International Journal of Software Engineering and Its Applications
Vol.12, No.3 (2018), pp.29-38
http://dx.doi.org/10.21742/ijseia.2018.12.3.03

Calculation of the Generalized Estimations in Sets of Features and their Interpretation

Madrakhimov Shavkat

Department of Algorithms and Programming Technologies at the National University of Uzbekistan, Tashkent
mshavkat@yandex.ru

Abstract

A problem of searching the regularities (properties) on the basis of mapping results of the objects descriptions is studied on a subset of features on a numerical axis. A method is proposed for partitioning the features into sets by constructing graphs of connectedness of features. Generalized estimates by each set of features are interpreted as a latent quantitative feature for describing objects in the two-class recognition problem. The detected value characterizing regularity by the latent feature is regarded as a value of a linguistic variable and interpreted.

Keywords*: quantitative and nominal features, latent features, generalized estimates, linguistic variable, object stability, agglomerative hierarchical algorithm, hidden regularities*

1. Introduction

Discovering hidden regularities in databases is one of the most important problems of data mining. The revealed regularities are a source of new knowledge, which are used to explain the decision-making process in various subject areas. Experts have an opportunity through the intellectual analysis of data to put forward and verify their hypotheses concerning the problem under consideration. According to the results of the analysis, the mined (extracted) knowledge can be presented in the form of logical rules, and in some cases it can be written in the form of an analytical representation of symbols.

In this paper the problem of reducing the dimensionality of the features space by mapping objects descriptions on the numeric axis in defined sets of features is considered [1-4]. The result of the mapping of each set of features is interpreted as a latent (obviously immeasurable) quantitative feature for describing objects in the two-class recognition problem. The sets of values of the latent features are divided into disjointed intervals, the boundaries of which are calculated by a special criterion. The set of test values in [0,1] is interpreted in terms of fuzzy logic [5]. For the sample described by nominal features, the potential possibility for detecting regularities is substantially limited. This is due to the fact, that in the framework of the theory of measurement, properties of relations of the features nominal values are defined as weak scale.

The belonging of latent quantitative features to strong measurement scales significantly increases the possibilities for detecting regularities by their values.

The process of extracting new knowledge is offered in the form of following stages:

– partition of the initial set of features into disjoint groups and formation of the set of latent features on their base;

Received (May 8, 2018), Review Result (August 22, 2018), Accepted (September 4, 2018)

ISSN: 1738-9984 IJSEIA

– description of objects by group of different types of features mapping into numerical axis and interpretation of mapping results as the latent features;

– partition of the values of latent feature into disjointed intervals by a special criterion which takes an optimal value in [0,1];

– the values of the criterion are interpreted in terms of the subject domain by using the fuzzy logic apparatus.

For the goal of the research there are three problems that are to be solved:

– to form subsets of features from the initial set of features as directed by the domain specialist or by using the rule of hierarchical agglomerative grouping;

– through the calculation of generalized estimates (latent feature) for subsets of different types of features, mapping the descriptions of objects into numerical axis;

– partition of the mapping results into disjointed intervals and their interpretation.

The problem of extracting knowledge by the result of sequential decision, described above, is studied for the first time.

2. Statement of the Problem

We consider the recognition problem in a standard formulation. Objects of the sample $E_0 = \{S_1,\ldots,S_m\}$ belong to one of the classes K_1 or K_2 $(E_0 = K_1 \bigcup K_2)$ and each object is described using n different type of features $X(n) = (x_1,\ldots,x_n)$, ξ of which are measured in interval scales, and $n-\xi$ are measured on the nominal. We denote the set of indexes of quantitative features through I and its nominal features by J and, here $|I|+|J| = n$.

Let on the sample E_0 the algorithm A is defined to map descriptions of objects from E_0 to numerical axis for a group of different type features $X(k) \subset X(n)$, $1 < k \le n$ and the criterion for dividing result of the mapping into two disjoint intervals, where a set of optimal values for them belongs to [0,1];

We need to perform the following actions:

– mapping description of objects by a set of features $X(k), 1 < k \le n$ into numeric axis through the algorithm *A*;

– to obtain the optimal partition of mapping results into two intervals by the criterion *R*.

