SEARCH, PROCESSING, AND APPLICATION OF LOGICAL REGULARITIES OF CLASSES

Yury Zhuravlev, Vladimir Ryazanov

***Abstract**: A model of the type of estimates calculation, based on the systems of logical regularities of classes (LRC), to solve supervised classification problems is considered. The basic definitions are given. Two approaches for processing sets of LRC (based on the construction of the shortest logical class descriptions and clustering sets of LRC) are described. Different ways to use LRC are considered: based on LRC sets classification, construction of various logical descriptions of classes, the calculation of informative features, logical correlations of features, minimization of the feature space, assessment of outliers.*

***Keywords**: classification, pattern recognition, the calculation of estimates, logical regularity of class.*

***ACM Classification Keywords**: I.2.4 ARTIFICIAL INTELLIGENCE Knowledge Representation Formalisms and Methods – Predicate logic, I.5.1 PATTERN RECOGNITION Models – Deterministic, H.2.8 Database Applications, Data mining.*

1. Introduction

The standard task of supervised classification by precedents was considered with n features, l disjoint classes $K_1, K_2, ..., K_l$ and m reference objects $X = \{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_m\}$ (training sample) [Zhuravlev, 1978]. The notation $\tilde{K}_i = X \cap K_i, i = 1, 2, ..., l,$ and assumption $\tilde{K}_i \neq \varnothing$, $i = 1,2,...,l$ were used. Arbitrary object $\mathbf{x} \in \bigcup_{i=1}^{l} K_i$ is identified by its description in the form of feature vector $\mathbf{x} = (x_1, x_2, ..., x_n)$. For simplicity, we assume that $x_i \in R$ (binary-valued and k - valued features are considered as a special case of real-valued). When the training sample analysis, we will often (especially without specifying) write simply K_i implying that we consider always in training $\tilde{K}_i$.

2. Logical Regularities of Classes and Basic Definitions.

Consider the following set of elementary predicates that depend parametrically on unknown $\Omega_1, \Omega_2 \subseteq \{1, 2, ..., n\}, \mathbf{c}^1, \mathbf{c}^2 \in R^n$ [Zhuravlev et al, 2006; Ryazanov, 2007]. We will use the notation $(x \le a) = \begin{cases} 1, & x \le a, \\ 0, & \text{otherwise.} \end{cases}$

Definition 1. Predicate

$$P^{\Omega_1, \mathbf{c}^1, \Omega_2, \mathbf{c}^2}(\mathbf{x}) = \&_{j \in \Omega_1} (c_j^1 \le x_j) \&_{j \in \Omega_2} (x_j \le c_j^2) \quad (1)$$

is called the logical regularity of class (LRC) K_λ, $\lambda = 1,2,...,l$, if

1. $\exists \mathbf{x}_t \in \tilde{K}_\lambda : P^{\Omega_1, \mathbf{c}^1, \Omega_2, \mathbf{c}^2}(\mathbf{x}_t) = 1,$
2. $\forall \mathbf{x}_t \notin \tilde{K}_\lambda : P^{\Omega_1, \mathbf{c}^1, \Omega_2, \mathbf{c}^2}(\mathbf{x}_t) = 0,$
3. $P^{\Omega_1, \mathbf{c}^1, \Omega_2, \mathbf{c}^2}(\mathbf{x}) = \underset{\{P^{\Omega_1^*, \mathbf{c}^{*1}, \Omega_2^*, \mathbf{c}^{*2}}(\mathbf{x})\}}{extr} \Phi(P^{\Omega_1^*, \mathbf{c}^{*1}, \Omega_2^*, \mathbf{c}^{*2}}(\mathbf{x}))$, where Φ - predicate quality criterion.

The predicate (1), satisfying only the first two constraints, is called admissible predicate of this class.

The predicate (1), satisfying only the first and third restrictions, is called partial logical regularity of class K_λ.

The set $N(P^{\Omega_1, \mathbf{c}^1, \Omega_2, \mathbf{c}^2}) = \{\mathbf{x} \in R^n : c_j^1 \le x_j, j \in \Omega_1; x_j \le c_j^2, j \in \Omega_2\}$ will be called the interval of predicate (analogue to intervals of elementary conjunctions in the algebra of logic).

Example of interval of LRC with $\mathbf{x}_t \in \tilde{K}_\lambda$ is presented in Figure 1. Here black marks marked objects satisfying this LRC.

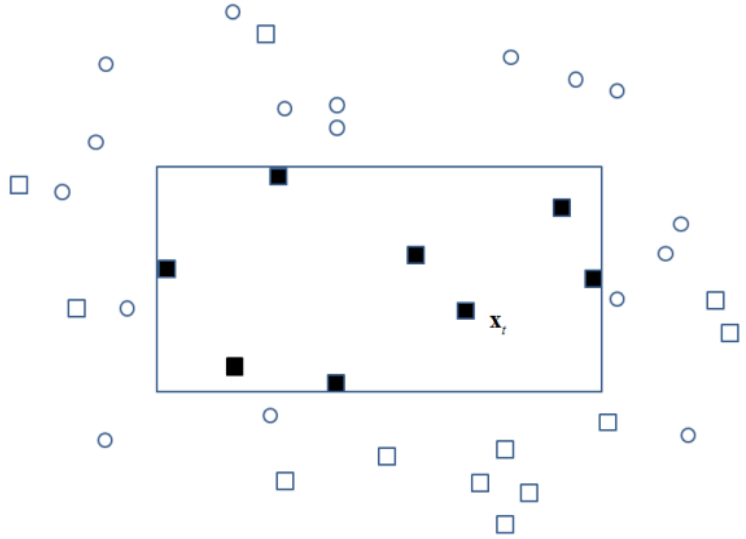


Figure 1. Example of LRC with $\mathbf{x}_t \in \tilde{K}_\lambda$

Two predicates $P^{\Omega_1,\mathbf{c}^1,\Omega_2,\mathbf{c}^2}(\mathbf{x})$, $P^{\Omega_3,\mathbf{c}^3,\Omega_4,\mathbf{c}^4}(\mathbf{x})$ are called equivalent if $P^{\Omega_1,\mathbf{c}^1,\Omega_2,\mathbf{c}^2}(\mathbf{x}_t) = P^{\Omega_3,\mathbf{c}^3,\Omega_4,\mathbf{c}^4}(\mathbf{x}_t), t = 1,2,...,m$.

