Combinatorial Analysis of the Solvability Properties of the Problems of Recognition and Completeness of Algorithmic Models. Part 2: Metric Approach within the Framework of the Theory of Classification of Feature Values

I. Yu. Torshin[a,*] and K. V. Rudakov[a,b,]**

[a]*Moscow Institute of Physics and Technology, Institutskii per. 9, Dolgoprudny, Moscow oblast, 141700 Russia*
[b]*Dorodnicyn Computing Centre, Federal Research Center "Informatics and Control," Russian Academy of Sciences, ul. Vavilova 40, Moscow, 119333 Russia*
*e-mail: tiy1357@yandex.ru
**e-mail: rudakov@ccas.ru

Abstract—The properties of solvability/regularity of problems and correctness/completeness of algorithmic models are fundamental components of the algebraic approach to pattern recognition. In this paper, we formulate the principles of the metric approach to the data analysis of poorly formalized problems and hence with obtain metric forms of the criteria of solvability, regularity, correctness, and completeness. In particular, the analysis of the compactness properties of metric configurations allowed us to obtain a set of sufficient conditions for the existence of correct algorithms. These conditions can be used for assessment of the quality of the methods of formalization of the problems for arbitrary algorithms and algorithmic models. The general schema proposed for the data analysis of poorly formalized problems includes the criteria in the cross-validation form and can assess not only the quality of formalization, but also the extent of overtraining pertaining to the procedures of generation and selection of feature descriptions.

Keywords: algebraic approach, metric analysis, theory of classification of feature values, compact metric spaces, clustering, combinatorial theory of solvability

DOI: 10.1134/S1054661817020110

1. INTRODUCTION

An adequate formalization of problems (i.e., generation of feature descriptions and of classes of objects) is a necessary condition for the application of the constructs of the algebraic approach to recognition [1−8] as well as of the rest of the methods of intellectual data analysis [9−15]. Formalization consists in defining a function $\varphi : X \rightarrow Q$ that transforms the set of initial descriptions of objects X into a set of precedents Q. A one-to-one correspondence between the elements of the sets X and Q corresponds to the satisfiability of the *strong form of the axiom of correspondence* [9].

Within the present series of papers, we carry out an analysis of poorly formalized problems by two complementary approaches: *factorization*, i.e., reduction of a problem to a binary form and *metrization*, i.e., introduction of metrics in the feature sets and feature descriptions of objects. In [12], on the basis of the factorization approach, we obtained combinatorial criteria of *solvability* and *regularity* of recognition problems, of *correctness* of algorithms, and of *completeness* of algorithmic models.

Suppose given a formalization method φ, a set of initial information I_i, a set of final information I_f, a space of "admissible" descriptions of objects $J_{\mathrm{ob}} \subseteq I_i \times I_f$, and a set of precedents $\varphi(a) \subseteq J_{\mathrm{ob}}$. The set $\varphi(a)$ consists of objects $q_i = (m_i, \iota_i)$, where m_i corresponds to the feature values $q_i[1], \ldots, q_i[k], \ldots, q_i[n]$, $q_i[k] \in I_k$ (I_k is a set of values of the k-th feature), and ι_i contains information $q_i[n+1], \ldots, q_i[n+l]$ on the membership of objects in each of the l classes. Then, for a given feature selection mask $\chi = (\gamma_1, \ldots, \gamma_k, \ldots, \gamma_n)$, $\gamma_k \in [0,1]$, and a factorization method $\{\delta_k(v_1, v_2) : I_k^2 \rightarrow [0,1], k = 1, \ldots, n\}$ (verification of the membership of the values v_1 and v_2 of the kth feature in the same equivalence class of feature values), we formulated the solvability and regularity criteria for the corresponding problem [12]:

- *solvability criterion of problem $Z(\varphi(a))$:*

$$(1) \quad \underset{\varphi(a)}{\forall} \, q_1, q_2 : \iota_1 \neq \iota_2 \Rightarrow \underset{1..n}{\exists} \, k : \neg \delta_k(q_1[k], \, q_2[k]) \wedge \gamma_k = 1.$$

- *regularity criterion of problem $Z(\varphi(a))$:*

(2) $\underset{\varphi(a)}{\forall} q_1, q_2 : \underset{1..n}{\exists} k : \neg\delta_k(q_1[k], q_2[k]) \wedge \gamma_k = 1.$

For a given algorithm $A_h : I_i \rightarrow I_f$ and a given vector of parameters $\theta \in R^{n'}$ of the algorithm, which reflects the "internal settings" of the algorithm, we formulated the *correctness criterion of the algorithm* [12]:

(3) $A_h(m_i, \theta, \chi): \underset{\varphi(a)}{\forall} (m_i, \iota_i) : A_h(m_i, \theta, \chi) = \iota_i.$

For a given regular sampling operator $\hat{\zeta}_r$, $\hat{\zeta}_r X = \{a_1, a_2, ..., a_k, ... a_{|\hat{\zeta}X|} | a_k \subset X, \quad Z(\varphi(a_k))$ is regular$\}$, and a parametric algorithmic model $M_A[\hat{\Theta}] = \{(A_h(\theta_h), \theta_h), A_h(\theta_h) | A_h(\theta_h) : I_i \rightarrow I_f\}$ ($\hat{\Theta}$ is a method of calculating the vectors of parameters θ_h that is used in the given model), we formulated a *completeness criterion of an algorithmic model* [12]:

(4) $\underset{\hat{\zeta}_r X}{\forall} a \underset{M_A[\hat{\Theta}]}{\exists} A_h(\theta_h, \chi_h): (\underset{\varphi(a)}{\forall}(m_i, \iota_i) : A_h(m_i, \theta_h, \chi_h) = \iota_i).$

For the above-described criteria, we obtained appropriate combinatorial functionals $r_1(\varphi(a), \chi)$, $r_2(\varphi(a), \chi)$, $r_3(\varphi(a), A_h)$, and $r_4(\hat{\zeta}_r X, M_A[\hat{\Theta}])$, which characterize the "degree" of satisfiability of the criteria for specific $\varphi(a)$, χ, A_h, $\hat{\zeta}_r X$, and $M_A[\hat{\Theta}]$.

We also obtained cross-validation forms of the criteria, which include testing over a set of samples $\hat{\zeta} X$ (in the case of the completeness criterion of models, the testing is performed over a regular $\hat{\zeta}_r X$):

• *solvability criterion of problems* $Z(\varphi(a))$, $a \in \hat{\zeta} X$:

(1.1) $\underset{\hat{\zeta}X}{\forall} a, b, a \neq b \underset{\varphi(a)}{\forall} q_1, q_2 : \iota_1 \neq \iota_2 \quad \Rightarrow$ $\underset{1..n}{\exists} k : \neg\delta_k(q_1[k], q_2[k]) \wedge \gamma_k(\varphi(b)) = 1,$

• *regularity criterion of problems* $Z(\varphi(a))$, $a \in \hat{\zeta} X$:

(2.1) $\underset{\hat{\zeta}X}{\forall} a, b, a \neq b \underset{\varphi(a)}{\forall} q_1, \quad q_2 : \underset{1..n}{\exists} k : \neg\delta_k(q_1[k],$ $q_2[k]) \wedge \gamma_k = 1,$

• *correctness criterion of algorithm* A_h:

(3.1) $\underset{\hat{\zeta}X}{\forall} a \neq b \underset{\varphi(a)}{\forall} (m_i, \iota_i) : A_h(m_i(\varphi(a)), \quad \hat{\Theta}\varphi(b),$ $\chi(\varphi(b))) = \iota_i,$

• *completeness criterion of a homogeneous parametric algorithmic model* $M_A[\hat{\Theta}]$:

(4.1) $\underset{\hat{\zeta}_r X}{\forall} a \underset{\hat{\zeta}_r X}{\exists} b \neq a | \underset{M_A[\hat{\Theta}]}{\exists} A_h : (\underset{\varphi(a)}{\forall}(m_i, \iota_i) : A_h(m_i,$ $\hat{\Theta}\varphi(b), \chi(\varphi(b))) = \iota_i).$

The homogeneity of a model $M_A[\hat{\Theta}]$ consists in the equality of the range of values and the dimension of the vector of parameters θ_h, each of which is calculated by a unique method $\hat{\Theta}$ by some set of precedents. The expressions "$\theta_h = \hat{\Theta}\varphi(b)$" and "$\chi_h = \chi(\varphi(b))$"

indicate that the vector of parameters and the feature selection mask are calculated on the basis of the set of precedents $\varphi(b)$ and in fact correspond to the "learning" of the algorithm $A_h(\theta_h, \chi_h)$.

Criteria (1)–(4) and (1.1)–(4.1) make it possible to formulate various methods for the calculation of feature selection masks (for example, on the basis of the dead-end property of masks with respect to the solvability/regularity criteria [13–15]) on the basis of one or another set of precedents $\varphi(a)$. The masks $\chi(\varphi(a))$, $\chi(\varphi(a)) = (\gamma_1(\varphi(a)), ..., \gamma_k(\varphi(a)), ..., \gamma_n(\varphi(a)))$, $\gamma_k \in [0,1]$, thus obtained imply, first, that the features are ordered by "informativity" (so that the smaller values of index k correspond to the greater informativity) and, second, that there exist two classes of features: informative ($\gamma_k = 1$) and noninformative ($\gamma_k = 0$) ones.

Therefore, the results of [13–15] can be considered as ingredients of a theory pertaining to *classification of feature values,* which, in fact, has been developed during the analysis of poorly formalized problems [9–12]. Within the framework of such a theory, each feature takes a finite set of values on any finite sample of objects. These values of features have both the initial description, which is represented in the set X (the statement of the problem in the problem domain), and a description in the form of a subset of objects of the sample under test and a subset of Cartesian products of subsamples of objects from pairwise different classes.

Within this theory, methods for calculating masks $\chi()$ on the basis of the set of precedents $\phi(a)$ represent a special variant of algorithms, which, just as the sought algorithms $A_h(\theta_h, \chi_h)$, can also be characterized from the viewpoint of the "accuracy of classification," "generalizing ability," etc. Therefore, the combinatorial functionals $r_{1c}(\hat{\zeta}X)$, $r_{2c}(\hat{\zeta}X)$, and others obtained in [12] and corresponding to the above cross-validation forms of the criteria of algebraic approach indeed characterize the "overfitting" of the algorithms for calculating the masks $\chi()$ (i.e., feature selection procedures). In the present study, we formulate the principles of the metric approach to the analysis of the satisfiability of the criteria of the algebraic approach to pattern recognition.

2. THE PRINCIPLES OF THE METRIC APPROACH TO THE ANALYSIS OF POORLY FORMALIZED PROBLEMS

The experimental and cross-validation analysis of the above-described criteria and the practical investigation of poorly formalized problems within the factorization approach requires that a few preliminary steps should be made. First, one should define a formalization method ϕ, which is essentially required for constructing the sets of precedents on the basis of the given set of samples $\hat{\zeta} X$. Second, one should introduce

some methods $\delta_k()$ that define equivalence classes of the feature values. Third, one should define methods for estimating the informativity of the features.

An alternative approach to factorization is *the metric approach to the data analysis of poorly formalized problems*, which is based on the introduction of certain metrics over the sets of features and over the sets of feature descriptions of objects (i.e., metrization of a problem).

The metric approach significantly differs from the factorization approach. On any finite sample of objects, an arbitrary feature takes a finite set of values each of which can be considered as a binary (boolean) feature. Naturally, such a trivial factorization allows the application of the solvability/regularity and of the correctness/completeness criteria. However, the number of binary features in this case significantly increases with the size of the samples. Moreover, the binary features thus obtained turn out to be essentially interdependent. As a result, one can expect that the indices of "accuracy" and "generalizing ability" of the algorithms that perform the classification of the values of such features will be much lower than it would be expected.

Therefore, the metric approach (within which the concepts of compactness [10] and density [11] of sets of points in a metric space and of classes as certain "compact" sets of points are introduced) seems to be more promising. The metric approach allows one to take into account the interdependence of the feature descriptions and to identify the groups ("condensations") of feature descriptions, which, in a sense, are more "informative" than some individual features or feature values.