3. Computation of the Stability of Sample's Objects

The concept of the stability of the object naturally follows from the procedure of calculating the optimal neighborhood of the *k* nearest neighbors, and in this case the stability is understood as a qualitative indicator of the object, describing its $E_0 = \{S_1,\ldots,S_m\}$ location among the objects of the same class [1,2]. As an algorithm, offered in this paper, the measure of similarity between the objects $S_0 = (x_{a1},...,x_{an})$ and $S_b = (x_{b1},...,x_{bn})$ of the sample E_0 is calculated by the metric of Juravlyev as following:

$$\rho(S_a,S_b) = \sum_{i\in I}|x_{ai}-x_{bi}| + \sum_{j\in J}\begin{cases}1, & x_{ai} \ne x_{bi},\\ 0, & x_{ai} = x_{bi}.\end{cases} \quad (1)$$

In Eq. (1) the values of the quantitative features are normalized to [0,1]. In order to exclude exhaustive search of informative features subsets for each object $S_d \in K_p, d = 1,...,m, p = 1,2$ in the set of $X(k)$, we built no decreasing sequence

$$S_{d_0}, S_{d_1}, ..., S_{d_{m-1}},\ S_{d_0} = S_d \tag{2}$$

ccorresponding to the objects of the sample E_0, satisfying the inequality $\rho(\mathrm{S}_{d_i}, \mathrm{S}_d) \le \rho(\mathrm{S}_{d_{i+1}}, \mathrm{S}_d), \mathrm{i} = 0,...,\mathrm{m}-2$ by the metric (1).

We introduce notations u_p^1, u_p^2 for the number of representatives $K_p, p = 1,2$ in the interval $\left[c_1^d, c_2^d\right], \left(c_2^d, c_3^d\right]$ accordingly. Here $c_1^d = 0, c_2^d = \rho(\mathrm{S}_d, \mathrm{S}_{d_h})$ and $c_3^d = \rho(\mathrm{S}_d, \mathrm{S}_{d_{m-1}})$, h is the order number in the sequence (2).

The definition of the interval boundary is based on the criterion that each of intervals $[c_1, c_2], (c_2, c_3]$ contains the value of the distance in the metric (1) of the representatives of only one class and its extreme value for the subset $X(k)$ is calculated as:

$$\left(\frac{\sum_{i=1}^{2} u_i^1 (u_i^1 - 1) + u_i^2 (u_i^2 - 1)}{\sum_{i=1}^{2} |K_i| (|K_i| - 1)} \right) \left(\frac{\sum_{d=1}^{2} \sum_{i=1}^{2} u_i^d \left(|K_{3-i}| - u_{3-i}^d \right)}{2 |K_1| |K_2|} \right) \to \max_{c_1 < c_2 < c_3} \tag{3}$$

For the intervals built by the criterion (3), we define $\alpha_1(\mathrm{p}) = \left| \left\{ S_{d_i} \in K_p \mid \rho\left(S_d, \mathrm{S}_{d_i}\right) \in [c_1, c_2] \right\} \right|$, $\alpha_2(\mathrm{p}) = \left| \left\{ S_{d_i} \in K_{3-p} \mid \rho\left(S_d, \mathrm{S}_{d_i}\right) \in [c_1, c_2] \right\} \right|$, $\theta_1(\mathrm{p}) = \alpha_1(\mathrm{p}) / \left|K_p\right|$ and $\theta_2(\mathrm{p}) = \alpha_2(\mathrm{p}) / \left|K_{3-p}\right|$. The stability of the object $S_d \in K_p, p = 1,2$ by subset of the features $X(k)$ is calculated as following:

$$U(S_d, X(k)) = \theta_1(p)(1 - \theta_2(p)). \tag{4}$$

4. Computation of the Generalized Estimations of Sample Objects

The generalized estimation is a value of combined indices, obtained on the base of numerical analysis of opposition like - *"patient – practically healthy", "hard worker-loafer", "rich-poor", "excellent student-debtor student", "generous- greedy"* and so on [3]. The data about oppositions may be presented as the sample objects of two disjoint classes at the formalization of the problem. The meaning of computing the estimation leads to explain the process of decision, for example, patients diagnosis by the value of the degree of disease severity.

