Scientific Research Publishing

Journal of Intelligent Learning Systems and Applications, 2019, 11, 65-75
https://www.scirp.org/journal/jilsa
ISSN Online: 2150-8410
ISSN Print: 2150-8402

On the Matrices of Pairwise Frequencies of Categorical Attributes for Objects Classification

Vladimir N. Shats

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

How to cite this paper: Shats, V.N. (2019) On the Matrices of Pairwise Frequencies of Categorical Attributes for Objects Classification. *Journal of Intelligent Learning Systems and Applications*, **11**, 65-75.
https://doi.org/10.4236/jilsa.2019.114004

Received: May 28, 2019
Accepted: September 27, 2019
Published: September 30, 2019

Abstract

This paper proposes two new algorithms for classifying objects with categorical attributes. These algorithms are derived from the assumption that the attributes of different object classes have different probability distributions. One algorithm classifies objects based on the distribution of the attribute frequencies, and the other classifies objects based on the distribution of the pairwise attribute frequencies described using a matrix of pairwise frequencies. Both algorithms are based on the method of invariants, which offers the simplest dependencies for estimating the probabilities of objects in each class by an average frequency of their attributes. The estimated object class corresponds to the maximum probability. This method reflects the sensory process models of animals and is aimed at recognizing an object class by searching for a prototype in information accumulated in the brain. Because these matrices may be sparse, the solution cannot be determined for some objects. For these objects, an analog of the *k*-nearest neighbors method is provided in which for each attribute value, the class to which the majority of the *k*-nearest objects in the training sample belong is determined, and the most likely class value is calculated. The efficiencies of these two algorithms were confirmed on five databases.

Keywords

Categorical Attributes, Classification Algorithms, Invariants of Matrix Data, Data Processing

1. Introduction

The solution to the classification problem is reduced to calculating a function that divides a training sample (TRS) into classes and simultaneously obtains an acceptable classification accuracy for a test sample (TS). In most existing me-

thods, algorithms for calculating these functions have considerable computational complexity [1] [2] [3]. In previous work [4], the method of invariants (MI) was proposed, where this function is a linear combination of the simplest functions of the values of each feature that qualitatively simplifies the computation algorithm. It was shown in [5] that the MI corresponds to sensory process models of animals, which aim to recognize an object's class by searching for a prototype in the information accumulated in the brain.

The MI proceeds from the fact that in classification problems, the accuracy of the data plays a special role since the objects, their descriptions, and their classes are correlated, and each type of entity has a randomness component. Therefore, a given data matrix is just one of possible random realization of the matrices that form the set of invariants with respect to the class. This approach is consistent with the concept proposed by L. Zadeh, which says that for most manually solved tasks, high accuracy is not required because the brain perceives only a "trickle of information" about the external world [6]. Moreover, for systems whose complexity exceeds a certain threshold, accuracy and practical sense are almost mutually exclusive characteristics.

In the MI, the range of attribute values after randomization, accompanied by an introduction of an additive component that follows a uniform distribution, is divided along each attribute into equal numbers of intervals, within which the feature values are assumed to be equiprobable. All objects falling within the interval receive an index of the corresponding attribute equal to the interval number.

For each index, one can find lists of numbers of TRS objects of a certain class and then calculate the frequencies of the indices. With some error, these frequencies will be the same for the objects in the TRS and the TS because both samples belong to the same general population. Therefore, it is possible to estimate the probability of the individual attributes of any object in each class. Then, using the simplest formula of the total probability, estimate the probability of an object having a specific set of feature values. Finally, the class of the object is determined based on the maximum likelihood principle.

There is an obvious analogy between indices and categories, the values of which can always be described by a finite sequence of integers 1, 2... Therefore, the MI serves as the basis for this article, in which two algorithms are proposed: one implements the simplest version of the MI developed for quantitative attributes, and the other more fully takes the features of categorical attributes into account.

The efficiency of the new algorithms was tested on five databases [7].

2. Assumptions and Preliminary Assumptions

The article is devoted to solving classification problems for which all attributes are categorical. The solution is based on two MI assumptions:

- The data matrix has a set of invariants with respect to a class of objects.
- Object classes differ in the attribute probability distributions.

