A Recognition Method Based on Collective Decision Making Using Systems of Regularities of Various Types

O. V. Senko[a] and A. V. Kuznetsova[b]

[a] *Dorodnicyn Computing Centre, Russian Academy of Sciences, ul. Vavilova 40, Moscow, 119333 Russia*

[b] *Emanuel Institute of Biochemical Physics, Russian Academy of Sciences, ul. Kosygina 4, Moscow, 117977 Russia*

e-mail: senkoov@mail.ru, azfor@narod.ru

Abstract—A new recognition method that implies weighted voting on systems of "syndromes," i.e., subregions of the attribute space where objects of one class dominate, is given. It is a modified version of statistically weighted syndromes developed previously. To find syndromes, it searches for optimal partitions within several models of different levels of complexity. Syndromes to be included in the final set used in collective decision making are selected by the criterion for the partitioning degree of classes and by the parameter related to the complexity of the partitioning model involved. The weighted voting procedure can be interpreted as the convex correction of sets of predictors. The generalizing potential of such procedures is discussed. Experimental results of comparing the given method with the previous version (SWS) and alternative techniques are presented. To estimate the efficiency, several criteria are used, including a way to analyze recognition accuracy on the totality of all possible decision rules (ROC analysis).

Key words: recognition, collective decision making, optimal partition, convex correction.

DOI: 10.1134/S1054661810020069

1. INTRODUCTION

We consider a recognition problem in its typical statement. We assume that objects of an entire assembly Ω that is the union of nonoverlapping classes K_1, ..., K_l are described by the attributes $X_1, ..., X_n$. In addition, we assume that there is a sample of objects $\tilde{S}_o = \{s_1 = (\alpha_1, \mathbf{x}_1), ..., s_m = (\alpha_m, \mathbf{x}_m)\}$ independently taken from Ω, where $\mathbf{x}_j = (x_{1j}, ..., x_{nj})$ is the vector of values of variables $X_1, ..., X_n$ for the object s_j, α_j is the integer indicator showing the class the object s_j, $j = \overline{1, m}$ belongs to. Use the learning sample $\tilde{S}_o$ to find the algorithm that would ensure the most accurate recognition of arbitrary objects from Ω over some subset of attributes from $X_1, ..., X_n$. In this work, we consider the recognition method that evolves the statistically weighted syndrome (SWS) method first proposed in [6]. SWS implies collective decision making over systems of "syndromes," that is, subregions of the attribute space where objects of one class dominate. It belongs to a rather wide class of collective methods, including the known models such as the test algorithm [2], KORA-type algorithms [3, 4], and methods employing logical regularities [1, 5]. SWS was successfully applied to solve a number of problems of biomedical diagnostics [6–8].

Received January 11, 2010

2. SWS

At its initial stage, the SWS constructs the set of "syndromes" $\tilde{Q}_j$ for each K_j using the learning information $\tilde{S}_o$. A syndrome is a subregion of the attribute space such that the fraction of objects of at least one class differs significantly from the fraction of objects of this class included in the entire sample or in the neighboring subregions. Suppose the object s^* is recognized using the vector description $\mathbf{x}^*$ that belongs to the intersection of the syndromes $q_1, ..., q_r$ of the system $\tilde{Q}_j$. The estimate of the object s^* for the class K_j is calculated as the prediction of the indicator function $I_j(s)$ of the class K_j at the point $\mathbf{x}^*$. The prediction (estimate) is calculated by statistically weighted voting [9] by the formula

$$\Gamma_j(s^*) = \frac{\sum_{i=1}^{r} w_i^j \nu_i^j}{\sum_{i=1}^{r} w_i}, \quad (1)$$

where ν_i^j is the fraction of the class K_j in the sample $\tilde{S}_o$ with the vector descriptions $\mathbf{x}$ from the "syndrome" q_i, and w_i is the so-called weight of the i-th "syndrome." SWS uses weights obtained analytically under the maximum condition of the special quasi-likelihood functional [9] and calculated by the formula

$$w_i = \frac{m_i}{m_i\kappa_i + 1}\frac{1}{\hat{d}_i}, \qquad (2)$$

where $\hat{d}_i$ is the squared sample variance of the indicator function of the class K_j on the objects $\tilde{S}_o$ with the descriptions **x** from the "syndrome" q_i. The relation d_i^c/d_i is the determinacy coefficient κ_j, where d_i^c is the squared variance of the function $\xi_j(\mathbf{x}) = E[I_j|q(\mathbf{x})]$ inside the "syndrome" q_i. The determinacy coefficient shows the degree of dependence of the indicator function on the variables X inside the syndrome q_i. If $\kappa_j = 1$, the variables X define the value $I_j(s)$ completely. If $\kappa_j = 0$, the indicator random function $I_j(s)$ is independent of the variables X. To calculate the coefficient κ_j, one can apply an iterative procedure that gradually becomes self-consistent. However, numerous experiments on real data showed that rather satisfactory results are obtained for $\kappa_j = 1$, which is the value of κ_j used to implement the method.

Once estimate (2) is calculated, the standard decision rule is applied; i.e., the object s^* is attributed to the class $K_{j_{\max}}$, where $\Gamma_{j_{\max}}(s^*)$ is the maximum value of the estimates $\Gamma_1(s'), \ldots, \Gamma_l(s')$.

Optimal partitions of intervals of admissible values of prognostic attributes are used to search for "syndromes." Thus, for an arbitrary attribute X_i, we search for the optimal partition R on T_R of subintervals $a_1, \ldots, a_{T_R}$. The partition is considered optimal if the quality functional

$$F_g(R, \tilde{S}_o, K_j) = \frac{1}{\nu_0^j(1-\nu_0^j)}\sum_{t=1}^{T_R}(\nu_l^j - \nu_0^j)^2 m_t, \qquad (3)$$

where m_t is the number of objects in $\tilde{S}_o$, for which $X_i \in a_t$, ν_t^j is the fraction of objects K_j among the objects $\tilde{S}_o$ with $X_i \in a_i$, and ν_0^j is the fraction of objects K_j in the entire learning sample, attains its maximum on it.

