Methods for Recognition and Prediction Based on the Voting Procedures[1]

V. V. Ryazanov, O. V. Sen'ko, and Yu. I. Zhuravlev

Computer Center, Russian Academy of Sciences, ul. Vavilova 40, Moscow, 117967 Russia
e-mail: Rvv@ccas.ru

Abstract—The algorithms for pattern recognition and prediction based on the procedures of voting over the systems of logical regularities are considered. A logical regularity is defined as a certain subdomain in a multidimensional space of prognostic variables; the objects of one of the classes to be recognized are assumed to be prevalent in such a subdomain. The methods aimed at searching for the logical regularities and based on performing the optimal partitions of the domains of admissible values of individual features and on the search for optimal neighborhoods of standards from the learning sample in the subspaces of feature descriptions are considered. The methods for constructing the voting procedure (including the method based on optimization of the likelihood function of a special type) are discussed.

INTRODUCTION

Consider the problem of the recognition of objects belonging to L classes, i.e., $K_1, \ldots, K_L$. In this case, initial learning information or learning sample may be represented as a set of objects or standards $\tilde{S}_0 = \{(\bar{\alpha}_1, I(S_1)), \ldots, (\bar{\alpha}_n, I(S_m))\}$, where $\bar{\alpha}_j = (\alpha_j^1, \ldots, \alpha_j^L)$ is the indicator vector of the object S_j or $\alpha_j^i = 1$ if $S_j \in K_i$ or $\alpha_j^i = 0$ otherwise. Typically, the description $I(S_j)$ of the object S_j is the vector $\bar{x}_j = (x_{j1}, \ldots, x_{jn})$ of the values of the features $X_1, \ldots, X_n$. It is assumed that the classes $K_1, \ldots, K_L$ are disjoint.

At the stage of learning, we use initial learning information to construct the set of so-called logical regularities in the multidimensional space of features R^n. By logical regularity, a subdomain is meant that belongs to the space R^n and contains the objects of a single class. We may also use a loose definition when the objects of several classes are allowed to be included in a logical regularity; however, in this case, the objects of one class are presumed to be prevalent.

We assume that it is required to recognize the object S^* with the description $\bar{x}^*$ belonging to the logical regularities $Q_1, \ldots, Q_k$ from $\tilde{Q}_0$. In this case, the assessment of the object S^* as belonging to the class K_l is calculated as the weighted sum

$$\Gamma_l(S^*) = \left[\sum_{t=1}^{k} \gamma_{tl} \mathrm{wei}_{tl}\right] / \left[\sum_{t=1}^{k} \mathrm{wei}_{tl}\right]. \tag{1}$$

Here, γ_{tl} is the assessment of the logical regularity Q_t as belonging to the class K_l and wei_{tl} is the weight of the logical regularity Q_t in the class K_l. There are several methods for calculating the quantities wei_{tl} and γ_{tl}. For example, we may assume that $\gamma_{tl} = 1$ if the number of objects belonging to the class K_l in the logical regularity Q_t exceeds the number of objects belonging to any other class; otherwise, we assume that $\gamma_{tl} = 0$. The quantity γ_{tl} may be also taken as equal to the fraction of objects belonging to the class K_l among all objects included in the logical regularity Q_t.

There are also several methods for calculating the weights wei_{tl}. For example, it may be assumed that the weight of a logical regularity is equal to the number of objects of the learning sample this sequence contains. The method for calculating the weights of base sets and based on the optimization of a special maximum-likelihood function is outlined in Section 3.

In the case of the problem of predicting the values of stochastic function Y taking the values from an interval in the real axis, learning information $\tilde{S}_0$ can be represented as a set of objects $\{(y_1, \bar{x}_1), \ldots, (y_n, \bar{x}_m)\}$, where $y_1, \ldots, y_m$ are the values of the function Y measured at the points $\bar{x}_1, \ldots, \bar{x}_m$. As in the case of the pattern recognition problem, a set of logical regularities $\tilde{Q}_0$ is constructed.

