ANALYSIS AND SYNTHESIS OF A SELF-LEARNING INFORMATION-EXTREME DECISION SUPPORT SYSTEM

Shelehov Igor
Ph.D. in Technical Science, Associate Professor, Sumy National Agrarian University, Ukraine
ORCID ID: 0000-0003-4304-7768

One of the main trends in the modern development of scientific and technological progress across all sectors of the socio-economic sphere is the intellectualization of creative activity through the modeling of human cognitive processes in decision-making using computer-based systems.

The use of intelligent technologies in designing and manufacturing products that meet global standards is particularly relevant for the Ukrainian industry. Today, any complex product is not competitive unless it incorporates an intelligent component, and this trend will only intensify. Therefore, leading global manufacturers of modern automated control systems (ACS) have realized that focusing on producing high-precision technological equipment without an intelligent component does not allow for effective control of poorly formalized processes under a priori uncertainty and resource constraints. For example, a key requirement for using such equipment is strict incoming quality control of raw materials. However, for raw materials of natural origin, which are widely used in chemical, metallurgical, food, and other industries, fulfilling this requirement is complicated due to unresolved technical issues in real-time monitoring. The intellectualization of ACS for managing complex, weakly formalized technological processes enables them to acquire adaptability based on self-learning and pattern recognition.

At the same time, scientific and methodological issues related to improving the functional efficiency of self-learning ACS remain insufficiently explored. The main shortcomings of existing self-learning methods in ACS hinder their practical application, including ignoring the overlap of recognition classes in the feature space and the lack of optimization of feature dictionaries during the learning process. This leads to low accuracy in recognizing the functional states of technological processes and the need for excessively large, information-redundant feature dictionaries.

These methodological and theoretical challenges in analyzing and synthesizing highly efficient intelligent ACS highlight the relevance of developing mathematical models for classification-based control, evaluating functional efficiency, and optimizing spatiotemporal parameters of system operation, including feature dictionary parameters. Additionally, it underscores the need for corresponding information technology tools for designing a self-learning decision support system (DSS) operating in a factor cluster analysis mode.

The enhancement of the functional efficiency of automation in controlled technological processes distributed in space and time is associated with the development of scientific and methodological foundations for designing and implementing adaptive automated control systems (ACS) based on self-learning and automatic classification [1–6]. The main characteristics of such systems include:

- The presence of a weakly formalized controlled process, characterized by non-stationarity, implicativity, multi-criteria nature, and the influence of uncontrolled factors.
- The presence of fuzzy input data due to arbitrary (non-zero) initial conditions of the functioning of the learning ACS in a monitoring mode under conditions of a priori uncertainty, informational, and resource constraints.
- Closed feedback loops allow operation in a mode that compensates for internal and external disturbances to stabilize the controlled process.
- The presence of local automatic control systems using analog and digital PID controllers.
- The presence of a decision support system that can operate in two separate (or combined) modes: training, in which error-free decision rules are formed based on a training sample, and examination, where decisions are made regarding the functional state of the system and the generation of control commands.

● A multi-threaded operational mode leads to asynchronous information processing, which complicates the application of traditional mathematical modeling methods for self-learning decision support systems.

● The ability to self-assess functional efficiency.

● Knowledge orientation (the ability to acquire and infer knowledge).

● The use of modern electronic computing systems with high-capacity long-term memory and rapid data processing.

● Ergonomics and the ability to function in an interactive mode.

● Object-oriented system design methodology.

An analysis of the current state and development trends in the synthesis and analysis of self-learning ACS indicates intensive research and implementation of design methods based on artificial neural networks [6–8]. There is ongoing improvement in real-time learning and classification algorithms [9], natural language processing [10], image and signal recognition [11], and more. Among the primary applied tasks solved using neural networks are financial forecasting [12], intelligent data processing [13], system diagnostics [14], network operation monitoring [15], data encryption [16, 17], and others. The main unresolved challenges in this area include:

● The problem of interpretability of weight coefficients is associated with the difficulty of interpreting the meaning of input signal intensity and weight coefficients.

● The problem of interpretability of the transfer function (complexity in interpreting and justifying the additivity of the argument and the form of the neuron activation function).

● The dimensionality problem leads to a "combinatorial explosion" when determining the structure of neural connections, selecting weight coefficients, and choosing a transfer function.

● The problem of linear separability, since neuron excitation takes exclusively Boolean values (0 or 1) [18].

● The interpretability issue reduces the quality of obtained results, the dimensionality problem imposes significant constraints on network capacity and structural complexity, and the linear separability problem necessitates the use of complex multi-layer networks even for relatively simple tasks. A logical step toward resolving the latter situation is developing and implementing control systems based on fuzzy artificial neural networks [19].

● The insufficient efficiency and modelability of known statistical classification methods have driven the intensive development of fuzzy classification methods, initially formulated by L. Zadeh [19]. These methods have primarily been applied in hybrid ACS operating under uncertainty and the influence of multiple uncontrolled factors [20, 21]. However, issues related to evaluating functional efficiency and optimizing the learning process of ACS remain largely unexplored.

One of the key directions in designing intelligent DSS, which are an integral part of adaptive ACS for distributed technological processes, is the development of highly efficient machine learning algorithms that enable the construction of error-free decision rules based on high-dimensional training matrices [22]. Solving this problem within deterministic or statistical approaches, which form the foundation of this research field, is complicated due to the model-based nature of learning algorithms, making them unsuitable for practical applications. Ignoring class overlap in recognition and the absence of algorithms for optimizing the learning process based on a direct criterion of functional efficiency indicate the incompleteness of this learning approach for control systems. The deterministic-statistical approach to analyzing and synthesizing self-learning DSS aims to integrate the advantages of deterministic and statistical methods while overcoming their limitations [3, 23, 24]. The trend toward synthesizing optimal self-learning DSS algorithms through qualitative analysis of existing classical approaches and their modifications, considering the current level of computing technology development and additional practical application requirements, manifests in the latest promising developments in this field, one of which is the information-extremal intelligent technology.

The core idea of analysis and synthesis methods for self-learning DSS within the framework of information-extremal intelligent technology (IEI technology) is optimizing structured spatiotemporal parameters of system operation by transforming similarity relations in a fuzzy partitioning of the feature space into equivalence relations during the learning process. Optimization of system parameters is performed through a hierarchical iterative transitive closure procedure on dynamic mappings within corresponding optimization circuits, aiming to find the global maximum of the informational criterion of functional efficiency within its working domain. The construction of an error-free classifier based on a training matrix, following the principles of optimal control duality [25], reduction [26], and information maximization [3, 4], is performed in a discrete sub-perceptual space through permissible transformations of an a priori fuzzy unimodal distribution of pattern realizations to fit them into an optimal recognition class container reconstructed in a radial basis. This transformation of the a priori distribution of pattern realizations is achieved by deliberately modifying feature values.

A key factor determining the development of artificial intelligence technologies is the growth rate in computing power. The increase in modern computer productivity, combined with improvements in algorithm quality, enables the practical application of previously developed theoretical research. A similar transformation is occurring in intelligent DSS for automating complex technological processes, which strive to use relatively simple yet computationally intensive adaptive behavior algorithms.

Methods of IEI technology for the analysis and synthesis of a self-learning (self-adaptive) automated control system (ACS), integrated with an intelligent decision support system, are based on maximizing the system's informational capacity by introducing additional informational constraints under conditions of a priori uncertainty, fuzzy data, and resource limitations. Further development of IEI technology has led to the creation of several methods that complement and extend the capabilities of the fundamental method—the functional-statistical testing method – and enable its efficient application to solving practical problems in the automation of spatially and temporally distributed technological processes. This is achieved by automatically classifying their functional states under conditions of a priori uncertainty and generating optimal control messages for the user-operator.

The key conceptual principles of the information-extreme method for the analysis and synthesis of self-learning DSS are as follows:

● The criterion of functional efficiency for a self-learning ACS is directly linked to a direct assessment of the system's informational capacity, which uniquely determines the functional efficiency of the DSS.

● The self-learning process is conducted within a deterministically statistical framework and involves constructing relatively simple, error-free decision rules based on a multidimensional training matrix. These rules allow obtaining a full probabilistic assessment of the current functional state of a technological process in an examination mode, i.e., directly in an operational mode, approximating the limiting value.

● The self-learning process takes place under conditions of fuzzy compactness of image realizations, implying the intersection of recognition classes, which is inherent in the practical tasks of automating distributed technological processes. It consists of a targeted iterative multi-cycle optimization of the spatiotemporal functioning parameters to successively approach the global maximum of the informational criterion of functional efficiency, computed within the working (permissible) domain of its function definition to its ultimate maximum, which determines the construction of an error-free classifier.

● The use of logarithmic statistical informational measures, which, according to A.N. Kolmogorov [27], possess the property of compressing the volume of a sample random sequence without losing statistical regularities, allows the use of representative training samples whose volume is an order of magnitude smaller than those required for calculating statistics in multidimensional statistical analysis [28].

● The object-structured design principles embedded in IEI technology [29] enable the inheritance and refinement of its methods, fostering their development within the framework of solving the problem of informational synthesis for a broad class of self-learning ACS.

Thus, IEI-technology methods exhibit a certain degree of universality in designing self-learning ACS, enabling both general and specific tasks of their informational synthesis to be addressed.

Figure 1 illustrates the fundamental principles underlying the information-extreme method for the analysis and synthesis of self-learning DSS. Moreover, the method is based on well-known principles of the systems approach and pattern recognition [3, 30, 31], as well as object-oriented design [29].