For categorical attributes, the number of values or levels n that individual objects can take is an important characteristic of the problem. In real tasks for quantitative attributes, the value of n_q, as a rule, considerably exceeds that of n_c —the corresponding value for categorical attributes. According to the theory proposed by C. Shannon, the information volume per value of a feature increases in proportion to the value of $\log_2(n_q)/\log_2(n_c)$. Therefore, in tasks involving categorical features, the "information load" of the data often increases several fold. This circumstance manifests in an increase in the number of objects of different classes that have the same attribute values. This reduces the difference between the attribute frequencies for objects of different classes, which can lead to an increase in the number of classification errors.

However, categorical attributes also have "favorable" features. The probability of an object of a certain class is an unknown function of its attributes, which takes into account the interrelations among all the attributes. Usually, this function is nonlinearly dependent on the attribute values of the object. This relationship is indirectly taken into account in the accepted assumption of the MI, since the frequencies of attribute indices are calculated for a particular class of objects. Then, this dependence becomes linear, which greatly simplifies algorithm's calculation. One algorithm takes the same approach for categorical attributes whose values are, as noted above, analog indices.

The second algorithm considers the peculiarities of categorical attributes in a different way and is based on a new solution to the question of attribute relationships. Usually, the relationship between random variables is estimated using the Pearson correlation coefficient or the rank correlation coefficient. However, in the framework for this method, we are interested in the frequencies of attribute values that take a relatively small number of values. The paper further shows that pairwise frequencies of features allow an approximate assessment of the relationship between the features of objects of the same class (note that, as a rule, only a weak correlation exists between the categorical features of objects in the same class).

However, pairwise frequencies do not allow the determination of the class of TS objects if no object has the same combination of attribute values in the TRS. To classify objects, this algorithm uses an analog of the k-nearest neighbors method: the object is assigned to a class for which the total number of the k-nearest neighbors of the TRS' objects for each attribute are maximized.

3. Two Algorithms for Solving the Classification Problem

3.1. Statement and Basic Algorithm

Let the vectors $X_k, k \in (1, N)$ describe the values of categorical attributes objects, which form the TRS $\{(\boldsymbol{X}_s, y_s) \mid s \in (1, M)\}$, where y is the vector of the object class labels, M is the number of objects, and missing data are excluded. Without loss of generality, we assume that the values of the attributes $X_k, k \in (1, N)$ and classes (possibly after preliminary encoding) belong to the sets of integers $j \in (1, n_k)$ and $i \in (1, C)$, respectively, where $n_k \ll M$ is the number of values

of attribute X_k and C is the number of classes. The problem is to classify the TS objects.

We denote s objects by $\boldsymbol{x}_s = \left(x_{s1}, \cdots, x_{sN}\right)^{\mathrm{T}}$ and the data matrix by $\boldsymbol{Q} = \left\|x_{sk}\right\|_{M \times N}$. Consider the algorithm for the basic MI (algorithm 1). Using matrix $\boldsymbol{Q}$, we find lists $\omega_i = \left\{s \mid s \in \left(1, M\right), y_s = i\right\}$ of numbers of objects of class $i \in \left\{1, C\right\}$. The sample probability of objects in class i determines the obvious dependence:

$$p_i\left(\boldsymbol{x}\right) = p\left\{X_1 = x_{s1}, \cdots, X_N = x_{sN} \mid s \in \omega_i, k \in \left(1, N\right)\right\}. \tag{1}$$

This dependence allows finding objects whose attribute value $x_k = j$. Let $r_{kj} \geq 0$ denote the number of such objects. Then, the frequency of a value j for an attribute k of the TRS object of class i equals $\left(f_{kj}\right)_i = r_{kj} / l_i$, where $l_i = \left|\omega_i\right|$.

Object $\boldsymbol{x}$ arises as a result of appearances of each attribute k with the corresponding value j. Since these events form a complete group of incompatible events, the total probability formula gives an estimate of the probability that an object belongs to class i:

$$p_i\left(\boldsymbol{x}\right) = \frac{1}{N} \sum_{k=1}^{M} \left(f_{kj}\right)_i, \tag{2}$$

where j is the value of attribute k for object $\boldsymbol{x}$.