Let us assume that we want to predict the value of Y at a certain point $\bar{x}$ belonging to the logical regularities

[1] This work was supported by the Russian Foundation for Basic Research (project nos. 97-01-00495 and 99-01-00433) and the International Science Foundation (grant no. 96-952).

Received July 16, 1999

$Q_1, \ldots, Q_k$ from $\tilde{Q}_0$. We first use the voting procedure to calculate the estimate $\vartheta_{vot}(\bar{x})$ as

$$\vartheta_{vot}(\bar{x}) = \left[\sum_{i=1}^{k} \tilde{y}_i \mathrm{wei}_i\right] / \left[\sum_{i=1}^{k} \mathrm{wei}_i\right], \quad (2)$$

where $\tilde{y}_i$ is the value of function Y averaged over all objects $\tilde{S}_0$ belonging to Q_i; i.e., $\tilde{y}_i = \left[\sum_{\bar{x}_j \in Q_i} y_j\right] / m_i$. The predicted value $\vartheta(\bar{x})$ is calculated by a simple linear transformation of the estimate obtained by voting: $\vartheta(\bar{x}) = \alpha\vartheta_{vot}(\bar{x}) + \beta$.

The coefficients α and β are determined from a learning sample by the method of least squares. The weights wei_i in (2) are calculated by the method to be considered in Section 3. The problem of recognizing the patterns for two disjoint classes K_1 and K_2 may be treated as the problem of approximation of binary indicator function belonging to the class K_1.

1. THE METHOD FOR SEARCHING FOR LOGICAL REGULARITIES BASED ON PARTITIONING THE DOMAINS OF ADMISSIBLE VALUES OF INDIVIDUAL FEATURES

As experience shows, it is quite sufficient in many cases to use the procedure for partitioning the ranges of admissible values of individual features in searching for logical regularities. In this case, partitioning of the range of admissible values of the feature X_i is performed over a set of the pairs $\{(\bar{\alpha}_1, x_{1i}), \ldots, (\bar{\alpha}_m, x_{mi})\}$ for recognizing the patterns, or over a set of the pairs $\{(y_1, x_{1i}), \ldots, (y_m, x_{mi})\}$ for stochastic approximation of continuous functions.

Let r be an element in the partition of the range of admissible values of the feature X_i. A set of points of multidimensional feature space R^n with the component $X_i \in r$ is referred to as the logical regularity formed by the element r.

Let $\tilde{R}$ be a partition including the elements $r_1, \ldots, r_k$ and let $q_1, \ldots, q_k$ be the logical regularities formed by such elements. The function Y defined on the logical regularity q_l may be approximated by the average value $\tilde{y}_l = \sum_{x_{jl} \in r_l} y_j / m_l$, where m_l is the number of objects of the learning sample with $X_i \in r_l$.

The functional of the quality of partition F_1 is specified as a normalized sum of the squared deviation of the prediction from the true values over all of the objects of the learning sample. In this case, the predicted value for the object s_j having $X_i \in r_l$ is calculated as the value averaged over all objects that belong to the learning sample, differ from s_j, and have $X_i \in r_l$; thus, we have

$$F_1 = 1 - \left\{\sum_{l=1}^{k} \sum_{x_{jl} \in r_l} [(y_j - (\tilde{y}_l m_l - y_j))/(m_l - 1)]^2\right\} / D,$$

where $D = \sum_{j=1}^{m} (y_j - \tilde{Y})^2$, with $\tilde{Y} = \sum_{j=1}^{m} y_j / m$.

The other functional F_2 is equal to the sum of squared deviations of the averaged quantities $\tilde{y}_1$ from $\tilde{Y}$, with these deviations being multiplied by the number of objects in the corresponding elements; i.e., we have

$$F_2 = \sum_{l=1}^{k} (\tilde{y}_l - \tilde{Y})^2 m_l.$$

An important property of a partition is its complexity. By the complexity of a partition, we mean the number of elements it contains.