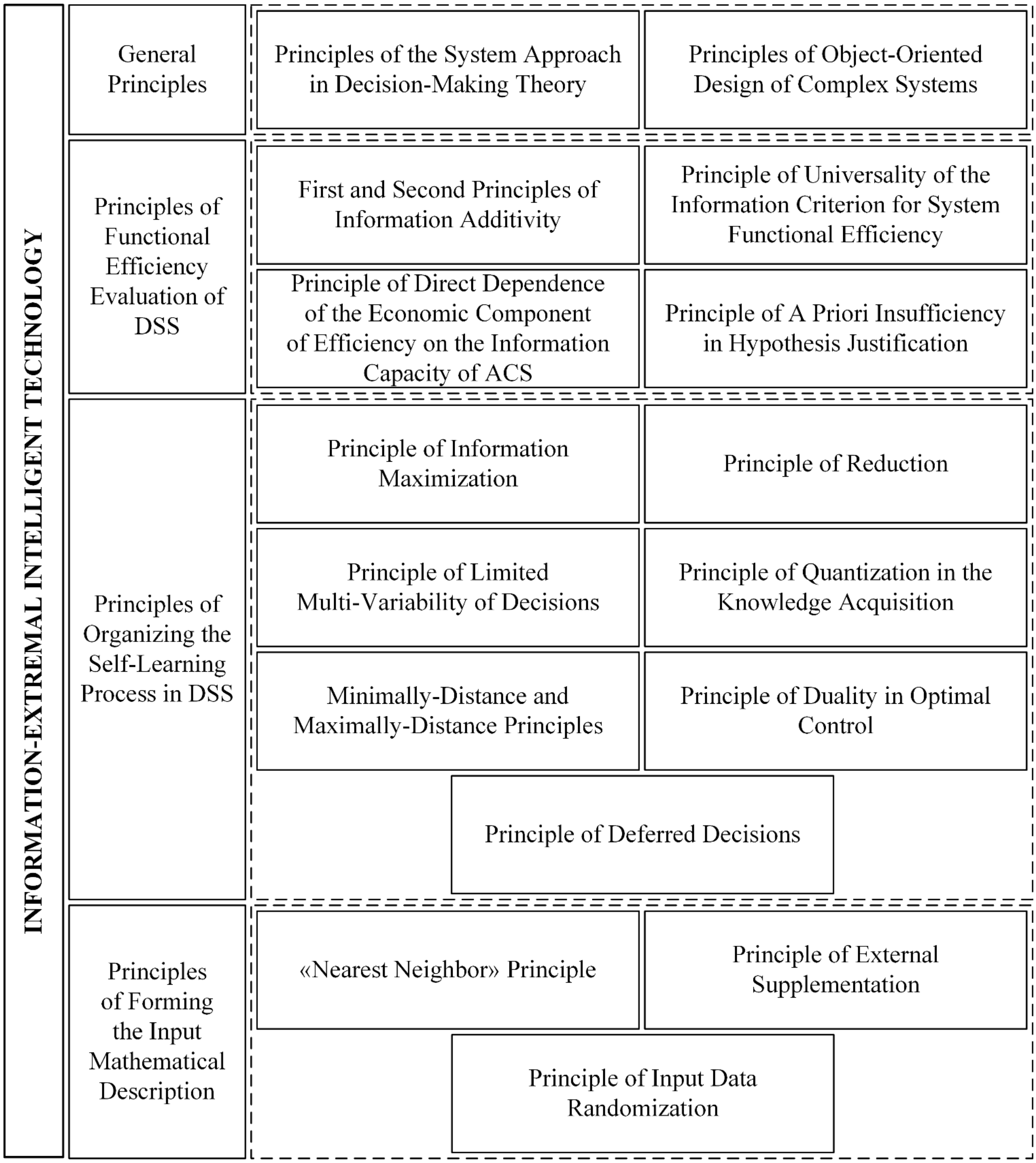

Fig. 1. *Structure of the fundamental principles of the information-extreme method for the analysis and synthesis of self-learning DSS.*

The information maximization principle, which is substantiated by the extremality of sensory perception of an image, was experimentally proven by P. K. Anokhin [2]. This principle is implemented by introducing additional constraints that increase the diversity of objects.

The principle of generality of the information criterion of functional efficiency of the system directly follows from the logical-gnoseological aspect of the nature of information, which is an attribute of control and management processes. This principle determines the expediency of using the information criterion to assess the functional efficiency of the system. From a cybernetic point of view, the efficiency of the ACS functioning is determined by the information indicator of the degree of correspondence of the control to its functional state. Since dynamic changes in its state characterize the system's functioning, then gnoseologically, the informational nature of the criterion of functional efficiency is determined by the diversity of data, characteristics, functional states, and modes of the system during its intended use.

The first and second principles of information additivity. The first principle of information additivity allows, within the framework of the syntactic approach, to evaluate the value of information through its quantitative characteristics [32]. The second principle of information additivity allows for determining the operating range of values of the information criteria of functional efficiency that satisfies the requirements: the greater the amount of information about the recognized images, the greater the reliability of the decisions made.

The principle of reduction of separate functions. This principle consists in the purposeful simplification of a hypothetically existing best separate function of a complex form into a separating function of a more straightforward form, the implementation of which does not entail significant computational costs. The work [33] substantiates the principle of reduction. It shows that the problem of synthesizing complex separating hypersurfaces should be replaced by the problem of synthesizing a feature space of lower dimensionality, in which images can be separated even by a linear classifier. Such a general formulation is justified if the redundant feature dictionary contains features with different informational loads. Reducing their space for a dictionary of informative features can lead to decreased recognition reliability due to information loss. The way out of this situation is to extend the principle of reduction to a hypothetical - best in the informational sense, but complex in form - separating function, provided a simultaneous purposeful transformation of both the distribution of image realizations and the parameters of the separating function of a relatively simple form.

The principle of quantization of the knowledge acquisition process. This principle is implemented in the method of functional-statistical tests through step-by-step accumulation of knowledge during the learning process.

The principle of direct dependence of the economic component of efficiency on the information capacity of the ACS. The principle is that the maximization of the information IFE, which evaluates the functional efficiency of the system, leads to the minimization of average costs during its operation. Thus, the system's economic efficiency is determined by its information capacity.

The principle of a priori insufficiency of hypothesis justification (Bernoulli-Laplace principle). A priori information is incomplete in assessing the efficiency of the ACS functioning. Therefore, according to the Bernoulli-Laplace principle, adopting equiprobable hypotheses is justified. Implementing this principle requires the ACS to make decisions under the worst, in the statistical sense, conditions of its functioning. This guarantees that improving the system's operating conditions will not reduce its functional efficiency but rather increase it.

The principle of composition. The essence of this principle is that the mandatory elements of the mathematical model of the learning process according to the MFCS are the mapping of the universe W of ACS tests onto the set of values of the information criterion E: g: $W \rightarrow E$ and the mapping of the set E onto the set of system functioning parameters G: f:$E \rightarrow G$, where $f \circ g$:$W \rightarrow G$ is determined under the condition $w \in W$. Thus, the condition of complete composition must be satisfied in the MFCS: set E is common to all sequences of mapping sets used in iterative procedures for optimizing the system functioning parameters, which directly or indirectly affect the functional efficiency of the learning ACS.

The "nearest neighbor" principle guarantees the maximum asymptotic complete recognition probability, considered the classifier's potential complete probability. According to the compactness hypothesis (crisp or fuzzy), this principle excludes the influence of realizations of distant classes on the geometric parameters of the containers being optimized at the ACS learning stage.

The principle of randomization (reducing to randomness) of input data. This principle allows, along with the deterministic characteristics of the ACS functional state, to consider random realizations of recognition images, making it possible to evaluate the learning process's accuracy characteristics and calculate the system's information capacity.

The principle of limited multivariance of decisions made. Iterative procedures for optimizing the learning process are based on this principle since the algorithmic information synthesis of the ACS means generating a limited number of possible decision options, which are evaluated during the learning process according to the information IFE.

The principles for evaluating the functional efficiency of control systems justify the appropriateness of using informational criteria of functional efficiency from both a logical-gnoseological and cybernetic perspective. They define the essential properties of information as a measure of the diversity of functional states and system modes when used for its intended purpose.

The principles governing the organization of learning and examination processes detail the methods for knowledge presentation and accumulation during training and their application during the examination mode. Additionally, they justify simplifying the structure of generalized descriptors (concepts) of class realizations and the methods for their construction. These principles lay the foundation for optimizing the redundant feature vocabulary, formulating the notion of informativeness, and determining its relationship with the reliability of decision-making in self-learning DSS.

The principle of a priori fuzzy compactness of class realization vectors ensures the transition from a traditional a priori crisp partitioning of classes, which is model-based, to its fuzzy variant, characteristic of practical control problems for complex organizational-technical objects and technological processes.

The principles for forming a training sample allow determining such samples' representative volume, statistical stability, and homogeneity.

Thus, IEI technology is based on well-known principles of automatic classification and decision theory and specific principles reflecting the informational nature of optimal functioning in a self-learning intelligent DSS.

According to the ideology of IEI technology, permissible transformations of a priori fuzzy partitioning in the feature space are carried out within a discrete sub-perceptual space, enabling:

- Enhanced noise immunity of analysis and synthesis methods for self-learning ACS, applied to weakly formalized, spatially and temporally distributed technological processes.
- Targeted transformation of a priori fuzzy partitioning in the feature recognition space into a crisp equivalence partitioning of classes, allowing the construction of an error-free classifier based on the training sample.

Let's consider the binary feature recognition space Ω_B, a subset of the Hamming space with the cardinality *Card* $\Omega_B=2^N$, where N is the number of recognition features. In classification analysis, the i-th feature X_i is treated as a random variable, the values of which form a repeated sample $\{x_i^{(j)} \mid j=\overline{1,n}\}$ of size n from the population. Then, the set of reflected properties of the m-th functional state of the control system and the relationships between its elements, which can be defined in the recognition feature space as a specific region, will be considered a recognition class (pattern) $X_m^o,\ m=\overline{1,M}$. The set of such recognition classes $\{X_m^o \mid m=\overline{1,M}\}$ forms the alphabet of recognition classes.

In functional-statistical testing within the IEI technology, natural, simulation, or directly performed tests during the operation of the ACS are considered using Bernoulli schemes, during which the informational capacity of the system is evaluated, and a decision is made regarding the sufficiency of their execution. Further, the term "testing" refers specifically to functional-statistical testing.

The deterministic-statistical approach to decision-making requires the setting of both normalized (operational) and control tolerances for the features. Let X_1^o be the base class, which characterizes the maximum functional efficiency of a learnable ACS, i.e., it is the most desirable for the recognition system. Then, the normalized tolerance field $\{\delta_{H,i} \mid i=\overline{1,N}\}$ is such that the feature value X_i lies within it with probability p_i =1 or p_i=0, provided that the ACS functional state relates to the base class X_1^o. The control tolerance field $\{\delta_{K,i} \mid i=\overline{1,N}\}$ is such that the feature value X_i lies within it with probability $0 < p_i < 1$, provided that the ACS functional state relates to the base class X_1^o.

In IEI technology, control tolerances are introduced to randomize the decision-making process; as for completeness, both deterministic and statistical characteristics of the controlled process must be used. The domain of definition for the control tolerance system is the corresponding system of normalized tolerances.

As the vector representation of a pattern $x_m^{(j)} \in X_m^o$, we consider a binary random structured vector, which is the j-th row of the binary training matrix $\left\| x_{m,1}^{(j)} \right\|,\ m=\overline{1,M},\ i=\overline{1,N},\ j=\overline{1,n}$:

$$x_m^{(j)} = < x_{m,1}^{(j)}, \ldots, x_{m,i}^{(j)}, \ldots, x_{m,N}^{(j)} >,\ j=\overline{1,n_{\min}},$$

where $x_{m,i}^{(j)}$ is the i-th coordinate of the vector, which takes a unit value if the feature X_i is within the tolerance field $\delta_{K,i}$, and a zero value otherwise; $n_{\min}$ is the minimum number of trials ensuring the representativeness of the training sample.

Since, in the case of a normal distribution of pattern realizations, the hypothesis of compactness (either strict or fuzzy) of pattern realizations is justified, introducing the concept of a "container" in IEI technology is warranted. This concept represents an approximated approximation of the "precise" complex closed separating hypersurface of a recognition class, which is reconstructed at each training step in the radial basis of the feature space as a regular geometric figure or a combination of several regular geometric figures. In this context, the geometric center of the container can be determined by any method. One way to define the geometric center of the container of class X_m^o is to form the reference vector representation of the pattern $x_m \in X_m^o$, which corresponds to the mathematical expectation of the random realization vectors $\{x_m^{(j)}\}$ of class X_m^o. The structure of the binary reference vector of class X_m^o is given by:

$$x_m = < x_{m,1},\ \ldots,\ x_{m,i},\ \ldots,\ x_{m,N} >,\quad m=\overline{1,M},$$

where $x_{m,i}$ is the i-th coordinate of the vector, which takes a unit value if the feature X_i falls within the normalized tolerance field $\delta_{H,i}$ and a zero value otherwise.