Two intervals $N(P^{\Omega_1,\mathbf{c}^1,\Omega_2,\mathbf{c}^2})$, $N(P^{\Omega_3,\mathbf{c}^3,\Omega_4,\mathbf{c}^4})$ are called equivalent if $N(P^{\Omega_1,\mathbf{c}^1,\Omega_2,\mathbf{c}^2}) \cap X = N(P^{\Omega_3,\mathbf{c}^3,\Omega_4,\mathbf{c}^4}) \cap X$.

Definition 2. The following criterion

$$F(P^{\Omega_1,\mathbf{c}^1,\Omega_2,\mathbf{c}^2}(\mathbf{x})) = \left|\{\mathbf{x}_i : \mathbf{x}_i \in \tilde{K}_\lambda, P^{\Omega_1,\mathbf{c}^1,\Omega_2,\mathbf{c}^2}(\mathbf{x}_i) = 1\}\right|$$

will be called as the standard quality criterion of the predicate of class K_λ.

Definition 3. A logical regularity of class (LRC) $P^{\Omega_1,\mathbf{c}^1,\Omega_2,\mathbf{c}^2}(\mathbf{x})$ is called minimal if there is no such equivalent LRC $P^{\Omega_3,\mathbf{c}^3,\Omega_4,\mathbf{c}^4}(\mathbf{x})$ that $N(P^{\Omega_3,\mathbf{c}^3,\Omega_4,\mathbf{c}^4}) \subset N(P^{\Omega_1,\mathbf{c}^1,\Omega_2,\mathbf{c}^2})$.

3. Finding of Logical Regularities with Standard Quality Criteria.

Consider the general problem of finding LRC with standard quality criteria. The problem will be solved under the restriction $\mathbf{x}_t \in \tilde{K}_\lambda$ where this object will be called the support object. First of all, we note the following. If $P^{\Omega_1,\mathbf{c}^1,\Omega_2,\mathbf{c}^2}(\mathbf{x})$ is LRC than $P^{\mathbf{c}^1,\mathbf{c}^2}(\mathbf{x})$ is also LRC. For simplicity, we omit the indices $\Omega_1 = \Omega_2 = \{1,2,\ldots,n\}$. There is enough to supplement the parameters $\mathbf{c}^1, \mathbf{c}^2$ on features, not included in Ω_1, Ω_2, by always running conditions $\min\limits_{\mathbf{x}_i \in K_\lambda} x_{ij} = c_j^1 \le x_j, j \notin \Omega_1$ or $x_j \le c_j^2 = \max\limits_{\mathbf{x}_i \in K_\lambda} x_{ij}, j \notin \Omega_2$. LRC $P^{\Omega_1,\mathbf{c}^1,\Omega_2,\mathbf{c}^2}(\mathbf{x})$ and $P^{\mathbf{c}^1,\mathbf{c}^2}(\mathbf{x})$ are equivalent. We shall seek "minimum" LRC (stretched on some subsets of K_λ).

Let $D_i^1 = \{d_{i1}^1, d_{i2}^1, \ldots, d_{iu_i}^1\}$ and $D_i^2 = \{d_{i1}^2, d_{i2}^2, \ldots, d_{iv_i}^2\}$ are all monotonically decreasing / increasing the choices of values of left and right borders of predicates (1) for the training set (possible values of features of objects of $\tilde{K}_\lambda$), respectively.

We assume further for simplicity that the first order of the training sample objects (and only they) belong to the class in question $\tilde{K}_\lambda = \{\mathbf{x}_1, \mathbf{x}_2, \ldots \mathbf{x}_h\}$.

We construct the numerical matrix $\mathbf{M} = \begin{pmatrix} \mathbf{B}_1^1\mathbf{B}_1^2\mathbf{B}_2^1\mathbf{B}_2^2\ldots\mathbf{B}_n^1\mathbf{B}_n^2 \\ \mathbf{C}_1^1\mathbf{C}_1^2\mathbf{C}_2^1\mathbf{C}_2^2\ldots\mathbf{C}_n^1\mathbf{C}_n^2 \end{pmatrix}_{m\times N}$,

$N = \sum\limits_{i=1}^{n}(u_i + v_i)$,

$\mathbf{B}_i^1 = (b_{ij}^{1q})_{h\times u_i}$, $q = 1,2,\ldots,h,\ i = 1,2,\ldots,n$, $j = 1,2,\ldots,u_i$,

$\mathbf{B}_i^2 = (b_{ij}^{2q})_{h\times v_i}$, $q = 1,2,\ldots,h,\ i = 1,2,\ldots,n$, $j = 1,2,\ldots,v_i$,

$\mathbf{C}_i^1 = (c_{ij}^{1q})_{(m-h)\times u_i}$, $q = 1,2,\ldots,m-h,\ i = 1,2,\ldots,n$, $j = 1,2,\ldots,u_i$,

$\mathbf{C}_i^2 = (c_{ij}^{2q})_{(m-h)\times v_i}$, $q = 1,2,\ldots,m-h,\ i = 1,2,\ldots,n$, $j = 1,2,\ldots,v_i$, где

$$b_{ij}^{1q} = \begin{cases} 1, & x_{qi} \geq d_{ij}^{1}, \\ 0, & \text{otherwise}, \end{cases} \qquad b_{ij}^{2q} = \begin{cases} 1, & x_{qi} \leq d_{ij}^{2}, \\ 0, & \text{otherwise}, \end{cases}$$

$$c_{ij}^{1q} = \begin{cases} 1, & x_{(h+q)i} < d_{ij}^{1}, \\ 0, & \text{otherwise}, \end{cases} \qquad c_{ij}^{2q} = \begin{cases} 1, & x_{(h+q)i} > d_{ij}^{2}, \\ 0, & \text{otherwise}. \end{cases}$$