An increase in the partitioning complexity results in an increase in the value of the quality functional. Every so often, however, an increase in accuracy of approximation based on learning information constitutes, in fact, a mere fitting procedure and does not yield an increase in the accuracy of prediction for new objects. One of the possible ways to optimize the complexity is based on the estimate of the stability of the partition boundaries in the case of slight variations in the learning sample. Let $b_1, \ldots, b_k$ be the boundaries to a certain optimal partition R constructed on the basis of learning information $\tilde{S}_0$. We denote the boundaries defined on the basis of the sample $\tilde{S}_0$ without the object s_j by $b_1^j, \ldots, b_k^j$. The following functional F_{st} may be used as a measure of stability of the boundaries:

$$F_{st} = 1 - \sum_{i=1}^{k} \sum_{j=1}^{m} (b_i^j - b_j)^2 / (kD\backslash).$$

In actual calculations, we seek a partition that maximizes the quality functionals F_1 or F_2, with the value of the functional F_{st} exceeding a threshold value specified beforehand.

2. A MODEL OF PRECEDENT-BASED LOGICAL REGULARITIES

For each object $S_\mu \in K_\lambda$ from the learning table, we seek an "optimal" pair (ω, ε), with $\omega = \{i_1, i_2, \ldots, i_k\}$, and $\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n) \geq 0$, that specifies the neighborhood of the learning object in the subspace of feature descriptions. This problem is multiextremal. Therefore, we may seek a set of local optimal solutions and use them in the voting procedures, by analogy with representative samples. Each such a pair (ω, ε) defines the predicate

$$P_i(S) = \&(a_{\mu j} - \varepsilon_j \le x_j \le a_{\mu j} + \varepsilon_j) \qquad (3)$$

that possesses the following properties:

(a) $P_i(S_\tau) = 1$ for a certain reference object $S_\tau \in K_\lambda$;

(b) $P_i(S_\tau) = 0$ for all reference objects $S_\tau \notin K_\lambda$; and

(c) $\phi(P_i) = \max$, where ϕ is a certain criterion of optimality.

Each predicate $P_i(S)$ will be referred to as the logical regularity of the corresponding class. In this section, we consider the problem of searching for optimal subsets of the parallelepiped type in the space of admissible feature descriptions of the objects S_t, $I(S_t) = (a_{t1}, a_{t2}, \ldots, a_{tm})$, of the class K_j. We now introduce additional designations and definitions. By the feature neighborhood of the object S_t, we mean the subset $O_\lambda(\varepsilon_\lambda) = \{x: a_{t\lambda} - \varepsilon_\lambda \le x_\lambda \le a_{t\lambda} + \varepsilon_\lambda\}$. Henceforth, we omit the superfluous indices, but we always keep in mind that we deal with the object S_t. The feature neighborhoods are assumed to be equivalent if $O_\lambda(e_l) \cap I_0 = O_\lambda(e_2) \cap I_0$.

By the system $\tilde{O}(\omega, \varepsilon)$ of feature neighborhoods the set $\{O_\lambda(\varepsilon_\lambda), \lambda \in \omega\}$ is meant, where $\omega = \{i_1, i_2, \ldots, i_k\} \subseteq (1, 2, \ldots, n)$ is a subset of features and $\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n) \ge 0$ is the vector of numerical parameters.

The system $\tilde{O}(\omega, \varepsilon)$ is referred to as the separating system for the class K_j if the relation $\forall S_v \notin K_j$ is valid: $S_v \notin \cap O_\lambda(\varepsilon_\lambda)$, with $\lambda \in \omega$.

The number set $\Sigma_\lambda = \{\sigma_{i\lambda} = |a_{i\lambda} - a_{t\lambda}|, i = 1, 2, \ldots, m\}$ $(\lambda = 1, 2, \ldots, n)$ is put into correspondence with the following numerical sequences $\tilde{N}_\lambda = \sigma_{\lambda 1}, \sigma_{\lambda 2}, \ldots, \sigma_{\lambda h}$, $h = h(\lambda)$:

(I) $0 \le \sigma_{\lambda i} < \sigma_{\lambda i+1}$, $i = 1, 2, \ldots, h-1$;

(II) $\forall e \in \Sigma_\lambda$, $\exists i$, $1 \le i \le h$: $e = \sigma_{\lambda i}$.