The metric approach suggests the construction and the analysis of the properties of conjugate pairs of metrics (ρ_π, ρ_q), where ρ_π is a metric on the space of features and ρ_q is a metric on the space of objects. Depending on the goals of the study, this pair of metrics can be defined either on the basis of initial feature descriptions of objects in a set X or on the basis of feature descriptions in the already constructed set of precedents $\varphi(a)$, $a \in \hat{\zeta}X$. Note that, in practice, it is expedient to define the metric ρ_q as a function of ρ_π whereas the definition of ρ_π in terms of ρ_q is of less practical interest.

The metrics (ρ_π, ρ_q) can be defined by purely heuristic methods (for example, by introducing some linear space on the vectors of features, some probability measures, and so on). When introducing heuristic metric functions, the latter may turn out to be semimetrics ($\rho_\pi \geq 0$, $\rho_q \geq 0$). The formalism developed for the analysis of metric configurations can also be applied to both metrics and semimetrics.

In one of the previous papers [9], during the analysis of the formalization of problems, the authors obtained a theoretical substantiation for introducing the metric ρ_π. Therefore, below, by a "metric" we mean precisely ρ_π defined on the space of features (unless it is stated that the discourse concerns the metric ρ_q).

3. ON THE GENERATION OF "DERIVATIVE" FEATURES OVER THE SET OF INITIAL DESCRIPTIONS OF FEATURES

In [9], we have shown that the presence of feature descriptions of objects in the set of initial descriptions X is equivalent to the definition of a system of unordered subsets of X, $\pi(X) = \{\varnothing, a_1, a_2, \ldots, a_i, \ldots, a_j, \ldots, |a_i \subset X, a_i \neq a_j\}$. Assuming that all subsets $\pi(X)$ are either "open" or "clopen," *the topology of feature descriptions* $\tau(X)$ corresponding to X is defined as $\tau(X) = \{\breve{b} | b \subseteq \underline{\tau}(X)\}$, where $\underline{\tau}(X) = \{\hat{a} | a \subseteq \pi(X)\}$, "˘" denotes the union of elements of the set of subsets $\pi(X)$, and "ˆ" is the intersection of elements. If the set $\pi(X)$ satisfies the necessary and sufficient condition for the existence of the topology $\tau(X)$ over the set X, $\breve{\pi}(X) = X$, then $\pi(X)$ represents a *prebase* of $\tau(X)$, and the set $\underline{\tau}(X)$ is a *base* of $\tau(X)$.

The elements of $\tau(X)$ may be subsets of each other, so that a natural ordering by inclusion arises between these subsets, that corresponds to a partial order relation defining the existence of a corresponding *lattice* $L(\tau(X))$ each element $\upsilon \in L(\tau(X))$ of which uniquely corresponds to a certain element $u \in \tau(X)$.

Numerical features are arranged along the maximal chains of a boolean lattice, so that the "motion" from the minimal element of the chain to the maximal element of the chain corresponds to enumeration of the values of the numerical feature, from minimal to maximal value. *Quantiles of a numerical feature* correspond to subchains of fixed length. *Binary (boolean) features* correspond to bundles of maximal chains, each of which includes three nodes of the lattice: the zero element $\varnothing$, a corresponding subset of X (the "central node"), and the identity element I. "*Categorial*" features correspond to *antichains* of a boolean lattice.

When introducing an isotonic (monotonic) estimate $v[]$ in a boolean lattice $L(\tau(X))$ ($v : L(X) \to R^+$, $\forall a, b : v[a] + v[b] = v[a \wedge b] + v[a \vee b]$), the function $\rho_\pi(x, y) = v[x \vee y] - v[x \wedge y]$ is a metric that forms a metric space $M(\tau(X)) = (\tau(X), \rho_\pi(x, y))$. A *subspace* M' of a metric space $M(\tau(X))$, $M' \subseteq M(\tau(X))$, is a metric space constructed over the corresponding subset of the topology $\tau(X)$; i.e., $\tau' \subseteq \tau(X)$ satisfies $M' = M(\tau') = (\tau', \rho_\pi(x, y)) \subseteq M(\tau(X)) = (\tau(X), \rho_\pi(x, y))$.

When the strong form of the axiom of correspondence is satisfied, we say that the set of initial descriptions of objects X is *regular* if the corresponding set of precedents Q is regular (see the Introduction).

Theorem 1. *For a regular X and under the strong form of the axiom of correspondence, the topology $\tau(X)$, the lattice $L(\tau(X))$, and the metric space $M(\tau(X))$ are isomorphic to each other and contain subsystems that are isomorphic to an arbitrary prebase $\pi(X)$.*

Proof. The isomorphism between $\tau(X)$ and $L(\tau(X))$, $\tau(X)$ and $M(\tau(X))$, and $L(\tau(X))$ and $M(\tau(X))$ is obvious by construction, since both the lattice and the metric space are constructed over elements of one and the same topology $\tau(X)$. In addition, the values of the metric for an arbitrary pair of points in $M(\tau(X))$ are calculated on the basis of pairs of elements. For a regular X, the topology $\tau(X)$ is a discrete topological space, and the lattice $L(\tau(X))$ isomorphic to $\tau(X)$ is an atomically generated boolean lattice (see Theorems 2 and 3 in [9]). In this case, the set $\tau(X)$ contains an arbitrary subset of X. Hence, any element of the system of unordered subsets $\pi(X)$ is uniquely assignable to the corresponding element of $\tau(X)$ (since $\pi(X) \subseteq \tau(X)$). Since the sets $\tau(X)$ are pairwise different for regular X, each element of a subset of $\tau(X)$ corresponding to the mapping is also uniquely assignable to a certain element of $\pi(X)$. Accordingly, there exists a sublattice in $L(\tau(X))$ and a subspace in $M(\tau(X))$, related by an isomorphism to a prebase of $\pi(X)$. The theorem is proved.

Corollary 1. The topology $\tau(X)$ *and the lattice* $L(\tau(X))$ *are isomorphic, so that the bijective operator* $\hat{\lambda} : 2^{\tau(X)} \to 2^{L(\tau(X))}$ *is defined.*

Since the subsets of the set X are elements of the topology $\tau(X)$ and the lattice $L(\tau(X))$, $\hat{\lambda}$ can be considered as the identity operator.

Corollary 2. *There exists a bijection between* $L(\tau(X))$ *and* $M(\tau(X))$, *so that the operator* $\hat{\kappa}$, $\hat{\kappa} : 2^{L(\tau(X))} \to 2^{M(\tau(X))}$, *and its inverse,* $\hat{\kappa}^{-1}$, *are defined.*

Corollary 3. *Suppose that, for the kth feature with the range of values* I_k, *the injective operator* $\hat{\varsigma}_k : I_k \to L(\tau(X))$ *is defined. Under the hypothesis of the theorem, to each object from X, there corresponds an ordered subset of elements of the lattice* $L(\tau(X))$, *the intersection of all nonempty elements of this set containing a single object.*

Under the hypothesis of the theorem, each element of the set of initial descriptions X uniquely corresponds to a certain element of the set Q; i.e., an object in the set $x_i \in X$ corresponds to a vector of heterogeneous feature descriptions $(q_i[k])$, $k = 1, ..., n$. Take an arbitrary, ith, object from a regular X. The operator $\hat{\varsigma}_k$ maps every value of the kth feature of the ith object, $q_i[k]$, to the corresponding element of $L(\tau(X))$, $\hat{\varsigma}_k q_i[k]$, so that the initial description $x_i \in X$ corresponds to the ordered set $\varsigma_i = (\hat{\varsigma}_k q_i[k])$. If the set

$\varsigma_i = \mathbf{I}\hat{\varsigma}_k q_i[k]$, $\hat{\varsigma}_k q_i[k] \neq \varnothing$ contains more than one object, then this contradicts the condition of regularity of X. If $\hat{\varsigma}_i$ is empty, this corresponds to $x_i \notin X$, which is certainly false.

Thus, the regularity of X is the most important condition imposed on the set of initial descriptions since the regularity guarantees the existence of an isomorphism between $\pi(X)$ and the corresponding subsystems $\tau(X)$ and $L(\tau(X))$ and, hence, the applicability of the constructions of the formalism developed.

An important corollary to Theorem 6 from [9] is the fact that, under the substitution of the height of a lattice element for $\nu[\,]$, we obtain Frechet−Nikodym metric. The metrics over elements of the lattice $L(\tau(X))$ can also be introduced heuristically: for example, by the analysis of "interactions" between different types of features (numerical−numerical, binary−numerical, and so on) [9].

The metric space $M(\tau(X)) = (\tau(X), \rho_\pi)$ contains points corresponding to specific values of features and to all possible combinations of feature values represented in the descriptions of objects from X. Under the regularity of X, $\tau(X)$, $L(\tau(X))$, and $M(\tau(X))$ also contain elements corresponding to an arbitrary classification problem with two classes C^+ and C^- (since lattice $L(\tau(X))$ is boolean and topology $\tau(X)$ represents a discrete topological space). Naturally, the metric ρ_π allows one to calculate the distances between the points corresponding to the classes of objects and the points corresponding to some combinations of feature values.

The earlier developed methods of analysis of the properties of compactness [10] and of point density [11] in metric spaces that arise during the formalization of recognition problems can be used when certain *metric configurations* (ρ-configurations, symmetric matrices of pairwise distances (ρ_{ij}) are defined). Note that, in [9−11], ρ-configurations are also represented as triangular semimatrices of distances, which are elements of the corresponding metric cone.

According to Theorem 1 and its corollaries, the definition of a certain subset $\psi \subseteq L(\tau(X))$, $|\psi| = n_0$, corresponds to the choice of a metric subspace $M(\psi) \cong M(\tau(X))$ isomorphic to a certain element of the space of matrices $\mathbf{P}^{n_0} = R^{n_0} \times R^{n_0} \times ... \times R^{n_0} = (R^{n_0})^{n_0}$. The isomorphism between $M(\psi)$ and the ρ-configuration $(\rho_{ij}) \in \mathbf{P}^{n_0}$ suggests the existence of the inverse $\hat{\rho}_{n_0}^{-1}$ of the operator $\hat{\rho}_{n_0} : \{M(\psi) | \psi \subseteq L(\tau(X)), |\psi| = n_0\} \to \mathbf{P}^{n_0}$ so that a *metric configuration* $\rho(\psi) = \hat{\rho}_{n_0} M(\psi)$ is defined. Denote the ith row of the matrix $\rho(\psi)$ as $\rho[i]$ and the distance between the ith and jth points of the ρ-configuration $\rho(\psi)$ as ρ_{ij}, or as $\rho(i, j)$.

The set ψ generating the ρ-configuration $\boldsymbol{\rho}(\psi)$ may contain elements of $L(\tau(X))$ corresponding to a *prebase* of $\tau(X)$ (the "initial" set $\pi(X)$), to the *base* of $\tau(X)$ (the set $\underline{\tau}(X)$), to some subsets of the "most informative" values of features, etc. In the analysis of a certain recognition/classification problem, ψ also contains elements of $L(\tau(X))$ that correspond to the classes C^+ and C^-.