Formulas (1) and (2) yield a class probability estimate for the TRS objects. Since TRS and TS belong to a single general population, the formula also determines the frequencies of the TS objects. According to the maximum likelihood principle, the calculated class of the object $\boldsymbol{x}$ is

$$I\left(\boldsymbol{x}\right) = \arg\max_{i \in (1, C)} p_i\left(\boldsymbol{x}\right). \tag{3}$$

3.2. Features of the Model of Probability Density Objects

Essentially, the MI is based on the assumption that a class of objects can be recognized by the probability distribution of its attributes. According to (2), the probability $p_i\left(x\right)$ received its point estimate equal to the average frequencies attributes of object $\boldsymbol{x}$ of class i. Thus, the empirical frequency distribution of features is transformed into the frequency distribution of objects. Therefore, the MI considers the average composition of the attribute distribution as a probability distribution for objects of a particular class.

We investigate the characteristics of this distribution in the case of two attributes that have typical forms of attribute frequency distributions. Our analysis showed that the distributions of each attribute can be considered a sample of the theoretical distributions described by unimodal laws, the maximum of which is located in the middle and the "tails" of the distribution.

Consider the following task. Let objects have two categorical attributes, the values of which describe random variables Y and Z with probability densities $\varphi_y\left(y\right) = \dfrac{b}{a^2 + y^2}$ and $\varphi_z\left(z\right) = c + d * z + g * z^2$, respectively, where $y \in \left(0, n\right)$,

$z \in (0, n)$, and a, b, c, d, g, h and n are parameters. From formula (2), a random variable $U = (Y + Z)/2$ is the composition of Y and Z, which simulates the total distribution of the objects. We are interested in the features of this distribution.

Note that the functions $\varphi_y(y)$ and $\varphi_z(z)$ allow us to obtain an analytical solution for the distribution composition of the above types of attributes. Since these functions determine the corresponding density distribution, their parameters are related by the following:

$$\int_0^n \varphi_Y(y)\mathrm{d}y = 1, \quad \int_0^n \varphi_Z(z)\mathrm{d}z = 1.$$

Obviously, $U = \tilde{Y} + \tilde{Z}$, where $\tilde{Y} = Y/2$ and $\tilde{Z} = Z/2$ are random variables [8]. Given that density $\varphi_{\tilde{y}}(\tilde{y}) = \varphi_y(\mu(\tilde{y}))\mu'(\tilde{y})$, $\mu(\tilde{y}) = 2\tilde{y}$, we obtain $\varphi_{\tilde{y}}(\tilde{y}) = \dfrac{2b}{a^2 + 4\tilde{y}^2}$. Similarly, we find that $\varphi_{\tilde{z}}(\tilde{z}) = 2 * (c + 2 * d * \tilde{z} + 4 * g * \tilde{z}^2)$.

The density $\varphi_U(u)$ is a convolution of the functions $\varphi_{\tilde{Y}}$ and $\varphi_{\tilde{z}}(\tilde{z})$:

$$\varphi_U(u) = \int_0^u \varphi_{\tilde{Y}}(\tilde{y})\varphi_{\tilde{Z}}(u - \tilde{y})\mathrm{d}\tilde{y}.$$

The range of u is divided into segments: $0 \le u \le n/2$ and $n/2 < u \le n$. Because $\tilde{z} \ge 0$, the lower and upper limits of the integrals are equal to 0 and u for the first segment and $u - n/2$ and u for the second segment, respectively. Then, we can obtain the formula for calculating the density:

$$\varphi_U(u) = 4 * b \sum_{q=1}^{3} A_q w_q(u),$$

where $A_1 = c + 2 * d * u + 4 * g * u^2$, $A_2 = -(2 * d + 8 * g * u)$, $A_3 = 4 * g$

$$w_q(u) = \begin{cases} \int_0^u \dfrac{\tilde{y}^{q-1}}{a^2 + 4 * \tilde{y}^2}\mathrm{d}\tilde{y} & 0 \le u \le n/2 \\ \int_{u-n/2}^u \dfrac{\tilde{y}^{q-1}}{a^2 + 4 * \tilde{y}^2}\mathrm{d}\tilde{y} & n/2 < u \le n \end{cases}$$

Sub-integral functions are tabulated and not given for the abbreviated entries.

We performed calculations were performed for a wide range of parameters. The results are illustrated in **Figure 1**, where the density φ_u is determined for the case in which the density φ_z follows a normal distribution, and the φ_y distribution is close to hyperbolic. The figure shows that with respect to the curve φ_z, the ordinates of curve φ_u increase in the region of high values of density φ_y and decrease in sections with low values. Consequently, the function φ_u does not follow a normal distribution. However, confidence intervals of continuous random variables can be estimated only for normal distributions.

