IOP Conference Series: Materials Science and Engineering

PAPER • OPEN ACCESS

Improving the efficiency of information processing based on the entropy-parametric approach

To cite this article: V G Polosin *et al* 2019 *IOP Conf. Ser.: Mater. Sci. Eng.* **537** 022015

IOP Conf. Series: Materials Science and Engineering **537** (2019) 022015 doi:10.1088/1757-899X/537/2/022015

Improving the efficiency of information processing based on the entropy-parametric approach

V G Polosin[1], O E Bezborodova[2,5], O N Bodin[3], A A Trofimov[3] and A G Ubiennykh[4]

[1] Medical Cybernetics and Computer Science Department, Penza State University, 40, Krasnay Str., Penza 440026, Russia
[2] Technospere Safety Department, Penza State University, 40, Krasnay Str., Penza 440026, Russia
[3] Information and Measuring Equipment and Metrology Department, Penza State University, 40, Krasnay Str., Penza 440026, Russia
[4] Information Computing Systems Department, Penza State University, 40, Krasnay Str., Penza 440026, Russia

E-mail: oxana243@yandex.ru

Abstract. The paper contains an analysis of the task of establishing quantitative information reflecting the properties of the observed object. The authors of the article showed that in order to operate, both controls instruments and measuring instruments, it is necessary to form a "Euclidean measure" for the property of the object at the current time. The article describes an independent method for estimating the uncertainty of the state of an object by calculating the entropy of its observed properties. In particular, it is shown that the entropy potential is proportional to the measure of Euclidean space. A description is given of the method for determining the shape of an approximating function for sampling results based on a comparison of the difference between Euclidean measures and measures of the information space formed for entropy potential.

1. Introduction

The practical purpose of the functioning of any engineering device is to convert energy or information. The task of any control system in the most general sense is the processing of information about the current operating mode of a controlled object and the development on the basis of this control action, that ensuring the current operating mode of the object approaches the specified mode [1, 2]. Information processing is necessary for the formation of the control action aimed at changing the state of the system. Information processing means are all kinds of computing devices for converting signals in accordance with a certain algorithm. Biological systems are of particular importance that is a chaotic system in which the probabilistic properties are due to its internal structure. In a biological system, the object activity is interrelated with the production of entropy, which characterizes the disorder of a biological object [3]. Probability distributions of time intervals and data disorder contain information about the chaotic properties of an object. In this regard, the processing of probabilistic and informational characteristics of a biological object, reflecting its deterministic and chaotic properties, is an actual problem. [4].

IOP Conf. Series: Materials Science and Engineering **537** (2019) 022015 doi:10.1088/1757-899X/537/2/022015

2. Organization of information processing in measurement and control systems

The organization of measurement processes consists in comparing the properties of an object of research with a measure of the same property using technical means that are included in the structure of control systems [5]. Structures of technical means of measurement and control contain specific devices for comparing the parameters received from the object with the specified value of the measure of the same value. In both cases, *the Euclidean measure* of the difference between the value of a quantity coming from the object of observation or control and the measure value of the same homogeneous quantity is formed at the output of the comparison devices. Depending on the purpose of the device, the measure of difference Δx is used to control the object when organizing the process of control or to form measure of measurement device. Thus, the scheme of organization of measurement and control processes contains an object of observation and a device for forming a measure based on an estimate of the difference Δx between similar properties of the object and the measure.

3. Metric space used for object control

In metrology, the concept of measure is defined as the measuring instrument used to reproduce and store the value of physical quantity [6]. A more general representation contains the mathematical definition of the measure as numeric function that associates with each set, from a certain family of sets, a non-negative number [7]. In mathematical analysis, the measure is the way to assign the number to each suitable subset of that set, intuitively interpreted as its size [8]. Moreover, the function ρ is called the measure if it satisfies the property of non-negativity and additivity. An example of the mathematical measure is the Lebesgue measure on a Euclidean space, which assigns the many dimensional volume of Euclidean geometry to suitable subsets of the n-dimensional Euclidean space $\mathbf{R}^n$. Technically, the measure is the function that assigns the non-negative real number to certain of a set X.

In the development of theoretical and practical models, such philosophical categories as space, expressing the order of coexistence of individual objects, and the time that determines the order of change of events are used as design tools. During the measurements, a metric space is used, in which the distance between any pair of elements of this space is determined. Among the known metric spaces, one should distinguish the Euclidean space. There, the operation of scalar multiplication of vectors is given. The “Euclidean meter” is defined as the space metric, equal to the distance between its two points. The formula for calculating the n-dimensional Euclidean measure is

$$\rho = \left(\sum_{i=1}^{n} \left(x_i - y_i \right)^2 \right)^{1/2} \tag{1}$$

where x_i, y_i are the coordinates of points in space.