Theorem 2. *For a regular X, the set of mappings $\bar{\Omega} = \{\hat{\varpi} : 2^{M(\tau(X))} \to M(\tau(X))\}$ is nonempty.*

Proof. Elements of the lattice $L(\tau(X))$ are given by subsets of the set X, which are also elements of the topology $\tau(X)$. For a regular X, $\tau(X)$ forms a discrete topological space and contains any subset of X, and there exists a bijection between the elements of $\tau(X)$ and the boolean lattice $L(\tau(X))$. Accordingly, an arbitrary subset $\psi \subseteq L(\tau(X))$ uniquely corresponds to a set of subsets of X which are also represented by elements of the lattice $L(\tau(X))$ and by points of $M(\tau(X))$, such that $\bar{\Omega} = \{\hat{\varpi} : \{\hat{\kappa}\psi \,|\, \psi \in L(\tau(X)), \hat{\kappa}\psi \,\tilde{\subseteq}\, M(\tau(X))\} \to \hat{\kappa}L(\tau(X))\}$. Introduce an *operation* $\hat{C}$ for the *aggregation of feature values*, $\hat{C} : 2^{L(\tau(X))} \to L(\tau(X))$. Such is, for example, the operation of union of the subsets of X corresponding to the elements of the set ψ with the formation of the set $\breve{\psi} \subseteq X$, or the operation of intersection of sets in ψ with the formation of $\hat{\psi} \subseteq X$, and so on. Since the topology $\tau(X)$ is discrete and the lattice $L(\tau(X))$ is boolean, the operation $\breve{\psi}$, as well as the operation $\hat{\psi}$ (just as a result of any other operation $\hat{C}$), will be valid both in $\tau(X)$ and in $L(\tau(X))$. Since the operator κ is bijective, an arbitrary subset $e \,\tilde{\subseteq}\, M(\tau(X))$ can be mapped to a corresponding element $M(\tau(X))$ by the operation $\hat{\kappa}\hat{C}\hat{\kappa}^{-1}e$, so that the mappings $\hat{\varpi}$ exist, at least for the considered examples of the operator $\hat{C}$. The theorem is proved.

Corollary 1. *A subset of points of an arbitrary ρ-configuration over a regular X can be uniquely assigned a single element of $L(\tau(X))$ and a single element of $\tau(X)$.*

Corollary 2. *Suppose that, in an arbitrary ρ-configuration $\boldsymbol{\rho}(\psi)$, $\psi \subseteq L(\tau(X))$, n subsets of the set ψ and an operator $\hat{C}$ are defined by a system of sets $S_n(\psi) = (\psi'_1, \ldots, \psi'_p, \ldots, \psi'_n \,|\, \psi'_p \subseteq \psi)$. Then the ρ-configuration $\boldsymbol{\rho}(\psi)$ can be uniquely transformed into an n-dimensional metric configuration.*

Corollary 3. *A mapping $\hat{\varpi} \in \bar{\Omega}$ can be applied to all points of the configuration $\boldsymbol{\rho}(\psi)$ if and only if $S_n(\psi)$ contains the covering of ψ, i.e., $\bar{S}_n(\psi) = \psi$.* This is obvious because the operator κ is bijective.

Theorem 2 and its corollaries show that the ρ-configuration $\boldsymbol{\rho}(\psi)$ generated over $\psi \subseteq L(\tau(X))$ can be transformed into a certain "derived" n-dimensional ρ-configuration $\boldsymbol{\rho'}(\psi)$ if (1) the set of initial descriptions X is regular, (2) n subsets ψ are defined, and (3) the operation of *aggregation of feature values C* is defined.

Recall that the elements of a set ψ correspond to certain values of features or combinations such values. Therefore, the mappings $\hat{\varpi} \in \bar{\Omega}$ actually correspond to the *procedures of generation of some "derived" feature descriptions* on the basis of the collections of feature values defined by a system of subsets from $S_n(\psi)$ of the subset $\psi \subseteq L(\tau(X))$.

The subsets comprising the system $S_n(\psi)$, on one hand, can be defined empirically from some expert considerations. On the other hand, one can apply the methods of the analysis of the properties of compactness and density [10, 11] in order to define the system $S_n(\psi)$. An essential condition imposed on $S_n(\psi)$ is the covering of the entire set ψ by the elements of $S_n(\psi)$ (Corollary 3 to Theorem 2).

4. METRIC FORMS OF THE CRITERIA OF SOLVABILITY, REGULARITY, CORRECTNESS, AND COMPLETENESS

Consider a two-class problem in which each object from the set of precedents Q belongs either to a class $C^+ \subseteq Q$ or to a class $C^- \subseteq Q$ so the conditions $C^+ \cup C^- = Q$ and $C^+ \cap C^- = \varnothing$ are satisfied. For a regular X, a boolean lattice $L(\tau(X))$ always contains elements "C^+" and "C^-" that correspond to the definitions of the classes C^+ and C^-, $N^+ = |C^+|$, $N^- = |C^-|$. In an atomically generated boolean lattice $L(\tau(X))$, $N^+!$ chains of length N^+ are incident to the "C^+" element and correspond to various combinations of feature values represented in the initial descriptions of objects. The computation process of a correct algorithm can be envisioned as a "motion" along one of these chains.

Suppose that a certain algorithm makes υ errors on objects from the class C^+. This corresponds to the existence of $N^+!/(N^+ - \upsilon)!$ vertices of the lattice to each of which $N^+!$ chains are incident. Therefore, taking into account that in real-world problems N^+ is measured in millions of objects, the complexity of the lattice $L(\tau(X))$ precludes seeking the correct algorithms for solving the problem simply by complete enumeration of the lattice elements.

The introduction of sets $\psi \subseteq L(\tau(X))$, of a system of subsets $S_n(\psi)$, and of mappings $\hat{\varpi} \in \bar{\Omega}$ is essential for finding tractable statements of poorly formalized problems and allows one to reduce significantly the extent of enumeration of the elements of the lattice $L(\tau(X))$ when searching for correct algorithms. For example, when searching for the conjunctions of the

feature values of complexity m in the method of logical rules in the "initial" ρ-configuration $\boldsymbol{\rho}(\psi)$ consisting of n_0 points, one needs to enumerate $C_{n_0}^m$ combinations of feature values. On the contrary, in the derived configuration $\boldsymbol{\rho}'(\psi)$, enumeration of only $C_n^m \ll C_{n_0}^m$ combinations will be required, $n < n_0$. When analyzing specific problems, it often turns out that $n \ll n_0$, which allows one to reduce the extent of enumerative search by orders of magnitude.

To estimate the collections of feature descriptions within the theory of classification of feature values developed here, one applies the solvability/regularity and correctness/completeness criteria. Consider interrelations between the properties of the ρ-configurations under test and the satisfiability of these criteria.

In the formalism developed, the elements of the lattice are given by subsets of the set of objects X. Suppose that a given set $\psi \subseteq L(\tau(X))$, $\psi = \{u_1, \ldots, u_j, \ldots, u_{n_0}, u_{n_0+1}\}$, $u_j \in L(\tau(X))$, $|\psi| = n_0 + 1$, contains, along with the elements corresponding to some feature values $(u_j, j = 1, \ldots, n_0)$, the element $u_{n_0+1} = \hat{\lambda}C^+$ corresponding to the set of all objects of the class C^+. Define a "primary" ρ-configuration $\boldsymbol{\rho}_0(\psi) = \hat{\rho}_{n_0+1}M(\{\hat{\lambda}^{-1}u_j\})$. Note that the distance between $n_0 + 1$ points of the ρ-configuration $\boldsymbol{\rho}_0(\psi)$ corresponding to the class C^+ and an arbitrary jth point of $\boldsymbol{\rho}_0(\psi)$ can take, generally speaking, the zero value (i.e., $\boldsymbol{\rho}_0(\psi)$ may correspond to a semimetric).

Suppose given an operator $\hat{\varpi} \in \bar{\Omega}$, a set $\psi \subseteq L(\tau(X))$, and a system of subsets $S_{n+1}(\psi)$ of the set ψ such that $S_{n+1}(\psi)$ also contains an element $\hat{\lambda}C^+$, $S_{n+1}(\psi) = (\psi_1', \ldots, \psi_p', \ldots, \psi_n', \hat{\lambda}C^+)$, $\psi_p' \subseteq \psi$, $p = 1, \ldots, n, \forall p \leq n : \psi_p' \cap \hat{\lambda}C^+ = \varnothing$. On the basis of the set $S_{n+1}(\psi)$ and the operation $\hat{\varpi} \in \bar{\Omega}$ ($\hat{\varpi} = \hat{\kappa}\hat{C}\hat{\kappa}^{-1}$), which includes the operator of aggregation of feature values $\hat{C}$ ($\hat{C} = \hat{\kappa}^{-1}\hat{\varpi}\hat{\kappa}$), we form a "derivative" ρ-configuration $\boldsymbol{\rho}_o'(\psi)$. To this end, we apply $\hat{C}$ to every element of $S_{n+1}(\psi)$ so that $\boldsymbol{\rho}_o'(\psi) = \hat{\rho}_{n+1}M(\{\hat{\lambda}^{-1}\hat{C}\psi_p'\})$, $p = 1, \ldots, n+1$. Naturally, the $n+1$-dimensional $\boldsymbol{\rho}_o'(\psi)$ also contains the point corresponding to C^+. Just as $\boldsymbol{\rho}_0(\psi)$, the configuration $\boldsymbol{\rho}_o'(\psi)$ may correspond to a semimetric.

Theorem 3. *Suppose given a ρ-configuration $\boldsymbol{\rho}(\psi)$ (this may be a "primary" $\boldsymbol{\rho}_0(\psi)$ or a "derived" $\boldsymbol{\rho}_o'(\psi)$ configuration) containing a point $\hat{\kappa}\hat{\lambda}C^+$, $\hat{\lambda}C^+ \in L(\tau(X))$, represented by a row with index $i(C^+)$*

($\rho[i(C^+)]$). Then the fulfillment of the condition $\underset{\boldsymbol{\rho}(\psi)}{\exists} \rho[j]$, $j \neq i(C^+) : \rho(i(C^+), j) < \varepsilon$, $\varepsilon = \min \rho_{ij}, \varepsilon > 0$, is sufficient for the existence of at least one correct algorithm for solving the corresponding solvable problem $Z(C^+ \cup C^-)$.

Proof. A metric space $M(\tau(X))$ is generated by a metric that is a function of an isotonic estimate $v[]$ over elements x, y of the lattice $L(\tau(X))$ (see [9]). For $\rho_\pi(x, y) \approx 0$, we obtain $v[x \vee y] \approx v[x \wedge y]$, which, due to the isotonicity of the estimate, implies $x \approx y$. In the case of finite lattices, the condition $\rho_\pi(x, y) \approx 0$ is transformed into $\rho_\pi(x, y) = 0$ (since, in finite lattices, the values of the metric are generally discrete, with a finite step size). Thus, the condition $\rho_\pi(x, y) < \varepsilon$, where $\varepsilon > 0$ is a minimal nonzero distance in $\boldsymbol{\rho}(\psi)$, implies the identity of the sets x and y.

Next, an arbitrary ρ-configuration is defined as a composition of operators $\hat{\rho}_n$ and $M()$, $\boldsymbol{\rho}(\psi) = \hat{\rho}_nM(\psi)$. Due to the existence of inverse operators of $\hat{\rho}_n$ and $M()$, an arbitrary point of $\boldsymbol{\rho}(\psi)$ corresponds to a certain element of the set ψ, which, in turn, corresponds to a certain combination of feature values in the initial descriptions of objects X or to the class C^+. Hence, the jth element of $\boldsymbol{\rho}(\psi)$, $j \neq i(C^+)$, which satisfies $\rho(i(C^+), j) < \varepsilon$, corresponds to a combination of feature values that (1) allows one to distinguish all the objects of the class C^+ and (2) distinguishes none of the objects of the class C^-, i.e., is a test. Within, for example, the method of logical rules, this combination of feature values can be represented as an appropriate disjunctive normal form and corresponds to a correct algorithm. If one can construct at least one correct algorithm for solving a problem $Z(C^+ \cup C^-)$, then the problem $Z(C^+ \cup C^-)$ is certainly solvable. The theorem is proved.

Corollary 1. *The fulfillment of the hypothesis of the theorem for all elements of a regular set of samples $\hat{\zeta}_r X$ is equivalent to the existence of at least one complete algorithmic model.*

This is obvious from the definition of the completeness of an algorithmic model (see the Introduction).