An arbitrary neighborhood $O_\lambda(\varepsilon_\lambda)$ is equivalent to the neighborhood $O_\lambda(e)$, where $e = \max\{\sigma_{\lambda i}: \sigma_{\lambda i} \le \varepsilon_\lambda, i = 1, 2, \ldots, h\}$. Therefore, it is sufficient to consider only the finite subsets $\{O_\lambda(\varepsilon_\lambda), \varepsilon_\lambda \in \tilde{N}_\lambda\}$ of the parametric families $\{O_\lambda(\varepsilon_\lambda), \varepsilon_\lambda \ge 0\}$.

We assume that

$$\theta_{\lambda i}(\varepsilon_\lambda) = \begin{cases} 0, & |a_{i\lambda} - a_{t\lambda}| \le \varepsilon_\lambda, \\ 1, & |a_{i\lambda} - a_{t\lambda}| > \varepsilon_\lambda; \end{cases}$$

$$\gamma_{\lambda i}(\varepsilon_\lambda) = \begin{cases} 1 - \theta_{\lambda i}, & \alpha(S_i) = \alpha(S_t), \\ \theta_{\lambda i}, & \alpha(S_i) \ne \alpha(S_t), \end{cases}$$

where $\alpha(S_i) = (\alpha_1(S_i), \alpha_2(S_i), \ldots, \alpha_l(S_i))$ and

$$\alpha_v(S_i) = \begin{cases} 1, & S_i \in K_v, \\ 0, & S_i \notin K_v. \end{cases}$$

We refer to the functionals f: $\{O_\lambda(\varepsilon_\lambda)\} \longrightarrow R^+$ and F: $\{\tilde{O}(\omega, \varepsilon)\} \longrightarrow R^+$ as the local and global functionals, respectively.

We introduce the following family of local functionals:

$$f(O_\lambda(\varepsilon_\lambda); \pi_0, \pi_1, \ldots, \pi_l), \quad \pi_i \ge 0, \quad i = 1, 2, \ldots, l,$$

$$\pi_v > 0, \text{ if } \alpha_v(S_t) = 1, \quad \pi_v - \text{const}:$$

$$f(O_\lambda(\varepsilon_\lambda)) = \sum_{\mu=1}^{l} \pi_\mu \sum_{i=m_{\mu-1}}^{m_\mu} (1 - \gamma_{\lambda 1}(\varepsilon_\lambda)) + \pi_0.$$

Let $H(\lambda) = \{1, 2, \ldots, h(\lambda)\}$. By $D = \{\tilde{O}(\omega, \varepsilon)\}$ we denote a set of all systems $\tilde{O}(\omega, \varepsilon)$ that are the separating systems for the class K_j. We assume that $F(\tilde{O}(\omega, \varepsilon)) = \sum_{\lambda \in \omega} f(O_\lambda(\varepsilon_\lambda))$.

The problem of searching for the optimal system of neighborhoods that acts as a separating system for the class K_j can be formulated as follows:

$$F(\tilde{O}(\omega, \varepsilon)) \longrightarrow \min, \quad \tilde{O}(\omega, \varepsilon) \in D. \qquad (4)$$

Let $E_\lambda = \{e_{\lambda u}, u = 1, 2, \ldots, \kappa(\lambda)\}$ be a certain subsequence of $\tilde{N}_\lambda$.

We introduce the matrix $B = \|\beta_{vi}(e_{\lambda u})\|_{\mu \times \eta}$, where $\mu = m - m_j + m_{j-1}$, $\eta = \sum \kappa(\lambda)$, $\lambda = 1, 2, \ldots, n$; and

$$\beta_{vi}(e_{\lambda u}) = \begin{cases} 0, & |a_{t\lambda} - a_{t\lambda}| \le e_{\lambda u} \\ 1, & |a_{\tau\lambda} - a_{t\lambda}| > e_{\lambda u}. \end{cases} \qquad (5)$$