Consider a set of vectors $\{< x_{ij}^{1}, x_{ij}^{2} >\}$, $< x_{ij}^{1}, x_{ij}^{2} > =$ $< x_{11}^{1}, x_{12}^{1}, ..., x_{1u_1}^{1}, x_{11}^{2}, x_{12}^{2}, ..., x_{1v_1}^{2}, x_{21}^{1}, x_{22}^{1}, ..., x_{2u_2}^{1}, x_{21}^{2}, x_{22}^{2}, ..., x_{2v_2}^{2}, ..., x_{n1}^{1}, x_{n2}^{1}, ..., x_{nu_n}^{1}, x_{n1}^{2}, x_{n2}^{2}, ..., x_{nv_n}^{2} >$ with restrictions

$$x_{ij}^{1}, x_{ij}^{2} \in \{0,1\}, \sum_{j=1}^{u_i} x_{ij}^{1} = 1, \sum_{j=1}^{v_i} x_{ij}^{2} = 1, \quad i=1,2,...,n. \tag{2}$$

We associate the choice of units in $< x_{ij}^{1}, x_{ij}^{2} >$ to the parameter values $c_i^1 \in D_i^1, c_i^2 \in D_i^2, i = 1,2,...,n$. The set of all predicates of possible boundaries D_i^1, D_i^2 is in a one-to-one correspondence with the set of binary vectors of data, so we will use also the notation $F(< x_{ij}^{1}, x_{ij}^{2} >)$ as an entry for the standard optimality criterion. We form the following systems of inequalities and equalities.

$$f_q^c(< x_{ij}^{1}, x_{ij}^{2} >) = \sum_{i=1}^{n} (\sum_{j=1}^{u_i} c_{ij}^{1q} x_{ij}^{1} + \sum_{j=1}^{v_i} c_{ij}^{2q} x_{ij}^{2}) \geq 1, q = 1,2,...,m-h \ , \tag{3}$$

$$f_q^b(< x_{ij}^{1}, x_{ij}^{2} >) = \sum_{i=1}^{n} (\sum_{j=1}^{u_i} (b_{ij}^{1q} - 1) x_{ij}^{1} + \sum_{j=1}^{v_i} (b_{ij}^{2q} - 1) x_{ij}^{2}) = 0, q = 1,2,...,h. \tag{4}$$

LRC search problem can be formulated as the following discrete optimization problem:

Task Z:

$F(< x_{ij}^1, x_{ij}^2 >)$ =< number of executed equations in (4)> → max,

with restrictions (2-3).

We associate to the task **Z** similar problem **ZC** with respect to real variables.

Task ZC:

< number of executed equations in (4)> → max,

with restrictions (3), (5-6)

$$x_{ij}^1 \geq 0\,,\ i = 1,2,...,n\,,\ j = 1,2,...,u_i\,, \tag{5}$$

$$x_{ij}^2 \geq 0\,,\ i = 1,2,...,n\,,\ j = 1,2,...,v_i\,. \tag{6}$$

Let $Q = \{q : \sum_{i=1}^{n}(\sum_{j=1}^{u_i}(b_{ij}^{1q} - 1)x_{ij}^{\circ 1} + \sum_{j=1}^{v_i}(b_{ij}^{2q} - 1)x_{ij}^{\circ 2}) = 0, q = 1,2,...,h\}$

where $< x_{ij}^{\circ 1}, x_{ij}^{\circ 2} >$ is some solution of the problem **ZC.**

Let of feature number $i = 1,2,...,n$ is fixed and $p = \min\{j : b_{ij}^{1q} = 1, \forall q \in Q\}$,

$r = \min\{j : b_{ij}^{2q} = 1, \forall q \in Q\}$. Let $x_{ip}^{*1} = 1, x_{ij}^{*1} = 0, j \neq p,$

$x_{ir}^{*2} = 1, x_{ij}^{*2} = 0, j \neq r.$ Vector $< x_{ij}^{*1}, x_{ij}^{*2} >$ is defined after performing similar operations for $i = 1,2,...,n$.

Theorem. Vector $< x_{ij}^{*1}, x_{ij}^{*2} >$ is the solution of the problem **Z.**

This theorem provides a basis for creating an algorithm of search of LRC (1) with standard quality criteria:

1. Calculation of the support object.

2. Calculation of sets D_i^1, D_i^2.
3. Problem **ZC** solving and finding solutions {Q, $< x_{ij}^{*1}, x_{ij}^{*2} >$}, {Q} of the problem **Z**.
4. Search of all minimal LRC is carried out by repetition these calculations for all objects of class, taken as support objects.

Currently combinatorial (accurate), relaxation (approximate), and genetic algorithms established for finding LRC of given training table [Kovshov et al., 2008].

4. Processing Sets of LRC

4.1. Construction of LRC of minimum complexity

Let found $P^{\mathbf{c}^1,\mathbf{c}^2}(\mathbf{x})$. Let us consider the logical regularities search equivalent to the $P^{\mathbf{c}^1,\mathbf{c}^2}(\mathbf{x})$, but with minimal complexity. For example, the task is to find equivalent for $P^{\mathbf{c}^1,\mathbf{c}^2}(\mathbf{x})$ the LRC $P^{\Omega_1,\mathbf{c}^1,\Omega_2,\mathbf{c}^2}(\mathbf{x}) = \underset{j\in\Omega_1}{\&}(c_j^1 \le x_j) \underset{j\in\Omega_2}{\&}(x_j \le c_j^2)$, for which $|\Omega_1| + |\Omega_2|$ is minimal.