Let us explain the meaning of Theorem 3. Indeed, the assertion of the theorem can be applied to both "primary" $\boldsymbol{\rho}_0(\psi)$ and "derived" $\boldsymbol{\rho}_o'(\psi)$ configurations. From the practical point of view, however, the presence of a "feature" among the primary feature descriptions that is actually identical to the sought class C^+ is a degenerate case and is not encountered in real-world poorly formalized problems. When such an element is found in a derived" $\boldsymbol{\rho}_o'(\psi)$, this means that the operations undertaken (a choice of the system of

subsets $S_{n+1}(\psi)$, aggregation of feature values, etc.) allow one to distinguish such a chain among all the chains of $L(\tau(X))$ that is incident to the vertex C^+ of the lattice and, accordingly, contains all combinations of features necessary for constructing a correct algorithm for the class C^+. In this sense, the condition $\underset{\rho(\psi)}{\exists}\,\rho[j],\ j \neq i(C^+) : \rho(i(C^+),\ j) < \varepsilon$ is a constructive criterion for assessing the "quality" of the generated sets of feature descriptions.

When analyzing real data of poorly formalized problems, the cases when $\rho(i(C^+), j) < \varepsilon$ are not encountered. Nevertheless, the distances from the point $i(C^+)$ to other points of the ρ-configuration $\boldsymbol{\rho}(\psi)$ under test characterize the "quality" or "informativity" of $\boldsymbol{\rho}(\psi)$ with respect to the class C^+ the greater is the distance between an arbitrary point and the point $i(C^+)$, the lesser is the "informativity" of the feature value, corresponding to the arbitrary point in view.

Therefore, we can assume that the points from $\boldsymbol{\rho}(\psi)$ that correspond to the most "informative" features are located in some "neighborhood" of the point $i(C^+)$. Define a *parametric neighborhood of the ith point*, $\overline{O}(i, r)$, as a subset of rows of the matrix $\boldsymbol{\rho}(\psi)$ chosen on the basis of the correspondence of the distances $\rho(i(C^+), j)$ to a value of a certain parameter r, i.e., the *radius of the neighborhood*. In the simplest case, this is a closed spherical neighborhood of the ith point, $\overline{O}(i, r) = \{\rho[j] | \rho(i, j) \leq r,\ \rho[i], \rho[j] \in \boldsymbol{\rho}(\psi)\}$. For any definition of the neighborhood, the parameter r always has a certain *maximal value* $r_{\max}$ such that $\overline{O}(i, r_{\max}) = \boldsymbol{\rho}(\psi)$, whose existence is obvious from the finiteness of $\boldsymbol{\rho}(\psi)$. When $\boldsymbol{\rho}(\psi)$ contains a point $i(C^-)$, i.e., a point corresponding to a class of objects that is a complement of C^+ in the boolean lattice $L(\tau(X))$, one can define $r_{\max}$ as $r_{\max} = \rho(i(C^+), i(C^-))$.

Theorem 4. *Suppose given a regular set of descriptions of objects X. For given set $\psi \subseteq L(\tau(X))$, ρ-configuration $\boldsymbol{\rho}(\psi)$ containing a point with index $i(C^+)$, and method of choice of the parametric neighborhood $\overline{O}(i, r)$, the solvability* (1), *regularity* (2), *correctness* (3), *and completeness* (4) *criteria can be represented as functions of the radius of the neighborhood of the point $i(C^+)$.*

Proof. Based on Theorem 3 and the definitions of criteria (1)−(4), we obtain "neighborhood" formulations of these criteria. To this end, we should define operations ϕ and δ_k and a method for calculating a mask $\chi = (\gamma_k)$.

A configuration $\boldsymbol{\rho}(\psi)$ consisting of n points, whether $\boldsymbol{\rho}_0(\psi)$ or $\boldsymbol{\rho}'_o(\psi)$, is generated by a set $\psi \subseteq L(\tau(X))$, $|\psi| = n$, whose elements are given by various subsets of the set of objects X. Each such subset, corresponding to a point in $\boldsymbol{\rho}(\psi)$, is uniquely defined by a characteristic function over the set X (naturally, the set of values of the characteristic function is $[0,1]$). Hence, any point of $\boldsymbol{\rho}(\psi)$ (including $i(C^+)$) can be considered as a point representing some boolean (binary) feature (this feature may not even be represented in the initial feature description of $\pi(X)$). Therefore, we define a *set of initial information* $I_i = [0,1]^n$ and a function $\delta_k(v_1, v_2) = (v_1 = v_2)$.

We will consider the problem in the two-class form, i.e., $\boldsymbol{\rho}(\psi)$ contains a single point, which corresponds to some class, the point $i(C^+)$. The criteria formulated below are generalized to the case of a problem with many classes simply by introducing additional points into $\boldsymbol{\rho}(\psi)$ that correspond to other classes and testing the criteria for each class. Therefore, we define a *set of final information* as $I_f = [0,1]$.

For a regular X, an isomorphic *set of precedents* $\varphi(a)$ for an arbitrary sample $a = \{x_i, i = 1...|a|\} \in \hat{\zeta}X$ consists of objects $q_i = (m_i, \iota_i)$, $m_i \in I_i$, $\iota_i \in I_f$, i.e., $m_i = (\alpha_1^i, ..., \alpha_k^i, ..., \alpha_n^i)$, $\alpha_k^i \in [0,1]$, where $\alpha_k^i = (x_i \in c_k | c_k \in \psi)$, $\iota_i = (x_i \in C^+)$. Accordingly, φ, the *operator of formalization of the problem*, is defined as $\varphi(a) = \{((x_i \in c_k | c_k \in \psi), (x_i \in C^+)) | x_i \in a\}$, where $a \in \hat{\zeta}X$ is the sample considered. For a given neighborhood of the point $i(C^+)$ with radius r, $\overline{O}(i(C^+), r)$, a feature selection mask $\chi = (\gamma_k)$ is calculated as $\chi = (\gamma_k = (\rho[k] \in \overline{O}(i(C^+), r)))$.

Then, substituting the operations described into definitions (1)−(4), we obtain corresponding parametric criteria whose satisfiability depends on the radius of a neighborhood in the metric configuration $\boldsymbol{\rho}(\psi)$:

• *solvability criterion of problem $Z(\varphi(a))$*,

(1.2) $\exists r \leq r_{\max} : (\underset{\varphi(a)}{\forall}\, q_1, q_2 : \iota_1 \neq \iota_2 \Rightarrow \underset{1..n}{\exists}\, \rho[k] \in \overline{O}(i(C^+), r)) : (x_1 \in c_k | c_k \in \psi) \neq (x_2 \in c_k | c_k \in \psi))$,

• *regularity criterion of problem $Z(\varphi(a))$*,

(2.2) $\exists r \leq r_{\max} : (\underset{\varphi(a)}{\forall}\, q_1, q_2 : \underset{1..n}{\exists}\, \rho[k] \in \overline{O}(i(C^+), r)) : (x_1 \in c_k | c_k \in \psi) \neq (x_2 \in c_k | c_k \in \psi))$,

• *correctness criterion of algorithm $A_h(m_i, \theta, \chi)$*,

(3.2) $\exists r \leq r_{\max} : \underset{\varphi(a)}{\forall}\, (m_i, \iota_i) : A_h(m_i, \theta, ((\rho[k] \in \overline{O}(i(C^+), r)))) = \iota_i$,

• *completeness criterion of algorithmic model $M_A[\hat{\Theta}]$*,

(4.2) $\exists r \leq r_{\max} : \underset{\hat{\zeta}_r X}{\forall} a \underset{M_A[\hat{\Theta}]}{\exists} A_h : (\underset{\varphi(a)}{\forall}(m_i, \ \iota_i) : A_h(m_i,$

$\theta_h, ((\rho[k] \in \bar{O}(i(C^+), r)))) = \iota_i).$

Each of the criteria (1.2)−(4.2) can easily be represented as a function with the range of values [0,1], that depends on the parameter r. The theorem is proved.

Corollary 1. *For* $\varphi(a)$, $|\varphi(a)| = N$, *testing the satisfiability of criteria* (1.2)−(4.2) *is equivalent to calculating the values of combinatorial functionals, each of which is a function of the radius* r.

Take, for example, the solvability criterion (1.2). The functional corresponding to (1.2) is expressed as follows:

$$r_{12}(\varphi(a), r) = \frac{2}{N(N-1)} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} (\iota_i \neq \iota_j \Rightarrow \underset{1..n}{\exists} \rho[k]$$

$$\in \bar{O}(i(C^+), r)) : (x_i \in c_k \mid c_k \in \psi) \neq (x_j \in c_k \mid c_k \in \psi)).$$

Then the solvability of $Z(\varphi(a))$ for some value of $r = r'$ is equivalent to $r_{12}(\varphi(a), r') = 1$. In the same way as in [12], we define the combinatorial functionals $r_{22}(\varphi(a), r)$, $r_{32}(\varphi(a), A_h, r)$, and $r_{42}(\hat{\zeta}_r X, M_A[\hat{\Theta}], r)$.

Corollary 2. *When criteria* (1.2)−(4.2) *are satisfiable for* $r = r_{\max}$, *one can calculate some minimal values of the radii for which the criteria are still satisfiable.*

In the case of the criterion (1.2), such a minimal radius is defined as $r_s = \min_{r' \leq r_{\max}} r' : (\forall r < r' : r_{12}(\varphi(a), r) = 0) \wedge (\forall r \geq r' : r_{12}(\varphi(a), \ r) = 1)$. Conventionally, this value of the radius, r_s, can be called the "solvability radius." The "regularity radius" r_r, the "correctness radius" r_c, and the "completeness radius" r_f are calculated analogously.

Corollary 3. *When criteria* (1.2)−(4.2) *are satisfied, for given problem,* $\rho(\psi)$, *and algorithmic model,* $r_s \leq r_r \leq r_c \leq r_f \leq r_{\max}$.

Criterion (2.2) does not contain the congruences $\iota_1 \neq \iota_2$ existing in (1.2), so that a greater number of pairs of objects are compared in (2.2) than in (1.2). To distinguish a greater number of pairs of objects, one may need a greater number of distinguishing features. A still greater number of distinguishing features may be needed to find correct algorithms and, even more so, complete algorithmic models. Hence, the values of the corresponding radii grow.

Corollary 4. *Suppose that the point* $i(C^-)$ *corresponding to the class* C^- *opposite to* C^+ *in a two-class statement of the problem is also represented in* $\rho(\psi)$. *Then, the values of the functionals* $r_{12}(\varphi(a), r)$, $r_{22}(\varphi(a), r)$, $r_{32}(\varphi(a), A_h, r)$, *and* $r_{42}(\hat{\zeta}_r X, M_A[\hat{\Theta}], r)$ *and the radii* r_s, r_r, r_c, *and* r_f *may be significantly different for* $i(C^+)$ *and* $i(C^-)$.

Depending on the choice of $\psi \subseteq L(\tau(X))$, the ρ-configuration $\rho(\psi)$ may contain features that are "positive" for C^+ (i.e., features that are more frequent in objects of the class C^+ correspond to the points that are closer to the point $i(C^+)$) and, for example, may not contain features that are "negative" for C^+ (that correspond to points that are closer to the point $i(C^-)$). Then, naturally, for the same value of the radius, the neighborhoods of the points $i(C^+)$ and $i(C^-)$ will contain significantly different numbers of points corresponding to distinguishing features.

Thus, the satisfiability of the solvability/regularity and correctness/completeness criteria depends on the radius of the neighborhood of the point $i(C^+)$, corresponding to the class of objects considered in the recognition problem, in the ρ-configuration $\rho(\psi)$. The cross-validation forms of criteria (1.2)−(4.2), similar to criteria (1.1)−(4.1), are obtained from (1.2)−(4.2) by introducing a certain *feature generation operator* $\hat{\psi}$ that calculates a set ψ generating $\rho(\psi)$ on the basis of some set of precedents $\varphi(b)$. Then the ρ-configuration $\rho(\hat{\psi}\varphi(b))$ is a result of "learning" on the set $\varphi(b)$, and the satisfiability of criteria (1.2)−(4.2) is tested on a test set $\varphi(a)$, $a, b \in \hat{\zeta} X$, $a \neq b$. The methods of introducing the operator $\hat{\psi}$ will be considered separately.