From the analysis, it should be noted that the composition distributions of individual attributes result in a poorly predictable distribution for certain classes of objects. Thus, the effectiveness of the various MI algorithms depends on the data characteristics for a particular task and can be tested only empirically.

3.3. Algorithm 2

Algorithm 1 reduces the MI assumption that the individual classes of objects are

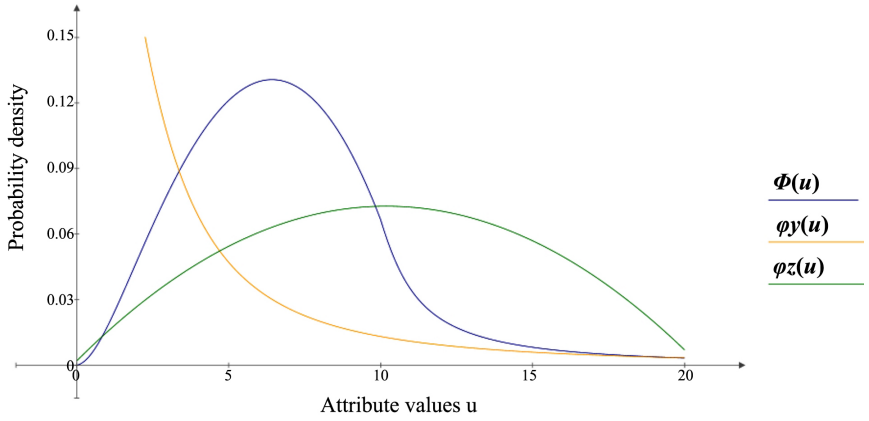

Figure 1. Density curves of random variables Y, Z and U ($\varphi y, \varphi z$ and Φ correspond to φ_y, φ_z and φ_u).

from different distributions of features to classify objects according to the frequencies of the categorical attribute values. However, another variant of the approximate realization of this assumption is also possible.

For any type of attribute, the probability of an arbitrary object $\boldsymbol{x} = (x_1, \cdots, x_N)^{\mathrm{T}}$ of class i is determined by the following relation:

$$P(\boldsymbol{x}) = p_1(x_1)\, p_2(x_2 \mid x_1)\, p_3(x_3 \mid x_1, x_2) \cdots p_N(x_N \mid x_1, \cdots, x_{N-1}), \tag{4}$$

where $p_k(x_k \mid x_1, \cdots, x_{k-1})$ is the conditional probability of an attribute X_k at values $x_1, \cdots, x_{k-1}$ of attributes $X_1, \cdots, X_{k-1}$. Here, $p_1(x_1)$ is found by formula (1).

Consider the features of this dependence for categorical attributes. Here, the elements of the set of Cartesian products of the attributes X_k and X_{k+1}, $k \in (1, N-1)$ are ordered pairs: $(\tilde{x}_k, \tilde{x}_{k+1})$, where $\tilde{x}_k \in \{1, n_k\}$ and $\tilde{x}_{k+1} \in \{1, n_{k+1}\}$. Let $e^i_{k,k+1}$ be the number of objects of class i whose attributes correspond to a pair $(\tilde{x}_k, \tilde{x}_{k+1})$. Then, the frequency of the pair $f^i_{k,k+1} = e^i_{k,k+1} / l_i$ gives the sample estimate probability of the pair for object $\boldsymbol{x}$ of class i. The set of frequencies defines a matrix

$$R^i_{k,k+1} = \left\| f^i_{gw} \right\|_{n_{k+1} \times n_k},$$

constructing a matrix of pairwise frequencies (MPF) for the attributes k and $k+1$ for the TRS objects of class i. There are $N-1$ MPFs for each class. According to the concept formed by the above matrix, we can define the properties of the TRS and TS objects. Then, from formula (4), we obtain the approximate dependence for estimating the probability that object $\boldsymbol{x}$ belongs to class i

$$P_i(\boldsymbol{x}) = p_1(x_1)\, f^i_{x_1,x_2}\, f^i_{x_2,x_3} \cdots f^i_{x_{N-1},x_N}. \tag{5}$$