Consider the next problem of integer linear programming.

$$\sum_{j=1}^{n}(y_j^1 + y_j^2) \to \min,$$

$$\sum_{j=1}^{n}(1-(c_j^1 \le x_{ij}))y_j^1 + \sum_{j=1}^{n}(1-(x_{ij} \le c_j^2))y_j^2 \ge 1, \forall \mathbf{x}_i \notin \tilde{K}_\lambda, \qquad (7)$$

$$y_j^1, y_j^2 \in \{0,1\}.$$

The set of all unit components of the solution $(y_1^1, y_2^1, ..., y_n^1, y_1^2, y_2^2, ..., y_n^2)$ uniquely determines the corresponding subsets of features Ω_1, Ω_2.

4.2. Construction of logical description of the classes of minimum complexity

Let for the class K_λ the set $P_t(\mathbf{x}), t \in T$ was calculated.

Definition 4. The logical description of class K_λ be called logical sum $D_\lambda(\mathbf{x}) = \underset{t\in T}{\vee} P_t(\mathbf{x})$.

Definition 5. The shortest logical description of class K_λ be called logical sum $D_\lambda^s(\mathbf{x}) = \bigvee_{t \in T' \subseteq T} P_t(\mathbf{x})$, where $|T'| \to \min$ and function $D_\lambda^s(\mathbf{x})$ is equal to $D_\lambda(\mathbf{x})$ with given training sample.

Next, we consider that $P_t(\mathbf{x}), t \in T$ is the set of all minimal logical regularities, corresponding LRC found.

Definition 6. The logical sum $D_\lambda^m(\mathbf{x}) = \bigvee_{t \in T^* \subseteq T} P_t(\mathbf{x})$ is called the minimum logical description of the class, in which $\sum_{t \in T^* \subseteq T} (|\Omega_{t1}| + |\Omega_{t2}|) \to \min_{T^*}$, as a function $D_\lambda^m(\mathbf{x})$ the same as $D_\lambda(\mathbf{x})$ in the training sample.

Logic (shortest, minimal) class descriptions are analogous representations of partial Boolean functions in the form of reduced disjunctive normal form (shortest, minimal), and geometric images of logical regularities of classes are analogous to the maximum intervals.

The task of searching the shortest logical descriptions formulated as a problem for covering:

$$\sum_{t \in T} y_t \to \cdot \min, \tag{8}$$

$$\sum_{t \in T} P_t(\mathbf{x}_i) y_t \geq 1, \cdots \forall \mathbf{x}_i \in K_\lambda, \cdot y_t \in \{0,1\}. \tag{9}$$

The task of searching minimal of logic descriptions formulated as a problem for covering with the other functional and a set T:

$$\sum_{t \in T} (|\Omega_{t1}| + |\Omega_{t2}|) y_t \to \cdot \min, \tag{10}$$

with constraints (9).

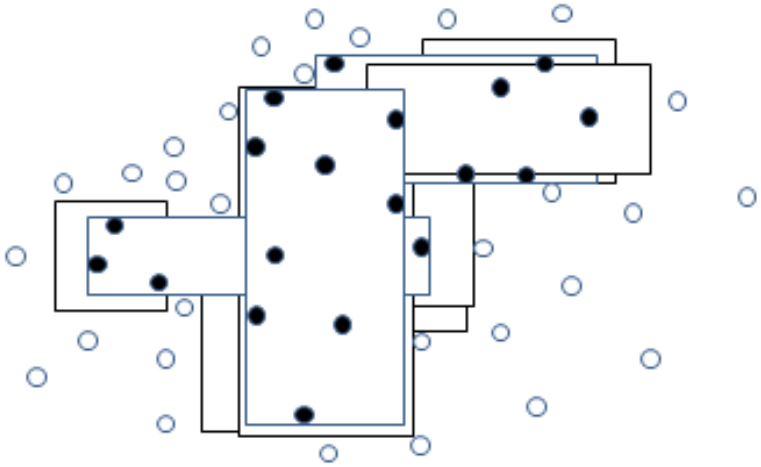


Figure 2. The points of a class are covered with a variety of LRC

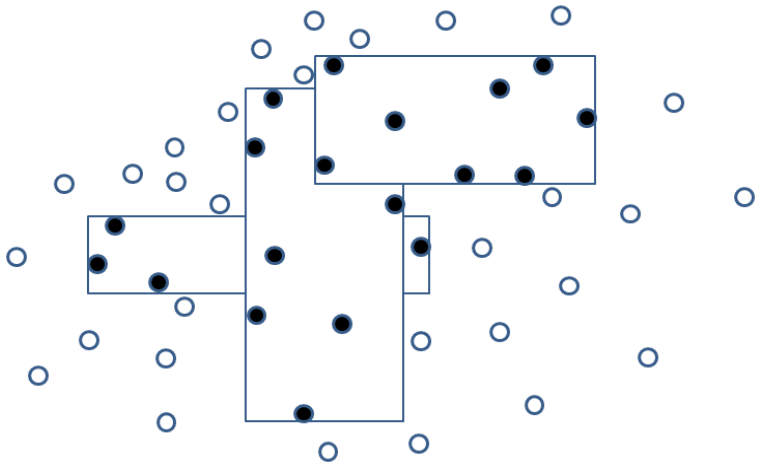


Figure 3. Shortest covering class corresponding to Figure 2

Note that the source can contain a plurality of equal or similar elements, "degenerate" solutions corresponding to the local maximum with small absolute values, the power sets of data can be quite large. At the same time, shortest and minimal logical descriptions are not redundant subsets expressing both the basic properties of data sets and the properties of the classes themselves.

Definition 7. Logical complexity (compactness) of classes are called values:

1. $\psi_1(K_j)$=< *the number of conjunctions in* $D_j^s(\mathbf{x})$>, $j = 1,2,...,l$;
2. $\psi_2(K_j)$=< *the number of variables in* $D_j^m(\mathbf{x})$>, $j = 1,2,...,l$.

The magnitude of $\Phi(X)=\sum_{i=1}^{l}\psi(K_i)$ is called the logical task complexity, if ψ is a criterion of logical complexity of class.