The critical SWS parameter is the number of subintervals in the partitioning models involved. Practice has shown that the greater the number of elements in the partition, the higher the instability, given relatively small sample sizes (less than 100–200 objects). Therefore, the model complexity must be adjusted depending on the problem. The functional $F_{\text{ins}}(R, \tilde{S}_o, X_i, K_j)$, i.e., the squared ratio between the averaged variation of boundaries of optimal partitions and the sample variance X_i, is used as the SWS parameter to control the complexity of the partitioning model of the attribute X_i. We assume that $b_1^g, \ldots, b_{T_{R-1}}^g$ are the boundaries of the optimal partitions constructed using the learning sample with the g-th object excluded, and $\hat{b}_t = \frac{1}{m}\sum_{t=1}^{m} b_t^g$ is the mean position of the t-the boundary. Then, we can write the boundary instability functional as

$$F_{\text{ins}}(R, \tilde{S}_o, X_i, K_j)$$
$$= \left[\frac{1}{T_R - 1}\frac{1}{m}\sum_{g=1}^{m}\sum_{t=1}^{T_R-1}(b_t^g - \hat{b}_t)^2\right] \Big/ \sum_{k=1}^{m}(x_{ik} - \hat{x}_i)^2.$$

The functional $F_{\text{ins}}(R, \tilde{S}_o, K_j)$ is used to choose the optimal number of boundaries in the partitioning model and select attributes. The user sets the threshold δ_{ins} for the value $F_{\text{ins}}(R, \tilde{S}_o, X_i, K_j)$. If $F_{\text{ins}}(R, \tilde{S}_o, X_i, K_j) < \delta_{\text{ins}}$ for some attribute, we move to a simpler model with fewer boundary points. If the simplest model with one boundary point turns out to be unstable, the attribute X_i is excluded from the recognition process.

Note that SWS uses the threshold value for the functional $F_g(R, \tilde{S}_o, K_j)$ to select attributes. SWS uses "one-dimensional" syndromes that correspond to subintervals of partitions and various "two-dimensional" syndromes that are mutual intersections of "one-dimensional" syndromes.

Decision rule. Estimates of the recognized object s^* for the classes $K_1, \ldots, K_l$ are calculated by formula (1). However, an additional decision rule is needed to attribute s^* to the classes. A simple decision rule that attributes an object to a class turns out to be efficient when fractions of objects of different classes in the learning sample are sufficiently close. When the content of classes differs significantly, the simple decision rule tends to attribute most objects or even all of them to big classes. In the latter case, recognition-based analysis turns out to be useless. This is the reason SWS uses the normalized decision rule that attributes an object to the class with a maximal normalized estimate. Normalized estimates are calculated by the formula $\hat{\Gamma}_j(s^*) = \frac{\Gamma_j(s^*)}{\sum_{i=1}^{L}\Gamma_i(s^*)}$. With the normalized decision rule, one can attain a close accuracy of recognition for all classes.

3. STATISTICALLY WEIGHTED VOTING AS A CONVEX CORRECTION PROCEDURE

We can represent the estimate calculated by formula (1) as

$$\Gamma_j(s') = \sum_{i=1}^{r} c_i \nu_i^j, \qquad (4)$$

where $c_{i'} = \dfrac{w}{\sum_{i'=1}^{r} w_{i'}} > 0$, $\sum_{i'=1}^{r} c_{i'} = 1$. We can treat the value ν_i^j as the prediction of the indicator function $I_j(s)$. Therefore, estimate (1) can be interpreted as convex correction of predictions $\nu_1^j, ..., \nu_r^j$ that correspond to particular syndromes.

Convex correction procedures (CCPs) possess a number of interesting properties related to their generalizing potential.

Let Ω be the space of events related to individual objects to be predicted. Each element (object) ω of the space Ω is matched with the value of the predicted variable Y and the vector $\mathbf{x}(\omega)$ of independent variables $X_1, ..., X_n$ used to calculate predictions of Y. Predictions of the variable Y are assumed to be calculated by some algorithms (predictors) $z_1, ..., z_r$ preliminarily trained by the samples belonging to the probabilistic space $\Omega_t = \Omega \times ... \times \Omega$, which is the Cartesian product of t spaces Ω. For the sake of simplicity, we use the same designations for predictors and predictions calculated by them. Since the prediction depends on the learning sample $\omega_t \in \Omega_t$ and the vector of independent variables, we can write the predictor z_* as $z_*(\mathbf{x}, \omega_t)$. In what follows, we use the latter, more detailed form along with the reduced one, if needed.

The convex correction procedure over the predictors $z_1, ..., z_r$ calculates the prediction $\hat{Z}_{\text{conv}}(z_1, ..., z_r, c_1, ..., c_r) = \sum_{i'=1}^{r} c_{i'} z_{i'}$, where $c_{i'} \geq 0$, $i' = \overline{1, r}$, $\sum_{i'=1}^{r} c_{i'} = 1$. Let $\delta_{\text{conv}} = E_\Omega[(Y - \hat{Z}_{\text{conv}})^2]$ be the mathematical expectation of the squared CCP error and $\delta_{i'} = E_\Omega[(Y - z_{i'})^2]$ be the mathematical expectation of the squared error of an individual predictor, $i' = \overline{1, r}$.

Theorem 1. *If the inequalities* $c_{i'} \geq 0$, $i' = \overline{1, r}$, $\sum_{i'=1}^{r} c_{i'} = 1$ *hold, the inequality*

$$\delta_{\text{conv}} \leq \sum_{i'=1}^{r} c_{i'} \delta_{i'}$$

holds.

Proof. One can easily show that $\sum_{i=1}^{r} c_i [Y - z_i]^2 = \sum_{i=1}^{r} c_i [Y - Z_{\text{conv}} + Z_{\text{conv}} - z_i]^2 = \sum_{i=1}^{r} c_i [Y - Z_{\text{conv}}]^2 + \sum_{i=1}^{r} c_i [Z_{\text{conv}} - z_i]^2 = [Y - Z_{\text{conv}}]^2 + \sum_{i=1}^{r} c_i [Z_{\text{conv}} - z_i]^2$.