Calculation of generalized estimates occurs by aggregating, as a rule, the raw features. The results of aggregation can be considered as the values of new latent features in the description of permissible objects. The calculation of generalized estimates of objects is made relative to individual classes. The need to reduce the solution to a two-class recognition problem with objects from K_t and $CK_t = E_0 \setminus K_t, t = \overline{1,l}$ is due to the fact that:

- any generalized estimate (indicator) is relative. The objects of each class are contrasted with objects of opposite classes (for example, the class of sick and dead from influenza and the class of practically healthy people);
- there are no classes of analytic functions for restoring dependencies in the space of different-types of features.

Computation of the generalized evaluations is done by means of aggregation, as a rule, of the initial features. The results of aggregation may be considered as values of the new

latent features in the description of permissible objects. The generalized estimation may be used for displaying the relationship between objects of two classes in different-type features space on the numerical axis.

To calculate the estimation of any permissible object $S=(x_1,...,x_n)$ by the subset of features $X(k)$ the following functional is used:

$$R(S)=\sum_{i\in X(k)}^{n} w_i t_i\left(x_i-c_1^i\right)/\left(c_2^i-c_0^i\right) \tag{5}$$

where c_0^i, c_1^i and c_2^i are the boundaries of the interval by (2) of ordered sequence of i-th feature, $t_i\in\{-1,1\}$ and the vector $T=(t_1,...,t_n)$ is defined by following condition:

$$\min_{S_p\in K_1} R(S_p)-\max_{S_p\in K_2} R(S_p)\to\max. \tag{6}$$

Searching the solution of the multiextremal problems by (6) is realized using the stochastic optimization algorithm. Step-by-step implementation of this algorithm is as follows:

Step 1. A choice of the number of iteration $k,\left[\frac{n}{2}\right]\le k\le n$, *iter*=0, $T_{\max}=\overbrace{(1,...,1)}^{n}$, $Z_{\max}=-n$;

Step 2. *iter= iter+1*. $\Omega=\{1,...,n\}$. A choise of the initial value of the vector $T=(t_1,...,t_n)$ for the next iteration. Computation of the values of $R(S_j)(S_j=(x_{j1},...,x_{jn}))\ \forall S_j\in E_0$ and $Z=\min_{S_j\in K_1} R(S_j)-\max_{S_j\in K_2} R(S_j)$;

Step 3. For $\forall i\in\Omega$ computation of the value (with the replacement of t_i by $-t_i$) $Z_i=\min_{S_j\in K_1} R^*(S_j)-\max_{S_j\in K_2} R^*(S_j)$, where $R^*(S_j)=R(S_j)-2t_i w_i\left(x_{ji}-c_1^i\right)/\left(c_2^i-c_0^i\right), j=\overline{1,m}$;

Step 4. $Z_p=\max_{i\in\Omega} Z_i$. If $Z_p>Z$, then $Z=Z_p$, $\Omega=\Omega\setminus p$, $R(S_j)=R(S_j)-2t_p w_p\left(x_{jp}-c_1^p\right)/\left(c_2^p-c_0^p\right), j=\overline{1,m}, t_p=-t_p$, and go the the Step 3;

Step 5. If $Z>Z_{\max}$ then $Z_{\max}=Z$, $T_{\max}=T$;

Step 6. If $iter<k$ then go to 2;

Step 7. Printing $Z_{\max}$, $T_{\max}$.

Steps of the algorithm from 2 to 4 represent the calculation of local maxima for different initial values of the elements of the vector T. The maximum value $Z_{\max}$ among the local maxima and the corresponding values of the elements of the vector $T_{\max}$ (step 5) is chosen as the solution of the problem by condition (6).

For objects with the description in different-type features space, additionally, a definition of the weights of nominal features and their gradations is required.