Thus, for the class X_m^o the container $K_m^o \subset X_m^o$ within the framework of IEI technology serves as its "transparent" shell. The "transparency" of this shell means that the containers may overlap when constructing a fuzzy partitioning of the feature space into classes.

Since the main task of the training stage of the ASC under IEI technology is to construct decision rules by partitioning the feature space into recognition classes in an optimal way in the informational sense, the evaluation of precise characteristics of the learning process becomes crucial. In [3], the average asymptotic (or upper bound) complete reliability of decision-making in the trained DSS is defined as:

$$\overline{P}_{np}^{*} = \frac{1}{M}\sum_{m=1}^{M} P_{np,m}^{*} = \frac{1}{ML}\sum_{m=1}^{M}\sum_{l=1}^{L} D_{l,m}^{*}, \tag{1}$$

where $P_{np,m}^{*} = \sum_{l=1}^{L} p(\mu_l) D_{l,m}^{*}$ is the asymptotic complete reliability of recognizing realizations of class X_m^o; $D_{l,m}^{*}$ represents the extreme asymptotic values of the l-th recognition reliability of realizations of class X_m^o, which defines the maximum E_m^* of the informational criterion of functional efficiency of the learning process; L – is the number of statistical hypotheses.

In formula (1), $p(\mu_l)$ is the unconditional probability of accepting the statistical hypothesis μ_l. According to the principle of insufficient reason (Bernoulli-Laplace principle of insufficient grounds), in the absence of additional information about the hypotheses $\{\mu_l\}$ it is reasonable to assume they are equally probable, i.e., $p(\mu_l) = 1/L$.

The efficiency of an Automated Decision Support System (ADSS) depends on its spatial-temporal operating parameters—information support characteristics that influence the system's functional efficiency. The operational parameters optimized during the training process will be referred to as training parameters. As an optimization criterion for the training process within the IEI-technology framework, any statistical informational criterion of functional efficiency can be used, as it serves as a natural measure of the diversity of recognition classes and simultaneously acts as a functional parameter of the accuracy characteristics of a self-learning distributed integrated ADSS.

Important operational parameters of ADSS include the parameters of the feature dictionary. The optimization of these parameters within the IEI-technology framework is based on the concept of informativeness, both of individual features and their groups. In this context, informativeness is considered as the degree of influence of a feature on the optimization criterion of the training process, which reflects the functional efficiency of ADSS.

Thus, within the IEI-technology framework, four main groups of features can be distinguished:

- "Informative" features, whose presence in the dictionary increases the value of the criterion of functional efficiency;
- "Non-informative" features, which do not affect the functioning of the self-learning ADSS;
- "Interfering or misleading" features, whose inclusion in the feature dictionary leads to a loss of efficiency in ADSS training;
- "Latent" features, whose "hidden nature" is due to their low frequency of occurrence, do not exceed the selected decision threshold.

Let us consider the formulation of the problem of information-extreme synthesis of a self-learning ADSS operating in the factorial cluster analysis mode.

Let the known a priori alphabet be given in the general case of fuzzy recognition classes, characterizes of functional states of the technological process. The training matrix of the «object-property» type $\| y_{m,i}^{(j)} \|,\ i=\overline{1,N},\ j=\overline{1,n},$ where N, n – are the number of recognition and testing features, respectively, and a priori, in the general case, redundant in the informational sense feature dictionary $\Sigma^{|N|}$. Let the base class X_1^o characterize the most desirable functional state. It is necessary to:

1) For an a priori classified fuzzy partition $\tilde{\Re}^{|M|}$, construct an optimal partition in the sub-perceptual discrete feature space Ω_B through permissible transformations (here and further in work, in the informational sense) that correspond to a clear equivalence partitioning of the classes $\Re^{|M|}$:

$$\left(\forall X_m^o \in \tilde{\Re}^{|M|}\right)\left[X_m^o \neq \varnothing, m=\overline{1,M}\right] \tag{2}$$

$$\left(\exists X_m^o \in \tilde{\Re}^{|M|}\right)\left(\exists X_c^o \in \tilde{\Re}^{|M|}\right)\left[X_m^o \neq X_c^o \to X_m^o \cap X_c^o \neq \varnothing,\ m,c=\overline{1,M}\right] \tag{3}$$

$$\left(\forall X_m^o \in \tilde{\Re}^{|M|}\right)\left(\forall X_c^o \in \tilde{\Re}^{|M|}\right)\left[X_m^o \neq X_c^o \to KerX_m^o \cap KerX_c^o = \varnothing\right] \tag{4}$$

$$(\forall X_m^o \in \Re^{|M|})\left(\Sigma_M^{(i)} \in \Omega_B\right)(E_m \in G_E)[if\ E_m^* = \max_{\{k\}} E_m\ then\ \Sigma_M^* := \Sigma_M^{(i)},\ i=\overline{1,N}] \tag{5}$$

$$\left(\forall X_m^o \in \tilde{\Re}^{|\Lambda|}\right)\left(\forall X_c^o \in \tilde{\Re}^{|\Lambda|}\right)\left[X_m^o \neq X_c^o \to (d_m^* < d(x_m \oplus x_c))\,\&\,(d_c^* < d(x_m \oplus x_c))\right] \tag{6}$$

$$\bigcup_{X_m^o \in \tilde{\Re}} X_m^o \subseteq \Omega_B, \tag{7}$$

where $KerX_m^o$ – the core of class X_m^o; $KerX_l^o$ – the core of class X_l^o, the closest neighbor to class X_m^o; $\Sigma_M^{(i)}$ – the current feature dictionary for the a priori alphabet of classes, which contains *i* features, $i=\overline{1,N}$; E_m – informational criterion of functional efficiency for the training of the self-learning decision support system; G_E – the range of criterion of functional efficiency values; Σ_M^* – the optimal feature dictionary for M classes; d_m^* – the optimal radius of the container $K_m^o \subset X_m^o$; $d(x_m \oplus x_c)$ – the inter-center code distance for classes X_m^o and X_c^o; d_c^* the optimal radius of the container $K_l^o \subset X_c^o$.

2) At the stage of DSS training in the factorial cluster analysis mode, construct, using permissible transformations in the sub-perceptual discrete feature space Ω_B, the optimal open, clear partition of class equivalence $\Re^{|\Lambda|}$, $\Lambda > M$, under the condition:

$$\left(\forall X_m^o \in \tilde{\Re}^{|\Lambda|}\right)\left[X_m^o \neq \varnothing, m=\overline{1,\Lambda}\right] \tag{8}$$

$$\left(\exists X_m^o \in \tilde{\Re}^{|\Lambda|}\right)\left(\exists X_c^o \in \tilde{\Re}^{|\Lambda|}\right)\left[X_m^o \neq X_c^o \to X_m^o \cap X_c^o \neq \varnothing, c=\overline{1,\Lambda}\right] \tag{9}$$

$$\left(\forall X_m^o \in \tilde{\Re}^{|\Lambda|}\right)\left(\forall X_c^o \in \tilde{\Re}^{|\Lambda|}\right)\left[X_m^o \neq X_c^o \to KerX_m^o \cap KerX_c^o = \varnothing\right] \tag{10}$$

$$(\forall X_m^o \in \Re^{|\Lambda|})\left(\Sigma_\Lambda^{(i)} \in \Omega_B\right)(E_m \in G_E)[if\ \overline{E}^* = \max_{\{k\}} \overline{E}\ then\ \Sigma_\Lambda^* := \Sigma_\Lambda^{(i)},\ i=\overline{1,N}] \tag{11}$$

where $\Sigma_\Lambda^{(i)}$ – the current feature dictionary for the new class alphabet $\{X_m^o\}^{|\Lambda|}$, which contains *i* feature, $i=\overline{1,N_{\Lambda-1}^*}$; $\overline{E} = \frac{1}{\Lambda}\sum_{m=1}^{\Lambda}\sum_{i=1}^{N} E_{m,i}$ – the averaged value of the training criterion of functional efficiency of the DSS; Σ_Λ^* – the optimal feature dictionary for Λ classes;

$$\left(\forall X_m^o \in \tilde{\Re}^{|\Lambda|}\right)\left(\forall X_c^o \in \tilde{\Re}^{|\Lambda|}\right)\left[X_m^o \neq X_c^o \to (d_m^* < d(x_m \oplus x_c))\,\&\,(d_c^* < d(x_m \oplus x_c))\right] \tag{12}$$

$$\bigcup_{X_m^o \in \tilde{\Re}} X_m^o \subseteq \Omega_B \tag{13}$$

In equation (11), $E_{m,i}$ – represents the training criterion of functional efficiency values of the system for recognizing class $X_m^o,\ m=\overline{1,\Lambda}$ for the current feature dictionary $\Sigma_\Lambda^{(i)}$.

3) At the examination stage, i.e., directly in the working mode of DSS, assess the current functional state of the controlled technological process, and when changing the power of the class alphabet, form a representative training matrix $\| x_{m,i}^{(j)} \mid m=1,\Lambda; i=1,N_{\Lambda-1}^*; j=\overline{1,n} \|$, where $N_{\Lambda-1}^*$ – is the power of the previous optimal dictionary $\Sigma_{\Lambda-1}^*$.

4) Retrain the DSS for the class alphabet $\{X_m^o\}^{|\Lambda|}$ with feature dictionary optimization.

5) In case of inconsistency of the current functional state with class X_1^o, implement a correction operator.

Thus, the specificity of the DSS self-learning task within the framework of IEI technology lies in combining the factorial cluster analysis task with the assessment of feature informativeness and the optimization of feature dictionary parameters through a multi-cycle structuring based on the parameters of the iterative procedure functioning, in search of the global maximum of the self-learning information criterion of functional efficiency in the working (admissible) domain of its function definition.

Let us consider a mathematical model of an information-extremal DSS that learns with a constant cardinality of the feature dictionary. The mathematical model should include, as a mandatory component, an input mathematical description, which we will present in the form of a set-theoretic structure:

$$\Delta_B = \langle G, T, \Omega, Z, Y, V, F, \Phi, \Pi, \mathrm{H} \rangle,$$

where $F: G \times T \times Z \to \Omega$ – is the operator of the feature space formation; $\Pi: G \times T \times \Omega \to Z$ the transition operator, reflecting the mechanism of state changes under the influence of internal and external disturbances; $\mathrm{H}: G \times T \times \Omega \times Z \to V$ – is the operator of a transition to a new type of decision rules.