4.3. Processing LRC sets using a cluster analysis

The overall idea presented in [Gupal et al., 2015] and consists of the following. When processing the sets of vectors we can solve the problem of clustering in the 2, 3, ... clusters. For each cluster is calculated its "standard" (for example, the sample mean vector). The resulting system of "standards" is taken as a result of the processing of the initial set of precedent vectors. In this case, the clustering objects are functions LRC. We obtain the new predicates as a result of the clustering of the set LRC and calculations for each cluster, which in general are "partial" LRC. These predicates are evaluated. Thus, we can calculate and estimate a given number of "sufficiently" different partial logical regularities using the initial set of LRC. Figures 4 and 5 are examples. On Figure 4 points of a class (the "black circles" class) covered by system of LRC. Figure 5 shows the same example, but shows only two intervals. Intervals are substantially different and cover mostly large number of points of this class.

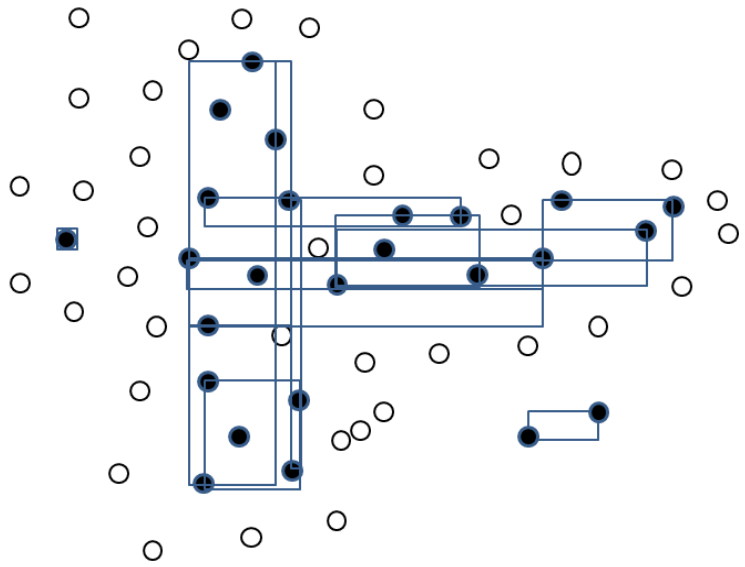


Figure 4. Class points are covered with a large number of intervals

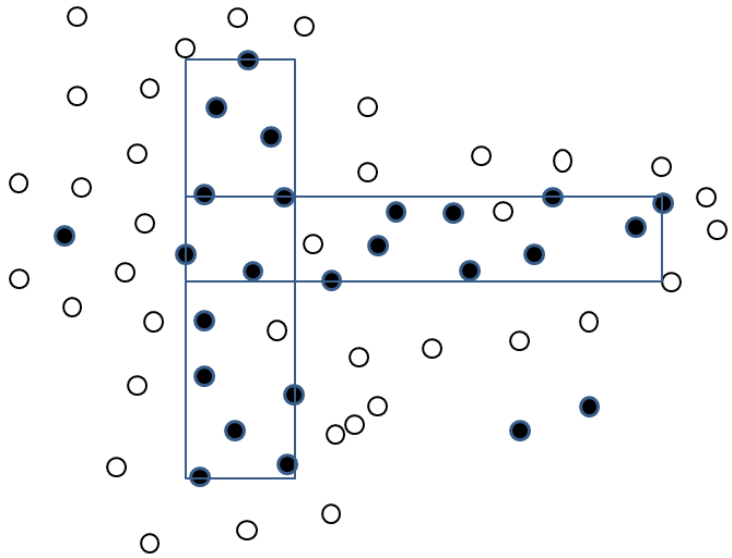


Figure 5. "Substantial" number of class points are covered by two intervals containing perhaps a small number of elements of another class

To implement this idea, we need to create a method of clustering a set of functions (predicates) and to calculate the "standard" for a variety of functions forming the cluster of functions. The set of predicates of each class should be weighted.

We put in a one-to-one correspondence of each $P_i^{\mathbf{c}^1,\mathbf{c}^2}(\mathbf{x})$ from $\mathbf{P}_t$ the binary vector $\mathbf{z}_i$ as follows: $P_i^{\mathbf{c}^1,\mathbf{c}^2}(\mathbf{x}) \Leftrightarrow \mathbf{z}_i = (z_{i1}, z_{i2}, ..., z_{ih}), z_{ij} \in \{0,1\}, j = 1,2,...,h$. Here $h = |K_t|$, vector $\mathbf{z}_i$ marks the original objects of study in the class K_t in which the predicate $P_i^{\mathbf{c}^1,\mathbf{c}^2}(\mathbf{x})$ is equal to one. The weight $y(\mathbf{z}_i)$ of each vector $\mathbf{z}_i$ (and of the corresponding LRC $P_i^{\mathbf{c}^1,\mathbf{c}^2}(\mathbf{x})$) is equal to the share of objects of class K_t, for which this LRC is equal to 1.

So, the initial problem is reduced to the clustering of the set of binary vectors $\mathbf{z}_i = (z_{i1}, z_{i2}, ..., z_{ih})$ with known weights $y(\mathbf{z}_i)$ and calculating the standard of each cluster. As a basic method of clustering we will take a method based on the minimization of variance [Duda et al, 2000].

Let the number of clusters l is fixed. We formulate clustering on l clusters by minimizing the variance criterion as follows:

$$J(\mathrm{K}) = \sum_{i=1}^{l} \sum_{\mathbf{z}_t \in K_i} y_t \left\| \mathbf{z}_t - \mathbf{m}_i \right\|^2 \to \min_{\mathrm{K}} , \tag{11}$$

where $y_t = y(\mathbf{z}_t)$, $\mathrm{K} = \bigcup_{i=1}^{l} K_i, K_i \cap K_j = \varnothing, i \neq j, i, j = 1, 2, ..., l$, $\mathbf{m}_i = \frac{\sum_{\mathbf{z}_t \in K_i} y_t \mathbf{z}_t}{\sum_{\mathbf{z}_t \in K_i} y_t}$.