In developing the means of measurement and control, the “Euclidean measure” is used to form a control action. As disadvantages of such a system, it should be noted that it is not possible to control and correct random effects on the measured values. To eliminate these shortcomings, observation of the magnitude for a long time is usually used, followed by statistical processing of the observation results.

Effective monitoring and control of objects is possible only if random disturbances are taken into account. Classical methods of research and analysis of the state of objects are based on the description of random effects using correlation functions and spectral density functions. These methods require considerable time, the use of complex equipment, the use of additional material means (the use of reagents), which makes it difficult to apply classical methods of analysis in the construction of control and measurement systems. Under such conditions, observation and control of object parameters has to be carried out in conditions of uncertainty about the acting perturbations that is similar to the “information vacuum” [9].

IOP Conf. Series: Materials Science and Engineering **537** (2019) 022015 doi:10.1088/1757-899X/537/2/022015

4. Measure of the space of elementary events

4.1. Euclidean measure of the events space

One of the typical tasks of data processing, received from a controlled object, is to build a functional change in the signal over time, subject to random disturbances. One of the typical tasks of data processing, received from a controlled object, consist to build a functional change in the signal over time, if damage occurs from random perturbations. For these purposes, a sample of the measurement results of the signal during a specified time interval is formed. Then the function of the approximation is determined on the basis of the data obtained. The parameters of this function contain information about the state of the object. The conformity assessment of the approximation function to a sample of measurement results is carried out by comparing measures that formed for sampling measurement results and for discrete sampling of the approximating function that was be found at the same points in time.

To analyze data that was received from the control object during a period of time t signal value array is formed. A sample of the volume N of values of a random variable ξ is a set of values (x_1, x_2, …x_N) of a random characteristic obtained as a result of independent measurements at different points in time. After the measurement has taken place, the Sampling of random values is a vector of dimension N with independent components. When analyzing random variables, the scatter of a sample of values is estimated using sample variance. A non-negative number equal to the square root of the sample variance multiplied by (N – 1) corresponds to the "Euclidean measure" for the sampling space of current control results centered relative to the mathematical expectation.

$$\rho_c(x,\ m_x) = \left(\sum_{i=1}^{N}\left(x_i - m_x\right)^2\right)^{1/2} \tag{2}$$

where m_x is the expected value of the controlled parameter.

Thus, the product of the variance by the number N of the experimental results is a Euclidean measure between the sample vector of a random variable and the expectation vector, which is used in circuits to control an object or to form a measure when solving a measurement problem.

Expression (2) defines the Euclidean measure of the sample for the current control results. By analogy, a measure of comparison is determined for apriority known control function.

$$\rho_a(y,\ m_y) = \left(\sum_{i=1}^{N}\left(y_i - m_y\right)^2\right)^{1/2} \tag{3}$$

where y_i is the sample of values obtained by substituting time series counts in the control function.

The difference of measures formed in the probabilistic space of elementary events allows us to compare the sample of measurement results and the sample of the approximating function. The formula for the difference of measures (2) and (3) is:

$$\Delta_{\rho\ p} = \rho_a(y,\ m_y) - \rho_c(x,\ m_x) \tag{4}$$

When processing the values of a controlled parameter of an object, this uncertainty is formed as the difference of Euclidean measures calculated in the probabilistic space of elementary events for sampling a random variable of the object and for sampling an approximating function with a priori known form. The difference (4) reflects the uncertainty of the state of the object, the minimization of which is achieved by organizing negative feedback in the control systems of the object or the means of measurement for the formation of a control action. In continuous processes, this uncertainty is formed on the basis of the difference in the Euclidean measures specified by the device for forming the measure and the actual values of the properties of the physical object.

As the main drawback of the method of organizing the difference between the Euclidean measures of an object or a device for forming a measure, one should note the invariance of the a priori given form of the control function. An error in the choice of the form of the function leads to large values of the uncertainty of the “Euclidean measure” in the probabilistic space of elementary events, comparable to the size of the measure.

A significant reduction in the uncertainty of the state of an object is achieved by choosing the form of the control function based on an additional assessment of its properties. To determine the form of functions, the methods of statistical evaluation of initial and central moments of high order have been used, that are used to select the forms of statistical distributions [10, 11, 12, 13]. Another approach to analyzing the form of functions is to compare the informational entropy of the control function and the informational entropy of the distribution result. This approach also found application in modern literature [14, 15, 16, 17].

4.2. Entropic measure of the events space

The entropy of the probability $H(y)$ of the distribution of the monitored parameter y is an independent way to assess the level of uncertainty of the state of an object [18, 19, 20].

$$H(y) = -\int_{-\infty}^{\infty} f(y) \ln f(y) dy \tag{5}$$

where $f(y)$ is the distribution density of the values of the observed property.