In formula (5), $f^i_{x_k,x_{k+1}}$ is the element of matrix $R^i_{k,k+1}$ that corresponds to the frequency of the attribute pair values k and $k+1$ of an object in class i. The estimated class of this object is determined by an analog of formula (3):

$$\tilde{I}(\boldsymbol{x}) = \arg\max_{i \in (1,C)} P_i(\boldsymbol{x}). \tag{6}$$

3.4. Improving the Accuracy of Algorithm 2

From formula (5), it follows that $P_i(\boldsymbol{x})=0$ if one of the factors $f^{i}_{x_k,x_{k+1}}=0$. Such a case occurs when there is no object with the same attribute value among the TRS objects of class *i*. The total number of possible combinations of attribute values is $v=n_1n_2\cdots n_N$ and, as a rule, $v\gg l_i$. Therefore, MPFs often contain zero elements and can be sparse.

If $P_i(\boldsymbol{x})=0$ for all *i*, then uncertainty arises, since formula (5) "does not work". Note that when applying algorithm 1, such situations are practically excluded. The MI serves as the basis for eliminating this uncertainty, since it assumes that many data matrices exist that are invariant with respect to a class of objects. It can be assumed that in the case of invariant transformations, the relative position of the attribute values of TRS objects will be preserved near the singular points corresponding to the attribute values of an "undefined" object. Consequently, we can use the idea underlying the *k*-nearest neighbor method to solve classification problems.

We assume that the "undefined" object has a class to which most of the *k*-nearest TRS objects belong. Since the concept of distance between objects is not defined in the MI, we will evaluate the "proximity" for each attribute value of an "undefined" object.

Let Z be a set of TS objects for which the class could not be determined using formula (5) and object $\boldsymbol{z}=(z_1,\cdots,z_N)^{\mathrm{T}}\in Z$. The goal is to find TRS objects of class *i* whose attributes X_k are in *h* neighborhoods of z_k, $k\in(1,N)$. The numbers of these objects form the set $D=\{t\,|\,|x_{tk}-z_k|\le h, t\in\omega_i\}$, and their frequency is $T_{ik}(z_k,h)=\dfrac{|D|}{|\omega_i|}$. Having calculated the frequencies, we can find the average frequency $\overline{T}_i(\boldsymbol{z},h)$ of all the attributes of object $\boldsymbol{z}$ in class *i*. Then, the calculated class of object $\boldsymbol{z}$ is equal to

$$\overline{I}(\boldsymbol{z},h)=\max_{i\in(1,C)}\overline{T}_i(z,h), \tag{7}$$

where *h* is a parameter whose domain is the set of integers $\{1,\cdots,\tilde{n}\}$, where $\tilde{n}=\min(n_k)$.

Let $\mathbf{1}_i(\boldsymbol{z},h)$ be an indicator of class *i* that equals 1 when the calculated class is not equal to the class of object $\boldsymbol{z}$ and 0 otherwise. Then, the number of incorrectly classified objects in the set Z will be equal to

$$F(h)=\sum_{\boldsymbol{z}\in Z}\mathbf{1}_i(\boldsymbol{z},h). \tag{8}$$

The calculated value of parameter *h*, denoted by $\tilde{h}$, and the corresponding value $\overline{I}(\boldsymbol{z},h)$ can be found via the minimum number of "undefined" objects:

$$\tilde{h}=F(h)\to\min_{h\in(1,\tilde{n})}. \tag{9}$$

4. The Effectiveness of New Algorithms

The MI serves as a general conceptual framework for formulas (1)-(3) and (4)-(9), which respectively define algorithms 1 and 2 for solving the classification

problem. The effectiveness of the algorithms was studied with five databases from the UCI repository; the objects in these databases, the objects had only categorical features. The characteristics of the bases given in Table 1 that cover rather wide ranges of values for the numbers of objects (267 - 20,000), features (3 - 22) and classes (2 - 26).

The dependencies in (3) and (5) are applicable not only for the TS but also for the TRS. Therefore, we calculated the test error rate, f_c, and the training error rate, f_l. All the calculations were performed on the basis of the cross-validation procedure. The database was divided into 10 datasets of approximately equal size. The first 9 datasets were used as the TRS, and the remaining dataset was used for testing. This procedure was applied 10 times. Consequently, for each base, a sequence of 10 pairs of TRS and TS variants was considered. For each partitioning variant $m \in (1,10)$, we calculated the error rates f_{cm} and f_{lm}.

