

# Classification Based on Invariants of the Data Matrix

Vladimir N. Shats

St. Petersburg, Russia  
Email: vlshats@hotmail.com

**How to cite this paper:** Shats, V.N. (2017) Classification Based on Invariants of the Data Matrix. *Journal of Intelligent Learning Systems and Applications*, 9, 35-46. <https://doi.org/10.4236/jilsa.2017.93004>

**Received:** May 16, 2017

**Accepted:** August 21, 2017

**Published:** August 24, 2017

Copyright © 2017 by author and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

---

## Abstract

The paper proposes a solution to the problem classification by calculating the sequence of matrices of feature indices that approximate invariants of the data matrix. Here the feature index is the index of interval for feature values, and the number of intervals is a parameter. Objects with the equal indices form granules, including information granules, which correspond to the objects of the training sample of a certain class. From the ratios of the information granules lengths, we obtain the frequency intervals of any feature that are the same for the appropriate objects of the control sample. Then, for an arbitrary object, we find object probability estimation in each class and then the class of object that corresponds to the maximum probability. For a sequence of the parameter values, we find a converging sequence of error rates. An additional effect is created by the parameters aimed at increasing the data variety and compressing rare data. The high accuracy and stability of the results obtained using this method have been confirmed for nine data set from the UCI repository. The proposed method has obvious advantages over existing ones due to the algorithm's simplicity and universality, as well as the accuracy of the solutions.

## Keywords

Artificial Intelligence, Classification Algorithms, Granular Computing

---

## 1. Introduction

The classification problem is the central problem in machine learning, and methods for solving it are dealt with in a considerable number of research papers, which is constantly growing. Nevertheless, the analysis of modern methods, which is adequately described in [1] [2] [3], shows that some essential particulars of this task have not been considered.

The problem regards the set of real objects, the patterns of which are represented

by the feature vectors. The composition of features that describes the objects is, to a certain extent, random, and in some cases, the list can be changed or shortened. In addition, feature values contain random errors of measurement or observation. The influence of the inevitable uncertainty in the relationship between the real object and its model (pattern) is further increased, since the given information is divided between the object and its model, and neither of them is fully defined.

However, the existing methods have been unable to consider these factors as they use mathematical tools within the framework of formalization of pure mathematics. Such approaches have another drawback: in solving the problem, one must proceed from the assumption of existence of a metric in the feature space and a probability density function for the objects of each class.

At the same time, all objects have feature values that are very similar or equal, so the object classes differ in the probability density functions for features, but not for objects. Yet only a generalized function can accurately consider the discontinuous densities of the features. Therefore, each of the existing methods is applied in a restricted area whose boundaries can be established, as a rule, only experimentally.

Studies in recent decades regarding the principles of information processing in complex systems open up new possibilities for solving the problem and eliminate these gaps in the theory. Most of these works were triggered by soft computing theory and were based on the concept of an organism as a granular system. It is assumed that the system consists of holons or granules, which simultaneously represent a single entity and its part in the larger system on different hierarchical levels [4] [5] [6] [7]. Objects, subsets of programs or intelligent agents are examples of such entities. Here, the key issue is the operation for forming and separating granules.

Granular computing is a paradigm of research in the field of artificial intelligence. It covers multiple process modeling concepts of information processing in various hierarchical systems, as well as new approaches to learning with fuzzy databases [8] [9]. In this respect, the paradigm has common roots with the methods of machine learning. In the general theory of pattern recognition [10], the remaining unfinished, objects are “split” into elements of multiple hierarchical levels linked by combinatorial relations, and the simplest standard blocks are at the lowest level. Recently, granulation has been used to solve the problem of classification based on two existing machine-learning methods [11].

The present work is based on the concept of soft computing L. Zadeh, according to which the human mind considers the “comprehensive inaccuracy of the real world” [12] [13]. The author supposes that flows of inaccurate and partially correct information enters the brain through visual, auditory, tactile and other sensors, and the brain selects only the information that is related to a specific task. This information corresponds to the original information “with a minimum degree of accuracy”.

This concept is reflected in the proposed method, which builds upon a new

approach to feature description. The range of each feature value is divided into  $n$  equal intervals. Each interval can contain anywhere from zero to several objects forming granules. The length of the granules randomly depends on the ordinal interval, which is called index of the corresponding feature. Then, each object will be uniquely described by its index vector, and the data matrix will be transformed into an index matrix. For each index, we can calculate the ratio between the number of objects from a certain class and the total number of objects, which defines the index frequencies and provides the sampling density estimate. These results allow us to establish the class of any object and the value of the parameter  $n$  that provides an acceptable margin of error.