Therefore, we can write the CCP prediction error for Y for the arbitrary object $\omega \in \Omega$, which is $[Y(\omega) - Z_{\text{conv}}(\omega)]^2$, as

$$\sum_{i=1}^{r} c_i [Y(\omega) - z_i(\omega)]^2 - \sum_{i=1}^{r} c_i [Z_{\text{conv}}(\omega) - z_i]^2.$$

Hence,

$$[Y(\omega) - Z_{\text{conv}}(\omega)]^2 \leq \sum_{i=1}^{r} c_i [Y(\omega) - z_i(\omega)]^2. \quad (5)$$

Applying the mathematical expectation to the left-hand and right-hand sides of inequality (5), we obtain the conclusion of the theorem. The theorem is proved.

All predictors are assumed to be trained by the samples ω_t that belong to the probabilistic space Ω_t. We introduce the following designations:

$\hat{z}_i(\mathbf{x}) = E_{\Omega_t}[z_i(\mathbf{x}, \omega_t)]$ for an arbitrary vector of prognostic variables,

$\Delta_i = E_{\Omega_2} E_\Omega[(Y - z_i)^2]$ for the generalized error of the predictor z_i,

$\Delta_i^{\text{bias}} = E_\Omega\{[E_\Omega(Y|\mathbf{x}) - \hat{z}_i(\mathbf{x})]^2\}$ for the bias component of the generalized error of the predictor z, and

$\Delta_i^{\text{var}} = E_\Omega E_{\Omega_t}\{[\hat{z}_i(\mathbf{x}) - z_i(\mathbf{x}, \omega_t)]^2\}$ for the variational component of the generalized error of the predictor z.

For the arbitrary predictor z_*, the known expansion $\Delta_* = E_\Omega[Y - E_\Omega(Y|\mathbf{x})]^2 + \Delta_*^{\text{var}} + \Delta_*^{\text{bias}}$ holds. The first summand is the noise component independent of the form of the predictor.

Theorem 2. *For the rational values of the coefficients* $c_{i'}$ ($i' = \overline{1, r}$) *and if the conditions* $c_{i'} \geq 0$ *are met for* $i' = \overline{1, r}$, $\sum_{i'=1}^{r} c_{i'} = 1$, *the following inequalities hold:*

$$\Delta_{\text{conv}}^{\text{bias}} \leq \sum_{i'=1}^{r} c_{i'} \Delta_{i'}^{\text{bias}}, \quad (6)$$

$$\Delta_{\text{conv}}^{\text{var}} \leq \sum_{i'=1}^{r} c_{i'} \Delta_{i'}^{\text{var}}, \quad (7)$$

$$\Delta_{\text{conv}} \le \sum_{i'=1}^{r} c_{i'} \Delta_{i'}. \quad (8)$$

Proof. To prove inequality (8), it is enough to integrate inequality (5) over the space Ω_t. To prove inequality (6), we can repeat the proof of Theorem 1, replacing Y with $E_\Omega(Y|\mathbf{x})$ and $z_i(\mathbf{x}, \omega_t)$ with $\hat{z}_i(\mathbf{x})$. Inequality (7) for arbitrary rational coefficients $c_{i'}$ that satisfy the conditions can easily be obtained from the inequality for the variational error of the arithmetical mean prediction obtained in [11]. In [11], the inequality $\Delta_{\text{conv}}^{\text{var}} \le \sum_{i'=1}^{r} c_{i'} \Delta_{i'}^{\text{var}}$ was shown to hold for $c_{i'} = 1/r$, $i' = \overline{1, r}$. If the coefficients $c_{i'}$ are rational, we can always choose the natural numbers $m_1, \ldots, m_r$ such that $c_{i'} = m_i / \sum_{i'=1}^{r} m_{i'}$. We duplicate each of predictors z_i m_i times. As a result, we obtain the extended system $\tilde{Z}_{\text{ex}}$ consisting of $\sum_{i'=1}^{r} m_{i'}$ predictors. For the convex correction procedure over $\tilde{Z}_{\text{ex}}$ that calculates the collective prediction in the form $\frac{\sum_{i'=1}^{r} m_{i'} z_{i'}}{\sum_{i'=1}^{r} m_{i'}} = \sum_{i'=1}^{r} c_{i'} z_{i'}$, the inequality

$$\Delta_{\text{conv}}^{\text{var}}(\tilde{Z}_{\text{ex}}) \le \frac{\sum_{i'=1}^{r} m_{i'} \Delta_{i'}^{\text{var}}}{\sum_{i'=1}^{r} m_{i'}} = \sum_{i'=1}^{r} c_{i'} \Delta_{i'}^{\text{var}}$$

holds. This proves inequality (7). Theorem 2 is proved.

The choice of the prediction method can affect only two components of the error, the variational and bias components. The value of the bias component depends on the degree of consistency in the form between the approximating functions of the model involved and the real dependencies. For instance, the bias component is small for linear dependences approximated by linear models [11]. The value of the variational component largely depends on the consistency between the complexity of the model involved and the data volume used to adjust it. SWS uses rather simple one-dimensional partitioning models to construct "syndromes," ensuring their high stability. By Theorem 2, applying the convex correction to sets of "syndromes" can only increase stability as compared to the mean stability for the set. Thus, total high learning stability is ensured. Convex correction procedures also help reduce the bias component.

4. THE MULTIMODEL STATISTIC SYNDROME METHOD

4.1. SWS Disadvantages and Ways to Overcome Them

SWS is highly efficient in many problems. However, its obvious drawback is that it fails when one-dimensional partitions are not sufficient to detect data regularities. When solving such problems, it is natural to search for regularities within more diverse families that include partitions of higher dimension and a more sophisticated form. More sophisticated models always allow attaining a higher partitioning degree of objects from different classes on the learning sample. However, more sophisticated approximation tools can often make the detected regularities lose stability and the variational components of errors increase. In fact, only one of its forming variables can provide for the high partitioning power of the complex regularity. The second variable can increase the partitioning power insignificantly, by actually approximating random deviations [10]. This leads to the problem of choosing a more sophisticated model that approximates, wherever possible, true measurements with minimal loss of stability.