When substantiating the hypothesis of fuzzy compactness, which occurs in practice, let's consider an a priori fuzzy partitioning $\tilde{\Re}^{|M|} \subset \Omega$. We apply the operator θ of admissible transformations of the input mathematical description of the DSS into the binary feature space Ω_B for the purpose of fuzzy factorization of the feature space: θ: $Y \to \tilde{\Re}^{|M|}$. Let the classification operator: $\tilde{\Re}^{|M|} \to I^{|l|}$ test the main statistical hypothesis about the belonging of realizations $\{x_m^{(j)} \mid j=\overline{1,n}\}$ to class X_m^o, where $I^{|l|}=\{\gamma_1,\gamma_2,...,\gamma_l\}$ – is the set of hypotheses. By evaluating statistical hypotheses, the operator γ: $I^{|l|} \to \Im^{|q|}$ forms a set of accuracy characteristics $\Im^{|q|}$, where $q = l^2$ is the number of accuracy characteristics. The operator $\phi: \Im^{|q|} \to E$ calculates the set of values of the information criterion of functional efficiency, which is a function of the accuracy characteristics. The optimization loop of the geometric parameters of the partitioning $\tilde{\Re}^{|M|}$ by searching for the maximum of the criterion of functional efficiency of learning to recognize realizations of class X_m^o is closed by the operator r: $E \to \tilde{\Re}^{|M|}$. Then, the categorical model in the form of a diagram of mapping sets involved in the learning process according to the basic algorithm within the IEI technology, for an a priori fuzzy partitioning, has the form shown in Figure 2 [3].

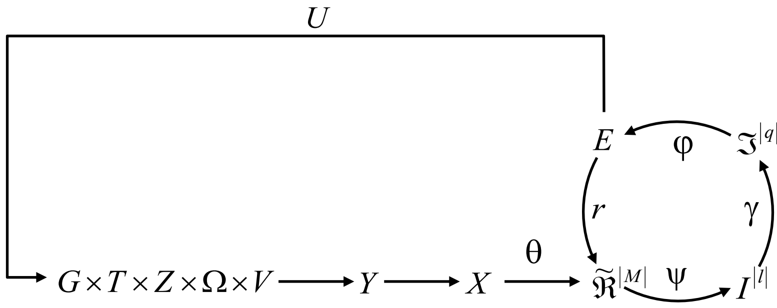

Fig. 2. *Diagram of Mapping Sets for the Basic Learning Algorithm*

where

G space of input signals (factors) acting on the DSS;

T set of time instances for information acquisition;

Ω recognition feature space;

Z – space of possible states;

V – set of types of decision rules;

Φ: $G\times T\times\Omega\times Z\rightarrow Y$ – operator for forming the sample set Y operator for forming the sample set;

Y sample set (input training matrix $\| y_{m,i}^{(j)} \|$);

X sample set that forms a binary training matrix $\| x_{m,i}^{(j)} \mid i=\overline{1,N},\ j=\overline{1,n} \|$, analogous in structure to the input training matrix $\| y_{m,i}^{(j)} \|$;

$U: E \rightarrow G\times T\times\Omega\times Z\times V$ – operator that regulates the learning process and allows optimizing the parameters of its plan, which determine, for example, the volume and structure of tests, the order of consideration of recognition classes, and so on.

The composition of the admissible transformation operator $\theta=\theta_1\circ\theta_2$ in Figure 2 consists of operator θ_1, which forms a sample binary set X – the input, in the general case, real-valued binary training matrix $\| x_{m,i}^{(j)} \|$ of the "process-property" type, and operator θ_2, which restores the optimal partitioning of the feature space into equivalence classes during the DSS learning process.
Among the learning parameters that significantly affect the classifier's reliability, one can consider the control tolerance fields $\{\delta_{k,i} \mid i=\overline{1,N}\}$ for feature values, the selection levels $\{\rho_m\}$ of the coordinates of the reference binary vectors, the quantization step in time τ of image realizations, the parameters of the feature dictionary $\Sigma^{\{N\}}$, environmental influence parameters, and others.

The diagram of mapping sets used in the examination within the IEI technology is shown in Figure 3 [3].

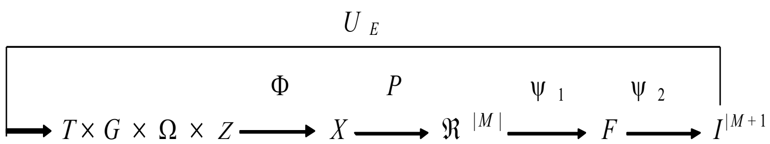

Fig. 3. *Diagram of mapping sets during DSS operation in examination mode*

In the diagram (Figure 3), the operator Φ_E maps the universe of tests onto a sample set X, which forms a binary examination matrix $\| x_i^{(j)} \mid i=\overline{1,N},\ j=\overline{1,n} \|$, analogous in structure and formation parameters to the binary training matrix $\| x_{m,i}^{(j)} \|$.

Within IEI technology, the iterative optimization of the DSS self-learning process will be carried out using the information criterion of functional efficiency, a functional characteristic of accuracy. The iterative process of optimizing the geometric parameters of the partition $\tilde{\Re}^{|M|}$, according to the diagram in Figure 2, is implemented by the operator r: $E \rightarrow \tilde{\Re}^{|M|}$ by searching for the maximum of the criterion of functional efficiency:

$$E_m^* = \max_{\{d\}} E_m\left(d\right), \tag{13}$$

where $\{d\}$ is the set of learning steps of the DSS to recognize realizations of class X_m^o.

As is known, the basic idea of IEI technology lies in changing the values of features in the sub-perceptual space through admissible transformations. One such transformation is the optimization of a control tolerance system on features, which, within IEI technology, consists of selecting such a control tolerance system from the term set D that iteratively approximates the value of the global maximum of the information optimization criterion E in the working (admissible) domain of its function to its largest (limit) value. Figure 2.6 shows a categorical model of control tolerance optimization in DSS learning [3].

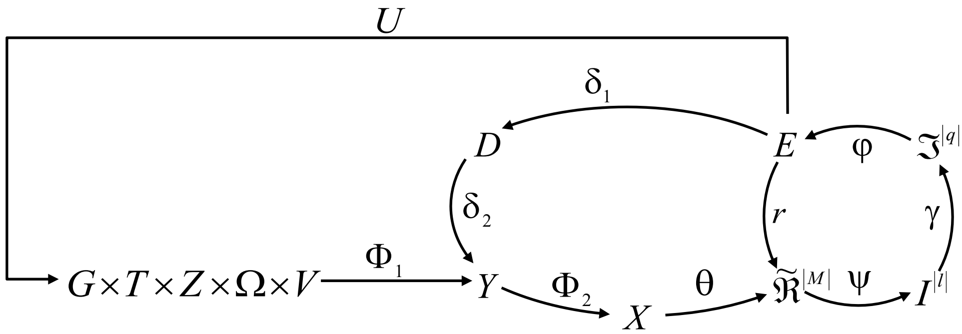

Fig. 4. *Diagram of mapping sets in the optimization of control tolerance system using IEI technology*

In Figure 4, the operators δ_1 and respectively evaluate the influence of the optimized parameter on the functional efficiency of the DSS and regulate the iterative optimization process.

Figure 5 shows the contour of operators that directly optimize the control tolerance system and includes the contour of optimization of the geometric parameters of the partition $\tilde{\Re}^{|M|}$.

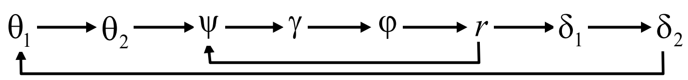

Fig. 5. *Contour of optimization of control tolerances on features*

Thus, the optimization of the geometric parameters of class containers is carried out at each step of control tolerance system optimization and is an internal cycle of the information-extremal algorithm for DSS learning.

The mathematical model of feature dictionary optimization (feature selection) in IEI technology can be described as an additional optimization contour in the learning algorithm. The modified mathematical model can be represented as a corresponding diagram of mapping sets (Figure 6):

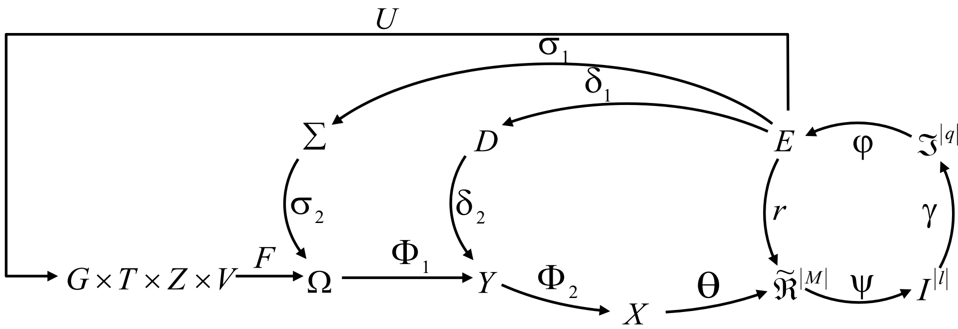

Fig. 6. *Diagram of mapping sets in the process of optimizing the recognition feature dictionary in IEI technology*

The contour of feature dictionary optimization is shown in Figure 7.

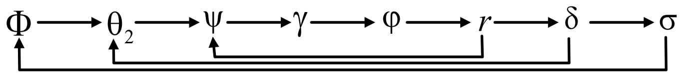

Fig. 7. *Contour of feature dictionary optimization*

In Figure 7, the operator $\sigma = \sigma_1 \circ \sigma_2 : E \to \Omega$ changes the feature space Ω according to the corresponding algorithm for optimizing the feature dictionary. For the obtained current version of the dictionary Σ in the learning process, its parameters are optimized using either the basic or control tolerance system optimization algorithm, the structural diagrams shown in Figures 5 and 6, respectively. In this case, the optimization of the feature dictionary is carried out by an iterative procedure of searching for the maximum of the objective function using the algorithm:

$$\Sigma^* = \arg\max_{\Sigma\in\Omega}\{\max_{G_\delta}\{\max_{\{k\}} EK_k\}\}, \qquad (14)$$

where EK_k – some generalized objective function calculated at the *k*-th step of DSS learning and includes both the information criterion of functional efficiency, the calculation of which is a feature of IEI technology and additional conditions (e.g., the minimum dimension of the feature space, etc.), which are characteristic of the corresponding feature selection algorithm; G_δ – the domain of admissible values of the control tolerance field; $\{k\}$ – the set of learning steps.

Additional conditions for calculating the objective function indicate the existence of auxiliary contours of feature dictionary optimization, which are related to other optimized parameters of DSS functioning. Considering this, the previous mapping diagram (Figure 3) takes the form shown in Figure 8, where the dashed-dotted arrows indicate possible additional operators for dictionary optimization that use the features of optimizing other parameters of the learning DSS. In this case, the operator h_D deletes a group of features that do not change the criterion of functional efficiency in the control tolerance system optimization process, provided that such optimization was carried out for each feature sequentially.