The effectiveness of adopted approach is explained by the fact that a given set of data defines a hierarchically organized system in which there are relations of the whole and part between its elements: features, objects and classes. The mechanism of operation of such a system is determined by frequencies of interaction of granules in accordance with the simplest formulas of probability theory.

Note that the uncertainty of initial data of the problem is taken into account indirectly under all transformations. They lead to random changes in the description of any object, but the relation between the object and its class remains the same. Therefore, we can assume that the granulation is based on an approximate calculation of the data matrix invariants whose role, with known error, is played by the index matrices.

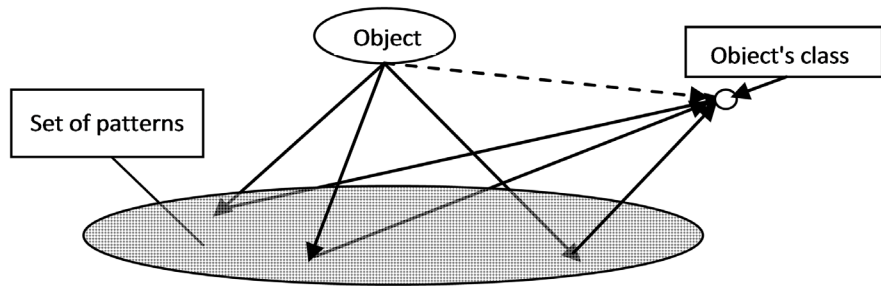
The article summarizes and develops previously completed studies [14] [15]. The method has been cross-checked on nine data sets from the UCI repository [16].

## 2. Transformation of Original Information

The article is devoted to the classification problem, in which training and control samples of real objects  $x \in \mathbf{X}$  are considered. They are described with errors by feature vectors  $q = (q_1, \dots, q_{Mk})^T$  and a distribution of objects into  $Mi$  disjoint classes. Objects have features of quantitative, categorical or mixed types, and there are no missing data. The task is to find an object classification rule for the training sample and evaluate its applicability to the control sample.

To identify objects and their patterns  $q$ , we will use the sequence of numbers  $s = 1, \dots, t, t+1, \dots, Ms$ , where  $t$  and  $Ms$  are the number of objects in the training sample and the total number of objects in the combined sample, respectively. Then,  $x^s, q^s$  and  $q_k^s$  will represent an object  $s$ , its pattern and the  $k$ th feature, respectively. Let  $\mathbf{Q}$  denote the data matrix for the combined sample, and  $\omega_i$  be the list of object numbers of class  $i, 1 \leq i \leq Mi$ .

This problem relates to the field of artificial intelligence and has specific relevant peculiarities. Here we are dealing with two entities of the arbitrary object number  $s$ : a real-world object  $x^s$  and information representing its pattern  $q^s$  that contains errors. It follows that pattern, specified in the task, is one of the realizations of the random variable (see. **Figure 1**), and there is a set of patterns of its possible values. Each of the vectors of this set could replace the corres-



**Figure 1.** Scheme of interrelations between the real object, its class and the set of its patterns.

ponding row of the matrix  $\mathbf{Q}$ . Note that these observations are consistent with the concept of soft computing.

Hence, the set  $\mathbf{x}$  would be represented by another data matrix, and there are infinite set such matrices. These matrices share a common property: they can be regarded as invariants of the matrix  $\mathbf{Q}$  in relation to the class of objects, although the concept of an invariant is defined for deterministic quantities. Therefore, it is advisable to find invariants that simplify the algorithm for solving the problem.

For the implementation of the relevant mapping, we will consider the given data set as a system that processes multi-level data with a hierarchical structure (Feature(Object(Class))). We need to find an approximate mapping of this set into another set such that the result of partitioning it into subsets could be estimated using the frequencies of the features.

Let us make two general remarks concerning the calculation of granules.

1) It should be noted that the frequencies of the features do not depend on the technique used to identify the feature values, such as magnitude, number or any other identifiers. On this basis, we can transform the system of reference for the features value and use any types of features, including mixed, in the calculations.

In particular, for a non-quantitative feature, we should establish relationship of partial order on the set of options for its values under the arbitrary rule of their numbering. The value  $q_k^s$  will be equal to the number of the corresponding options for the sequence numbers  $1, 2, \dots$ . Let us assume that such transformations of non-quantitative features are already incorporated into the matrix  $\mathbf{Q}$ .

2) The training set is designed to reduce the level of uncertainty of our knowledge of the properties of objects of each class. The uncertainty is measured by the value of information entropy, and it is obvious that increasing the entropy of each feature improves the solution quality. The maximum value of entropy is equal to  $\log N$ , where  $N$  is the number of options for the feature values. It follows that the accuracy of the results increases with increasing  $N$ , which characterizes the data variety. The conclusion is consistent with the law of requisite variety [17], which is in line with the theory that the adaptation of an organism requires the perception of the maximum amount of information about the possible effects of the environment.