We denote by p the number of gradation of the feature $r\in J$, g_{dr}^t is the number of values of t-th $(1\le t\le p)$ gradation of r-th feature in the description of the objects of class K_d, l_{dr} is the number of gradation of r-th feature in K_d. The difference on the r-th feature between classes K_1 and K_2 is defined as the value of

$$\lambda_r = 1 - \frac{\sum_{t=1}^{p} g_{1r}^t g_{2r}^t}{|K_1||K_2|}. \quad (7)$$

A value of β_r, the degree of uniformity of classes K_1 and K_2 on the r-th feature (similarity measure in class), is calculated according to the formulas:

$$D_{dr} = \begin{cases} (|K_d| - l_{dr} + 1)(|K_d| - l_{dr}), p > 2, \\ |K_d|(|K_d| - 1), p \le 2; \end{cases}$$

$$\beta_r = \begin{cases} \dfrac{\sum_{t=1}^{p} g_{1r}^t (g_{1r}^t - 1) + g_{2r}^t (g_{2r}^t - 1)}{D_{1r} + D_{2r}}, D_{1r} + D_{2r} > 0, \\ 0, \qquad D_{1r} + D_{2r} = 0. \end{cases} \quad (8)$$

Using the formulas (7), (8) the weight of nominal feature $r \in J$ is determined as

$$v_r = \lambda_r \beta_r. \quad (9)$$

It is easy to verify that the set of weight values of nominal and quantitative features, calculated by Eqs. (3) and (8) belong into the interval $[0,1]$. It is obvious that for the set of numbers identifying as p gradation of the nominal feature, it is always possible one-to-one mapping into the set $\{1,...,p\}$. Taking into account such mapping for the object $S = (x_1,...,x_n)$, a contribution of the feature $x_i = j$, $i \in J$, $j \in \{1,...,p\}$ in generalized estimation is determined by the

$$\mu_i(j) = v_i \left(\frac{\alpha_{ij}^1}{|K_1|} - \frac{\alpha_{ij}^2}{|K_2|} \right), \quad (10)$$

where $\alpha_{ij}^1, \alpha_{ij}^2$ are the numbers of the values in the j-th gradation of the i-th feature in the classes K_1 and K_2, respectively, v_i is a weight of the i-th feature, calculated on the basis of Eq. (9).

The generalized estimation $R(S_a)$ (a latent feature) [2] of the object $S_a = (x_{a1},\ldots,x_{an}) \in E_0$ is calculated as

$$R(S_a) = \sum_{i \in I} w_i t_i (x_{ai} - c_2^i) / (c_3^i - c_1^i) + \sum_{i \in J} \mu_i (x_{ai}). \quad (11)$$

The results of calculating the generalized estimates

$$R(S_1), R(S_2), ..., R(S_m)$$

are considered as the values of a new latent, quantitative features of y in the description of sample objects E_0.

5. The Algorithm of Grouping the Features

The need for the use of an agglomerative hierarchical grouping of features get up by the following reasons:

- starting from some (initially unknown) dimension of the initial features space, the relationship of proximity between objects becomes blurred;

- the structure of relationship between objects depends on the scale of the measurement of features and on the use of proximity measures.

The agglomerative hierarchical algorithm is used based on the similarity matrix values $B=\{b_{ij}\}$ when selecting the candidates for sequential grouping the features. The distance between the objects $S_a, S_b \in E_0$ by the features $\{x_i\ x_j\} \subset X(n)$ is calculated as followings:

$$\beta(i,j,S_a,S_b)=\begin{cases}\rho(S_a,S_b), & i \neq j\\ 0, & \text{otherwise.}\end{cases}$$

The b_{ij} element of the similarity matrix $B=\{b_{ij}\}_{n\times n}$ is a measure of the contribution of a pair of features $\{x_i, x_j\}$ in division of the sample E_0 into two classes K_1, K_2 and is determined as

$$b_{ij}=\begin{cases}\dfrac{\sum_{a=1}^{m}\sum_{b=1}^{m}\alpha(S_a,S_b)\beta(i,j,S_a,S_b)}{\sum_{p=1}^{2}|K_p|\left(m-|K_p|\right)}, & i \neq j\\ 0, & \text{otherwise.}\end{cases}$$

where $\alpha(S_a,S_b)=\begin{cases}0, & S_a\ S_b \in K_i, i=1,2,\\ 1, & S_a \in K_i\ S_b \in K_{3-i}.\end{cases}$