The f_{cm} and f_{lm} curves for different databases are shown in Figure 2 and Figure 3, respectively. The graphs are identified by an ordered pair *a_b*, where *a* is the first letter of the database name and *b* is the algorithm identifier. For these rates, the average values *E* and the standard deviations *St* are given in Table 1.

Database Car evaluation and Spect have no "undefined" objects; for them, the functions $F(h)$ were not calculated. Figure 4 depicts the curves Fb_h, Fh_h and

Table 1. Table of databases characteristics and calculation results.

Dataset	*M*	*N*	*C*	Algorithm 1				Algorithm 2			
				TS		TSR		TS		TSR	
				E	*St*	*E*	*St*	*E*	*St*	*E*	*St*
Breast Cancer	699	9	2	0.086	0.052	0.088	0.006	0.175	0.051	0.016	0.023
Car Evaluation	1728	6	4	0.546	0.036	0.536	0.006	0.076	0.015	0.069	0.003
Haberman's Survival	306	3	2	0.314	0.089	0.025	0.015	0.238	0.087	0.07	0.058
Letter Image Recognition	20,000	16	26	0.46	0.008	0.468	0.014	0.186	0.019	0.097	0.002
Spect	267	22	2	0.569	0.011	0.559	0.016	0.237	0.075	0.224	0.013

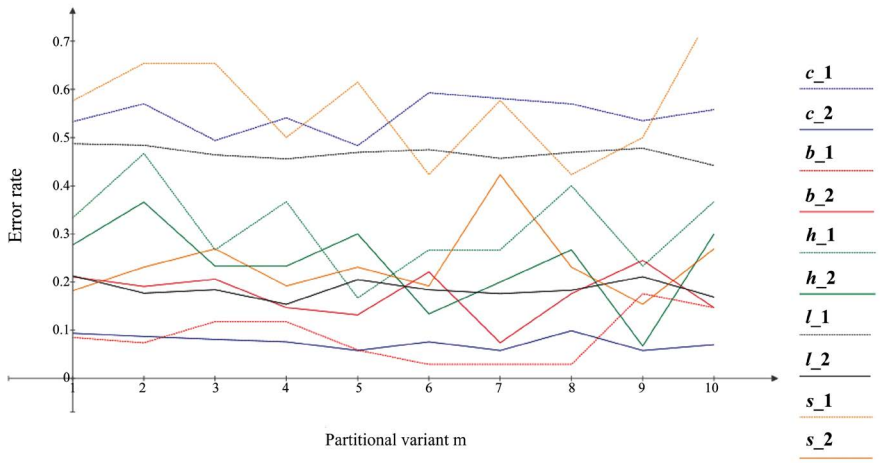

Figure 2. Frequency distributions of test errors f_{cm} for algorithms 1 and 2.

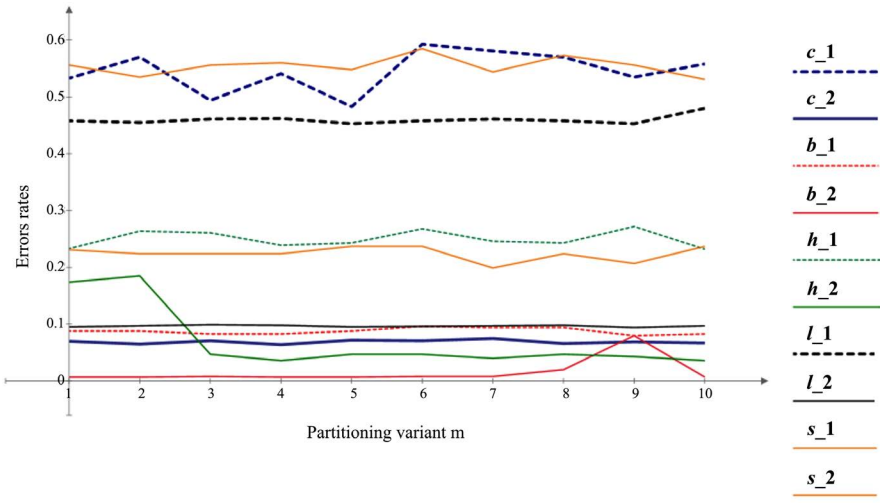

Figure 3. Frequency distributions of learning errors f_{lm} for algorithms 1 and 2.