One possible way to balance complexity and stability is to train using several families of partitions of different complexity levels simultaneously. "Syndromes" found using more sophisticated models are included in the output set only if they can improve significantly the partitioning degree of objects from different classes on the learning sample. We implemented such a strategy within the new recognition method, i.e., multimodel statistically weighted syndromes. The method employs the same collective decision-making procedure as SWS does, with the estimates calculated by formula (1). However, the method of multimodel statistically weighted syndromes (MSWS) combines several models to search for optimal partitions and selects syndromes against the model complexity. The MSWS decision rule coincides completely with the SWS rule as well.

4.2. Forming "Syndrome" Systems in MSWS

Partitioning models. We consider the method of multimodel statistically weighted syndromes employing the same voting procedure as in SWS with the difference that syndromes are searched by searching optimal partitions of one-dimensional or two-dimensional domains of the attribute space inside four a priori set models.

(a) The simplest one-dimensional model with one boundary point and two partition elements (model I); r_o^1 stands for the optimal partition for this model.

(b) The one-dimensional model with two boundary points and three partition elements (model II); r_o^2 stands for the optimal partition for this model.

(c) The two-dimensional model with two linear boundaries parallel to the coordinate axes and four elements (model III); r_o^3 stands for the optimal partition for this model.

(d) The two-dimensional model with one linear boundary arbitrarily oriented with respect to the coordinate axes and two elements (model IV).

r_o^4 stands for the optimal partition for model IV.

Within MSWS, we maximize the functional

$$F_l(R, \tilde{S}_o, K_j) = \frac{1}{\nu_0^j(1-\nu_0^j)} \max_{t \in T_R} [(\nu_t^j - \nu_0^j)^2 m_t]$$

rather than the functional $F_g(R, \tilde{S}_o, K_j)$ used in SWS. We use $F_l(R, \tilde{S}_o, K_j)$ rather than $F_g(R, \tilde{S}_o, K_j)$, since we need to ensure the quality of partitions with different numbers of elements to be compared properly. MSWS uses the threshold δ to select the detected regularities for the functional $F_l(R, \tilde{S}_o, K_j)$.

To decrease the retraining effect, the quality functional for more sophisticated partitioning models (II–IV) is multiplied by the penalty coefficient η. We consider the selection algorithm of the set $\tilde{Q}_j$ of syndromes for the class K_j in detail.

4.2.1. Selecting one-dimensional syndromes.

(1) Selection of one-dimensional syndromes found within model I.

We assume the optimal partition r_o^1 of the domain of admissible values of the attribute X_i to be found within model I and $F_l[r_o^1(X_i), \tilde{S}_o, K_j] > \delta$. Then, two syndromes formed by the boundary point b_i^1 of the partition r_o^1 will be included in the set $\tilde{Q}_j$.

We assume the optimal partition r_o^2 of the domain of admissible values of the attribute X_i to be found within model II and $F_l(r_o^2, \tilde{S}_o, K_j)*\eta > \delta$. Then, three syndromes formed by the boundary points b_i^{21} and b_i^{22} of the partition r_o^2 will be included in the set $\tilde{Q}_j$.

4.2.2. Selecting two-dimensional syndromes with boundaries parallel to the coordinate axes.

(1) Selection of two-dimensional syndromes found within models I and III.

We assume the two-dimensional optimal partition $r_o^3(X_{i'}, X_{i''})$ to be found for the pair of attributes $X_{i'}$ and $X_{i''}$ within model III, and $b_{i'}^3$ and $b_{i''}^3$ are the boundary points of the partition $r_o^3(X_{i'}, X_{i''})$ on the attributes $X_{i'}$ and $X_{i''}$, respectively.

They may differ from the boundary points $b_{i'}^1$ and $b_{i''}^1$ obtained for the same attributes within one-dimensional model I.

(a) If $F_l[r_o^3(X_{i'}, X_{i''}), \tilde{S}_o, K_j]*\eta > \delta$, four two-dimensional syndromes formed by the boundary point $b_{i'}^3$ for the attribute $X_{i'}$ and by the boundary point $b_{i''}^3$ for the attribute $X_{i''}$ will be included in the set $\tilde{Q}_j$.

(b) If $F_l[r_o^3(X_{i'}, X_{i''}), \tilde{S}_o, K_j]*\eta \le \delta$ and inequalities $F_l[r_o^1(X_{i'}), \tilde{S}_o, K_j] > \delta$ and $F_l[r_o^1(X_{i''}), \tilde{S}_o, K_j] > \delta$ hold, four two-dimensional syndromes formed by the boundary point $b_{i'}^1$ for the attribute $X_{i'}$ and by the boundary point $b_{i''}^1$ for the attribute $X_{i''}$ will be included in the set $\tilde{Q}_j$.

(c) If the inequality $F_l[r_o^2(X_{i'}), \tilde{S}_o, K_j]*\eta > \delta$ holds for only one of two attributes $X_{i'}$ and $X_{i''}$, for instance, $X_{i'}$, and the inequality $F_l[r_o^1(X_{i''}), \tilde{S}_o, K_j] > \delta$ holds, six syndromes formed by the boundary points $b_{i'}^{21}$ and $b_{i'}^{22}$ for the attribute $X_{i'}$ and by the boundary point $b_{i'}^1$ for the attribute $X_{i''}$ will be included in the set $\tilde{Q}_j$.

(d) If the inequalities $F_l[r_o^2(X_{i'}), \tilde{S}_o, K_j]*\eta > \delta$ and $F_l[r_o^2(X_{i''}), \tilde{S}_o, K_j]*\eta > \delta$ hold, nine syndromes formed by the boundary points $b_{i'}^{21}$ and $b_{i'}^{22}$ for the attribute $X_{i'}$ and by the boundary points $b_{i''}^{21}$ and $b_{i''}^{22}$ for the attribute $X_{i''}$ will be included in the set $\tilde{Q}_j$.