To implement the ideas in both of the comments, we will first calculate a matrix that is of a convenient form and has sufficient accuracy to represent information contained in the conditions of the task. To this end, we will sequentially apply randomization and indexing of information.

The randomization procedure is designed to transform the feature values  $q_k^s$  in the pseudo-feature values  $\tilde{q}_k^s = q_k^s + \alpha \cdot v_s$  for all  $k$ , where  $v$  is a random variable that is uniformly distributed in the interval  $[0, 1]$ , and  $\alpha \geq 0$  is a constant. In the case of high data variety, which often occurs for quantitative features,  $\alpha = 0$  is taken. As a result of the randomization, we obtain a matrix of pseudo-features  $\tilde{\mathbf{Q}}$  of size  $M_s \cdot M_k$ , in which all the elements of each column are different.

We assume that the interval  $[\tilde{q}_k^{\min}, \tilde{q}_k^{\max}]$  is the domain of pseudo-feature objects  $\tilde{q}_k$  from the set  $\mathbf{X}$ , where  $\tilde{q}_k^{\min}$  and  $\tilde{q}_k^{\max}$  are the minimum and maximum values of  $\tilde{q}_k$ , respectively. Let the points  $\tilde{q}_{k,m} = \tilde{q}_k^{\min} + (m-1)\Delta_k$  be placed in the interval with step  $\Delta_k = (\tilde{q}_k^{\max} - \tilde{q}_k^{\min}) / (n-1)$ , where  $m \in \{1, \dots, n\}$  with  $n \geq 2$ . The parameter  $n$  is an analogue of  $N$  and therefore is called the parameter of variety.

Now we can clarify the concept of the index. The value  $m = m_k^s$  is called the index feature  $k$  for object  $s$ , if  $\tilde{q}_k^s \in [\tilde{q}_{k,m}, \tilde{q}_{k,m+1})$ . Then, the number of indices does not exceed  $M_s$ , and index  $m = 1$  if  $\tilde{q}_k^s = \tilde{q}_k^{\min}$ . For any  $n$ , the object  $s$  will be approximately described in the index space by the index vector  $(m_1^s, \dots, m_{M_k}^s)^T$ , and the combined sample will be represented by the matrix  $\mathbf{Q}_n$  of size  $M_s \cdot M_k$ . For the sequence  $n = 2, 3, \dots$  we obtain a sequence of matrices  $\{\mathbf{Q}_n\}$ .

In these mappings, the initial data for the problem are significantly transformed and boundaries of not quantitative features are become fuzzy. The ongoing changes could be illustrated by the example of the vector of an object with dissimilar features:

$(red, 21.7, \dots)^T \rightarrow (5, 21.7, \dots)^T \rightarrow (5.1192031, 892031, \dots)^T \rightarrow (11, 75, \dots)^T$  at consecutive steps: before quantization, after quantization, after randomization and after indexation. An object can receive these values if the number options of feature 1 are not less than 5 at  $n \geq 75$ .

### 3. Design Formulas

It follows from the definition of index that  $m_k^s$  is equal to the integer part value of  $\tilde{q}_k^s / \Delta_k$  for any  $n$ , and hence is a random variable. In the general case, for each  $k$ , the group of objects of different classes with the same values of the ordered pair  $\{k, m\}$  falls within the interval  $[\tilde{q}_{k,m}, \tilde{q}_{k,m+1})$ . Taking into account the errors in the patterns, we introduce the assumption that it is possible to neglect the difference of feature values for objects pair. According to this assumption, the distribution of features in the domain will be piecewise constant on each interval  $[\tilde{q}_{k,m}, \tilde{q}_{k,m+1})$  and any object pair will match the index  $m$ .

Then we can partition the combined sample into subsets called granules that contain objects of ordered pairs. Of particular interest is a subset of training

sample objects of a certain class, which we will call information granules.

To calculate the composition of the granules, we establish on the set  $\{\tilde{q}_k^s | s \in (1, Ms)\}$  of partial order relations, at which the values of  $\tilde{q}_k^z$  on the axis object numbers  $z = z(s)$  will be arranged in non-decreasing order. The scheme for calculating indices and information granules of training sample objects for a feature varying from 12.5 to 15 at  $Ms = t = 60$  is revealed in **Figure 2**. It is shown that the composition of a granule has a stochastic nature.

As a result of these transformations, the values of features will be measured on a single scale with a division value equal to one index. Therefore, the matrix  $Q_n$  can be interpreted as an image of the combined sample, calculated by statistical modeling of its objects' measurements. Now we are able to move on to analyzing the distribution of frequencies for the pair  $\{k, m\}$ .