It can be shown that partition $\mathrm{K} = \{K_1, K_2, ..., K_l\}$ is a local optimal one if (12) is true for all pairs of clusters and for any $\hat{\mathbf{z}}$ of K_i

$$\frac{\sum_{K_i} y\hat{y}}{(\sum_{K_i} y - \hat{y})} \left\| \hat{\mathbf{z}} - \mathbf{m}_i \right\|^2 - \frac{\sum_{K_j} y\hat{y}}{(\sum_{K_j} y + \hat{y})} \left\| \hat{\mathbf{z}} - \mathbf{m}_j \right\|^2 \leq 0 \tag{12}$$

For simplicity of notation, $\sum y$ is used herein instead of $\sum_{\mathbf{z}_t \in K_i} y_t$, $\hat{y}$ - "weight" of vector $\hat{\mathbf{z}}$.

As a result, vectors $\mathbf{m}_i = (m_{i1}, m_{i2}, ...m_{ih})$ are calculated for each cluster. $m_{ij} \in \{0, \alpha_{i1}, \alpha_{i2}, ..., \alpha_{iu}\}, 0 \leq \alpha_{i\sigma} < \alpha_{i,\sigma+1} \leq 1$ are true by construction and the values $\alpha_{i\sigma}$ are calculated from the resulting clustering.

It offers two approaches to processing sets of LRC.

1. We solve the problem of clustering of LRC and vectors $\mathbf{m}_i, i = 1, 2, ..., l$ are calculated.

The vectors $\mathbf{m}_i^*, i = 1, 2, ..., l$ are accepted as a result of processing, which enter a set of the initial $\mathbf{P}_t$ and are closest to the respective $\mathbf{m}_i, i = 1, 2, ..., l$.

2. The standard of each cluster is a Boolean vector $\mathbf{b}_i = (b_{i1}, b_{i2}, ..., b_{ih}), b_{ij} \in \{0, 1\}$, that is a result of sampling the vector $\mathbf{m}_i, i = 1, 2, ..., l$, where $b_{ij} = \begin{cases} 1, & m_{ij} \geq \theta_i, \\ 0, & \text{otherwise} \end{cases}$ Here θ_i is selected from a finite set $D_i = \{0, \alpha_{i1}, \alpha_{i2}, ..., \alpha_{iu}\}$. Vector $\mathbf{b}_i$ corresponds for each choice of θ_i, choice of the optimal vector values is carried out by solving the problem of one-dimensional optimization $\psi(P(\theta_i)) \to \max_{\theta_i \in D_i}$.

There are various quality criteria ψ for partial logical regularities which corresponds to a choice of some θ_i. An example of a criterion $\psi(P(\mathbf{x}))$ may be, for example, the criterion $\psi(P(\mathbf{x})) = \sqrt{k_i} - \sqrt{n_i}$

[William et al., 1999], where k_i is a number of training objects of class K_i, on which the predicate $P(\mathbf{x})$ of this class (corresponding to the chosen θ_i) is performed. n_i is a number of training objects of other classes, where the predicate $P(\mathbf{x})$ is not satisfied. Practical illustration of the method on the data [Mangasarian et al, 1990] is given in [Gupal et al., 2015].

4.4. Informative features, logical correlations and minimization of feature space

Standard statement of recognition problem suggests that the initial information about the classes (training information) is given by sample of vectors of feature descriptions representing all classes. In many cases, the system of features is formed "spontaneously". It includes all parameters influencing the classification (at least hypothetically) and which can be calculated or measured. Regardless of the number of available features, initial system of features is usually the excessive. It may have the features that not affect on the classification. In some practical recognition problems, the calculation of the cost of the features can be significant and compete with the cost of losses for recognition. Solving of training problems with fewer features can also be more precise and the resulting solutions more sustainable. Thus, the solution of problems of feature space minimization is important in many ways.

Let us consider the problem of minimizing the of feature space in the following statement. Let there be a pattern recognition algorithms, the original feature space R^N of feature values $x_1, x_2, ..., x_N$ and quality criterion $f(A)$ of the algorithm *A*. Required to find a subspace of features $R^n, n \le N$ with features $x_{i_1}, x_{i_2}, ..., x_{i_n}$ with minimal n $(n \le N)$, for which $f(A) \ge f_0$, where f_0 is a some minimum acceptable accuracy of the recognition algorithm *A*, built according to the training data for the subspace [Vetrov et al, 2001].

Due to its combinatorial nature, methods of enumeration a large number of different feature subspaces are practically unrealizable, so sequential selection procedures of the features systems as subsystems of k from the k-1 feature are commonly used.

The problem of minimizing of feature space was considered for recognition models based on the voting on systems of the logical regularities.

Let P be a set of all minimal LRC of minimum complexity that was found from the training data, $N = |\mathrm{P}|$.

Definition 8. The value $wei(i) = N(i)/N$ is called the measure of informativity of feature $№i$ if $N(i)$ is the number of elements of the P containing the feature $№i$.

Let $N(i,j)$ is the number of simultaneous occurrences of features into one set of LRC. $Lcorr(i,j) = 1 - \dfrac{N(i,j)}{\min(N(i), N(j))}$ is called the logical correlation of features $№i$ and $№j$. We believe $Lcorr(i,j) \equiv 0$ if ***min***($N(i)$, $N(j)$)=*0,* because of a characteristics (i or/and j) "does not depend on" (this case arises, for example, if $x_i \equiv const$).

Consider the problem of finding clusters of features that have close correlation properties.

As a clustering algorithm for a given semimetric $r(i,j)$ (was used logical correlation) and a fixed number of classes has been used clustering procedure "hierarchical grouping" in which the distance between the clusters determined by the function

$$r(K_p, K_q) = \max_{i \in K_p,\, j \in K_q}(r(i,j)).$$

After finding the n clusters, the condensed subsystem of features includes the most informative initial features (no more than one from each cluster). As $r(K_p, K_q)$ the function $1 - Lcorr(i,j)$ was used.

