the corresponding $m$ using LSM.

Figure 3. Input signal after the outlier filtering process.

Figure 4. Correlation integral obtained on the basis of the integral risk indicator.
Figure 5. Dependence of the correlation dimension $D_c$ on the dimension of space.

As we can see from the graph in the figure 5, the transition to saturation begins from the points $m$ equal to 27-29. Calculations showed that for the presented data $D_c \approx 1.48$, for successful prediction of the behavior of the attractor of the integral risk of a production enterprise $m$ is taken to be approximately 31.

5. Conclusions
In the presented work, there is the application of the methods of stochastic dynamics to the analysis of the integral risk of industrial safety for an enterprise.

A method is developed for determining the minimum number of points of $m$ variables required to simulate the behavior of a dynamic system using Grassberger-Prokacchia algorithm with an accuracy satisfying the consumer. The method proposed in the article is implemented in the algorithm and the software module that implements it for predicting the dynamics of changes in the integral indicator of industrial safety risk of an enterprise. The introduction of the described method made it possible to automate the process of analyzing the initial time series and choosing the architecture of a neural network for modeling the dynamics of changes in the integral risk indicator.

To assess the effectiveness of forecasting with the selected indicator of points $m$, the values of the following parameters were analyzed:

- Mean Relative Error (MRE): relative error that indicates how large the absolute error is compared to the total size of the tested data. MRE result = 2.791%;
- Kullback-Leibler distance (KL), which shows the amount of information loss when replacing the true distribution with the calculated one. The smaller KL, the better coincidence of the distributions of the original and design vectors. KL result = 0.1658;
- $R^2$ is the coefficient of determination for the model takes values from 0 to 1. The closer the value of the coefficient to 1, the stronger the dependence. For acceptable models, it is assumed that the coefficient of determination should be at least 0.5. $R^2 = 0.6441$.

The metrics obtained reflect the good efficiency of using this solution in intelligent control systems to predict the dynamics of changes in the attractor of the integral risk of an industrial enterprise.

The research and proposals carried out in the article were implemented in the company "Russian Satellite Communications Company" in the innovative software complex for intelligent monitoring and control "Zodiac" for predicting industrial safety risks in a number of highly hazardous industrial facilities of the Fuel and Energy Complex of Russia [16].
References

[1] Boyes H, Hallaq B, Cunningham J and Watson T 2018 The industrial internet of things (IIoT): An analysis framework Comp. in Ind. 101 1-12

[2] Schwab K and Davis N 2018 Shaping the future of the fourth industrial revolution (Sydney: Currency)

[3] Takens F 1985 Dynamical systems and bifurcations (Berlin: Springer)

[4] Malinetskiy G G and Potapov A B 2002 Modern problems of nonlinear dynamics (Moscow: URSS)

[5] Golovko V A 2005 Neural network methods for processing chaotic processes VII All-Russ. Scient. and Techn. Conf. “Neur.” 43-91

[6] Takens F 1981 Detecting strange attractors in turbulence (Berlin: Springer)

[7] Loskutov A Yu 2002 Time series analysis (Moscow: Moscow State University)

[8] Piryonesi S M and El-Diraby T E 2020 Role of data analytics in infrastructure asset management: Overcoming data size and quality problems J. of Transp. Eng., Part B: Pav. 146(2) 04020022

[9] Lvovsky E N 1982 Statistical methods for constructing empirical formulas: Textbook (Moscow: High School)

[10] Levin S E, Nagibin S Ya and Shilov V V 2018 Monitoring of industrial safety of the fuel and energy complex National Security of Russia: Actual Aspects 30-6

[11] Moon F 1987 Chaotic Variations (New York: John Wiley &Sons)

[12] Franses P H 1998 Time series models for business and economic forecasting (Cambridge: Cambridge university press)

[13] Bezruchko B P and Smirnov D A 2006 Mathematical modeling and chaotic time series. Proceedings of higher educational institutions Appl. Nonl. Dyn. 14(1) 153-7

[14] Grassberger P and Procaccia I 1983 Characterization of strange attractors Phys. rev. lett. 50(5) 346-9

[15] Nicolis G 1990 Cognition of the complex (Moscow: Ripol Classic)

[16] Nagibin S Ya, Senatorov M Yu and Levin S E 2021 Intelligent monitoring, generalized control and security of critical infrastructure systems (Moscow: Ajax-Press)