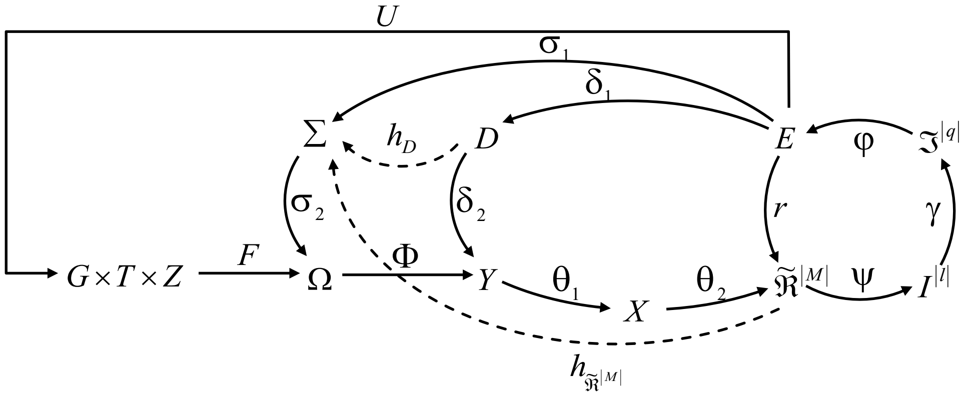

Fig. 8. *Diagram of mapping sets in the optimization of the feature dictionary using additional conditions*

The operator $h_{\tilde{\Re}^{|M|}}$ shown in Figure 8 checks which features were used to implement the maximum-distance or minimum-distance principles for optimizing the geometric parameters of the feature space partition. It should be noted that these contours only affect the dictionary optimization strategy since they can combine individual features into groups according to their influence on the learning DSS.

Figure 9 shows a categorical model of DSS functioning in the factorial cluster analysis mode with self-learning.

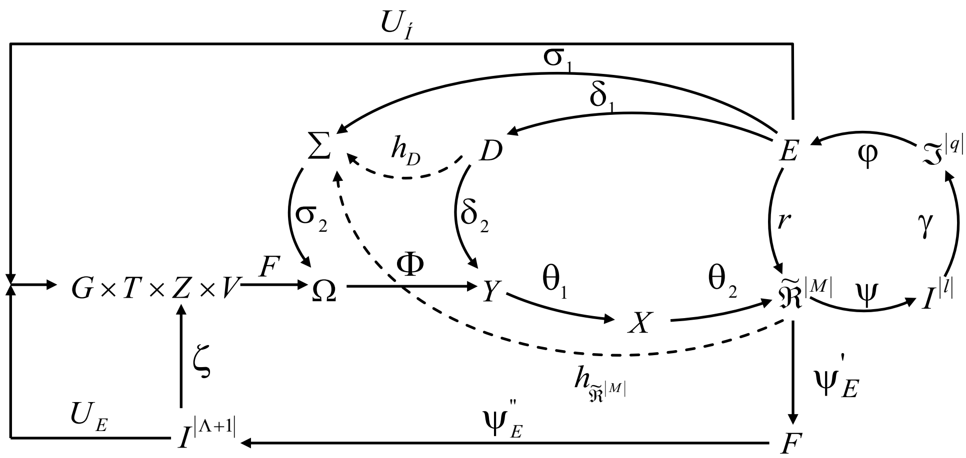

Fig. 9. *Diagram of mapping sets in the optimization of the feature dictionary in the factorial cluster analysis mode*

The difference between the mapping diagram (Figure 9) in the factorial cluster analysis mode and the above ones lies in the presence of parallel learning and examination contours. In this case, the operators U_H and U_E regulate the learning and examination processes, respectively, and the operator for classifying image realizations in the examination mode forms the composition $\psi_E = \psi_E^{/} \circ \psi_E^{//}$, where $\psi_E^{/}$ – is the operator for calculating the membership function of the image realization to the corresponding container; $\psi_E^{//}$ – is the operator for implementing decision rules.

The advantage of categorical models in the form of the above diagrams of mapping sets is that they allow, at the stage of system analysis of DSS, that learn (self-learn) in the factorial cluster analysis mode not only to establish the relationships between the elements of information support and data processing information flows but also significantly facilitate the development of system functioning algorithms.

A necessary and sufficient condition for the implementation of factorial cluster analysis in IEI technology is the fulfillment of the inequality:

$$\overline{\mu_m} = \frac{1}{n}\sum_{j=1}^{n}\mu_{m,j} \le c \quad (15)$$

,

where $\overline{\mu_m}$ the averaged membership function of the vector-realization of the recognized class to the container $K_m^0 \subset X_m^0$; c the threshold value that determines the acceptance of the hypothesis of "refusal" to classify $\gamma_{\Lambda+1} \in I^{|\Lambda+1|}$. $I^{|\Lambda+1|}$ is the set of hypotheses for an open alphabet, where $\gamma_{\Lambda+1}$ – is the hypothesis that allows the formation of a training matrix for a new class X_Λ^0 and, accordingly, the re-training of the system. In this case, for a hyperspherical container constructed in the radial basis of the feature space, which is acceptable for an unimodal distribution of class realizations, the geometric membership function, for example, for class X_m^o can have the form [3]

$$\mu_m = 1 - \frac{d(x_m^* \oplus x^{(j)})}{d_m^*} \quad (16)$$

,

where $d(x_m^* \oplus x^{(j)})$ - the code distance of the vector-realization $x^{(j)}$, being recognized from the vertex of the binary optimal reference vector $x_m^* \in \Re^{|\Lambda|*}$, determined in the learning process for the optimal strict partition $\Re^{|\Lambda|*}$; d_m^* – the optimal radius of the container of class X_m^o, calculated in the process of DSS learning.

Within the framework of factorial cluster analysis using IEI technology, the algorithm for aggregating a new class with a constant dictionary of recognition features consists of the operator forming an additional training matrix $\| x_\Lambda^{(j)} \|$, where $\Lambda = M + 1$, which consists of realizations of the examination matrix that yielded negative values of the membership function (16) for all classes. Upon reaching the required representativeness of the matrix $\| x_\Lambda^{(j)} \|$ the operator ς launches the process of re-training the DSS to construct a new partition of the feature space.

Thus, implementing factorial cluster analysis algorithms in modern ACS is necessary due to the low reliability of assessing the functional states of controlled, weakly formalized technological processes that occur under a priori uncertainty.

As a component of overall efficiency, functional efficiency determines the degree of correspondence between the system's functioning according to its working algorithm and the fulfillment of the assigned task according to the goal criterion. An important component of the goal criterion is the information criterion of the functional efficiency of the system's learning, which is a function of the accuracy characteristics of the decisions made by the system. The task of selecting and calculating the criterion of functional efficiency is a central problem in evaluating the functional efficiency of an intelligent DSS, for which the information approach is a priority in decision-making problems. Within IEI technology, two information measures have found wide application [3, 4]: the entropic measure of Shannon, which is an integral measure:

$$E_m^{(k)} = 1 + 0,5\left(\frac{\alpha^{(k)}}{\alpha^{(k)} + D_2^{(k)}} \log_2 \frac{\alpha^{(k)}}{\alpha^{(k)} + D_2^{(k)}} + \frac{D_1^{(k)}}{\beta^{(k)} + D_1^{(k)}} \log_2 \frac{D_1^{(k)}}{\beta^{(k)} + D_1^{(k)}} + \right.$$
$$\left. + \frac{\beta^{(k)}}{\beta^{(k)} + D_1^{(k)}} \log_2 \frac{\beta^{(k)}}{\beta^{(k)} + D_1^{(k)}} + \frac{D_2^{(k)}}{\alpha^{(k)} + D_2^{(k)}} \log_2 \frac{D_2^{(k)}}{\alpha^{(k)} + D_2^{(k)}} \right), \quad (17)$$

and the Kulback measure:

$$E_m^{(k)} = 0,5 \log_2 \left(\frac{D_1^{(k)} + D_2^{(k)} + 10^{-r}}{\alpha^{(k)} + \beta^{(k)} + 10^{-r}} \right) * \left[(D_1^{(k)} + D_2^{(k)}) - (\alpha^{(k)} + \beta^{(k)}) \right] \quad (18)$$

where $D_1^{(k)}$ – the first accuracycalculated at the k-th step of learning; $D_2^{(k)}$ – the second accuracy; $\alpha^{(k)}$ – the type I error; $\beta^{(k)}$ – the type II error; 10^{-r} – a sufficiently small number to avoid division by zero..

In the general case, the function's graph constructed according to (17) is a three-dimensional surface (Figure 2.12). It is symmetric concerning the bisector of the angle D_1OD_2, i.e., at the same values as the first and second certainties. In Figure 2.12, the second part of the graph is not shown for greater clarity.

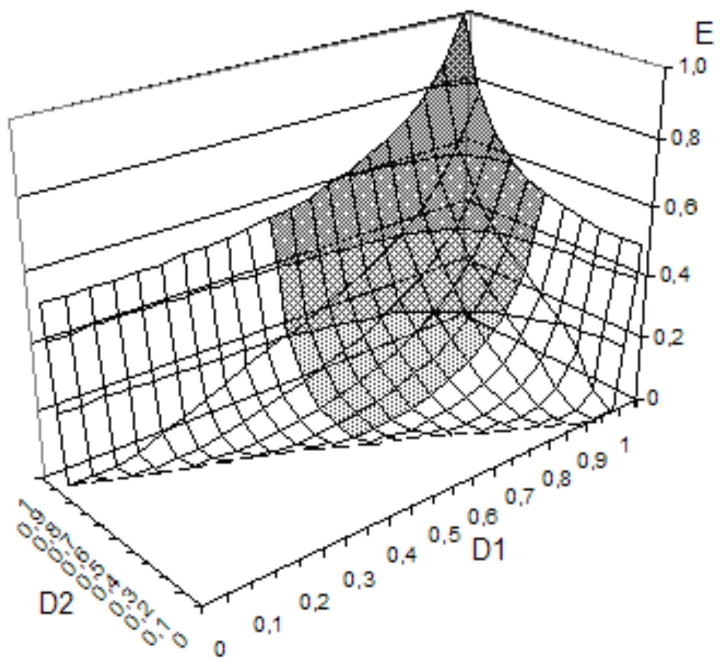

Fig. 10. *Graph of the dependence of criterion (17) on accuracy characteristics with a two-alternative decision*

The three-dimensional surface of the modified criterion $J = f(D_1, D_2)$, constructed according to formula (18), is shown in Figure 11.

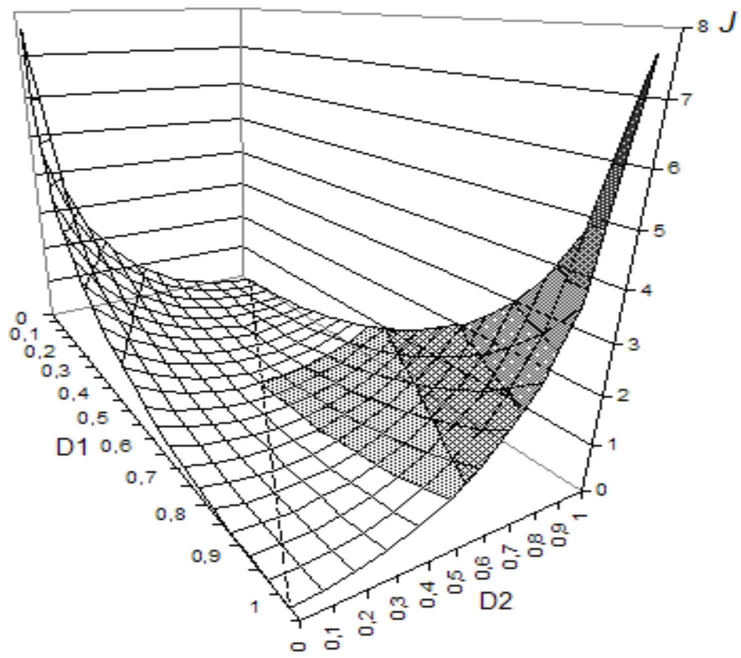

Fig. 11. *Graph of the dependence of criterion (18) on accuracy characteristics: the first and second accuracies*

As can be seen from Figures 10 and 11, the functions (17) and (18) are mutually non-unique. In practice, this shortcoming is eliminated by introducing a working (admissible) domain of the functions, in which the values of the accuracy characteristics – the first and second accuracies – must be greater than the corresponding type I and type II errors, i.e., $D_1 \geq 0.5$ and $D_2 \geq 0.5$. The working areas in Figures 10 and 11 are shown in the corresponding graphs in gray. Analysis of these graphs shows that as both the first and second certainties increase in the working area, the amount of information also increases, which is consistent with the second principle of information additivity [4].