Currently, the Rényi and Shannon entropy finds a lot of different applications in many fields modern science such as biology, medicine, geophysics, economics, computer science, electrical engineering, chemistry and physics. In wave mechanics, the Hirschman uncertainty, defined through the Shannon informational entropy, is often used as the uncertainty relation [21,22]. Heisenberg's uncertainty principle can be expressed as a lower bound on the sum of these entropies.

Since it is often impossible to directly estimate the entropy of a signal from experimental data, a priori known approximation is used for data processing by some distribution with specified characteristics, that limits the choice of its shape. Therefore, to assess the state of an object, a limited set of information characteristics is used that uniquely determine its entropy. The use of such characteristics as coordinates of entropy models of objects creates prerequisites for effective monitoring, forecasting and control [9, 23]. The problem of evaluation with a limited set of informational characteristics also discover in predicting the behavior of objects based on the use of fuzzy logic, the algebra of which is built on the elements of functions [14, 24].

In control and measurement systems, two similar characteristics of the uncertainty interval are used as a measure of the uncertainty of the state of an object: the entropy deviation [14, 25] and the entropy potential [9]. The Measure of the object uncertainty is set as half the range of a uniform distribution, having the same entropy as the distribution law of the observed parameter. The formula for calculating the object uncertainty interval for an arbitrary distribution law of the parameter y is

$$\Delta_e = \frac{1}{2} \cdot \exp(H(y)) \tag{6}$$

In instrument making, the uncertainty interval of an object is usually called the entropy error of measurement tools. In modern technology, the entropy error is used to evaluate the misinforming hindrance. Recently, another term for the interval of uncertainty of the state of the object is used in the development of control systems. The entropy potential is the kind of generalized parameter same as the characteristic of the level of instability of an object. The more the level of instability of an object increases, the greater will be the uncertainty of its state [9]. The expression for calculating the entropy potential for a sample of monitoring and control results at equal data grouping intervals Δx is [26].

$$\Delta_e = \frac{1}{2}\Delta x \cdot N \cdot \exp\left(-\frac{1}{N}\sum_{j=1}^{m} n_j \cdot \ln(n_j)\right) \tag{7}$$

The entropy potential similarly the standard deviation characterizes the uncertainty of the observed value. The entropy coefficient K_e is used as a characteristic of the instability of the observed object. It is equal to the ratio of the entropic potential to the standard deviation. The formula for calculating the entropic coefficient is

$$K_e = \frac{\Delta_e}{\sigma(x)} \tag{8}$$

The entropy coefficient K_e is in the range from 0 to 2,066. Due to the fact that the entropy potential is a scalar function defined in the whole set of values of the parameter y and is associated with the distribution spread parameters (RMS and variance).

$$\rho_a(y, m_y) = \frac{\Delta_e}{K_e} \cdot \left(N-1\right)^{1/2} \tag{9}$$

Its estimate can be postponed in the space of the Euclidean measure.

Information processing is the computational operation of systematically changing the content or presentation of information. For these purposes, it is convenient to use the differ*ence of measures formed in the space of elementary events.*

$$\Delta_{\rho\, e} = K_e \cdot \left(\rho_a(y, m_y) - \rho_c(x, m_x)\right) \tag{10}$$

Therefore, a distribution function with a known entropy coefficient is chosen a priori it is difference of measures (10) makes it possible to estimate the distribution parameters. If the form of the function matches the sample of monitoring results, the distribution coefficients are equal to each other. Then from expression (10) follows the proportionality between the entropy potential and "Euclidean measure". Obviously, the difference in the entropy potentials of the theoretical distribution function and the sampling of monitoring results reflects the difference in “information measures” (or “entropy measures”) in some other information space of statistical distributions associated with the Euclidean space of sampling results using the entropy coefficient.

4.3. Entropy–parametric measure of the events space

In the space of the entropy potential and the mean square deviation, the position of the object is conveniently characterized by the entropy–parametric potential, that is as a measure of uncertainty of the object, equal to the distance from the center of the coordinates of the space to the point of position of the object. The expression for the entropy – parametric potential is

$$\Delta_{\rho\, ep} = \left(\Delta_e^2 + \sigma^2\right)^{1/2} \tag{11}$$

Analysis of the results of information processing in the construction of measurement systems, management and control depends on the definition of the parameters of the approximating functions. In this case, the choice of the parameters of the approximating function is achieved by minimizing the difference between the entropy – parametric potentials calculated by formula (11) for a sample of measurement results and a sample of the approximating function.