Let  $l_{k,m}$  and  $l_{k,m}^i$  denote, respectively, the length of the granule  $\{k, m\}$  of objects of all classes of the training sample and only the class  $i$ . If no object of class  $i$  is in the interval  $m$ , then  $l_{k,m}^i = 0$ , and by definition, the respective values of  $m$  are not indices. It is clear that  $g_{k,m}^i = l_{k,m}^i / l_{k,m}$  is the frequency of the information granules  $\{k, m\}$  of class  $i$ , and at the same time, it is the sample estimate for the probability  $p_{k,m}^i$  of this granule as well as any of its objects. According to the problem statement, control and training samples belong to the single set. It follows that the objects of both samples have the same probability  $p_{k,m}^i$  and frequency  $g_{k,m}^i$ .

Since the appearance of each of the  $Mk$  features is an incompatible event, and these features form a complete group, the estimate of conditional probability that the object  $s$  belongs to class  $i$  is equal to

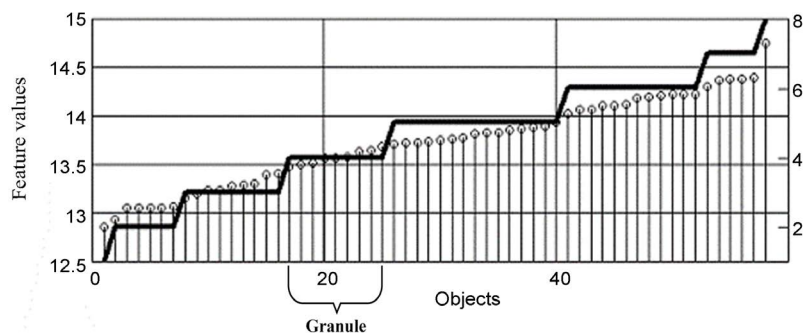
$$P(s \in \omega_i | q \in q^s) = \frac{1}{Mk} \sum_{k=1}^{Mk} p_{k,m}^i = \frac{1}{Mk} \sum_{k=1}^{Mk} g_{k,m}^i. \tag{1}$$

From (1) it follows that  $P(s \in \omega_i | q \in q^s)$  equals the average frequency of index  $m_s^k$  for the object  $s$  of class  $i$ . This frequency plays the role of a summarized object characteristic that considers the totality of feature values.

The estimation of the class of the object  $s$  is found using the obvious formula

$$I_s = \arg \max_{i \in (1, Mi)} P(s \in \omega_i | q \in q^s) \tag{2}$$

The relation  $i_s$  and  $I_s$  for the objects of the training and control samples,



**Figure 2.** The scheme for calculating the indices and information granule  $\{k, 4\}$  with  $n = 8$ .

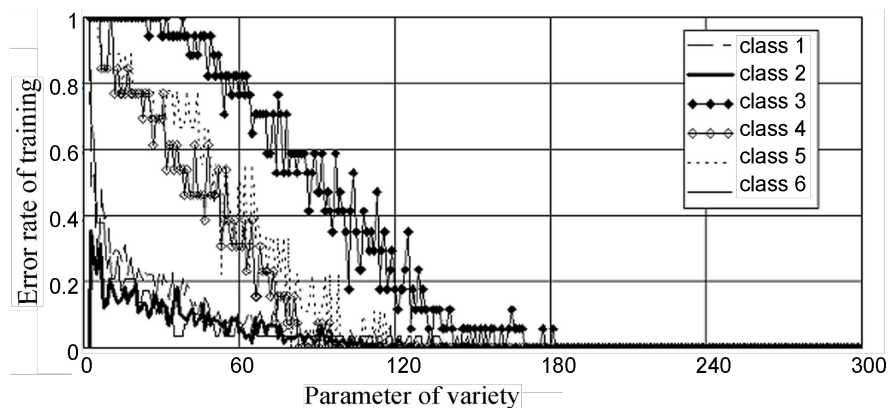
respectively, characterizes the algorithm’s ability to study and classify objects. The quality of teaching as a function of length of the training sample  $t$  is estimated by overage error rates  $(\delta_{i,t})_n$  for class  $i$  and by the overage rate for all classes  $\sigma_{t,n}$ . The classification accuracy is measured by the analogous frequencies  $(\theta_{i,t})_n$  and  $\Sigma_{t,n}$  for the control sample objects. The calculated value of the class  $I_s$  corresponds to some  $n$ , at which an acceptable accuracy of the solution is achieved.