The rule of hierarchical agglomerative grouping is based on the analysis of the relationship of objects stability defined by Eq. (6) to a set of initial features. The set $X(k)$ is considered to be formed, if the inclusion of another feature lowers the value of stability of the objects. Changing the stability of objects of the sample E_0 on the description from $X(k)$ to $X(k+1), 1 \le k < n$ $X(k) \subset X(k+1)$ is calculated as following [4]:

$$M\left(X(k),X(k+1)\right)=\frac{1}{m}\sum_{d=1}^{m}\begin{cases}1, & U\left(S_d,X(k)\right) \le U\left(S_d,X(k+1)\right),\\ 0, & \text{otherwise.}\end{cases}$$

The features grouping algorithm consists of the following steps:

Step 1. $l=0, \Omega=\{1,\ldots,n\}$;

Step 2. If $\Omega=\varnothing$, then go to Step 6, else $l=l+$, $G_l=\varnothing$;

Step 3. $b_{ij}=\max\limits_{u,v\in\Omega}\{b_{uv}\}_{n\times n}$, $G_l=\{i,j\}$, $\Omega=\Omega\setminus\{i,j\}$;

Step 4. If $\Omega=\varnothing$, then go to Step 6, else $Mon(p)=\max\limits_{p\in\Omega}\left\{M\left(G_l\ G_l\cup\{p\}\right)\right\}$;

Step 5. If $Mon(p)\ge 0.5$, then $G_l=G_l\cup\{p\}$ $\Omega=\Omega\setminus\{p\}$ and go to Step 4, else go to Step 2;

Step 6. End.

We calculate the generalized estimation $R_k(S_i)$ (a latent feature) of the objects $S_i \in E_0, i=\overline{1,m}$ by the group $G_k\ (1\le k\le l)$ and denote through $w(G_k)$ the optimal value of the criterion (4) by the sequence

$$R_k(S_1),\ldots R_k(S_m)$$

It is well known that existence of the monotonicity property in the calculation extends the possibility of using the standart recognition algorithms to the obtained results. Calculating the generalized estimations by formula (12), one can find that the sum of the

quantitative features depends on the random values t_i, the monotonicity property in the sequence of construction groups is satisfied only with the conditions $G_k \cap I = \varnothing (1 \le k \le l)$ and

$$w(G_p) \ge w(G_q) \; p < q .$$

The value $w(G_k) \in [0,1], 1 \le k \le l$ in terms of fuzzy logic can be used to describe the regularities in a natural language with the help of linguistic variables [5]. For example, membership $w(G_k)$ is one of the intervals $[0,0.3],(0.3,0.6],(0.6,1]$ can be interpreted as "*weak*", "*satisfactory*", "*strong*".

6. Computational Experiment

6.1. Experiment with Expert Specification of Subsets of Features

In the computational experiment, the data of a survey of physicians is analyzed on the base of medical products used for the treatment of patients with bronchial asthma [6-9]. The classification of the survey data was based on the answers of doctors on specialization - allergists, pulmonologists, therapists. Three variants of response grades for each drug are used (VEN-evaluation): 3 – vital and improving the quality of life (Vital – V); 2 – necessary (Essential – E); 1– secondary medicine need (non-essential – N). The sample consisted of 91 objects (expert physicians) which are represented by 102 gradations with the values of features of medicines.

On the basis of this sample, a series of hypotheses were tested by criterion (5), to describe its results a linguistic variable [5] "*Opinions of experts of allergists, pulmonologists, therapists on the VEN evaluation of a set of medications*" was used.

The "object-properties" table by the results of a survey of specialists - physicians (objects) on the medicines (properties) attribution in the VEN group is presented in Table 1.

Table 1. The Results of the Survey of Medical Specialists

№ doctors	Specialization of a doctor	1-medicines	2-medicines	…	n-medicines
1	Allergists	1	2		1
2	Pulmonologists	2	2		1
3	Therapists	1	1		2
…	…	…	…	...	…
m	Pulmonologists	2	3		1

We used the values of the linguistic variable presented in Table 2 to interpret the results of the analysis.