Here, the quantities $\beta_{vi}(e_{\lambda u})$ are obtained by comparing $I(S_t)$ with the reference description $I(S_\tau) \notin K_j$; the rows $\beta_i(e_{\lambda u})$ ($i = \sum_{l=0}^{\lambda-1} \kappa(l) + u$ and $\kappa(n)\kappa(0) = 0$) are ordered according to the following rules: $\lambda = 1$, $u = 1, 2, \ldots, \kappa(1)$; $\lambda = 2$, $u = 1, 2, \ldots, \kappa(2)$; ...; $\lambda = n$, $u = 1, 2, \ldots$. The row $\beta_i(e_{\lambda u})$ is assigned the following scalar weight:

$$p_{\lambda u} = f(O_\lambda(e_{\lambda u})) \ge 0. \qquad (6)$$

We now formulate the following problem:

$$\sum_{\lambda=1}^{n} \sum_{\mu=1}^{\kappa(\lambda)} p_{\lambda\mu} y_{\lambda\mu} \longrightarrow \min, \qquad (7)$$

$$\sum_{\lambda=1}^{n}\sum_{\mu=1}^{\kappa(\lambda)} \beta_{\nu i}(e_{\lambda\mu})y_{\lambda\mu} \geq 1, \quad \nu = m - m_j + m_{j-1}, \tag{8}$$

$$y_{\lambda\mu} \in \{0, 1\}. \tag{9}$$

Here, the coefficients are defined according to (5) and (6).

It was proved [2] that problem (4) may be reduced with polynomial complexity to the equivalent problem (7)–(9) in such a way that the subsequences E_λ ($\lambda = 1, 2, \ldots, n$) are irreducible, and the coefficients in (7) and (8) possess the properties

$$p_{\lambda\mu} > p_{\lambda\mu+1} \geq 0, \quad \beta_{\nu i}(e_{\lambda\mu}) \geq \beta_{\nu i+1}(e_{\lambda\mu+1}), \tag{10}$$

for a fixed λ.

The problem given by (7)–(9) with the properties of coefficients defined by (10) is known as the problem of integer linear programming with block–monotone columns of the constraint-matrix coefficients and the objective-functional coefficients (BMC-ILP) [2]. Since the coefficients are monotone, no more than a single value of $y_{\lambda\mu}$ can exist in the solution for each value of the subscript λ. The identity components of optimal solution are in one-to-one correspondence with the system of neighborhoods $\tilde{O}(\omega, \varepsilon)$ that serve as separating boundaries for the class K_j.

We note in the conclusion that $\tilde{O}(\omega, \varepsilon)$, $\omega = \{i_1, i_2, \ldots, i_k\}$ corresponds to the hyperparallelepiped $H_i = \{x\colon a_{ij_\nu} - \varepsilon_{ij_\nu} \leq x_{ij_\nu} \leq a_{ij_\nu} + \varepsilon_{ij_\nu}, \nu = 1, 2, \ldots, k_i\}$ that encompasses certain standards of only a single class K_j and has the object S_ν as the central element.

3. THE METHOD FOR STATISTICALLY WEIGHTED VOTING

We assume that the function Y depends stochastically on the parameters (features). We also assume that we have at our disposal learning information represented by a set of pairs $\tilde{S}_{in} = \{(y_1, \bar{x}_1), \ldots, (y_m, \bar{x}_m)\}$, where $y_1, \ldots, y_m$ are representations of the random function Y and $\bar{x}_1, \ldots, \bar{x}_m$ are the vectors of the values of features $X_1, \ldots, X_n$ corresponding to the specified representations of Y. The problem is posed to predict the value of the function Y for any admissible values of the features $X_1, \ldots, X_n$ by using the learning sample $\tilde{S}_0 \subseteq \tilde{S}_{in}$. In order to predict the values of Y, we suggest using the procedure of voting over the systems of logical regularities.

We assume that the σ-algebra S is defined in the feature space R^n and P is the probabilistic measure on (R^n, S). We assume that the system of logical regularities is given by $\tilde{Q}_0 \subseteq S$. We denote the minimal algebra containing $\tilde{Q}_0$ by $\tilde{Q}_0^1$. Let $\tilde{Q}_0^2$ stand for the set of elements $\tilde{Q}_0^1$ that cannot be represented by a union of other elements $\tilde{Q}_0^1$.