4.2.3. Selecting two-dimensional syndromes with linear boundaries arbitrarily oriented with respect to the coordinate axes. We assume the two-dimensional optimal partition r_o^3 to be found for the pair of attributes $X_{i'}$ and $X_{i''}$ within model IV. If the inequality $F_l[r_o^4(X_{i'}, X_{i''}), \tilde{S}_o, K_j]*\eta > \delta$ holds, two syndromes separated by the linear boundary arbitrarily oriented with respect to the coordinate axes will be included in the set $\tilde{Q}_j$.

5. REAL DATA EXPERIMENTS

5.1. Experimental Conditions

We held experiments to compare MSWS with SWS and alternative recognition algorithms, i.e., the q-nearest neighbors, the Fisher linear discriminant, and the support vector machine.

Table 1. Comparison results for the efficiency of different methods on the set of problems

Task		MSWS	SWS	LD	NN	SVM
1	m_{ex}	81	76	64	49	69
	f_{ex}	69.8%	65.5%	55.2%	42.2%	59.5%
	f_{ex}^{av}	68.7%	65.9%	49.3%	39.9%	54.5%
2	m_{ex}	55	51	48	44	51
	f_{ex}	68.8%	63.8%	60%	55%	63.8%
	f_{ex}^{av}	69.1%	63.7%	59.4%	55.3%	63.1%
3	m_{ex}	122	80	165	156	173
	f_{ex}	62.2%	40.8%	84.2%	79.6%	88.3%
	f_{ex}^{av}	63.5%	55.1%	49.3%	45.1%	50%
4	m_{ex}	43	43	46	40	45
	f_{ex}	71.7%	71.7%	76.7%	66.7%	75%
	f_{ex}^{av}	76%	69.5%	51%	51%	53.3%
5	m_{ex}	125	125	113	120	127
	f_{ex}	86.2%	86.2%	77.9%	82.8%	87.6%
	f_{ex}^{av}	86.4%	83%	69.8%	79.7%	81.4%

Table 1 gives the efficiency of MSWS compared to the q-nearest neighbors and the support vector machine for nine applied problems. We used the software implementation of the q-nearest neighbors and support vector machine and SWS found in the RECOGNITION software system. Table 1 uses the total fraction of correctly recognized objects $f_{ex} = \frac{\sum_{j=1}^{L} n_j^c}{\sum_{j=1}^{L} n_j}$ and the mean accuracy over classes $f_{ex}^{av} = \frac{1}{L}\sum_{j=1}^{L}\frac{n_j^c}{n_j}$, where n_j is the number of objects in the class K_j and n_j^c is the number of its correctly recognized objects, as the measure of efficiency. Recognition for each problem was done in the sliding mode control, which means that the learning sample is partitioned into the set of nonoverlapping subsamples consisting of the same, wherever possible, number of objects. Recognition of objects of each subsample is done by the algorithm trained by objects not included in this subsample. Within MSWS, we chose the threshold δ out of three values 3, 5, and 7. The experimental results include the maximal accuracy in the sliding control mode for these values.

For the q-nearest neighbors, we estimated the optimal number of neighbors during the training process. The number of nearest neighbors that maximize the estimate of the recognition accuracy in the sliding control mode was used as optimal.

We applied the version of the support vector machine that uses Gaussians as the kernel. The kernel size was chosen by the sliding control from the interval [3, 10].

Applied problems. We considered the following problems.

(1) Computer-based estimation of the severity of pneumonia. To train and estimate the recognition accuracy, we used a learning sample that included descriptions of 116 cases with values of 41 clinical parameters. For each case, the group of experts estimated the severity on a four-score scale. The severity estimation problem is reduced to a recognition problem with four classes.

(2) Diagnostics of malignant (melanoma) afflictions of parts of the skin using a set of 33 parameters that describes the respective images. The diagnostics problem is reduced to the recognition problem with three classes. The total number of objects in the learning sample is 80.

(3) Prediction of hysteromyoma relapse using 17 parameters describing the patient's condition. The problem is reduced to a recognition problem with two classes. The total number of cases included in the analyzed sample is 196.

(4) Prediction of hysteromyoma relapse using 69 immunological parameters is reduced to a recognition problem with two classes. The total number of cases included in the analyzed sample is 60. This problem stands out due to the high percentage of missed values.

Table 2. The MSWS efficiency for different values of the penalty coefficient η

Task		0.1	0.3	0.5	0.7	0.9
1	m_{ex}	73	73	81	80	78
	f_{ex}	62.9%	62.9%	69.8%	68.9%	67.2%
	f_{ex}^{av}	62.7%	61.6%	68.7%	69.0%	67.5%
2	m_{ex}	52	53	55	53	52
	f_{ex}	65%	66.3%	68.7%	66.2%	65%
	f_{ex}^{av}	64.8%	66.4%	69.1%	66.8%	65.7%
3	m_{ex}	99	98	122	123	124
	f_{ex}	50.5%	50%	62.3%	62.75%	63.2%
	f_{ex}^{av}	56.9%	56.6%	63.5%	63.8%	60.3%
4	m_{ex}	46	45	43	44	43
	f_{ex}	76.6%	75%	71.7%	73.3%	71.7%
	f_{ex}^{av}	79.1%	78.1%	76.0%	77.1%	76.0%
5	m_{ex}	125	125	123	123	123
	f_{ex}	86.2%	86.2%	86.2%	84.8%	84.8%
	f_{ex}^{av}	85.6%	86.4%	86.4%	85.5%	84.6%

(5) Autism diagnostics using 35 psychometrical parameters is reduced to a recognition problem with two classes. The total number of cases is 145.

(6) Prediction of viral hepatitis outcome using 20 parameters is reduced to a recognition problem with two classes. The sample contains 155 cases. The problem is taken from the open UCI Machine Learning repository (http://archive.ics.uci.edu/ml/).

(7) Diagnostics of oil well salinization is reduced to a recognition problem with three classes. A sample that corresponds to 114 experiments was used for training and testing in the sliding control mode. Five parameters were used to diagnose salinization.

(8) Prediction of daily radiation doses on the orbital station using a set of ballistic and geoheliophysical parameters.