#### 4. Existence and Accuracy of the Solution

We will consider the issue of convergence for the sequence of error rates. Let the granule  $\{k,m\}$  for a certain  $n$  contain an object  $s$  of the training sample of class  $i_s$  and some object  $w$ . Then the following inequality is satisfied:  $0 \leq |\tilde{q}_k^s - \tilde{q}_k^w| \leq \Delta_k$  is satisfied. If  $n \rightarrow \infty$ , then the step  $\Delta_k \rightarrow 0$  and  $\tilde{q}_k^w = \tilde{q}_k^s$ . In addition, the matrix  $\mathbf{Q}_n$  will be sparse, since an object falls into no more than  $M_s$  from an infinite number of intervals, and for the remaining intervals  $l_{k,m}^i = g_{k,m}^i = 0$ .

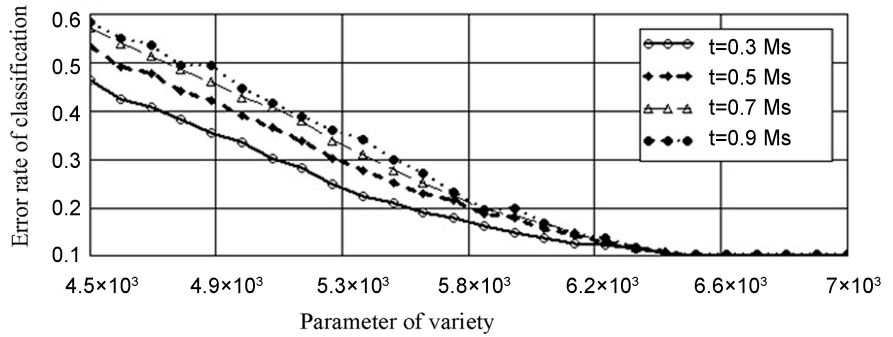
In this case for  $\alpha > 0$  we get that  $w = s$ , since all the values of  $\tilde{q}_k^s$  will be different. It follows that  $I_s = i_s$  and is achieved error-free training. If  $\alpha = 0$ , the granule  $\{k,m\}$  can contain only objects for which  $q_k^w = q_k^s$ . Since there cannot be two identical objects in a combined sample, this equality can be satisfied only for a narrow number of features. All the summands that define the average value of  $g_{k,m}^i$  for the object  $s$  will be greater than zero, while for an arbitrary object  $w$  all or only some of the summands are equal to zero. Here it is possible to have any relation average values of  $g_{k,m}^i$  for objects, and there is a risk of error. It is sufficient to take  $\alpha > 0$  to prevent this error.

Classes can differ significantly in the convergence of error rates of learning  $(\delta_{i,t})_n$  (see **Figure 3**). The process of convergence of  $\Sigma_{t,n}$  is illustrated in the graphs in **Figure 4**, which were calculated for the Abalone database for  $\alpha = 1$ . To evaluate the results of calculations, we note here that the number of classes  $M_i = 29$  and, according to the UCI, the accuracies of solutions obtained by various classical methods were equal to 33% - 65%.



**Figure 3.** The impact of the variety parameter  $n$  on the error rate of training  $(\delta_{i,j})_n$  for the data set Glass at  $t = Ms$ .





**Figure 4.** The impact of the variety parameter  $n$  on the error rate of classification  $\Sigma_{t,n}$  for different lengths of the training sample  $t$  of data set Abalone at  $\alpha = 1$ .

Nevertheless, the high accuracy of training does not guarantee acceptable classification accuracy since training is the first step in determining the class. Within the second step, it is necessary to evaluate the accuracies of solutions for the control sample objects based on relations (1) and (2), according to which any object for each  $n$  is characterized by the frequency  $g_{k,m}^i$ , if granules  $\{k,m\}$  exist. Otherwise,  $g_{k,m}^i = 0$ . Evidently, the accuracy of classification depends on the intersection lengths of the index sets for the same features in the training and control samples.

Here we face a problem of overtraining: if we restrict the value of  $n$ , it is possible to obtain a more accurate solution for the control sample, but the reliability of this result will be decreased because simultaneously the accuracy of learning will be lower. It is obvious that the simplicity of the algorithm reduces the computational complexity and severity of this problem, as well as allows us to estimate the effects of the characteristics of the matrix  $\mathbf{Q}$ .

The impact of the parameter  $\alpha$  can be observed in the example of the data set “Car evaluation”, where the variability of the data is low. Here objects are described by six features of nominal and ordinal types, with each feature possessing one of three or four values, and all objects of the same class having the same value for one of the features and only two variant values for other features. The task is complicated by the uneven distribution of the objects by class, since one class has 19 times more objects than the other does. Therefore, when  $\alpha = 0$ , there are significant errors even for high values of  $n$  ( $\sigma_{Ms,n} \approx 0.5$ ). The computational results for the case  $\alpha = 1$ , given in **Figure 5**, show that even in the task with such “bad” data practically error-free training and classification are achieved.