A qualitatively new result is obtained by applying the entropy – parametric potential to a random value of the difference of the vectors of a sample of measurement results and a sample of an approximating function. The uncertainty interval is a scalar function defined in the entire set of values of the monitored parameter and is associated with the parameters of the distribution scatter: the standard deviation and variance, entropy, etc. Mapping the measure of entropy uncertainty in the space of the Euclidean measure for experimental data and processing results is an effective tool for comparing them in the space of elementary events during processing based on the entropy-parametric approach. The use of informational and Euclidean measures in relation to the probabilistic space of elementary events allows us to construct an algorithm for finding the form of a function based on the intersection of the trajectory of the difference of informational and Euclidean measures.

5. References

[1] Gitis E I, Danilovich G A and Samoilenko V I 1968 *Technical Cybernetics* (Moscow: Soviet radio) p 488
[2] Lucas V A 2005 *Theory of Technical Systems Management* (Yekaterinburg: Publishing house of the Ural State Mining University) p 677
[3] Prigogine I R 2001 *Introduction to the thermodynamics of irreversible processes* (Izhevsk: RHD) p 160
[4] Glass L and Mackey M 1991 *From the clock to chaos. The rhythms of life* (Moscow: Mir) p 248
[5] Arzhayev D A, Bodin O N, Polosin V G, Nefediev D I and Ubiennykh A G 2016 Structural diagrams of the measuring devices of monitoring and control systems. *Measuring. Monitoring. Management. Control* **1(15)** 24-30
[6] Tsvetkov E I 2005 *The basis of mathematical metrology* (St. Petersburg: Polytechnic) p 510
[7] Vulikh B Z 1973 *Short course of the theory of functions of a real variable*. *Introduction to the theory of the integral* (Moscow: Nauka) p 352
[8] Stetz W Albert 2016 *Lectures on nonlinear mechanics and chaos theory* (World Scientific Publishing Co) p 142
[9] Lazarev V L 2005 Entropy approach to the organization of monitoring and management *Journal of Computer and Systems Sciences International* (Pleiades Publ., Ltd.) **6**(**44**) 893-9
[10] Fedorov M V 2002 The method of identification of the distribution of small samples *J. Ros. Chemical. Society of them. D I Mendeleev* **3(XIVI)** 9-11
[11] Prokhorov S A 2001 *Approximate analysis of random processes* (Samara: Samar. State Aerospace University) p 329
[12] Polosin V G 2014 Method of statistical analysis of electrocardiographic information Izvestiya *SFedU. Engineering sciences* (Taganrog: Publ. House of the SFedU) **10(159)** 120-6
[13] Hwang T Y and Huang P H 2002 On new moment estimation of parameters of the gamma distribution using it's characterization *Ann. Inst. Statis Math.* **54** 840-7
[14] Novitsky P V 1968 *Basics of the information theory of measuring devices* (Leningrad: Energy) p 248
[15] Polosin V G and Pershenkov P P 2014 Information-theoretic method for hypothesis testing with nonsymmetric distributions *Measurement Techniques* **12(56)** 1318-21
[16] Vöger V 2003 Information on the measured value as the basis for the formation of the probability density function *Measurement Techniques* **9** 3-9
[17] Tertychnaya S V and Polosin V G 2008 The study of the components of the source of radon on the basis of an analysis of the statistical results of measuring its volumetric activity *Proc. university. Volga region. Phys.-math. science.* (Penza) **4(8)** 71-8
[18] Shannon C E 1948 A mathematical theory of communication. *Bell Syst. Tech. J.* **27** 379-423
[19] Kavalerov G I and Mandelstam S M 1974 *Introduction to the information theory of measurements* (Moscow: Energy) p 376
[20] Shishkin I F 2010 *Theoretical metrology. Part 1. The general theory of measurement* (St. Petersburg) p 192
[21] Bialynicki-Birula I and Myciclski J 1975 Uncertainty Relations for Information Entropy in Wave Mechanics Commun. math. Phys. (Springer-Verlag GmbH) (44) 129-32

IOP Conf. Series: Materials Science and Engineering **537** (2019) 022015 doi:10.1088/1757-899X/537/2/022015

[22] Bialynicki-Birula I and Rudnicki L 2011 Entropic Uncertainty Relations in Quantum Physics *Statistical Complexity*. *Applications in Electronic Structure* Ed. Sed Kalids D. (Dordrecht Heidelberg London New York: Springer Science Business Media) 1-34
[23] Kolmogorov A N 1987 *Information Theory and Theory of Algorithms* (Moscow: Nauka) 305
[24] Kruglov V V, Dli M I, Golunov R Y 2001 *Fuzzy logic and artificial neural networks* (Moscow: Fizmatlit) p 225
[25] Novitsky P V and Zograf I A 1985 *Estimation of Measurement Error* (Leningrad: Energoatomizdat) p 248
[26] Polosin V G and Bodin O N 2015 Criteria for optimal control of a dynamic system *Proceedings of MIPT* (Dolgoprudny: Proceedings of MIPT) **3(27)** 131-9