For a given algorithm A_h (algorithmic model $M_A[\hat{\Theta}]$), the threshold values of the radii r_s, r_r, r_c, and r_f characterize, in part, the joint "quality" of the subsample of the feature values $\psi \subseteq L(\tau(X))$ and the metric ρ that generates a ρ-configuration $\rho(\psi)$. A practically more useful form of representation of data on the "quality" of $\rho(\psi)$ under test with respect to the recognition problem considered is given by the graphs $r_{12}(\varphi(a), r)$, $r_{22}(\varphi(a), r)$, $r_{32}(\varphi(a), A_h, r)$, and $r_{42}(\hat{\zeta}_r X, M_A[\hat{\Theta}], r)$. Moreover, quantitative estimates of such a joint "quality" can be obtained on the basis of (1) the empirical distribution function (EDF) constructed over the set of numbers $\gamma_{C^+} = \{\rho(i(C^+), j) \mid j = 1\ldots n, \ j \neq i(C^+)\}$ (in the present series of works, these EDFs are called the "i-spectra of an ρ-configuration" [11]) and by (2) the value of the minimum distance in the set γ_{C^+}.

5. METHODS OF GENERATION OF THE ρ_q-CONFIGURATIONS ON THE BASIS OF THE ρ_π-CONFIGURATIONS

Above we presented the results of the analysis of the solvability/regularity and correctness/completeness criteria from the viewpoint of the neighborhoods in

the ρ_π-configurations (i.e., ρ-configurations of the features). Below, we will show that these criteria of the algebraic approach can also be formulated through analysis of ρ_q-configurations, which reflect the distances between elements in the set of precedents (i.e., between the objects).

Of no less importance is that adequate methods for calculating the distances between pairs of objects are required to increase the efficiency of various "metric methods" of pattern recognition: the method of k-nearest neighbors, the estimate evaluation algorithm, the methods based on the introduction of linear spaces of certain dimension (SVM, RVM, etc.) [16].

One can propose several general approaches to the formation of ρ-configurations of objects by using the ρ-configuration of features. First, these approaches may be of totally heuristic character. Second, the approaches may be based on some elements of the suggested formalism (for example, on the analysis of ordered sets of lattice elements or unordered subsets of the set ψ that generates the ρ_π-configuration $\boldsymbol{\rho}(\psi)$). Here we focus on the second group of approaches.

First, within the formalism developed, *a primary feature description of an ith object can be represented by an n-dimensional ordered set of elements of the lattice $L(\tau(X))$ i.e., by the set $\varsigma_i = (\hat{\varsigma}_k x_i[k])$* (Corollary 3 to Theorem 1). Each element of ς_i is a subset of objects from X that corresponds to a value of the feature $x_i[k]$. In the case of numerical features, such a subset corresponds to the maximum element of the corresponding chain in $L(\tau(X))$ or to the complement of such an element and in the case of "categorial" features to an element of an antichain. Boolean features are represented simply by the vertices of the lattice ("central nodes" of the chain bundles, as it was mentioned previsouly). In the case of a "derived" $n + 1$-dimensional $\boldsymbol{\rho}'_o(\psi)$-configuration, information on objects can also be represented as an $n + 1$-dimensional ordered set ς'_i of subsets X whose elements correspond to "aggregated" binary features.

Representation of information on objects as the sets ς_i or ς'_i allows one to formulate the definitions of metrics ρ_q in different ways. Suppose that n-dimensional feature descriptions for objects q_1 and q_2 from a set of precedents Q are represented as the sets $\varsigma_1 = (\varsigma_k^1)$ and $\varsigma_2 = (\varsigma_k^2)$, $k = 1, \ldots, n$, $\varsigma_k^1, \varsigma_k^2 \in L(\tau(X))$. Since ς_1 and ς_2 are ordered and have identical dimensions, any method for calculating the metric distance ρ_q between q_1 and q_2 on the basis of the sets ς_1 and ς_2 necessarily involves (1) a method υ_k for estimating the pairwise matching of the kth elements of ς_k^1 and ς_k^2 and (2) a method Σ for combining the individual estimates $\upsilon_k(\varsigma_k^1, \varsigma_k^2)$. Thus, in the general case, the distance

between q_1 and q_2 is defined as $\rho_q(q_1, q_2) = \Sigma(\{\upsilon_k(\varsigma_k^1, \varsigma_k^2)\})$.

The "setting of the metric" in accordance with the set of precedents Q consists in the choice of the methods υ_k and Σ, the choice of the numerical parameters of these methods (if such parameters are used), and so on. For example, the method υ_k is defined for each kth position, thus allowing one to introduce such numerical parameters as the weights of features. The role of υ_k can be played by the metric ρ_π itself (since $\varsigma_k^1, \varsigma_k^2 \in L(\tau(X))$), by the product $\omega_k \rho_\pi$, where ω_k is the weight of the kth feature, etc., and the role of Σ can be played by a sum, sum of the squares, squared root of the sum of the squares, etc. A natural restriction on the operations υ_k and Σ is the requirement that the functional $\rho_q(q_1, q_2)$ obtained be a metric.

It is obvious that the calculation methods υ_k and Σ imply the introduction of some *linear vector space*. For example, the application of $\upsilon_k = (\varsigma_k^1 = \varsigma_k^2)$ and the use of summation as ς_k^1 is equivalent to the Hamming metric on the space of boolean vectors. When one applies $\upsilon_k = \rho_\pi(\varsigma_k^1, \varsigma_k^2)$ and Σ as a square root of a sum of squares, then one obtains the Euclidean metric in the n-dimensional vector space.

The use of "derived" ρ_π-configurations, $\boldsymbol{\rho}'_o(\psi)$, provides additional methods for generation of linear vector spaces on the space of objects. As a result of selection of feature values by $\varpi \in \bar{\Omega}$, the derived $\boldsymbol{\rho}'_o(\psi)$ contains elements each of which corresponds to some binary features and individual values of numerical or "categorial" features.

Suppose that set of subsets of X $\varsigma_0 = (\varsigma_{0,k})$, $k = 1, \ldots, n_0$, corresponds to the "primary" feature description of the object q, and $\varsigma' = (\varsigma'_{k'})$, $k' = 1, \ldots, n$, to a description corresponding to a "derived" ρ_π-configuration $\boldsymbol{\rho}'_o(\psi)$. Depending on the aggregation operator of feature values $\hat{C} = \hat{\kappa}^{-1}\hat{\varpi}\hat{\kappa}$, one can obtain various n-dimensional boolean or numerical vectors on the basis of the set ς'.

Suppose, for example, the occurrence frequencies of feature values in the primary description (i.e., $|\varsigma_k|/N$) are very small. Then it is expedient to define an operator $\hat{C}$ as the operation of union of sets, i.e., $\hat{C}\psi = \bar{\psi}$. In this case, some sets $\varsigma_{0,k}$ from the primary description ς_0 appear in some sets $\varsigma'_{k'}$. Then the pair of descriptions $(\varsigma_0, \varsigma')$ can be assigned an n-dimensional boolean vector $(\alpha_t | \alpha_t = (\exists_{\varsigma_0} \varsigma_{0,k} : \varsigma_{0,k} \subseteq \varsigma'_t)$, $t = 1, \ldots, n)$ or an n-dimensional numerical vector $(\beta_t | \beta_t = |\{\varsigma_{0,k} | \varsigma_{0,k} \subseteq \varsigma'_t, \varsigma_{0,k} \in \varsigma_0\}|, t = 1, \ldots, n)$, and

then one can introduce some metric over sets of the corresponding vectors.

Second, in addition to the above-described methods, metrics can be introduced by introducing linear spaces corresponding to the ρ-configuration under test. The generation and analysis of such spaces can be carried out with the use of ρ-networks [10], in which the distances of the points to vertices of the ρ-network serve as a basis for the formation of vectors of given dimension. This approach can be applied to introducing a linear space on the basis of both ρ_q-configurations and ρ_π-configurations. The generation of linear spaces on the basis of ρ-configurations is extremely important and promising direction of research.

Third, note that the representation of information on an object in the form of an ordered set of subsets of X allows one not only to introduce some ρ_q-metric, but also to *assess the "quality" of the feature description of the object*. Let $\varsigma = (\varsigma_k)$, $k = 1, ..., n$, be a feature description of object q, whether "primary" ς_0 or "derived" ς'. The empirical distribution function of distances ρ_π between the elements of ς characterizes the "widths" of a feature description of an object: whether the object is described by a single subset of strongly correlated features (small values of ρ_π), or the description of the object q represents "widely separated" weakly correlated features (high values of ρ_π). Pairwise "widely separated" features are similar to the concept of a basis of a linear space in which orthogonal vectors are also weakly correlated (i.e., the values of each coordinate of a vector may vary independently of the values of other vectors).

Fourth, one can introduce a *distance between objects on the basis of unordered subsets of set ψ that forms the ρ_π-configuration $\rho(\psi)$*, be it the "primary" $\rho_0(\psi)$ or the "derived" ρ_π-configuration $\rho'_o(\psi)$.

Theorem 5. *In the case of a regular set X, each object of X corresponds to a unique subset of the points of the ρ_π-configuration $\rho(\psi)$.*

Proof. An arbitrary n-dimensional ρ_π-configuration $\rho(\psi)$ is generated by the set of feature values $\psi = \{u_1, ..., u_k, ..., u_n\}$, $u_k \in L(\tau(X))$, $u_k \subseteq X$. By definition, a regular set of initial descriptions X is isomorphic to a regular set of precedents Q. Therefore, the membership of an arbitrary object $x_i \in X$ in a set u_k is equivalent to the existence of an object $q_i \in Q$ in the feature description that corresponds to the value of the features such that each element of the set $\varsigma = (\varsigma_k)$ describing the object x_i is either an empty set or one of elements of ψ. By Theorem 1, in the case of a regular set of precedents, the intersection of nonempty elements $\varsigma = (\varsigma_k)$ contains a single element, i.e., the object x_i itself; i.e., each object is described by a unique collection of sets ς_k. Nonempty elements of ς

correspond to the points of $\rho(\psi)$, so that each object is represented by a unique subset of points of a metric configuration. The theorem is proved.

Corollary 1. *For a given set ψ, define an operator $\psi_q : X \to 2^\psi$. In the case of regular X, ψ_q is injective.*

It is obvious that each object corresponds to a unique collection of elements, so that different elements of X are mapped to different elements of the boolean 2^ψ.

Corollary 2. *The cardinality of the symmetric distance between nonempty sets $\psi_q(q_1)$ and $\psi_q(q_2)$ corresponding to objects q_1 and q_2 is the Hamming metric.*

In terms of the present formalism, the Hamming metric is defined as the number of distinguishing binary features for a given pair of objects. It is the symmetric difference between the sets $\psi_q(q_1)$ and $\psi_q(q_2)$ that contains all the distinguishing features for the objects q_1 and q_2.

Corollary 3. *The expression $\underset{\varphi(X)}{\forall} q_1, q_2 : \iota_1 \neq \iota_2 \Rightarrow \psi_q(q_1) \neq \psi_q(q_2)$ represents the solvability criterion of problem $Z(\varphi(X))$.*

Corollary 4. *The expression $\underset{\varphi(X)}{\forall} q_1, q_2 : \psi_q(q_1) \neq \psi_q(q_2)$ represents the regularity criterion of problem $Z(\varphi(X))$.*

Corollary 5. *Suppose that an n-dimensional configuration $\rho(\psi)$ contains points $i(C^+)$ and $i(C^-)$ corresponding to the classes C^+ and C^-. Suppose given a method for defining a neighborhood of a point $\bar{O}(i, r)$ and r^+ and r^- such that $C^+ \subseteq \cup \hat{\kappa}^{-1} \hat{\rho}_n^{-1} \bar{O}(i(C^+), r^+)$ and $C^- \subseteq \cup \hat{\kappa}^{-1} \hat{\rho}_n^{-1} \bar{O}(i(C^-), r^-)$. Then the fulfillment of the condition $\bar{O}(i(C^+), r^+) \cap \bar{O}(i(C^-), r^-) = \varnothing$ is sufficient for the existence of a correct algorithm for solving the problem $Z(\varphi(X))$.*

The proposition is obvious because the condition corresponds to the presence of a set of strong "positive" and "negative" features such that the features form corresponding Zhuravlev's "tests" that cover the classes C^+ and C^-.