The predicted value of the function Y at an arbitrary point of the feature space is calculated from data on distribution of objects of the learning sample over logical regularities of the system $\tilde{Q}_0$. Let a certain point $\bar{x}$ belong to the set $q \in \tilde{Q}_0^2$. As the predicted value of the function Y at the point x, we use the following estimate of conditional mathematical expectation:

$$M(Y|q) = \frac{\int_q^{P(Q)} YP(d\omega)}{P(Q)}.$$

Definition. The function $\xi_u(\bar{x})$ at the point $\bar{x}$ is defined by the equality $\xi_u(\bar{x}) = M(Y|q)$, where q is the element belonging to $\tilde{Q}_0^2$ and containing the point $\bar{x}$.

It is easy to verify that, for any $Q \in \tilde{Q}_0$, we have $M(Y|Q) = M(\xi_u|Q)$.

Indeed,

$$M(Y|Q) = \frac{\sum_{q \subset Q} \int_q YP(d\omega)}{P(Q)}$$

$$= \frac{\sum_{q \subset Q} \int_q YP(d\omega)}{P(Q)} = M(\xi_u|Q).$$

The quantity V_j is defined as the value of representations of the function Y averaged over all the objects of the learning sample from the logical regularity Q_j or $V_j = \frac{1}{n_j}\sum_{\bar{x}_i \in Q_j} y_i$.

Let the point $\bar{x}$ belong to the intersection of the logical regularities $Q_1, \ldots, Q_p$. The distribution densities of the values $V_1, \ldots, V_p$ of the function ξ_u within the logical regularities $Q_1, \ldots, Q_p$ are approximated with the use of the normal law; i.e., we have

$$f_j^0(V_j) = N_j^0(V_j, M(Y|Q_j))$$

$$\sim \exp\left[-\frac{n_j(V_j - M(Y|Q_j))^2}{(2D(Y|Q_j))}\right];$$

$$f_j^1(\xi_u) = N_j^1(\xi_u, M(Y|Q_j))$$

$$\sim \exp\left[-\frac{(\xi_u - M(Y|Q_j))^2}{(2D(\xi_u|Q_j))}\right];$$

$$j = 1, \dots, p.$$

In the case where the number of objects in the base sets is large, the use of the normal distribution for approximating the densities $f_j^0(V_j)$ can be substantiated employing the central limit theorem.

It is also worth noting that the entropy functional $H(f) = \int (x - M)^2 f(x)dx$ specified at a set of probabilistic distributions with the fixed mathematical expectation $M = \int xf(x)dx$ and variance $D = \int (x - M)^2 f(x)dx$ attains a maximum over the normal distribution $f = \sqrt{(1/2\pi)}\exp(-(x - M)^2/(2D))$. From this standpoint, the normal distribution is closest to the uniform one. Therefore, we may use the normal distribution to approximate a certain probability distribution in the case where only the mathematical expectation and variance are known for the latter distribution.

The likelihood function L is specified as the product of the functions f_j^0 and f_j^1; thus, we have

$$L(z_0, \dots, z_p) \sim \prod_{j=1}^{p} N_j^1(z_0, z_j) N_j^0(V_j, z_j). \quad (11)$$

Here, the conditional mathematical expectations $M(Y|q)$, $M(Y|Q_1)$, ..., $M(Y|Q_p)$ are regarded as unknown parameters $z_0, \dots, z_p$, respectively. The point corresponding to the maximum of function (9) coincides with the position of the minimum of the function

$$\Im(z_0, \dots, z_p) = \sum_{j=1}^{p} [n_j (V_j - z_j)^2 (2D(Y|Q_j))^{-1} + (z_0 - z_j)^2 (2D(\xi_u|Q_j))^{-1}]. \quad (12)$$

The following equations constitute the necessary conditions for minimum of the function $\Im(z_0, \dots, z_p)$:

$$\frac{\partial\Im}{\partial z_0} = \sum_{j=1}^{p} b_j^0 (z_0 - z_j)$$

....