Thus, an analysis of the functions used as the criterion of functional efficiency of DSS learning in IEI technology shows their compliance with the basic requirements for such criteria:

- They are direct and objective criteria.
- They are mathematically computable and have a geometric meaning.
- They characterize the degree of correspondence of the system to its purpose and the economic suitability of its use.
- They are constructive in nature, i.e., they allow the development of methods for analyzing and synthesizing the control system.
- They are universal, i.e., capable of evaluating the functional efficiency of a wide-purpose control system.
- They are sensitive to changes in the functioning parameters and characteristics of the learning control system.
- They allow for optimizing the learning process of the learning control system to maximize its asymptotic complete certainty of recognition.
- They have a functional relationship with the accuracy characteristics of the learning control system.
- They evaluate the reliability of the learning control system.
- They allow for predicting changes in the functional efficiency and reliability of the adaptive learning ACS.

The construction of an error-free classifier in IEI technology based on the "nearest neighbor" principle is possible in a particular case, provided that all image realizations enter the corresponding container of the recognition class, which does not guarantee the necessary performance of machine learning, which can be considered as the ratio of the optimal coverage of recognition classes to the entire feature space. Therefore, in the general case, the study of the influence of the cardinality of both the feature dictionary and the alphabet of recognition classes on the effective and capable estimation of the asymptotic complete probability of correct decision-making $P_t^* = 0,5D_1^* + 0,5D_2^*$, де D_1^*, D_2^* – is the asymptotic (extreme) first and second accuracies of recognition of class X_m^o, realizations, calculated from the results of optimization learning, becomes of significant scientific and practical importance.

It is known that in a binary space, a hypercube approximates a hyperspherical container. For the purpose of generalization and convenience of constructing such a container, the existence of a pseudo-hypersphere that describes the hypercube, i.e., contains all its vertices, is permissible. This allows us to further consider such parameters of container optimization in the radial basis of the Hamming space as the reference vector, for example, $x_m \in X_m^o$, the vertex of which defines the geometric center of the container $K_m^o \in X_m^o$, and the radius of the pseudo-spherical container, which is determined by the formula:

$$d_m = \sum_{i=1}^{N} \left(x_{m,i} \oplus \lambda_i \right), \qquad (19)$$

where $x_{m,i}$ – the i-th coordinate of the binary reference vector x_m; λ_i – the i-th coordinate of a specific vector λ, whose vertices belong to the surface of the container $K_m^o \in X_m^o$.

For simplicity, the code distance (19) between vectors x_m and λ will be denoted as $d_m = d\left(x_m \oplus \lambda\right)$, and instead of the term "pseudospherical," we will use the term "hyperspherical" container.

Let d_0^*, d_1^* be the optimal radius of the class containers X_0^o and X_1^o respectively, and let $d_c = d\left(x_0^* \oplus x_1^*\right)$ be the code distance between their centers – reference vectors $x_0^* \in X_0^o$ and $x_1^* \in X_1^o$ respectively. Taking into account the specifics of the binary Hamming space, the following assumptions can be made:

1) The capacity of the binary Hamming space for the feature recognition dictionary $\Sigma^{|N|}$ equals 2^N.

2) The number of binary realizations at a code distance d ($0 \le d \le N$) from a binary vector x is given by

$$B_d^N(x) = C_N^d = \frac{N!}{d!(N-d)!}.$$

3) The number of binary realizations belonging to any container of class X_K^o with radius d_K ($0 \le d_K \le N$) is given by

$$B_{d,K}^{N}=\sum_{i=0}^{d_K} B_N^i=\sum_{i=0}^{d_K}\frac{N!}{i!\left(N-i\right)!}.$$

At the same time,

$$B_{d,K}^{N}=\sum_{i=0}^{d_K-1}\frac{N!}{i!\left(N-i\right)!}+\frac{N!}{d_K!\left(N-d_K\right)!}=B_{d-1,K}^{N}+B_d^N\left(x_K\right),$$

where $B_{d-1,K}^{N}$ is the number of realizations of an image in a container with radius $d-1$.

Thus, in the case of strict partitioning $\Re^{|M|}$ for M classes, that is, when $d_0^*+d_1^*<d_c$, the number of corresponding realizations belonging to the containers of classes X_0^o and X_1^o is given by:

$$B_{\Re^{|m|}}^{N}=\sum_{k=0}^{m} B_{d,k}^N=\sum_{k=0}^{m}\sum_{i=0}^{d_k^*}\frac{N!}{i!\left(N-i\right)!},$$

while the number of realizations outside these containers is given by:

$$B_{\overline{\Re}}^{N}=2^N-B_{\Re}^{N}.$$

Consider the case of fuzzy partitioning $\tilde{\Re}^{|2|}$ for two classes X_0^o and X_1^o, which overlap, meaning that when $d_0^*+d_1^*\geq d_c$ (see Figure 12).

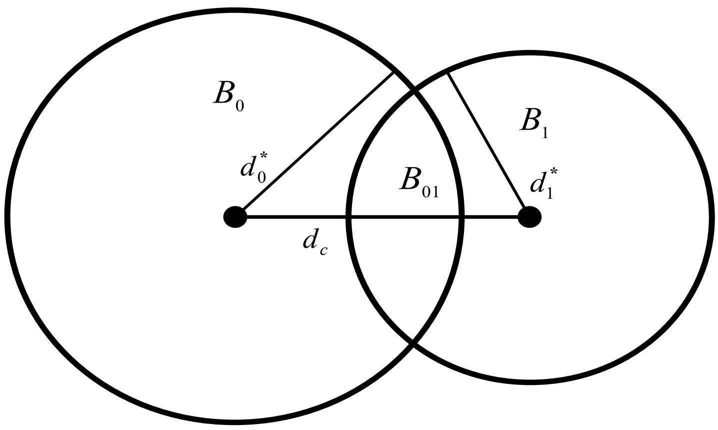

Fig. 12. *Geometric characteristics of class containers* X_0^o *i* X_1^o

The calculation of the value $B_{\tilde{\Re}^{|2|}}^{N}$ is complicated due to the presence of a region in the binary space $K_0^o \cap K_1^o$, where the class containers X_0^o and X_1^o overlap:

$$B_{\tilde{\Re}^{|2|}}^{N}=B_{\Re^{|2|}}^{N}-B_{\tilde{\Re}^{|2|}\left(X_0^o,X_1^o\right)}^{N}.$$

Lemma 1. The number of binary realizations located at a code distance d_0 from the binary vector x_0 and d_1 from the binary vector x_1 is zero if $d_0 + d_1 < d\left(x_0 \oplus x_1\right)$.

Proof. Let the binary vector x_0 be zero. Then, the vector x_1 contains $d\left(x_0 \oplus x_1\right)$ unit components and $N - d\left(x_0 \oplus x_1\right)$ zero components. A binary realization in Hamming space, located at a code distance d_0 from the binary zero vector x_0, contains d_0 unit components. The distance from this realization to the binary vector x_1 depends on the number of d_0 components that coincide with the $d\left(x_0 \oplus x_1\right)$ unit components of vector x_1. Then, the smallest code distance will occur when the number of such coincidences is maximally possible, i.e., it will be equal to $\left|d\left(x_0 \oplus x_1\right) - d_0\right|$. Thus, for $d_1 < \left|d\left(x_0 \oplus x_1\right) - d_0\right|$, which includes the case $d_0 + d_1 < d\left(x_0 \oplus x_1\right)$, the number of corresponding binary realizations is zero, which is proven.

Similarly, the largest code distance will occur when the number of such coincidences is minimally possible, i.e., it will be equal to $N - d\left(x_0 \oplus x_1\right) + d_0$, if $d\left(x_0 \oplus x_1\right) + d_0 \le N$, and – $\left|d\left(x_0 \oplus x_1\right) - d\right|_0$, if $d\left(x_0 \oplus x_1\right) + d_0 > N$. Thus, the number of corresponding binary realizations is also zero for the case when

$$d_1 > \begin{cases} N - d\left(x_0 \oplus x_1\right) + d_0, & if \quad d\left(x_0 \oplus x_1\right) + d_0 \le N; \\ \left|d\left(x_0 \oplus x_1\right) - d_0\right|, & if \quad d\left(x_0 \oplus x_1\right) + d_0 > N. \end{cases}$$

Thus, the code distance d_1, for which the number of corresponding binary realizations is non-zero, takes values in the interval $\left[d\left(x_0 \oplus x_1\right) - d_0; d\left(x_0 \oplus x_1\right) + d_0\right]$, if $d_0 \le N - d\left(x_0 \oplus x_1\right)$ and $d_0 \le d\left(x_0 \oplus x_1\right)$, or $\left[d\left(x_0 \oplus x_1\right) - d_0; d\left(x_0 \oplus x_1\right) + \left(N - d\left(x_0 \oplus x_1\right)\right) - \left(d_0 - \left(N - d\left(x_0 \oplus x_1\right)\right)\right)\right]$, i.e., $\left[d\left(x_0 \oplus x_1\right) - d_0; 2 \cdot N - d\left(x_0 \oplus x_1\right) - d_0\right]$, if $d_0 > N - d\left(x_0 \oplus x_1\right)$ and $d_0 \le d\left(x_0 \oplus x_1\right)$, or $\left[d_0 - d\left(x_0 \oplus x_1\right); d_0 + d\left(x_0 \oplus x_1\right)\right]$, if $d_0 \le N - d\left(x_0 \oplus x_1\right)$ and $d_0 > d\left(x_0 \oplus x_1\right)$, or

$\left[d_0 - d\left(x_0 \oplus x_1\right); \left(N - d\left(x_0 \oplus x_1\right)\right) - \left(d_0 - \left(N - d\left(x_0 \oplus x_1\right)\right)\right) + d\left(x_0 \oplus x_1\right)\right]$,

i.e. $\left[d_0 - d\left(x_0 \oplus x_1\right); 2 \cdot N - d\left(x_0 \oplus x_1\right) - d_0\right]$, if $d_0 > N - d\left(x_0 \oplus x_1\right)$ and $d_0 > d\left(x_0 \oplus x_1\right)$.