According to Theorem 5, on the basis of the sets $\psi_q(q_1)$ and $\psi_q(q_2)$ corresponding to the objects q_1 and q_2, one can also construct other metrics. Let us construct, for example, the distributions of minimum distances ρ_π from each element of $\psi_q(q_1)$ to the nearest element of $\psi_q(q_2)$ and, vice versa, from $\psi_q(q_2)$ to the nearest element in $\psi_q(q_1)$. Depending on the choice of the combination method of these distributions during calculating a certain functional, one can obtain different heuristic metrics for estimating the distances between objects.

Thus, within the present formalism, there are several different approaches to define metric distances between objects (ρ_q) on the basis of the distances between features (ρ_π) and, thus, the corresponding methods of generation of ρ_q-configurations on the basis of ρ_π-configurations.

6. ANALYSIS OF THE PROPERTIES OF ρ_q-CONFIGURATIONS AND THE CRITERIA OF SOLVABILITY, REGULARITY, CORRECTNESS, AND COMPLETENESS

It is clear from the aforesaid that, within the present formalism, a method of calculating ρ_q is a kind of "superstructure" over the method of calculating ρ_π. Approaches to assessing the "quality" of these ρ_q-configurations naturally result from the interrelation between the properties of ρ_q-configurations and the criteria of the algebraic approach.

In ρ_π-configurations, the classes of objects under study correspond to individual points of a configuration, while the objects themselves correspond, in the case of regular X, to certain subsets of points of the ρ_π-configuration. Conversely, the points of a ρ_q-configuration correspond to individual objects, while the classes to the subsets of points of the ρ_q-configuration.

By Theorem 5, each object q_i from a regular $\varphi(a)$ corresponds to a unique collection of points of a ρ_π-configuration, $\psi_q(q_i)$, whose elements uniquely correspond to the elements of a feature description in the form $\varsigma_1 = (\varsigma_k^1)$. The metric ρ_q for arbitrary objects q_1 and q_2 is defined as a functional over $\psi_q(q_1)$ and $\psi_q(q_2)$ or over descriptions ς_1 and ς_2. According to Corollaries 3 and 4 to the Theorem 5, the metric ρ_q should be defined in such a way that the *regularity criterion for ρ_q-configurations* (2.3) is satisfied. Taking into account the differences in the structure of expressions (1) and (2), one derives a *solvability/regularity criteria for ρ_q-configurations* (1.3):

(1.3) $\displaystyle \forall_{\varphi a)} q_1, q_2 : \iota_1 \neq \iota_2 \Rightarrow \rho_q(q_1, q_2) > 0,$

(2.3) $\displaystyle \forall_{\varphi(a)} q_1, q_2 : \rho_q(q_1, q_2) > 0.$

Let "$\rho_q^{\varphi_1}$" indicate that the method for calculating metric of ρ_q is set up with the use of the set of precedents φ_1 (as, for example, in the case of a metric of the form $\rho_q(q_1, q_2) = \Sigma(\{\upsilon_k(\varsigma_k^1, \varsigma_k^2)\})$, see above). Then criteria (1.3) and (2.3) imply the corresponding combinatorial functionals:

$$r_{13}(\varphi_1, \varphi_2) = \frac{2}{N(N-1)} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} (\iota_i \neq \iota_j$$

$$\Rightarrow \rho_q^{\varphi_2}(q_i, q_j) > 0), \quad q_i, q_j \in \varphi_1, \quad N = |\varphi_1|,$$

$$r_{13}(\varphi(a), \varphi(a)) = 1,$$

$$r_{23}(\varphi_1, \varphi_2) = \frac{2}{N(N-1)} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} (\rho_q^{\varphi_2}(q_i, q_j) > 0),$$

$$q_i, q_j \in \varphi_1, \quad N = |\varphi_1|,$$

$$r_{23}(\varphi(a), \varphi(a)) = 1.$$

Thus, the existence of pairwise differences between objects in the combinatorial forms of the solvability (1)/regularity (2) criteria and the presence of distinguishing features in the "neighborhood" forms of the criteria (i.e., the expressions (1.2) and (2.2)) corresponds to the existence of nonzero distances between the two objects in the corresponding ρ_q-configuration. Criteria (1.3) and (2.3) also represent obvious requirements on the generated metric on the set of objects, under which the collections of distances in ρ_q-configurations correspond to the preservation of the solvability/regularity of problems. The satisfiability of criteria (1.3) and (2.3) on "learning–control" pairs of samples can be controlled by calculating the values of the functionals $r_{13}(\varphi_1, \varphi_2)$ and $r_{23}(\varphi_1, \varphi_2)$.

Since conditions (3) and (4) contain vectors θ_h that set the parameters of the algorithm A_h, the analysis of *necessary and sufficient* criteria for the correctness of the algorithm and of the completeness of an algorithmic model on the basis of ρ_q-configurations cannot be carried out irrespective of a specific algorithm/model. However, some *sufficient* conditions for the existence of correct algorithms and of complete algorithmic models can be obtained without loss of generality on the basis of the analysis of the properties of compactness and density as applied to metric configurations.

In the present series of papers, the concept of compactness with respect to such discrete metric systems as ρ-configurations is constructed by analogy with the concept of compactness of metric spaces, which is used in functional analysis [10]. In particular, a generalization of the Heine–Borel–Lebesgue lemma to the case of metric spaces (a metric space is compact if and only if it is totally bounded and complete [17]) can be reformulated in some "discrete form," because, in a discrete and finite $M(\tau(X))$, there inevitably exists a certain minimum nonzero distance between points. Accordingly, the concepts of ε_0-boundedness, ε_0-completeness, and ε_0-compactness of discrete metric spaces are introduced. The introduction of these concepts allowed one to show the equivalence between the

combinatorial heuristic criterion of cluster and the concept of ε_0-compactness (Theorem 1 in [10]).

Definition 1 (*the heuristic criterion of a cluster*). *The points of an arbitrary cluster $K \subset M$ are closer to each other than to other points of a countable metric space M with metric ρ_{ij}:*

$$(5) \qquad \forall i \bigvee_{\substack{i \in K \\ k \in M \setminus K}} k \, \exists j \neq i : \rho_{ij} < \rho_{ik}, \qquad \rho_{ij} \geq \varepsilon,$$

$\varepsilon = \min \rho_{ij} | i \neq j$.

In other words, under condition (5) imposed on some set $K \subset M$, the set K is called a "cluster." Such criteria are widely used in various approaches to clustering points on the basis of distances. Definition (5) can be assigned a combinatorial functional $r_K(K, M)$ that characterizes the *degree of "ε-compactness" of the set $K \subset M$ of points in the metric space M, $r_K(K, M) = \frac{1}{N} \sum_{i=1}^{N} (\min_{j \notin K} \rho_{ij} > \min_{k \in K} \rho_{ik})$, i.e., $r_K(K, M) = 1$ under condition (5). Note that, under regularity criterion (2.3), ε is strictly greater than 0.

Theorem 6. *Suppose given a problem $Z(\varphi(a))$ with a class of objects C^+ and an N-dimensional ρ_q-configuration $\boldsymbol{\rho_q}$ corresponding to the problem $Z(\varphi(a))$. Then the condition $r_K(\varphi(C^+), \hat{\rho}_N^{-1} \boldsymbol{\rho_q}) = 1$ is sufficient for the existence of at least one correct algorithm for solving the problem $Z(\varphi(a))$.*

Proof. Under the condition $r_K(\varphi(C^+), \hat{\rho}_N^{-1} \boldsymbol{\rho_q}) = 1$ in an N-dimensional $\boldsymbol{\rho_q}$, for the class C^+ are defined the external boundary $B_{\text{ext}}(C^+) = \{a \notin C^+ | \exists b \in C^+ : \rho_{ab} = \min\{\rho_{bc}, c \notin C^+\}\}$ and the internal boundary $B_{\text{int}}(C^+) = \{b \in C^+ | \exists a \in B_{\text{ext}}(C^+) : \rho_{ab} = \min\{\rho_{bc}, c \in C^+\}\}$ so that $B_{\text{int}}(C^+) \cap B_{\text{ext}}(C^+) = \varnothing$ (see Corollaries 4–6 to Theorem 1 in [10]). Thus, between the points of the class C^+ and all the points of $\boldsymbol{\rho_q}$ that do not belong to C^+ (the class C^-), there is a certain "gap" by which (in corresponding N-dimensional vector space) one can draw a surface or surfaces that separate all points of C^+ from all points of C^-. The existence of such a separating surface or surfaces is equivalent to the existence of the corresponding correct algorithm. The theorem is proved.

Corollary 1. *A sufficient condition for the existence of a correct algorithm is as follows*:

$$(3.3) \; r_K(\varphi(C^+), \hat{\rho}_N^{-1} \boldsymbol{\rho_q}) = 1.$$

Corollary 2. *For $r_K(\varphi(C^+), \hat{\rho}_N^{-1} \boldsymbol{\rho_q}) = 1$, the fulfillment of the condition $r_K(\varphi(C^-), \hat{\rho}_N^{-1} \boldsymbol{\rho_q}) = 1$ is not necessary for the existence of a correct algorithm.*

Indeed, separating surfaces exist under the "compactness" of C^+ in $\boldsymbol{\rho_q}$, and the "compactness" of C^- is not necessary.

Corollary 3. *Suppose given a set of objects X in which the class C^+ is distinguished, and a regular sampling operator $\hat{\zeta}_r$. Denote a method of calculating an N-dimensional ρ_q-configuration $\boldsymbol{\rho_q}$ by the n-dimensional ρ_π-configuration $\boldsymbol{\rho_\pi}(\psi)$ as $\hat{\Lambda} : \mathbf{P}^n \to \mathbf{P}^N$, and the method of calculating $\boldsymbol{\rho_\pi}(\psi)$ by the set of precedents $\varphi(a)$, as $\boldsymbol{\rho_\pi}(\hat{\psi}\varphi(a))$. Let us formulate a sufficient criterion for the existence of a complete algorithmic model*:

$$(4.3) \; \underset{\hat{\zeta}_r X}{\forall} \, a : r_K(\varphi(C^+) \cap \varphi(a), \hat{\rho}_N^{-1} \hat{\Lambda} \boldsymbol{\rho_\pi}(\hat{\psi}\varphi(a))) = 1.$$

Corollary 4. *Let $r_{43}(\hat{\zeta}_r X) = \frac{1}{Y} \sum_{y=1}^{Y} r_K(\varphi(C^+) \cap \varphi(a_y), \hat{\rho}_N^{-1} \hat{\Lambda} \boldsymbol{\rho_\pi}(\hat{\psi}\varphi(a_y)))$, $Y = |\hat{\zeta}_r X|$. Then, $r_4(\hat{\zeta}_r X) = 1$ under criterion* (4.3).

It is obvious that the fulfillment of the condition of "compactness" of C^+ in a ρ_q-configuration $\boldsymbol{\rho_q}$ (3.3) by no means guarantees the existence of a unique separating surface and/or the uniqueness of a connected component of class C^+ (in the sense that two arbitrary points of class C^+ are never separated by the external boundary $B_{\text{ext}}(C^+)$ or the internal boundary $B_{\text{int}}(C^+)$) [10]. However, condition (3.3) guarantees that all points of the same class C^+ are in some "condensations", even if separated from each other by the boundary $B_{\text{ext}}(C^+)$ or $B_{\text{int}}(C^+)$. The "connectedness" criteria of the classes C^+/C^- are important auxiliary tools for estimating the multiplicity of possible solutions to problem $Z(\varphi(a))$, and obtaining these criteria presents a separate direction of research. Here we note that such criteria can be obtained, in particular, by the analysis of metric condensations [11] in ρ_q-configurations.

Thus, the introduction of the requirement of the "ε-compactness" of C^+ in a discrete metric space $\hat{\rho}_N^{-1} \boldsymbol{\rho_q}$ (which is obviously a more stringent requirement than the combinatorial criteria of correctness (3) and completeness (4)) allows one to obtain criteria of the possibility of *existence* of a correct algorithm (3.3) and, accordingly, a complete algorithmic model (4.3).