$$\frac{\partial\Im}{\partial z_p} = (b_p^0 + b_p^1) z_p - b_p^0 z_0 - c_p = 0. \quad (13)$$

Here,

$$c_j = \frac{n_j V_j}{D(Y|Q_j)}, \; b_j^0 = \frac{1}{D(\xi_u|Q_j)}, \; b_j^1 = \frac{n_j}{D(Y|Q_j)}. \quad (14)$$

It is easy to verify that system (11) has a unique solution

$$z_0 = M_0 = \frac{\sum_{j=1}^{p} c_j b_j^0 (b_j^0 + b_j^1)^{-1}}{\sum_{j=1}^{p} b_j^1 b_j^0 (b_j^0 + b_j^1)^{-1}},$$

....

$$z_p = M_p = \frac{(b_p^0 M_0 + c_p)}{b_p^0 + b_p^1}.$$

We propose that the quantities $M_0, M_1, \dots, M_p$ be regarded as the estimates of conditional mathematical expectations $M(Y|q)$, $M(Y|Q_2)$, ..., $M(Y|Q_1)$. As mentioned above, the estimate M_0 of the conditional mathematical expectation $M(Y|q)$ is used as a prediction $\phi(\bar{x})$ of the function Y at the point $\bar{x}$. As a result, we derived the following formula for the estimate $\phi(\bar{x})$ obtained by voting: $\phi(\bar{x}) = \frac{\sum_{i=1}^{p} V_j \mathrm{wei}_j}{\sum_{i=1}^{p} \mathrm{wei}_j}$.

Here, wei_j is the weight of the logical regularity Q_j. It is calculated from the formula $\mathrm{wei}_j = \frac{n_j}{n_j D(\xi_u|Q_j) + D(Y|Q_j)}$ or $\mathrm{wei}_j = \frac{n_j}{n_j \chi_j + 1} \frac{1}{D(Y|Q_j)}$, where the coefficient $\chi_j = \frac{D(\xi_u|Q_j)}{D(Y|Q_j)}$ defines the fraction of determinate (completely specified by predictable variables) component of variation of the function Y over the logical regularity Q_j. In actual representations, we used the values of χ_j equal to 1. The values of $D(Y|Q_j)$ were estimated on the basis of the learning sample.

REFERENCES

1. Zhuravlev, Yu.I., On Algebraic Approach to Solving the Problems of Recognition and Classification, *Probl. Kibern.*, Moscow: Nauka, 1978, no. 33, pp. 5–68.
2. Ryazanov, V.V., Recognition Algorithms Based on Local Optimality Criteria, *Pattern Recognit. and Image Anal.*, 1994, vol. 4, no. 2, pp. 98–109.
3. Dmitriev, A.N., Zhuravlev, Yu.I., and Krendelev, F.P., On Mathematical Principles of Classification of Objects and Phenomena, in *Diskretnyi analiz* (Discrete Analysis), Novosibirsk: Inst. Mat., Sib. Otd., Ross. Akad. Nauk, no. 7, pp. 3–11.

4. Baskakova, L.V. and Zhuravlev, Yu.I., A Model of Recognition Algorithms with Representative Samples and Systems of Support Sets, *Zh. Vychisl. Mat. Mat. Fiz.*, 1981, vol. 21, no. 5, pp. 1264–1275.

5. Sen'ko, O.V., The Algorothm of Prognosis, Based on the Procedure of Voting by System of Boxes on Multidimensional Space, *Pattern Recognit. and Image Anal.*, 1993, vol. 3, no. 3, pp. 283–284.

6. Sen'ko, O.V., Transformation of Continuous Features into a Discrete Form on the Basis of Optimization of the Approximation-Accuracy Functional, *Trudy 4 mezhdunarodnoi konferentsii PPIP-97* (Proc. 4th Int. Conf. PRIP-97), Minsk, 1997, pp. 18–20.

Vladimir V. Ryazanov. (See p. 553).

Oleg V. Sen'ko. (See p. 553).

Yuri I. Zhuravlev. (See p. 791).