(9) Prediction of tumor destruction using a set of 7 genetic parameters characterizing the density of karyons. The total number of cases is 77.

5.2. *MSWS Efficiency as Compared to SWS and Alternative Approaches*

Table 1 gives the results for MSWS as compared to SWS and alternative approaches. The upper cell of the table represents the total number of correctly recognized objects m_{ex}. The middle cell shows the total fraction of correctly recognized objects f_{ex}, while the lower cell represents the average fraction over the classes f_{ex}^{av}. The table also gives the average value of f_{ex}^{av} for all problems—$\hat{f}_{ex}^{av}$.

Table 1 shows that the MSWS accuracy is higher than the SWS accuracy in terms of the index f_{err}^{av} and is far better than the accuracy of statistical methods and the support vector machine for all problems, excluding Problem 7. However, the MSWS accuracy, in terms of the index f_{err}^{av} and f_{err} simultaneously, exceeds the accuracy of the other three methods only for problems 1, 2, and 8. In Problem 9, the MSWS accuracy turns out to equal the SWS accuracy, yet it exceeds the accuracy of the other alternative methods significantly. In problems 4, 5, and 6, the MSWS accuracy is slightly lower than the SVM and LD accuracy in terms of the index f_{err}, yet the excellence of the former in terms of f_{err}^{av} turns out to be significant. Finally, in Problem 3, the MSWS accuracy is significantly lower than the accuracy of alternative approaches in terms of f_{err}. However, they are almost inefficient in terms of the index f_{err}^{av}, i.e., the recognition accuracy in terms of f_{err}^{av} turns out to be less than 50%. Graphs of the interdependence of the recognition accuracy considered in the next section can give a more detailed idea on the relative efficiency of the methods.

5.3. *The MSWS Efficiency Depending on the Penalty Parameter η*

In this subsection, we deal with experimental study of how the penalty parameter influences the recognition accuracy estimate in the sliding control mode. We conducted studies to estimate the MSWS accuracy for

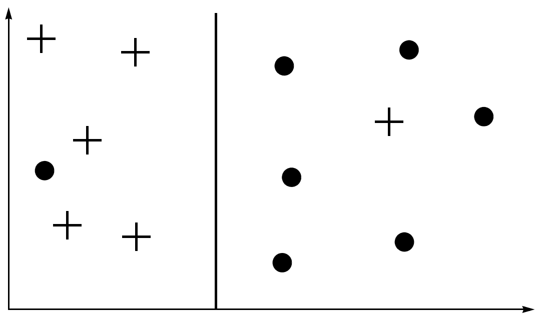

Fig. 1. Model I.

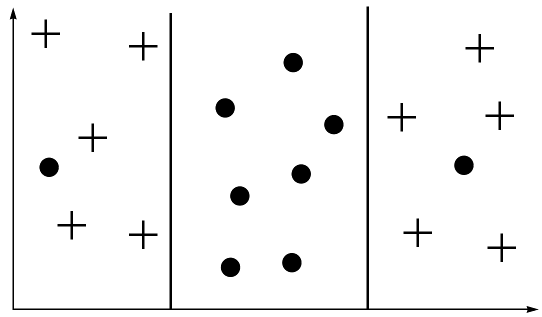

Fig. 2. Model II.

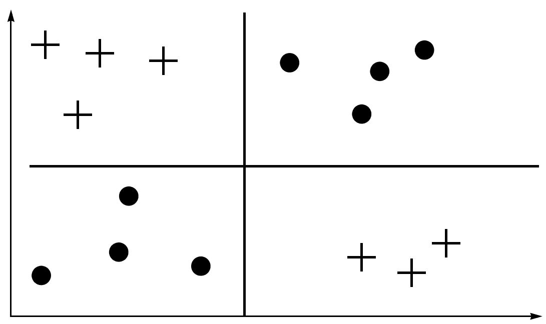

Fig. 3. Model III.

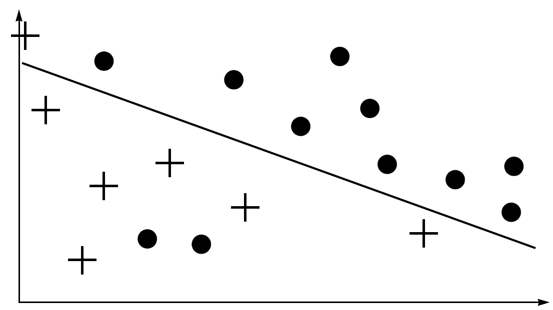

Fig. 4. Model IV.

the values of η 0.1, 0.3, 0.5, 0.7, and 0.9 and the parameter δ selected in the previous group of experiments.

Results given in Table 2 show that there is no trend in the way the recognition accuracy depends on the coefficient η common for all problems. In problems 4 and 9, the accuracy decreases slowly as η approaches 1. In Problem 3, on the contrary, the highest accuracy is attained for $\eta = 0.9$. However, the best results in most problems are attained when the value η is close to 0.5. Thus, the given results show that to achieve higher recognition accuracy, we need to use more sophisticated partitioning models and apply penalty coefficients to select the regularities found by these models.

6. TWO-DIMENSIONAL DIAGRAMS FOR COMMON RECOGNITION ACCURACY FOR TWO CLASSES

It is worth noting that to get the most complete information on the efficiency of recognition algorithms with two classes, one can use two-dimensional diagrams that describe common recognition accuracies of two classes for various threshold values for the estimates for classes. It is well known that any recognition algorithm can be represented as successive execution of the operator that calculates for each recognized object the vector of real estimates for classes and the decision rule that attributes objects to classes using the vector of estimates. The decision rule can be given by the threshold value δ_1 for the first class—the recognized object s is attributed to the class K_1 if the estimate $\gamma_1(s)$ for K_1 is greater than δ_1, and s is attributed to the class K_2 otherwise. Increasing the value δ_1, we monotonously decrease the percentage of correctly recognized objects of K_1 and increase the percentage of correctly recognized objects of K_2. For two classes, the estimate for one of them is usually the monotone function of the estimate for the other. Therefore, the set of decision rules with respect to thresholds for estimates for K_1 coincides with the set of decision rules with respect to thresholds for estimates for K_2. Unlike standard ROC diagrams that describe the dependence of the recognition accuracy of the class K_1 on the proportion of objects of the class K_2 erroneously attributed to the class K_1, we consider here the curves $D(f^1_{\text{ex}}, f^2_{\text{ex}})$ that describe the dependence of the proportion f^1_{ex} of correctly recognized objects of the class K_1 on the proportion f^2_{ex} of correctly recognized objects of the class K_2. Obviously, such approach implies that two classes are "equivalent." The important parameter that char-