The most important feature of the criteria (3.3) and (4.3) is that they contain neither a specific algorithm, nor the parameters of the algorithm; this allows one to carry out a cross-validation analysis of the satisfiability of these criteria for various methods of calculation of metric configurations $\boldsymbol{\rho_\pi}(\psi)$ and $\boldsymbol{\rho_q}$. In this case, the values of the functionals $r_K(C^+, \hat{\rho}_N^{-1} \boldsymbol{\rho_q})$ and $r_K(C^-, \hat{\rho}_N^{-1} \boldsymbol{\rho_q})$, which characterize the degree of the "ε-

compactness" of C^+ and C^-, allow one to qualitatively evaluate the expediency of a method of calculating $\boldsymbol{\rho_q}$.

7. CROSS-VALIDATION FORMS OF THE METRIC CRITERIA OF THE ALGEBRAIC APPROACH

Suppose given a set of initial descriptions X of N_0 initial descriptions of objects and a decomposition of X into two classes of objects, C^+ and C^-. A set of "admissible" descriptions of objects J_{ob} and a function φ calculating the set of precedents $Q = \varphi(X)$ are defined. For J_{ob} and φ, the strong form of the axiom of correspondence is postulated, the set X satisfying the regularity condition. Suppose that n_0 is the number of given feature descriptions of precedents in $Q = \varphi(X)$, and all I_k, the ranges of values of the kth feature, $k = 1, \ldots, n_0$, include a neutral element "Δ" corresponding to the indeterminacy of the value of a feature.

Then, the corresponding topology $\tau(X)$ and, hence, the lattice $L(\tau(X))$ can be defined [9]. Define a prebase $\pi(X)$ of the topology $\tau(X)$ by injective operations $\hat{\varsigma}_k : I_k \to L(\tau(X))$ such that $\pi(X)$ consists of subsets of X corresponding to feature values in the initial description, i.e., $\pi(X) = \{\hat{\varsigma}_k q_i[k] \mid q_i \in Q, k = 1, \ldots, n_0, i = 1, \ldots, N_0\}$. Define a *set of feature values* $\mathbf{I}(X) = \{(k, q_i[k]) \mid q_i \in Q\}$, $k = 1, \ldots, n_0$.

Cross-validation analysis implies the "learning" of an algorithm on a single sample $a \in \hat{\zeta} X$ and the "control" of the results of "learning" by the algorithm on another sample $b \in \hat{\zeta} X$. "Learning" in the case of the "neighborhood" criteria (1.2)–(4.2) implies the "setting" of some metric ρ_π by the corresponding set of precedents, while, in the case of the "compactness" criteria (1.3)–(4.3), it additionally implies the setting of the metric ρ_q.

Theorem 7. *If identical sets of feature values are defined for all elements of a given set of samples $\hat{\zeta} X$ over a regular set X, and aggregated feature descriptions generated for different samples are unambiguously assignable, then the results of cross-validation of the "neighborhood" criteria* (1.2)–(4.2) *can be represented as the distributions of values of the corresponding radii r_s, r_r, r_c, and r_f.*

Proof. Suppose that the *condition of identity of the sets of feature values* is satisfied for all $a \in \hat{\zeta} X$, i.e.,:

(6) $\underset{\hat{\zeta} X}{\forall} a, b : \mathbf{I}(a) = \mathbf{I}(b).$

Suppose that a set of subsets X $\psi(a) = \{\hat{\varsigma}_k q_i[k] \mid k = 1, \ldots, n_0, q_i \in \varphi(a)\}$ is constructed over the set $a \in \hat{\zeta} X$, $|\psi(a)| = n_1$. Under condition (6), for a regular X, the elements of an arbitrary pair of sets $\psi(a)$

and $\psi(b)$, $a, b \in \hat{\zeta} X$, are unambiguously assignable to each other and, accordingly, to certain elements of the prebase $\pi(X)$.

For a given method of calculation of the metric ρ_π and a given sample $a \in \hat{\zeta} X$, formulate a "primary" ρ_π-configuration $\boldsymbol{\rho}(\psi(a)) = \hat{\rho}_{n_1} M(\hat{\lambda}^{-1} \psi(a))$, $\psi(a) \subseteq L(\tau(X))$. Since the method of definition of the subsets $\psi(a)$ is such that, for any $a \in \hat{\zeta} X$, the elements of the systems of subsets $S_{n+1}(\psi(a), a) = (\psi'_1(a), \ldots, \psi'_p(a), \ldots, \psi'_n(a), \{a \cap \upsilon(C^+)\}_{n+1})$, $\psi'_p(a) \subseteq \psi(a)$, $\breve{S}_{n+1}(\psi) = \psi$, and $S_{n+1}(\psi(b), b)$ obtained are unambiguously assignable to each other, it follows that there exists a bijection $\tilde{\psi} : 2^{L(\tau(X))} \to 2^{L(\tau(X))}$ and the following *matching condition of aggregated feature descriptions* is satisfied:

(7) $\underset{\hat{\zeta} X}{\forall} a, b \ \exists \tilde{\psi} : \tilde{\psi}(\psi'_p(a), S_{n+1}(\psi(b), b)) = \psi'_p(b),$

$\tilde{\psi}^{-1}(\psi'_p(b), S_{n+1}(\psi(a), a)) = \psi'_p(a).$

When defining the operator $\hat{C} : \{\psi \mid \psi \subseteq L(\tau(X))\} \to L(\tau(X))$, under condition (7), the points of any two "derived" ρ_π-configurations $\boldsymbol{\rho}'(\psi(a), S_{n+1}(\psi(a), a)) = \hat{\rho}_{n+1} M(\hat{\lambda}^{-1} \{\hat{C} \psi'_p(a) \mid \psi'_p(a) \in S_{n+1}(\psi(a), a)\})$ and, accordingly, $\boldsymbol{\rho}'(\psi(b), S_{n+1}(\psi(b), b))$, $a, b \in \hat{\zeta} X$, are also unambiguously assignable. For short, we will use the notations $\boldsymbol{\rho}'(\psi(a)) = \boldsymbol{\rho}'(\psi(a), S_{n+1}(\psi(a), a))$ and $\boldsymbol{\rho}'(\psi(b)) = \boldsymbol{\rho}'(\psi(b), S_{n+1}(\psi(b), b))$.

Then the cross-validation forms of the "neighborhood" criteria (1.2)–(4.2) can be obtained from (1.1)–(4.1) by the substitution of the "aggregated" set of precedents $\varphi_a(a) = \{((x_i \in c_k \mid c_k \in \{\hat{C} \psi'_p(a) \mid \psi'_p(a) \in S_{n+1}(\psi(a), a)\}), (x_i \in C^+)) \mid x_i \in a\}$, the function $\delta_k(v_1, v_2) = (v_1 = v_2)$, and the mask $\chi(\varphi_a(b))$, $\chi = (\gamma_1, \ldots, \gamma_k, \ldots \gamma_n)$, $\gamma_k \in [0, 1]$, $\gamma_k = (\rho[k] \in \bar{O}(i(C^+), r))$, where the neighborhood of the point $i(C^+)$ is taken in the configuration $\boldsymbol{\rho}'(\psi(b))$, while the kth point of the configuration $\boldsymbol{\rho}'(\psi(b))$ corresponds to the point $\psi'_k(a) = \hat{C} \tilde{\psi}(\psi'_k(b), S_{n+1}(\psi(a), a))$ in the configuration $\boldsymbol{\rho}'(\psi(a))$ (the expressions are not presented due to obviousness and tediousness).

Since, by the hypothesis of the theorem, a bijective operator $\tilde{\psi}$ exists and the points of two arbitrary $\boldsymbol{\rho}'(\psi(a))$ and $\boldsymbol{\rho}'(\psi(b))$ are unambiguously assignable, then the features selected by the mask $\chi(\varphi_a(b))$ can also be applied to the objects of the set of precedents $\varphi_a(a)$. If conditions (6) and (7) are not satisfied, then the features obtained for different samples $a \in \hat{\zeta} X$ certainly do not correspond to each other; therefore,

there will be no question of cross-validation with regard to the selection of features.

Therefore, if (6) and (7) are satisfied, then, for every tested pair of samples (a,b), one obtains their own values of the radii $r_s(a,b)$, $r_r(a,b)$, $r_c(a,b)$, and $r_f(a,b)$ related to testing identical collections of features. Accordingly, the empirical distribution functions constructed over the sets of numbers $\{r_s(a,b)|a,b \in \hat{\zeta}X\}$, $\{r_r(a,b)|a,b \in \hat{\zeta}X\}$, $\{r_c(a,b)|a,b \in \hat{\zeta}X\}$, and $\{r_f(a,b)|a,b \in \hat{\zeta}X\}$ indeed characterize the satisfiability of the criteria in the cross-validation over the set of samples $\hat{\zeta}X$. The theorem is proved.

Conditions (6) and (7) in Theorem 7 are important conditions for the adequate performance of a cross-validation experiment. Satisfiability of the condition (6) can be technically implemented for sufficiently large sizes of samples and imposes an important constructive constraint on the choice of the sampling operator $\hat{\zeta}$: namely, it has to be chosen in such a way as to satisfy the cross-validation of the conditions (1.1)−(4.1). Condition (7), which guarantees correspondence between the elements of the set $S_{n+1}(\psi(a))$ for various $a \in \hat{\zeta}X$, defines the "determinacy" of the procedure of calculation of features for arbitrary samples from the set $\hat{\zeta}X$ and should be tested before the calculation of criteria (1.1)−(4.1) to select the most acceptable procedures of generation of $S_{n+1}(\psi(a))$.

The fulfillment of the conditions (6) and (7) is also important for the cross validation of the "compactness" criteria (1.3)−(4.3), since, as noticed earlier, the setting of the metrics ρ_q is performed on the basis of the calculations of the values of the metrics ρ_π. For example, suppose that, descriptions in the form of the sets $\varsigma_1 = (\varsigma_k^1)$ and $\varsigma_2 = (\varsigma_k^2)$, $k = 1, \ldots, n$, $\varsigma_k^1, \varsigma_k^2 \in L(\tau(X))$, are defined for objects q_1 and q_2. Then $\rho_q(q_1,q_2) = \Sigma(\{\upsilon_k(\varsigma_k^1,\varsigma_k^2)\})$, $\upsilon_k = \rho_\pi, \omega_k\rho_\pi, \ldots$, so that, in the case of the ρ_π-configuration $\boldsymbol{\rho}'(\psi(a)) = \boldsymbol{\rho}'(\psi(a), S_{n+1}(\psi(a), a))$, the distances in the ρ_q-configuration $\boldsymbol{\rho_q}(\boldsymbol{\rho}'(\psi(a)))$ are calculated as $\rho_q(\boldsymbol{\rho}'(\psi(a)))(q_1,q_2) = \Sigma(\{\upsilon_k(\boldsymbol{\rho}'(\psi(a)))(\varsigma_k^1(\psi(a)), \varsigma_k^2(\psi(a)))\})$.

Theorem 8. *The cross-validation testing of criteria* (1.3)−(4.3) *is possible if, for all elements of a given set of samples* $\hat{\zeta}X$, *the identical sets of feature values are defined, the aggregated feature descriptions generated for different samples are unambiguously assignable, and, for an arbitrary pair of samples* $a,b \in \hat{\zeta}X$, *the "aggregated" set of precedents* $\varphi_a(b)$ *contains a subset of elements that is equivalent to the set of precedents* $\varphi_a(a)$.

Proof. Consider two samples $a,b \in \hat{\zeta}X$ and the corresponding metric configurations $\boldsymbol{\rho_q}(\boldsymbol{\rho}'(\psi(a)))$ and $\boldsymbol{\rho_q}(\boldsymbol{\rho}'(\psi(b)))$. Under conditions (6) and (7), the feature descriptions of an arbitrary ith object in $\varphi_a(a) = \{q_i(a) = (m_i(a),\iota_i(a))\}$, $m_i(a) = m(\phi_1^i(a),\ldots,\phi_k^i(a),\ldots,\phi_{n_1}^i(a))$, and an arbitrary jth object in $\varphi_a(b) = \{q_j(b) = (m_j(b),\iota_j(b))\}$, $m_j(b) = (\phi_1^j(b),\ldots,\phi_l^j(b),\ldots,\phi_{n_1}^j(b))$, have identical dimension n_1. Moreover, there is a one-to-one correspondence between positions in $m_i(a)$ and $m_j(b)$, that is attributed to the existence of the bijection $\tilde{\psi}$ (7).