Now consider the joint impact of  $n$  and  $\alpha$  on the accuracy of solution. Let the objects  $s_1$  and  $s_2$  have the same value for feature  $q_k^{s_1} = q_k^{s_2}$ . Randomization should result in  $\tilde{q}_k^{s_1}$  and  $\tilde{q}_k^{s_2}$  belonging to different intervals of length  $\Delta_k$ , and these objects should be assigned different indices. Note that  $\Delta_k$  decreases with increasing  $n$  and  $\tilde{q}_k^{s_1} - \tilde{q}_k^{s_2} = \alpha(v_{s_1} - v_{s_2}) < \alpha$ . It follows that to preserve the randomization effect for large  $n$ , it is advisable to use larger values of  $\alpha$ .

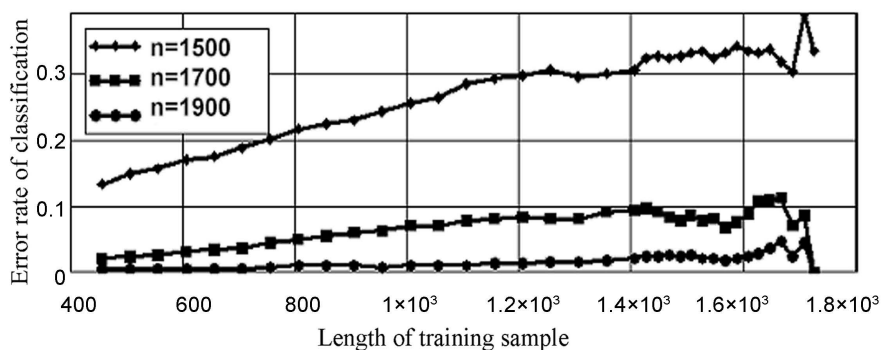


## 5. The Results of Solving Particular Tasks

The information about data sets from the repository UCI is given in **Table 1**, where  $\gamma$  is the ratio of the maximum to the minimum numbers of objects in different classes of the training sample. The tasks cover a wide range of values in terms of the numbers of objects (150-42121), features (3-16), classes (2-29), types of features (quantitative and mixed, including nominal, ordinal and integer) and the degree of unevenness of the object distributions between classes of the training sample  $\gamma$  (1-689).

The algorithm of the method is based on the data grouping that predetermines the decreasing influence of various kinds of outliers. Calculations have shown that the solution is stable for an infinite set of acceptable solutions: small oscillations of the parameters  $\alpha, t$  and  $n$  usually lead to insignificant changes in the error rate, and perceptible influence are rendered only by changes in  $n$  that are comparable with  $M_s$ . Therefore, for the parameter of variety, it is advisable to consider a “dimensionless” and more stable characteristic  $\bar{n} = n/M_s$ .

The calculations have confirmed that for sufficiently large  $\bar{n}$ , the errors are reduced with increases in the values  $\alpha$  and  $n$ . For  $t = 2/3 M_s$ , **Table 1** shows



**Figure 5.** The impact of the length of the training sample  $t$  of the data set Car evaluation on the error rate of classification  $\Sigma_{t,n}$  with different values of the parameter of variety  $n$ .

**Table 1.** Parameters of data sets.

Data set	$M_s$	$M_k$	$M_i$	Features	$\gamma$	$\alpha$	$\bar{n}$	$\bar{\Sigma}_{t,n}$
Abalone	4177	8	29	mixed	689	1	1.6	0.069
Adult <sup>b</sup>	42121*	14	2	mixed	3.2	10	1.5	0.034
Breast Cancer	699	9	2	integer	1.9	1	6.9	0.01
Car evaluation	1728	6	4	mixed	17.5	1	1.0	0.012
Glass	214	9	6	quant.	8.5	1	4.2	0.029
Haberman's Survival	306	3	2	integer	2.8	1	13,1	0.116
Iris <sup>a</sup>	150	4	3	quant.	1	0	4.0	0
Letter Image Recognition	20000	16	26	integer	1.1	1	2.6	0.076
Wine <sup>a</sup>	178	13	3	quant.	1.5	0	2.3	0.016

<sup>a</sup>Objects were numbered via a random number generator. <sup>b</sup>Objects with data errors are not considered.

the lower boundary of the value range of these parameters, under which the error rates are  $\sigma_{t,n} \leq 0.01$  and  $\Sigma_{t,n} \leq 0.05$ .