Lemma 2. The number of binary realizations located at a code distance d_0 from the binary vector x_0 and d_1 from the binary vector x_1 is zero if $d_1 = \left|d\left(x_0 \oplus x_1\right) - d_0\right| + 2p + 1$, where $p = 0, 1, 2, ...$

Proof. Let the binary vector x_0 – be zero. Then, the vector x_1 contains $d\left(x_0 \oplus x_1\right)$ unit components and $N - d\left(x_0 \oplus x_1\right)$ zero components. Consider the binary realization $x_{\min}$, zero components. Consider the binary realization x_1 $d_1 = \left|d\left(x_0 \oplus x_1\right) - d_0\right|$. It is clear that the nearest binary realization $x_{\min \pm 1}$, which differs from the given one, is characterized by the code distance $d_1 = \left|d\left(x_0 \oplus x_1\right) - d_0\right| \pm 1$. This is achieved by changing the value of one of the components of the given realization to its opposite, which simultaneously increases or decreases the code distance d_0 from $x_{\min\pm1}$ to the binary vector x_0 by one. Thus, the number of binary realizations located at code distance d_0 from the binary vector x_0 and d_1 from the binary vector x_1 is zero if $d_1 = \left|d\left(x_0 \oplus x_1\right) - d_0\right| \pm 1$. Binary realizations $\{x_{\min\pm2}$ characterized by the code distance $d_1 = \left|d\left(x_0 \oplus x_1\right) - d_0\right| \pm 2$, can be formed by changing the value of one of the unit components and one of the zero components of the realization $x_{\min}$ to their opposites without changing the code distance d_0. Applying similar reasoning for any element $\{x_{\min\pm2}\}$ allows us to establish that it is also impossible to form realizations based on them with code distances d_0 and $d_1 = \left|d\left(x_0 \oplus x_1\right) - d_0\right| \pm 3$. Analyzing the code distance interval d_1, previously defined, in this way, we obtain an additional set of values $d_1 = \left|d\left(x_0 \oplus x_1\right) - d_0\right| + 2p + 1$ ($p = 0, 1, ...$), for which the number of corresponding binary realizations is zero.

Let's present the structure of the binary Hamming space graphically (Figure 13).

Figure 13 shows the structure of a ten-dimensional binary space, in which containers (likely referring to Hamming spheres) of two intersecting classes, X_0^o and X_1^o, are reproduced, with radii $d_0^* = 5$, $d_1^* = 4$ respectively, and an intercenter distance $d\left(x_0 \oplus x_1\right) = 7$. In this case, the containers are presented in circles of the corresponding radius, from the centers of which diverge, depicted by a dotted line, circles of increasing radius from zero to ten, which characterize the dimension of the binary space.

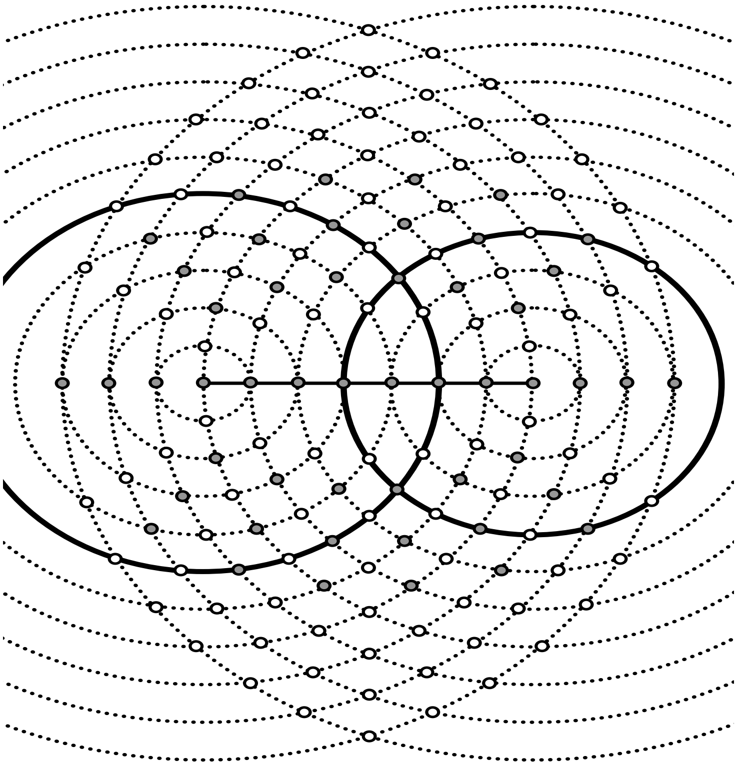

Fig. 13. *Structure of the binary Hamming space*

In Figure 13, the intersection points of these circles are marked with filled or empty circles depending on the number of binary realizations in them.

Lemma 3. The number of binary realizations at a Hamming distance d_0 from a binary vector x_0 and d_1 from a binary vector x_1 from a binary vector

$$n = \frac{d(x_0 \oplus x_1)!}{d_0!(d(x_0 \oplus x_1) - d_0)!},$$

if $d_1 = d(x_0 \oplus x_1) - d_0$ and $d_0 \le d(x_0 \oplus x_1)$.

Proof. Let the binary vector x_0 be the zero vector. Then, x_1 contains $d(x_0 \oplus x_1)$ unit components and $N - d(x_0 \oplus x_1)$ zero components. A binary realization characterized by the code distances from the lemma's condition contains d_0 unit components. The condition $d_1 = d(x_0 \oplus x_1) - d_0$ can be satisfied if and only if the coordinates of these unit components coincide with the coordinates of the unit components of the binary vector x_1. The number of such binary realizations will be equal to the number of combinations that can be formed from d_0 unit and d_1 zero components, i.e.,

$$\frac{(d_0 + d_1)!}{d_0!d_1!}$$

or

$$\frac{d(x_0 \oplus x_1)!}{d_0!(d(x_0 \oplus x_1) - d_0)!} = C(d(x_0 \oplus x_1), d_0).$$

Similarly, under the conditions $d_1 = d_0 - d(x_0 \oplus x_1)$ and $d_0 > d(x_0 \oplus x_1)$, the binary realizations contain $N - d_0$ zero components, whose coordinates coincide with the coordinates of the zero components of the binary vector x_1. The number of such binary realizations will be equal to the number of combinations that can be formed from $N - d_0$ zero and d_1 unit components, i.e.,

$$\frac{(N - d_0 + d_1)!}{(N - d_0)!d_1!},$$

or

$$\frac{(N - d(x_0 \oplus x_1))!}{(N - d(x_0 \oplus x_1) - d_1)!d_1!} = C(N - d(x_0 \oplus x_1), d_1).$$

Theorem. The number of binary realizations at a code distance d_0 from a binary vector x_0 and d_1 from a binary vector x_1 is equal to

$$n = \frac{d(x_0 \oplus x_1)!}{(d_0 - p)!(d(x_0 \oplus x_1) - d_0 + p)!} \cdot \frac{(N - d(x_0 \oplus x_1))!}{(N - d(x_0 \oplus x_1) - p)!p!},$$

where $p = \dfrac{d_1 + d_0 - d\left(x_0 \oplus x_1\right)}{2}$.

Proof. Let the binary vector x_0 be the zero vector. Then, x_1 contains $d\left(x_0 \oplus x_1\right)$ unit components and $N - d\left(x_0 \oplus x_1\right)$ zero components. Consider the binary realizations of the formation mechanism presented in Lemma 2. For realizations whose number is non-zero, the condition $d_1 = \left|d\left(x_0 \oplus x_1\right) - d_0\right| + 2p$, where $p =$ 0, 1, 2, ..., is possible if and only if the coordinates of $d_0 - p$ unit components simultaneously coincide with the coordinates of the unit components of the binary vector x_1 and the coordinates of p zero components coincide with the coordinates of the zero components of the binary vector x_1, if $d_0 \le d\left(x_0 \oplus x_1\right)$; or when simultaneously the coordinates of p zero components coincide with the coordinates of the unit components of the binary vector x_1 and the coordinates of $d_0 - d\left(x_0 \oplus x_1\right) + p$ unit components coincide with the coordinates of the unit components of the binary vector x_1, if $d_0 > d\left(x_0 \oplus x_1\right)$. The number of such binary realizations will be equal to

$$\begin{cases} \dfrac{d\left(x_0 \oplus x_1\right)!}{\left(d_0 - p\right)!\left(d\left(x_0 \oplus x_1\right) - d_0 + p\right)!} \cdot \dfrac{\left(N - d\left(x_0 \oplus x_1\right)\right)!}{\left(N - d\left(x_0 \oplus x_1\right) - p\right)!\,p!}, & if \quad d_0 \le d\left(x_0 \oplus x_1\right), \\ \dfrac{d\left(x_0 \oplus x_1\right)!}{\left(d\left(x_0 \oplus x_1\right) - p\right)!\,p!} \cdot \dfrac{\left(N - d\left(x_0 \oplus x_1\right)\right)!}{\left(N - d_0 - p\right)!\left(d_0 - d\left(x_0 \oplus x_1\right) + p\right)!}, & if \quad d_0 > d\left(x_0 \oplus x_1\right), \end{cases}$$

where

$$p = \frac{d_1 - \left|d\left(x_0 \oplus x_1\right) - d_0\right|}{2} = \begin{cases} \dfrac{d_1 - d\left(x_0 \oplus x_1\right) + d_0}{2}, & if \quad d_0 \le d\left(x_0 \oplus x_1\right), \\ \dfrac{d_1 + d\left(x_0 \oplus x_1\right) - d_0}{2}, & if \quad d_0 > d\left(x_0 \oplus x_1\right). \end{cases}$$

Thus, the number of binary realizations characterized by the code distances according to the conditions of the theorem is equal to

$$\frac{d\left(x_0 \oplus x_1\right)!}{\left(d_0 - p\right)!\left(d\left(x_0 \oplus x_1\right) - d_0 + p\right)!} \cdot \frac{\left(N - d\left(x_0 \oplus x_1\right)\right)!}{\left(N - d\left(x_0 \oplus x_1\right) - p\right)!\,p!},$$

where $p = \dfrac{d_1 + d_0 - d\left(x_0 \oplus x_1\right)}{2}$.

The distribution of realizations in the pseudo-sphere of the Hamming space with a radius equal to 10 code units is given in Table 1.