Table 2. Values of the Linguistic Variable

Interval	The value of the variable
[1..1]	*Absolutely do not match*
[0,7..1)	*Virtually do not match*
[0,5..0,7)	*Few match*
[0,2..0,5)	*Substantially coincide*
(0..0,2)	*Practically the same*
[0..0]	*Fully matched*

We studied the opinion of two groups of doctors K_1 and K_2, presented respectively by allergists $(|K_1|=23)$, pulmonologists and physicians $(|K_2|=68)$, on 3 specified sets of medications:

- the subset G_1 ={*Dexamethasone, Beclometasone, Prednisolone, Fluticasone, Triamcinolone, Budesonide, Methylprednisolone*} is medicines of the group "*Glucocorticoids*";

- the subset G_2 ={*Aminophylline, Salbutamol, Fluticasone propionate, Fenoterol, Theophylline, Orciprenaline, Fenoterol-ipratropium bromide, Fenoterol, Salmeterol* } is medicines of the group *"Antiallergic medications"*;

- the subsets G_3 ={*Midecamycin, Ceftazidime, Cefazolin, Ceftriaxone, Cefotaxime, Cefoperazone, Azithromycin*} is medicines of the group *"Antibiotics"*.

The set of values calculated by (4) and generalized estimations (features y^1, y^2, y^3) for the sets G_1, G_2, G_3 were analyzed by criterion (5).

Table 3. Results of the Computational Experiment by Sets of Features

Features set	The value (3) from generalized estimates	The value of the linguistic variable
G_1	0,55	*Few match*
G_2	0,51	*Few match*
G_3	0,36	*Substantially coincide*

In the computational experiment the truth of the hypothesis *"The opinions of doctors in group K_1 and K_2 on the use of sets of medicines χ^1, χ^2, χ^3 for the treatment of bronchial asthma essentially coincide*" was verified.

The results of the computational experiment are presented in Table 3.

As can be seen from Table 3, the validity of the hypothesis is confirmed only for the subset G_3 – "*Opinions of allergists with pulmonologists and physicians about separating medicines to VEN group are substantially coincides in the subset of drugs G_3*".

6.2. Experiment with a Defined Subset of Features

The data for the computational experiment were taken from [10]. A sample of 2126 fetal cardiotocograms (CTGs) were automatically processed and the respective diagnostic features measured. Each object of the sample is described by 36 features, 13 of which are nominal. CTG was classified according to the state of the embryo *Normal*, *Suspect* and *Pathologic* (*NSP*). The purpose of the experiment is to verify the truth of the statement that there is a subset of features of $X(k) \subset X(36)$, $1<k\le 36$, the compactness of the sampling objects is higher than in $X(36)$. The sample was divided on the base of the *NSP* into two classes K_1 and K_2, presented respectively by normal fetal condition $(|K_1|=1655)$, fetal condition suspected and pathological $(|K_2|=471)$.

As a result of applying the algorithm of hierarchical agglomerative grouping, the following 5 subsets of features can be presented:

the subset G_1 ={ *Tendency (histogram tendency)*, *CLASS (FHR pattern class code), SUSP (suspect pattern), FS (flat-sinusoidal pattern), E (no description), DS (severe*

decelerations), LD (no description), ALTV (percentage of time with abnormal long term variability), DP (prolongued decelerations), ASTV (percentage of time with abnormal short term variability), FM (foetal movement), Mean (histogram mean)};

the subset G_2 ={*B (REM sleep), Min (low freq. of the histogram), MSTV (mean value of short term variability), Mode (histogram mode), Median (histogram median), Variance (histogram variance), UC (uterine contractions)*};

the subset G_3 ={*e (end instant) , b (start instant), Width (histogram width), LBE (baseline value, medical expert), LB (baseline value), Max (high freq. of the histogram), Nmax (number of histogram peaks), AC (accelerations), Nzeros (number of histogram zeros)*};

the subset G_4 ={*A(calm sleep), AD(accelerative/declarative pattern (stress situation)), DE(declarative pattern (vagal stimulation)), DL (light decelerations), MLTV (mean value of long term variability)*};

the subset G_5 ={*D (active vigilance), C(calm vigilance)*}.

The Table 4 shows the values of linguistic variables for interpreting the level of compactness of sample objects according to the optimal value by criterion (5).

Table 4. The Value of a Linguistic Variable for Describing Compactness

Interval	The value of the variable
[1..1]	*Absolute compactness*
(0,55..1)	*High compactness*
[0,35..0,55]	*Average compactness*
[0,25..0,35)	*Significant compactness*
(0,1..0,25)	*Low index of compactness*
[0..0,1]	*Very low index of compactness*

The sets $G_1,...,G_5$ were analyzed by criterion (5) and the following optimal values are obtained.