The results were verified by 10-fold cross-validation, in which splits of combined sample  $\sigma_{t,n}$  and  $\Sigma_{t,n}$  for each variant were calculated for  $\alpha, t$  and  $\bar{n}$ , as listed in **Table 1**. It was found that all  $\sigma_{t,n} = 0$ . The average error rates of classification over the splits, denoted by  $\bar{\Sigma}$ , are provided in **Table 1**.

From **Table 1** it following that, in most cases,  $\bar{\Sigma}$  is smaller than the acceptable rate of 0.05, but for Haberman's Survival data set the norm is exceeded 2.3 fold. This situation is caused by the fact that the data set has the lowest data variety when the objects are described by three features of integer type and are divided into two classes. To maintain high accuracy, this property was compensated for by significantly increasing up  $\bar{n}$  to  $\bar{n} = 13.1$ . Calculations have shown that increasing  $\bar{n}$  by 6.2%, 12.5% and 25% causes the error rate  $\bar{\Sigma} = 0.116$  to decrease to 0.089, 0.069 and 0.063, respectively. Note that the average error rate of currently used methods exceeds 18% [18].

This analysis can greatly simplify the process of solving practical problems, in which, instead of a control sample, a test sample is specified, for which the distribution of classes is unknown. Now, we can use tabulated values of  $\alpha$  and  $\bar{n}$  as the first approximation, which is likely to be the final result. To verify this, from  $t$  objects of the training sample, we need to allocate  $\beta$  objects (for example,  $\beta = 2/3t$ ) and use them directly for training; the remaining objects can be used to monitor the solution accuracy.

Another important result was obtained by testing the method experimentally. A direct application of the above algorithm for the data set Adult gives the minimum values of  $\sigma_{t,n} = 0$  and  $\Sigma_{t,n} = 0.24$  at  $\beta = 2/3Ms$ . The relatively large number of errors can be explained by the fact that the variants of feature values are distributed very unevenly. In particular, one variant belongs to 92% of the objects, and each of the other options belongs to only 0.07% of objects, on average. In this case, because of the large total repeatability of objects with rare feature values, the probabilities of matching for pairs  $\{k, m\}$  of the training and control samples are decreased, which leads to an increased number of errors.

To mitigate of noted effects of rare information, we will introduce the additional procedure of compression, or pre-granulation, which is also based on the above considerations about the data reliability. Now, the values  $q_k^s$  of corresponding feature  $k$  are sorted in ascending order and divided into groups with the same feature values, which are identified by the sequence of variant numbers  $r_k = 1, 2, \dots, j_k$ , where  $j_k$  is the number of feature variants. Then, we check whether the group with  $r_k = 1$  contains at least  $\varepsilon \cdot Ms$  objects, and if not, then objects with  $r_k = 2$  are added to it, and the process continues until the condition is satisfied ( $\varepsilon \geq 0$  is design parameter). The value of  $q_k^s$  for all objects of the group, calculated in this way, is  $\tilde{r}_k = 1$ . Similar calculations are performed for subsequent values of  $r_k$  and  $\bar{r}_k$ . Obviously, in the general case  $\bar{j}_k \leq j_k$ .

After the additional transformation, the solution of task Adult has remained stable and errors are now closer to zero. The corresponding table data calculated

at  $\varepsilon = 0.01$ . Note that the error rates of 16 different solutions of the task, derived from the well-known algorithms, are in the range 0.14 - 0.21 [18].

This procedure has proven effective in addressing a number of other data sets, for example, Letter Image Recognition. It can also be observed as a way to reduce the values of  $\bar{n}$  to reduce the computation time.

## 6. Conclusions

The paper proposes a methodology for solution classification problems, the core of which is an approximate calculation of the invariants of data matrix. The new approach implements the concepts of soft computing and granulation and is biologically inspired. In essence, it reduces to the transformation of all measured scale features, in which the values of features, called indexes, are defined according to a single scale in the new units.

Developed methodology is based on procedures of randomization and indexation of the data set (and, in some cases, also a pre-granulation procedure), which generate an infinite sequence of index matrices. These matrices are invariants of the data matrix in relation to a class of objects. They provide error-free training and allow us to calculate the object class under the simplest formulas of total probability for any single type or mixed types of features.

The proposed method differs from existing ones by the universality and simplicity of the algorithm and, as a rule, almost an order of magnitude higher accuracy.

The obtained results go beyond the problem of classification and have independent significance for the solutions of the problems of data analysis. It can be expected that the method will receive a hardware implementation, and its extension to multi-level data will lead to the development of effective image recognition systems and information retrieval. The application of the method does not require mathematical education, which increases its innovative potential.