Table 1. *Distribution of image realizations in the pseudo-sphere of the Hamming space*

d_0 \ d_1	0	1	2	3	4	5	6	7	8	9	10
0	0	0	0	0	0	0	0	C_7^0	0	0	0
1	0	0	0	0	0	0	C_7^1	0	C_3^1	0	0
2	0	0	0	0	0	C_7^2	0	$C_7^1\,C_3^1$	0	C_3^2	0
3	0	0	0	0	C_7^3	0	$C_7^2\,C_3^1$	0	$C_7^1\,C_3^2$	0	C_3^3
4	0	0	0	C_7^4	0	$C_7^3\,C_3^1$	0	$C_7^2\,C_3^2$	0	$C_7^1\,C_3^3$	0
5	0	0	C_7^5	0	$C_7^4\,C_3^1$	0	$C_7^3\,C_3^2$	0	$C_7^2\,C_3^3$	0	0
6	0	C_7^6	0	$C_7^5\,C_3^1$	0	$C_7^4\,C_3^2$	0	$C_7^3\,C_3^3$	0	0	0
7	C_7^7	0	$C_7^6\,C_3^1$	0	$C_7^5\,C_3^2$	0	$C_7^4\,C_3^3$	0	0	0	0
8	0	$C_7^7\,C_3^1$	0	$C_7^6\,C_3^2$	0	$C_7^5\,C_3^3$	0	0	0	0	0
9	0	0	$C_7^7\,C_3^2$	0	$C_7^6\,C_3^3$	0	0	0	0	0	0
10	0	0	0	$C_7^7\,C_3^3$	0	0	0	0	0	0	0

Analysis of Table 1 shows that the Hamming space is not uniform. Furthermore, considering the property of combinations $C_N^{N-i} = C_N^i$, where $0 \le i \le N$, it can be stated that the structure of the binary space is symmetrical to the main and secondary diagonals of the table.

A detailed structure of the Hamming space, presented in Figure 9, is shown in Figure 14.

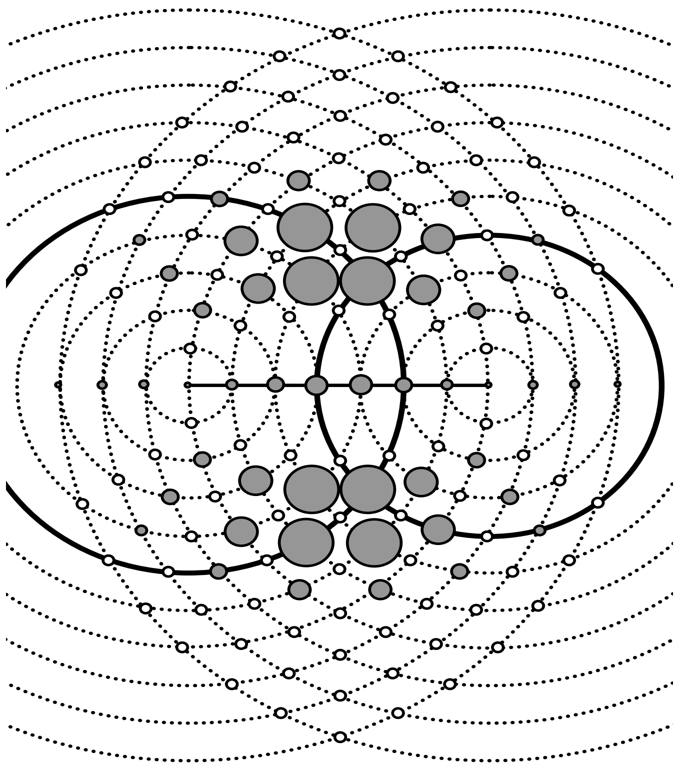

Fig. 14. *Detailed Structure of the Hamming Space*

Figure 14 shows the structure of a ten-dimensional binary space in which the intersecting containers of two classes, X_0^o and X_1^o, are depicted. The diameter of the filled circles corresponds to the number of binary realizations characterized by the respective distances from the centers of the depicted containers.

Let the number of binary realizations at a code distance d_0 from a binary vector x_0 and d_1 from a binary vector x_1 be denoted as $B^N_{x_0(d_0),x_1(d_1)}$. Then, the number of binary realizations that simultaneously belong to the container of class X_0^o with radius d_0^* ($0 \le d_0^* \le N$) and the container of class X_1^o with radius d_1^* ($0 \le d_1^* \le N$) is equal to

$$B^N_{\tilde{\mathfrak{R}}^{|2|}\left(X_0^o, X_1^o\right)} = \sum_{i=0}^{d_0^*} \sum_{j=0}^{d_1^*} B^N_{x_0(i),x_1(j)}.$$

Thus, the detailed analysis of the class partitioning structure in the Hamming space demonstrates the implicativeness and symmetry of the distribution of class vector-realizations within containers built on the radial basis of the feature space.

This chapter of the monograph, within the framework of IEI technology, presents the foundational principles of informational analysis and synthesis regarding the functioning of a self-learning decision support system capable of operating in factor cluster analysis mode with the optimization of the feature dictionary under the conditions of fuzzy data and resource constraints. The complex of logically interconnected categorical mathematical models within the IEI technology enables the analysis and synthesis of algorithms governing the functioning of self-learning decision support system in factor cluster analysis mode. The modifications of informational criteria developed within the IEI technology serve as general criteria for the evaluation of functional efficiency, as they characterize both the accuracy-related and geometric (distance-based) parameters of the decision rules of the DSS, which reconstructs containers in the radial basis of the recognition feature space during the learning process. The optimization of the feature dictionary is implemented through an iterative, parameter-structured, multi-cyclic procedure for seeking the maximum of the objective function, which is computed at the *i*-th step of the decision support system learning and encompasses both the informational criterion of functional efficiency and supplementary conditions that are characteristic of the corresponding feature selection algorithm. It is demonstrated that the structure of class partitioning in the Hamming space is characterized by implicatively and symmetry in the distribution of class realization vectors within the containers constructed in the radial basis of the feature space.

REFERENCES

1. Karabegović I., Banjanović-Mehmedović L. Industrial Robots: Design, Applications and Technology. New York : Nova Science Publishers, Incorporated, 2020. 463 p.
2. Korobiichuk I., Ladaniuk A. Technological Process. *Mechatronics 2019: Recent Advances Towards Industry 4.0.* 2019. Vol. 1044. P. 214–221. URL: https://doi.org/10.1007/978-3-030-29993-4_27.
3. Machine Training of the System of Functional Diagnostics of the Shaft Lifting / A. Dovbysh et al. *Problems of the Regional Energetics*. 2019. Vol. 2. P. 88–102. URL: https://doi.org/10.5281/zenodo.3367060.
4. Resilience and Resilient Systems of Artificial Intelligence: Taxonomy, Models and Methods / V. Moskalenko et al. *Algorithms*. 2023. Vol. 16, no. 3. P. 165. URL: https://doi.org/10.3390/a16030165.
5. A El-Yacoubi M., Vincent N., Kurtz C. Emerging Topics in Pattern Recognition and Artificial Intelligence. World Scientific Publishing Co Pte Ltd, 2024.
6. Ekman M. Learning Deep Learning: Theory and Practice of Neural Networks, Computer Vision, Natural Language Processing, and Transformers Using TensorFlow. Pearson Education, Limited, 2021. 752 p.

7. Migliaccio A., Iannone G. Systems Engineering Neural Networks. Wiley & Sons, Incorporated, John, 2023. 243 p.
8. Anastassiou G. A. Parametrized, Deformed and General Neural Networks. Cham : Springer Nature Switzerland, 2023. 854 p. URL: https://doi.org/10.1007/978-3-031-43021-3.
9. Sevakula R. K., Verma N. K. Improving Classifier Generalization: Real-Time Machine Learning Based Applications. Springer, 2022. 181 p.
10. Dyshel M., Lane H. Natural Language Processing in Action, Second Edition. Manning Publications Co. LLC, 2022. 550 p.
11. Bellanger M. Digital Signal Processing: Theory and Practice. Wiley & Sons, Incorporated, John, 2024. 397 p.
12. Sen J., Mehtab S. Machine Learning in the Analysis and Forecasting of Financial Time Series. Cambridge Scholars Publishing, 2023. 384 p.
13. Data-Driven Systems and Intelligent Applications / M. M. Ghonge et al. Boca Raton : CRC Press, 2024. 197 p. URL: https://doi.org/10.1201/9781003388449.
14. Recent Developments in Model-Based and Data-Driven Methods for Advanced Control and Diagnosis / ed. by D. Theilliol, J. Korbicz, J. Kacprzyk. Cham : Springer Nature Switzerland, 2023. 351 p. URL: https://doi.org/10.1007/978-3-031-27540-1.
15. Artificial Intelligence for Communications and Networks / ed. by X. Wang et al. Cham : Springer International Publishing, 2021. URL: https://doi.org/10.1007/978-3-030-90196-7.
16. Chalouf M.-A. Intelligent Security Management and Control in the IoT. Wiley & Sons, Incorporated, John, 2022. 300 p.
17. Intelligent Cybersecurity and Resilience for Critical Industries: Challenges and Applications / M. Zaydi et al. New York : River Publishers, 2025. 360 p. URL: https://doi.org/10.1201/9788770047746
18. Bishop C. M., Bishop H. Deep Learning: Foundations and Concepts. Springer International Publishing AG, 2023. 656 p.
19. New Horizons for Fuzzy Logic, Neural Networks and Metaheuristics / ed. by O. Castillo, P. Melin. Cham : Springer Nature Switzerland, 2024. 422 p. URL: https://doi.org/10.1007/978-3-031-55684-5.
20. Zhao G., Cui H. Cooperative Control of Multi-Agent Systems: A Hybrid System Approach. Springer, 2024. 251 p.
21. Lauer F., Bloch G. Hybrid System Identification: Theory and Algorithms for Learning Switching Models. Springer, 2019. 274 p.
22. Cohen M., Belta C. Adaptive and Learning-Based Control of Safety-Critical Systems. Cham : Springer International Publishing, 2023. 208 p. URL: https://doi.org/10.1007/978-3-031-29310-8.
23. Optimization and Computing using Intelligent Data-Driven Approaches for Decision-Making / I. Ali et al. Boca Raton : CRC Press, 2024. 228 p. URL: https://doi.org/10.1201/9781003536796.
24. Bandyopadhyay S. Decision Support System: Tools and Techniques. Taylor & Francis Group, 2023. 395 p.
25. Ma Z., Zou S. Optimal Control Theory: The Variational Method. Springer Singapore Pte. Limited, 2022. 363 p.
26. Model Order Reduction and Applications / M. Hinze et al. ; ed. by M. Falcone, G. Rozza. Cham : Springer Nature Switzerland, 2023. URL: https://doi.org/10.1007/978-3-031-29563-8
27. Vitányi P. M. B. An Introduction to Kolmogorov Complexity and Its Applications (Texts in Computer Science). Springer, 2019. 857 p.
28. Härdle W. K., Simar L. Applied Multivariate Statistical Analysis. 6th ed. Springer, 2024. 628 p.
29. Dathan B., Ramnath S. Object-Oriented Analysis, Design and Implementation. Cham : Springer Nature Switzerland, 2025. 478 p. URL: https://doi.org/10.1007/978-3-031-71240-1.
30. Kifle B. M. Artificial Intelligence: A Systems Approach from Architecture Principles to Deployment. MIT Press, 2024.
31. Stadler D., Hagmanns R. Pattern Recognition: Introduction, Features, Classifiers and Principles. de Gruyter GmbH, Walter, 2024. 348 p.
32. Sunoj S. M., G R., Nair N. U. Reliability Modelling with Information Measures. Taylor & Francis Group, 2022. 355 p.
33. L. Poort, L. A. L. Janssen, B. Besselink, R. H. B. Fey and N. V. D. Wouw, Abstracted Model Reduction: A General Framework for Efficient Interconnected System Reduction. IEEE Transactions on Control Systems Technology. 2025. P. 1-16. https://doi.org/10.1109/TCST.2025.3550027