Table 5. Th Results of Computational Experiment by Sets of Features

Features set	The value (3) from generalized estimates	The value of the linguistic variable
X	0,55	*Average compactness*
G_1	0,69	*Highest compactness*
G_2	0,26	*Significant compactness*
G_3	0,31	*Significant compactness*
G_4	0,35	*Average compactness*
G_5	0,07	*Very low index of compactness*

The results of the experiment (see Table 5) can be interpreted by the expression "*The highest compactness (separability) of objects of classes relative to the initial feature space* X *is manifested in the group of features* G_1".

7. Conclusions

Regularities in the sample are found by mapping descriptions of objects into numerical axis for specified subsets of features. The result of the mapping of each set is interpreted

as a latent quantitative feature for describing objects in the two-class recognition task. The set of values of the latent features are divided into disjointed intervals, the boundaries of which are calculated by a special criterion. The set of test values in [0,1] is interpreted in terms of fuzzy logic.

The results of a first computational experiment on the comparative assessment of medical specialists opinions by various categories on the importance of drugs groups in the treatment of bronchial asthma disease can be used in making expert decisions. The second experiment shows that the statement of the recognition problem for the CTG sample can be solved on a subset of the initial features, which give a relatively better separation of class objects of the sample.

The results can be applied to explain the decision-making process in pattern recognition problems in various subject areas.

References

[1] N. A. Ignatiev, “Generalized estimates and local metrics of objects in the intellectual analysis of data”, Monograph. - Tashkent: Publishing house "University", **(2014)**.
[2] N. A. Ignatiev and S. F. Madrakhimov, “Stability and generalized estimates of classified objects in a different type of features space//Comp”, Technology, Novosibirsk, vol. 16, no. 2, **(2011)**, pp.70-77.
[3] N. A. Ignatiev, “Computation of generalized indicators and intellectual analysis of data // Automation and telemechanics”, no. 5, **(2011)**, pp. 183-190.
[4] N. A. Ignatyev, S. F. Madrakhimov and D. Y. Saidov, “Stability of Object Classes and Selection of the Latent Features”, International Journal of Engineering Technology and Sciences (IJETS), vol. 7, no. 1, **(2017)** June, DOI: http://dx.doi.org/10.15282/ijets.7.2017.1.9.1071.
[5] V. V. Kruglov, M. I. Dli and G. R. Yu, “Fuzzy logic and artificial neural networks”, Textbook. - Moscow: Publishing House of Physical and Mathematical Literature, **(2001)**.
[6] N. D. Suyunov, “Bronchial asthma: epidemiological, medical and social and aspects of drug provision”, Medical Journal of Uzbekistan. - Tashkent, No. 2. **(2012)**, pp. 104-107.
[7] N. D. Suyunov, N. A. Ignatiev, S. F. Madrakhimov and G. M. Ikramova, “Pharmacoeconomic studies of medicinal provision of patients with bronchial asthma in Uzbekistan”, Pharmacia, Moscow, no 3, **(2012)**, pp. 33-36.
[8] N. D. Suyunov, S. F. Madraximov and G. M. Ikramova, “Analyzes of the medicine provision of patients with bronchial asthma”, News of Science and Education, NR 3 (27), Sheffield, Science And Education Ltd, **(2015)**, pp. 27-35.
[9] S. F. Madrakhimov and N. D. Suyunov, “Certificate of official registration of computer programs "Computation of the generalized estimate of medicinal products belonging to VEN groups”, No. DGU 02526. Agency for Intellectual Property of the Republic of Uzbekistan, -Tashkent, **(2012)** June 18.
[10] Dua, D. and Karra Taniskidou, E. UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science (**2017**).

Author

Madrakhimov Shavkat Fayzullayevich was born in 1960. He has graduated from the Faculty of Applied Mathematics of the National University of Uzbekistan in 1983. He holds a PhD degree in Physics and Mathematics and Associate Professor in specialty “Theoretical Foundations of Informatics”. He is author of more than 60 scientific publications in the field of artificial intelligence on the topic of data mining.