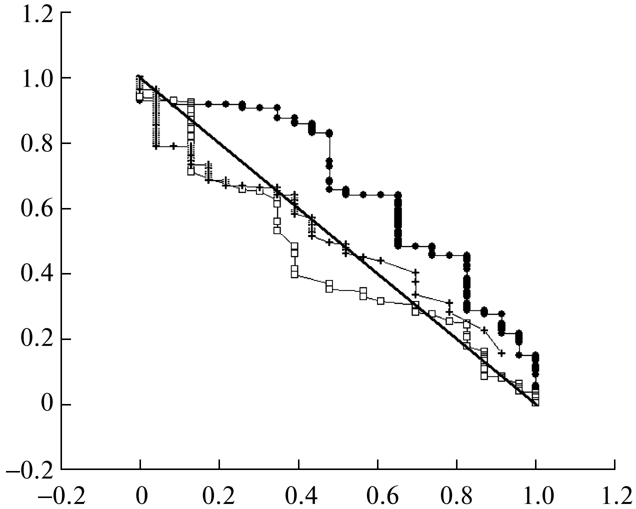

Fig. 5.

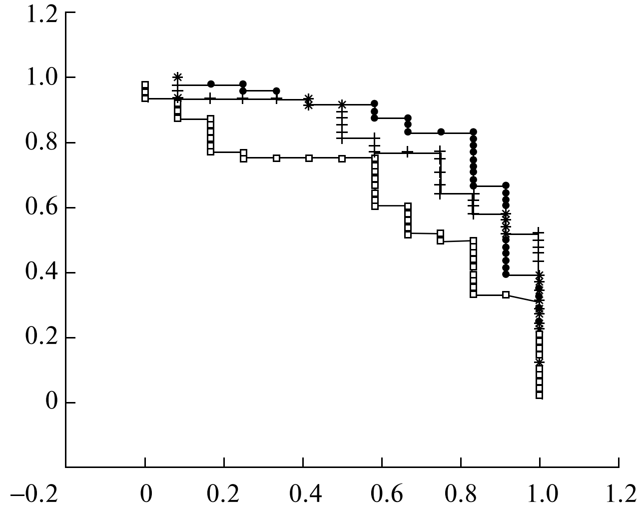

Fig. 6.

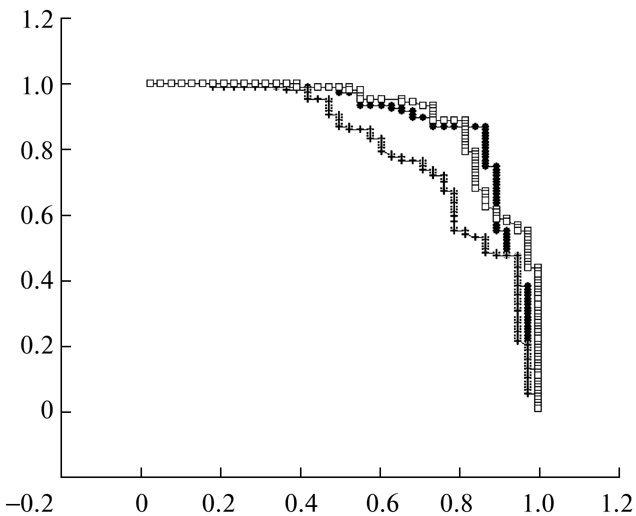

Fig. 7.

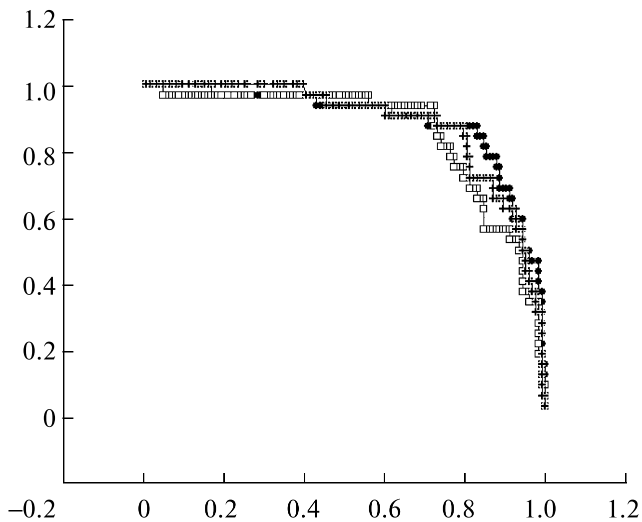

Fig. 8.

acterizes the curve $D(f^1_{ex}, f^2_{ex})$ is the average recognition accuracy f^{av}_{ex} for the threshold value that ensures minimal differences between f^1_{ex} and f^2_{ex}. We denote this accuracy by f^{equi}_{ex}. We held experiments to compare the efficiencies of MSWS, LD, and SVM in terms of the curves $D(f^1_{ex}, f^2_{ex})$ and indices f^{equi}_{ex}. The curves $D(f^1_{ex}, f^2_{ex})$ for problems 3–6 calculated in the mode are given in Figs. 1–4, respectively. To represent the curves that correspond to three of the compared methods, we use the following designations:

●—●—● MSWS,
+—+—+ LD,
□—□—□ SVM.

Figure 5 shows the curves $D(f^1_{ex}, f^2_{ex})$ for Problem 3. The obtained values f^{equi}_{ex} are 0.65 for MSWS, 0.48 for LD, and 0.46 for SVM.