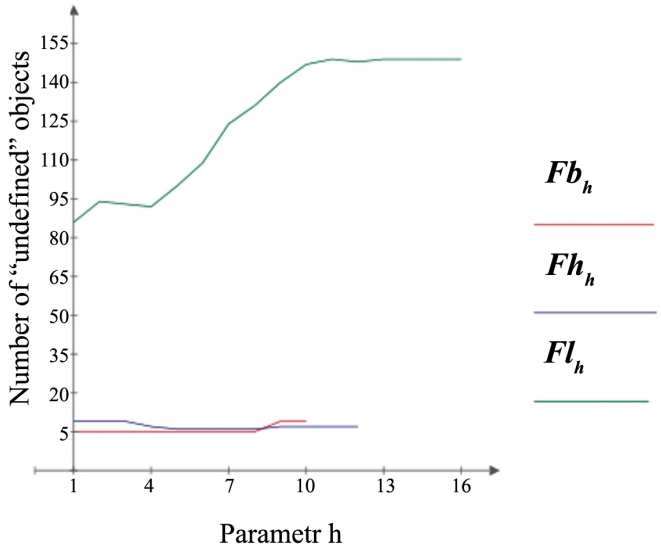

Figure 4. Graphs of the function $F(h)$ for Breast Cancer, Haberman's Survival and Letter Image databases.

Fl_h that reflect the features of these functions for the Breast Cancer, Haberman's Survival and Letter Image databases, respectively.

Below, we summarize the main results of the calculations:

1) With some exceptions, the error rate curves do not undergo drastic changes under the sequential changes in the composition of the TRS and TS objects under cross-validation. Both algorithms yield fairly stable results: in most cases, the error variances for TS and TRS are relatively small ($St/E<1$). The most stable results were obtained for algorithm 2, where $St/E<0.4$ for the TS. We note that the number of test errors is typically considerably higher than training errors.

2) Algorithm 2, as a rule, is much more accurate than algorithm 1. This is well illustrated in **Figure 2**, where almost all the dotted lines corresponding to algorithm 1 are concentrated in the upper part. The resulting conclusion is that con-

sidering the pairwise frequencies of attributes makes it possible to more accurately differentiate the latent properties of objects of different classes. For algorithm 2, the minimum values of the mean error E are 0.076 and 0.016 for the test and training samples, respectively.

3) In many cases, the introduction of the function $F(h)$ and a corresponding reduction in the number of "uncertain" objects can lead to significant increases in the efficiency of the MPF and in the accuracy of the solution.

We can conclude that these experiments confirm the operability of both algorithms.

5. Conclusions

The paper proposes two new algorithms based on the MI for classifying objects with categorical features. Both algorithms originate from the same assumption: that the objects in each class differ in attribute probability distribution, but both algorithms use different models to approximate the distributions. Under this assumption, an object class is defined by the individual frequencies of its attribute values rather than by the nonlinear functions of attributes values used in most existing methods. This characteristic explains the comparative simplicity of the proposed algorithms.

It has been established that along with the correlation between categorical attributes, for objects belonging to one class, a functional relationship exists between the attribute values, which is characterized by the frequencies of the pairwise attribute values. This set of frequencies forms an MPF, which is calculated for the TRS objects for each class and attribute. In one of the algorithms, the MPF is used in conjunction with an analog of the k-nearest neighbors method. This addition allows one to determine the class of a TS object when the TRS does not contain objects with the same combination of attribute values.

It can be expected that the MPF can also be applied to solve problems with quantitative attributes because the values (with some error) can be represented by integers corresponding to the data description with a coarser measuring scale.

An experimental examination has shown that algorithm 2, using the MPF, provides more reliable results than does algorithm 1.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

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. 2nd Edition, Springer, Berlin, 764.

[3] Murphy, K. (2012) Machine Learning. A Probabilistic Perspective. MIT Press, Cambridge, Massachusetts, London, 1098.

[4] 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

[5] Shats, V.N. (2018) The Classification of Objects Based on a Model of Perception. *Studies in Computational Intelligence*, **736**, 125-131.
https://doi.org/10.1007/978-3-319-66604-4_19

[6] 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 Science Publishing, Amsterdam, 3-18.

[7] Hogg, R.V., Tanis, E.A. and Zimmerman, D. (2015) Probability and Statistical Inference. 9th Edition, Pearson, London, 557.

[8] Asuncion, A. and Newman, D.J. (2007) UCI Machine Learning Repository. Irvine University of California, Irvine.