Figure 6 shows the variations of recognition accuracy of recognition models at two approaches to minimize the feature space on the example of the state of the ionosphere recognition problem [Sigillito et al., 1989]. Here, the black line represents the consistent screenings of less informative features, the gray line corresponds to minimizing the feature space according to the proposed algorithm in this paper. It is seen that the gray line is usually lower than the black.

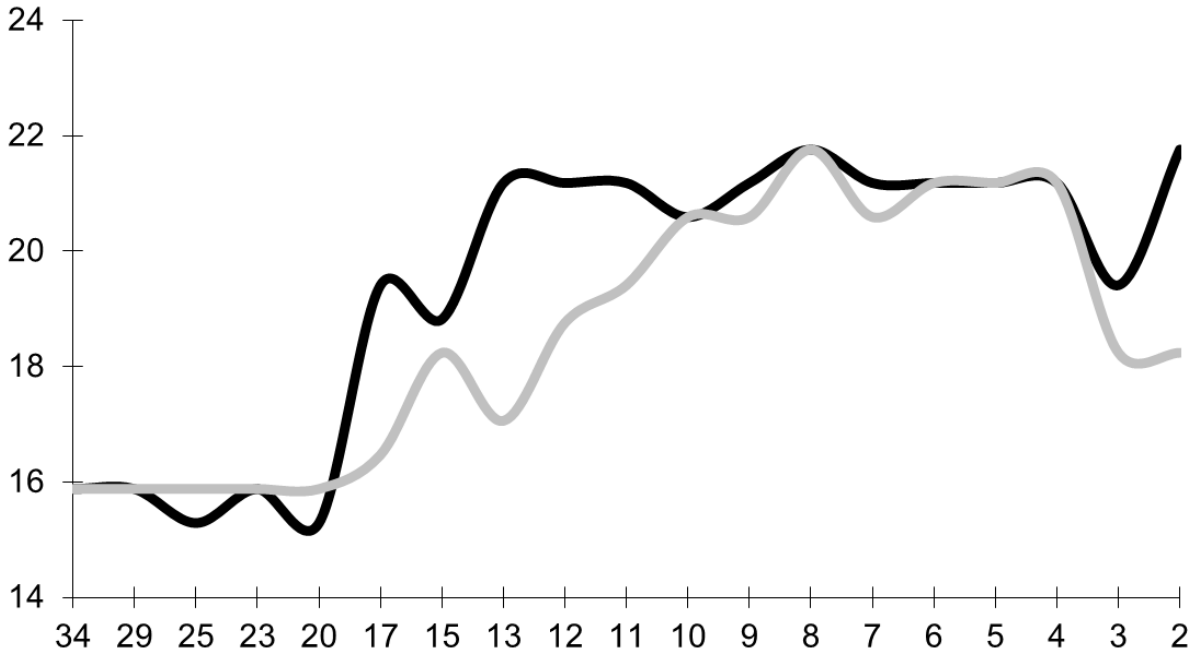


Figure 6. The dependence of the error rate of the number of features

4.5. The supervised classification procedures based on LRC

Calculation of estimates of the objects as a result of a simple voting on the found sets of LRC can lead to unsatisfactory results in recognizing of new objects [Lviv et al, 2015]. A similar effect can often be resolved by introducing the negatives LRC (which is equivalent to the use of " anti proximity") and the approximation of LRC by sigmoid functions.

Later, the case of two classes will be considered for simplicity. Suppose that the logical regularities $P_i^1(\mathbf{x}), i=1,2,...,m(1)$ are found for the first class and $P_i^2(\mathbf{x}), i=1,2,...,m(2)$ are found for the second class. Then estimation for the first class will be calculated by the formula $\Gamma_1(\mathbf{x}) = \sum_{i=1,2,..,m(1)} \alpha_i^1 P_i^1(\mathbf{x}) + \alpha_0^1 \overline{\bigvee_{i=1,2,..,m(2)} P_i^2(\mathbf{x})}$, and estimation for the second class will be calculated by the formula $\Gamma_2(\mathbf{x}) = \sum_{i=1,2,..,m(2)} \alpha_i^2 P_i^2(\mathbf{x}) + \alpha_0^2 \overline{\bigvee_{i=1,2,..,m(1)} P_i^1(\mathbf{x})}$. We will use a simple decision rule. Then the object classification will be on a sign of the following function:

$$f(\mathbf{x}) = \sum_{i=1,2,..,m(1)} \alpha_i^1 P_i^1(\mathbf{x}) + \alpha_0^1 \overline{\bigvee_{i=1,2,..,m(2)} P_i^2(\mathbf{x})} - \sum_{i=1,2,..,m(2)} \alpha_i^2 P_i^2(\mathbf{x}) - \alpha_0^2 \overline{\bigvee_{i=1,2,..,m(1)} P_i^1(\mathbf{x})}, \quad (13)$$

Here $P_i^1(\mathbf{x}), i = 1,2,\ldots,m(1)$ and $P_i^2(\mathbf{x}), i = 1,2,\ldots,m(2)$ are the logical regularities LRC of the first and second classes respectively, α_0^1, α_0^2, α_i^1, α_i^2 are the weighting coefficients. Object $\mathbf{x}$ belongs to the first class if $f(\mathbf{x}) > 0$ and belongs to the second class if $f(\mathbf{x}) < 0$. When $f(\mathbf{x}) = 0$ occurs a failure on the recognition or random classification.

Construction of the function $f(\mathbf{x})$ can be regarded as successive solution of two tasks:

1. Calculation of LRC $P_i^1(\mathbf{x}), i = 1,2,\ldots,m(1)$ and $P_i^2(\mathbf{x}), i = 1,2,\ldots,m(2)$, and the transition to the new $m(1)+m(2)+2$ - dimensional feature space of their values and the corresponding disjunction negation (with the sign "+" for the first class and "-" for the second).