Therefore, if $q_i(a)$ and $q_j(b)$ represent the same object from X, then, for every $\phi_k^i(a)$, there exists a renumbering method $p(k,a,b)$ such that $\phi_k^i(a) = \phi_{p(k,a,b)}^j(b)$ and, for any $\phi_l^j(b)$, $\phi_l^j(b) = \phi_{p(l,b,a)}^i(a)$. Hence, for identical $q_i(a)$ and $q_j(b)$, there exists a transformation $\hat{p}(a,b)$ such that $q_i(a) = \hat{p}(a,b)q_j(b)$ and $q_j(b) = \hat{p}(b,a)q_i(a)$. Let us define $\hat{p}(a,b)$ so that, if there is no corresponding element for $q_i(a)$ in $\varphi_a(b)$, then $\hat{p}(b,a)q_i(a) = \varnothing$. Then one can define a subset of elements of the aggregated set of precedents $\varphi_a(a)$ for which there exist corresponding elements in the set $\varphi_a(b)$, $\varphi'(a|b) = \{q \in \varphi(a) | \hat{p}(b,a)q \neq \varnothing\}$. Under condition (8), for all objects of the set $\varphi_a(a)$, there exist corresponding objects in the set $\varphi_a(b)$:

(8) $|\varphi'_a(a|b)|/|\varphi_a(a)| = 1$.

Accordingly, the use of $\varphi'_a(a|b) = \varphi_a(a)$ makes it possible to express the *cross-validation forms of the "compactness" criteria* (1.3)−(4.3).

• *Solvability criterion of problems*:

(1.4) $\quad \underset{\hat{\zeta}X}{\forall} a,b,a \neq b \quad \underset{\varphi'_a(a|b)}{\forall} q_1,q_2 : \iota_1 \neq \iota_2 \quad \Rightarrow$

$\rho_q(\boldsymbol{\rho}'(\psi(b)))(\hat{p}(b,a)q_1, \hat{p}(b,a)q_2) > 0$.

• *Regularity criterion of problems*:

(2.4) $\underset{\hat{\zeta}X}{\forall} a,b,a \neq b \underset{\varphi'_a(a|b)}{\forall} q_1, q_2 : \rho_q(\boldsymbol{\rho}'(\psi(b)))(\hat{p}(b,$

$a)q_1, \hat{p}(b,a)q_2) > 0$.

• *Sufficient condition of existence of a correct algorithm and a complete algorithmic model*:

(3.4) $\underset{\hat{\zeta}_r X}{\forall} a, b, a \neq b : r_K(\{\hat{p}(b,a)q_i | q_i \in \varphi'(a|b),$

$\varphi^{-1}(a) \in C^+\}, \hat{\rho}_N^{-1}\boldsymbol{\rho_q}(\boldsymbol{\rho}'(\psi(b)))) = 1$.

The theorem is proved.

Corollary 1. *Criteria* (1.4)−(3.4) *imply the corresponding combinatorial functionals that describe the satisfiability of each of the criteria on "learning" sets from* $\hat{\zeta}X$:

$$r_{14l}(\hat{\zeta}X) = \frac{1}{Y}\sum_{i=1}^{Y} r_{13}(\varphi_a(a_i),\varphi_a(a_i)),$$

$$r_{24l}(\hat{\zeta}X) = \frac{1}{Y}\sum_{i=1}^{Y} r_{23}(\varphi_a(a_i),\varphi_a(a_i)),$$

$$r_{34l}(\hat{\zeta}_r X) = r_{43}(\hat{\zeta}_r X), \quad Y = |\hat{\zeta}X|$$

and on "control" sets:

$$r_{14c}(\hat{\zeta}X) = \frac{1}{Y(Y-1)}\sum_{i=1}^{Y}\sum_{j=1,j\neq i}^{Y} r_{13}(\varphi_a(a_i),\varphi_a(a_j)),$$

$$r_{24c}(\hat{\zeta}X) = \frac{1}{Y(Y-1)}\sum_{i=1}^{Y}\sum_{j=1,j\neq i}^{Y} r_{23}(\varphi_a(a_i),\varphi_a(a_j)),$$

$$r_{34c}(\hat{\zeta}_r X) = \frac{1}{Y(Y-1)}\sum_{i=1}^{Y}\sum_{j=1,j\neq i}^{Y} r_K(\varphi(C^+) \cap \varphi_a(a_i),$$

$$\hat{\rho}_N^{-1}\hat{\Lambda}\rho(\hat{\psi}\varphi_a(a_j))).$$

Corollary 2. *The "overfitting" of the algorithms for the metrization of the feature descriptions by criteria* (1.4), (2.4), *and* (3.4) *is estimated as* $r_{14l}(\hat{\zeta}X) - r_{14c}(\hat{\zeta}X)$, $r_{24l}(\hat{\zeta}X) - r_{24c}(\hat{\zeta}X)$, *and* $r_{34c}(\hat{\zeta}_r X) - r_{43}(\hat{\zeta}_r X)$, *respectively.*

Just as in the case of the functionals $r_{1c}(\hat{\zeta}X)$, $r_{2c}(\hat{\zeta}X)$, etc., instead of $r_{14l}(\hat{\zeta}X)$, $r_{14c}(\hat{\zeta}X)$, etc., one can use empirical distribution functions over the corresponding sets, which are compared by the methods of nonparametric statistics.

Thus, the conditions (6)–(8) and the Theorems 7 and 8 provide constructive criteria for cross-validation testing of the satisfiability of the "metric" conditions of solvability, regularity, correctness, and completeness. The sets $S_{n+1}(\psi(a),a)$, required for calculating $\rho'(\psi(a)) = \rho'(\psi(a), S_{n+1}(\psi(a),a))$, $a \in \hat{\zeta}X$, can be established on the basis of the analysis of the metric "grains" and "condensations" in a ρ-configuration [11].

Note also that the analysis of the "ε-compactness" property can be carried out, in addition to the combinatorial criteria (3.3), (3.4), (4.3), and (5), also by the analysis of the metric condensations in ρ_q-configurations: one finds grains/condensations, calculates the distance of each condensation from the class C^+ using the metric ρ_π, the fraction of objects of C^+ in each condensation, and so on. The function σ defining the σ-isomorphism condition (which is necessary for the analysis of metric condensations [11]) is calculated by comparing the corresponding distances in the ρ-configurations constructed over different pairs of elements of the set of samples $\hat{\zeta}X$.

8. CONCLUSIONS

On the basis of the metric criteria of solvability and regularity of recognition problems, correctness of algorithms, and completeness of algorithmic models obtained in the present study, we have formulated a universal general approach to the data analysis of poorly formalized problems. An important advantage of the metric approach over the previously formulated factorization approach [12] is that the metric approach avoids the arbitrariness associated with the choice of factorization procedures of feature descriptions. In a sense, the factorization is performed "automatically": the metric configuration contains points corresponding to all possible values of features, which then are "classified" during the analyses of metric condensations, aggregation of feature values, and so on. The proposed algorithm for data analysis of poorly formalized problems first verifies a necessary and sufficient condition for the existence of a topology over the set of initial feature descriptions, chooses the most adequate sampling operator (first of all, on the basis of the satisfiability of the condition of identity of the sets of feature values). Then, under the condition of matching of aggregated feature descriptions (which guarantees the "determinacy" of the feature generation procedure used), the cross-validation testing of the criteria is performed. The quality of the formalization methods (including the methods of calculating metrics in the space of features and in the space of objects) is assessed by the combinatorial functionals obtained.

REFERENCES

1. Yu. I. Zhuravlev, "Correct algebras for sets of incorrect (heuristic) algorithms. Part 1," Kibernetika, No. 4, 5–17 (1977).
2. Yu. I. Zhuravlev, "Correct algebras for sets of incorrect (heuristic) algorithms. Part 2," Kibernetika, No. 6, 21–27 (1977).
3. Yu. I. Zhuravlev, "Correct algebras for sets of incorrect (heuristic) algorithms. Part 3," Kibernetika, No. 2, 35–43 (1978).
4. Yu. I. Zhuravlev, "On algebraic approach for solving recognition and classification problems," in *Problems of Cybernetics* (Nauka, Moscow, 1978), issue 33, pp. 5–68 [in Russian].
5. K. V. Rudakov, "On some universal limitations for classification algorithms," Zh. Vychisl. Mat. Mat. Fiz. **26** (11), 1719–1729 (1986).
6. K. V. Rudakov, "Universal and local limitations in heuristic algorithms correction," Kibernetika, No. 2, 30–35 (1987).
7. K. V. Rudakov, "Completeness and universal limitations in the problem on correcting the heuristic classification algorithms," Kibernetika, No. 3, 106–109 (1987).
8. K. V. Rudakov, "The way to apply universal limitations for researching classification algorithms," Kibernetika, No. 1, 1–5 (1988).
9. I. Yu. Torshin and K. V. Rudakov, "On the theoretical basis of metric analysis of poorly formalized problems

of recognition and classification," Pattern Recogn. Image Anal. **25** (4), 577−579 (2015).

10. I. Y. Torshin and K. V. Rudakov, "On metric spaces arising during formalization of recognition and classification problems. Part 1: Properties of compactness," Pattern Recogn. Image Anal. **26** (2), 274−284 (2016).

11. I. Yu. Torshin and K. V. Rudakov, "On metric spaces arising during formalization of problems of recognition and classification. Part 2: density properties," Pattern Recogn. Image Anal. **26** (3), 483−496 (2016).

12. I. Yu. Torshin and K. V. Rudakov, "Combinatorial analysis of the solvability properties of the problems of recognition and completeness of algorithmic models. Part 1: Factorization approach," Pattern Recogn. Image Anal. **27** (1), 16−28 (2017).

13. K. V. Rudakov and I. Yu. Torshin, "The way to analyze motifs informativeness on the base of solvability criteria in the problem on recognizing secondary protein structure," Inf. Ee Primen. **6** (1), 79−90 (2012).

14. Yu. I. Zhuravlev, K. V. Rudakov, and I. Yu. Torshin, "Algebraic criteria of local solvability and regularity as an instrument for researching amino acid sequences," Trudy Mosk. Fiz.-Tekhn. Inst. **3** (4), 45−54 (2011).

15. I. Yu. Torshin, "The study of the solvability of the genome annotation problem on sets of elementary motifs," Pattern Recogn. Image Anal. **21** (4), 652−662 (2011).

16. T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (Springer-Verlag, 2009).

17. A. N. Kolmogorov, *Selected Works. Probability Theory and Mathematical Statistics* (Moscow, 1986) [in Russian].

Translated by I. Nikitin

Ivan Yur'evich Torshin. Born 1972. Graduated from the Department of Chemistry, Moscow State University, in 1995. Received candidates degrees in chemistry in 1997 and in physics and mathematics in 2011. Currently is an associate professor at Moscow Institute of Physics and Technology, lecturer at the Faculty of Computational Mathematics and Cybernetics, Moscow State University, leading scientist at the Russian Branch of the Trace Elements Institute for UNESCO, and a member of the Center of Forecasting and Recognition. Author of 225 publications in peer-reviewed journals in biology, chemistry, medicine, and informatics and of 3 monographs in the series "Bioinformatics in Post-genomic Era" (Nova Biomedical Publishers, NY, 2006−2009).

Konstantin Vladimirovich Rudakov. Born 1954. Russian mathematician, Full member of the Russian Academy of Sciences, Head of the Department of Computational Methods of Forecasting at the Dorodnicyn Computing Centre, Federal Research Center "Informatics and Control," Russian Academy of Sciences, and Head of the Chair "Intelligent Systems" at the Moscow Institute of Physics and Technology.