## References

- [1] Bishop, C. (2006) Pattern Recognition and Machine Learning. Springer, Berlin, 738.
- [2] Hastie, T., Tibshirani, R. and Friedman, J. (2009) The Elements of Statistical Learning: Data Mining, Inference, And Prediction. Springer-Verlag, Berlin, 764. <https://doi.org/10.1007/978-0-387-84858-7>
- [3] Murphy, K. (2012) Machine Learning. A Probabilistic Perspective. MIT Press, Cambridge, London, 1098.
- [4] Zadeh, L. (1979) Fuzzy Sets and Information Granularity. In: Gupta, N., Ragade, R. and Yager, R., Eds., *Advances in fuzzy set theory and applications*, World Scientific Publishing, Amsterdam, 3-18.
- [5] Calabrese, M., Amato, A., Lecce, V. and Piuri, V. (2010) Holonic-Granularity Holonic Modelling. *Journal of Ambient Intelligence and Humanized Computing*, **1**, 199-209. <https://doi.org/10.1007/s12652-010-0013-3>
- [6] Lecce, V., Calabrese, M. and Martines C. (2015) Towards Intelligent Sensor Evolution: A Holonic-Based System Architecture. *The Sixth International Conference on Sensor Technologies and Applications, SENSORCOMM 2012*, Rome, 19-24 August, 2012, 248-254.

- [7] Yao, J., Vasiliacos, V. and Pedrycz, W. (2013) Granular Computing: Perspective and Challenges. *IEEE Transactions on Cybernetics*, **43**, 1977-1989. <https://doi.org/10.1109/TSMCC.2012.2236648>
- [8] Xu, W. and Li, W. (2016) Granular Computing Approach to Two-Way Learning Based on Formal Concept Analysis in Fuzzy Datasets. *IEEE Transactions Cybernetic*, **46**, 366-378. <https://doi.org/10.1109/TCYB.2014.2361772>
- [9] Li, J., Mei, C., Xu, W. and Qian, Y. (2015). Concept Learning via Granular Computing: A cognitive Viewpoint. *Information Sciences*, **298**, 447-467. <https://doi.org/10.1016/j.ins.2014.12.010>
- [10] Grenander, U. (1976) Lectures on Pattern Theory 1: Pattern Synthesis. Springer-Verlag, New York-Heidelberg-Berlin. <https://doi.org/10.1007/978-1-4612-6369-2>
- [11] Yao, J. and Yao, Y. (2002) A Granular Computing Approach to Machine Learning. *Proceedings of the 1st International Conference on Fuzzy Systems and Knowledge Discovery*, Singapore, 18-22 November 2002, 732-736.
- [12] Zadeh, L. (1973) Outline of a New Approach to the Analysis of Complex Systems and Decision Processes. *IEEE Transactions on Systems, MAN, and Cybernetics*, **1**, 28-44. <https://doi.org/10.1109/TSMC.1973.5408575>
- [13] Kruse, R., Borgelt, C., Klownnn, F., Steinbrecher, M. and Held, P. (2013) Computational Intelligence: A Methodological Introduction. Springer-Verlag, London, 490. <https://doi.org/10.1007/978-1-4471-5013-8>
- [14] Shats, V.N. (2015) On New Computing Technology in Machine Learning. *International Conference on Artificial Neural Networks "Neuroinformatics-2015"*, Moscow, 19-23 January 2015, **2**, 148-158.
- [15] Shats, V.N. (2016) Invariants of Matrix Data in a Classification Problem. *Stochastic Optimization in Informatics*, SPbGU, **12**, 17-32. <http://www.math.spbu.ru/user/gran/optstoch.htm>
- [16] Asuncion, A. and Newman, D.J. (2007) UCI Machine Learning Repository. Irvine University of California, Irvine.
- [17] Ashby, W.R. (1957) An Introduction to Cybernetics. Chapman and Hall, London, 294. <https://doi.org/10.1063/1.3060436>
- [18] Cios, K.J. and Kurgan, L. (2004) CLIP4: Hybrid Inductive Machine Learning Algorithm That Generates Inequality Rules. *Informing Science*, **163**, 37-83. <https://doi.org/10.1016/j.ins.2003.03.015>



**Submit or recommend next manuscript to SCIRP and we will provide best service for you:**

Accepting pre-submission inquiries through Email, Facebook, LinkedIn, Twitter, etc.

A wide selection of journals (inclusive of 9 subjects, more than 200 journals)

Providing 24-hour high-quality service

User-friendly online submission system

Fair and swift peer-review system

Efficient typesetting and proofreading procedure

Display of the result of downloads and visits, as well as the number of cited articles

Maximum dissemination of your research work

Submit your manuscript at: <http://papersubmission.scirp.org/>

Or contact [jilsa@scirp.org](mailto:jilsa@scirp.org)