2. Search of weighting coefficients by calculating a new feature space and separating hyperplane using the linear methods, for example, "Support Vector Machines", "Linear machine" or "Fisher's linear discriminant." We note that in the new feature space objects of the first class of training sample will correspond to the vectors of the form $\mathbf{y} = (\sigma_1, \sigma_2, \ldots, \sigma_{m(1)}, 1, 0, 0, \ldots, 0), \sigma_t \geq 0, \sum_{t=1}^{m(1)} \sigma_t > 0$. Objects of the second class are $\mathbf{z} = (0, 0, \ldots, 0, -\theta_1, -\theta_2, \ldots, -\theta_{m(2)}, -1), \theta_t \geq 0, \sum_{t=1}^{m(2)} \theta_t > 0$. So, classes are linearly separable in a given space. Should be noted that in the new feature space is possible to use other models.

4.6. Evaluation of outliers based LRC

Suppose that the shortest logical description $D_\lambda^s(\mathbf{x}) = \bigvee_{t \in T' \subseteq T} P_t(\mathbf{x})$ was calculated $D_\lambda^s(\mathbf{x}) = \bigvee_{t \in T' \subseteq T} P_t(\mathbf{x})$ for given training sample. Let $f(\mathbf{x}_i) = \sum_{t \in T' \subseteq T} y_t P_t(\mathbf{x}_i)$, where y_t is the weight of the corresponding logical regularity (eg, the number of objects that satisfy the given LRC). The values $f(\mathbf{x}_i)$ are ordered by an increase and normalize (for example, so that $\sum_{\mathbf{x}_i \in K_j} f(\mathbf{x}_i) = 1$). Minimum values of $f(\mathbf{x}_i)$ will be in accordance with the most unusual objects.

5. Conclusion

In the beginning we only have the training set. As a result of discrete analysis, we find sets of logical regularities (LRC) for each class. In fact, we find the conjunctions of intervals of attributes changes that characterize any class and does not hold for other classes of training objects. Found LRC are of independent interest for the practical user. What is the class? Previously, every class we were identifying with a set of its representatives. Now we can say that each class is characterized by a variety of some LRC (a variety of knowledge). It should be noted that we do not use any metric properties of objects. Features may be ordinal. Knowledge of informative features is not required. On the contrary, they can be estimated using the found sets of LRC. The article contains numerous possible applications of found sets for pattern recognition tasks. Further research in this area require the presence of cases of missing data, the linear relationships between variables, construction of optimal recognition procedures based on the found sets of LRC.

6. Acknowledgements

This work was supported by the Program of the Presidium of RAS №15 «Information, control and intelligent technologies and systems", Program №2 of Mathematical Sciences Department of RAS, RFBR № 14-01-00824_a, 15-01-05776_a, 15-51-05059 Arm_a.

Bibliography

[Zhuravlev, 1978] Yu.I. Zhuravlev. On the algebraic approach to solving the problems of recognition and classification. Problems of Cybernetics. M .: Nauka, 1978. Issure.33. pp. 5-68.

[Zhuravlev et al, 2006] Yu. I. Zhuravlev, V. V. Ryazanov, and O. V. Sen'ko, Recognition. Mathematical Methods. Program System. Practical Applications. Izd.vo "Fazis", Moscow, 2006,178 pp.

[Ryazanov, 2007] V.V.Ryazanov. Logical regularities in pattern recognition (parametric approach). Journal of Computational Mathematics and Mathematical Physics, T.47, №10, 2007, pp.1793-1808.

[Kovshov et al., 2008] N.V.Kovshov, V.L.Moiseev, V.V.Ryazanov. Algorithms for finding logical regularities in pattern recognition. Journal of Computational Mathematics and Mathematical Physics, T.48, 2008, N 2, pp. 329-344.

[Gupal et al., 2015] Anatoliy Gupal, Maxim Novikov, Vladimir Ryazanov. Processing sets of classes' logical regularities. International Journal "Information Theories and Applications", Vol. 22, Number 1, 2015, pp. 39-49.

[Duda et al, 2000] Duda, R. O., Hart, P. E., and Stork, D. G. : Pattern Classification. John Wiley and Sons. 2nd edition (2000).

[William et al., 1999] William W. Cohen and Yoram Singer Simple, Fast, and Effective Rule Learner, AAAI/IAAI 1999: 335-342.

[Mangasarian et al, 1990] Mangasarian O. L., Wolberg W.H.: "Cancer diagnosis via linear programming", SIAM News, Volume 23, Number 5, September 1990, pp 1 - 18.

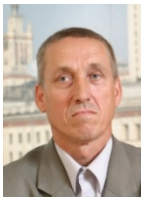
[Vetrov et al, 2001] D.P. Vetrov, V.V. Ryazanov. The minimization of feature space in pattern recognition. Reports of the 10 th All-Russian Conference "Mathematical Methods of Pattern Recognition (MMRO-10)", Moscow, 2001, 22-24.

[Sigillito et al., 1989] Sigillito, V. G., Wing, S. P., Hutton, L. V., \& Baker, K. B. (1989). Classification of radar returns from the ionosphere using neural networks. Johns Hopkins APL Technical Digest, 10, 262-266

Authors' Information

Yury Zhuravlev *– Deputy Director of Dorodnicyn Computing Centre, Federal Research Center "Computer Science and Control" of Russian Academy of Sciences, Russia, 119991 Moscow, Vavilov's street, 40; e-mail: zhur@ccas.ru*

Major Fields of Scientific Research: Discrete mathematics, Algebra, Pattern recognition, Data mining, Artificial Intelligence



Vladimir Ryazanov *– Heard of Department; Dorodnicyn Computing Centre, Federal Research Center "Computer Science and Control" of Russian Academy of Sciences, Russia, 119991 Moscow, Vavilov's street, 40; e-mail: rvv@ccas.ru, rvvccas@mail.ru*

Major Fields of Scientific Research: Pattern recognition, Data mining, Artificial Intelligence