Figure 6 shows the curves $D(f^1_{ex}, f^2_{ex})$ for Problem 4. The obtained values f^{equi}_{ex} are 0.83 for MSWS, 0.75 for LD, and 0.59 for SVM.

Figure 7 shows the curves $D(f^1_{ex}, f^2_{ex})$ for Problem 5. The obtained values f^{equi}_{ex} are 0.87 for MSWS, 0.74 for LD, and 0.81 for SVM.

Figure 8 shows the curves $D(f^1_{ex}, f^2_{ex})$ for Problem 6. The obtained values f^{equi}_{ex} are 0.84 for MSWS, 0.81 for LD, and 0.78 for SVM.

DISCUSSION

Figure 1 shows that the curves $D(f_{ex}^1, f_{ex}^2)$ for LD and SVM are close to the diagonal that connects the corner points of the unit square (0, 1) and (1, 0), which means they lack any recognition potential. At the same time, the MSWS curve is noticeably above the curve. The value $f_{ex}^{equi} = 0.65$ confirms the prognostic capabilities of the method. On most of the trajectory, the MSWS curve is farther from the diagonal than the alternative methods for Problem 4 as well. For problems 5 and 6, the curves for one of the alternative methods pass close to the MSWS curve, yet the latter is also farther from the diagonal on a significant part of the trajectory. The values of f_{ex}^{equi} turn out to be higher for MSWS in these problems as well.

CONCLUSIONS

We proposed a recognition model that calculates collective estimates using the system of subregions of the attribute space, where objects of one class, syndromes, dominate. We developed the method to form systems of syndromes that involves searching for optimal partitions for several models of different complexity and selecting syndromes taking into account the partitioning degree of classes by the respective optimal partition and the degree of complexity of the model involved. Theoretically, we showed high stability of the statistically weighted voting used to calculate collective estimates; this procedure can be treated as a convex combination of predictions calculated by individual syndromes. Theoretically, we showed that it is also possible to decrease, for the collective decision, the bias component of the recognition error as compared to prediction errors calculated by individual syndromes. The experiments carried out demonstrated high efficiency of the method for a number of applied problems with different dimensions and volumes of the learning sample. The MSWS efficiency was also confirmed by two-dimensional diagrams for the common recognition accuracy of two classes.

The experiments helped confirm that attaining higher recognition accuracy requires using more sophisticated partitioning models and applying penalty coefficients to select regularities found by these models.

ACKNOWLEDGMENTS

This study was supported by the Russian Foundation for Basic Research (grant nos. 08-01-90016 and 08-07-00437).

REFERENCES

1. Yu. I. Zhuravlev, V. V. Ryazanov, and O. V. Senko, *Recognition. Mathematical Methods. Program System. Applications* (Fuzis, Moscow, 2006), p. 176 [in Russian].
2. A. N. Dmitriev, Yu. I. Zhuravlev, and F. P. Krendelev, "On Mathematical Classification Principles for Objects and Phenomena," in *Sampling Analysis* (IM SO AN USSR, Novosibirsk, 1966), issue 7, pp. 3–11.
3. M. N. Vaintsvaig, "Learning Algorithms for Pattern Recognition "Kora"," in *Learning Algorithms for Pattern Recognition* (Sov. Radio, Moscow, 1973).
4. E. V. Dyukova, "Recognition Algorithms of "Kora" Type: Realization Difficulty and Metrical Properties," in *Recognition, Classification, Prediction (Mathematical Methods and Their Application)* (Nauka, Moscow, 1989), issue 2, pp. 99–125.
5. V. V. Ryazanov, "Logical Regularities in Recognition Problems (Nonparametric Approach)," Zh. Vychisl. Matem. Matem. Fiz., No. 10, 1793–1808 (2007).
6. V. A. Kuznetsov, O. V. Senko, A. V. Kuznetsova, et al., "Recognition of Fuzzy Systems by Method of Statistically Weighed Syndromes and Its Use for Immune and Hematologic Norm and Chronical Pathology," Chem. Phys. **15** (1), 81–100 (1996).
7. A. M. Jackson, A.V. Ivshina, O. V. Senko, A. V. Kuznetsova, et al., "Prognosis Intraversical Bacillus Calmette–Guerin Therapy for Superficial Bladder Cancer by Immunological Urinary Measurements: Statistically Weighted Syndromes Analysis," J. Urology **159** (3), 1054–1063 (1998).
8. A. V. Ivshina, J. George, O. V. Senko, B. Mow, T. C. Putti, J. Smeds, H. Norden, J. Bergh, E.T.-V. Liu, V. A. Kuznetsov, and L. D. Miller, "Genetic Reclassification of Histologic Grade Delineates New Clinical Subtypes of Breast Cancer," Cancer Res. **66**, 10 292–10 301 (2006).
9. O. V. Sen'ko, "The Use of a Weighted Voting Procedure on a System of Basic Sets in Prediction Problems," Comp. Maths. Math. Phys. **35**, 1249–1257 (1995).
10. O. V. Senko and A. V. Kuznetsova, The Optimal Valid Partitioning Procedures. Interstat Statistics on the Internet, Available from http://statjournals.net/ (Apr. 2006).
11. O. V. Senko, The Use of Collective Method for Improvement of Regression Modeling Stability, *InterStat. Statistics on the Internet,* Available from http://statjournals.net/ (June 2004).

Oleg V. Senko. Born in 1957. Graduated from Moscow Institute of Physics and Technology in 1981. Received candidate degree in 1990 and Doctoral (full doctor of physics and mathematics) degree in 2007. At present, senior researcher in Dorodnitsyn Computer Center, Russian Academy of Sciences. Scientific interests: data mining, mathematical models of pattern recognition, classification and forecasting, practical applications in medicine, and other fields.

Anna V. Kuznetsova. Born in 1961. Graduated from the Second Moscow Medical Institute in 1986. Received her candidate (candidate of biology) degree in 1990. At present, senior researcher in Emanuel Institute of Biochemical Physics. Scientific interests: pattern recognition, intellectual methods of data analysis, and developing of diagnostics and prediction algorithms in medicine and biological and medical applications.