3-бөлім | Раздел 3 | Section 3

Информатика | Информатика | Computer Science

IRSTI 28.23.01 DOI: https://doi.org/10.26577/JMMCS1302202611

A. Tursunov[1*], D. Narzullaev[1], Kh. Ilhamov[1], O. Anorov[2], B. Abdurakhmanov[1], Sh. Tajibaev[3,4]

[1]Tashkent Pharmaceutical Institute, Tashkent, Uzbekistan
[2]Tashkent State Transport University, Tashkent, Uzbekistan
[3]Humanities and Exact Sciences of Mamun university, Khorezm, Uzbekistan
[4]Karakalpak State University, Nukus, Uzbekistan
∗e-mail: tursunovr484za@gmail.com

ALGORITHMS FOR TRANSFORMING FEATURE TYPES FOR PROBLEMS STUDIES OF PAIRED INTERDEPENDENCIES OF FEATURES

The development and improvement of methods for digital processing of symbols, selection of informative symbols, preliminary processing of images and recognition of symbols, and their implementation in practice have been studied by scientists, but the development of methods and algorithms for preliminary processing of data, recognition of symbols and symbols, and selection of informative symbols, and the creation of data processing softwarebased on them have not been studied sufficiently. The article reviews computational algorithms for finding the optimal replacement of integer gradation levels of the number of permutations g used in determining the empirical order scale in the study of paired interrelationships of symbols.

Keywords: Transforming feature types, correlation analysis, experimental data table, feature types, optimization.

А. Турсунов[1*], Д. Нарзуллаев[1], Х. Илхамов[1], О. Аноров[2], Б. Абдурахманов[1], Ш. Тажибаев[3,4]

[1]Ташкент фармацевтика институты, Ташкент, Өзбекстан
[2]Ташкент мемлекеттік көлік университеті, Ташкент, Өзбекстан
[3]Мамун университетінің гуманитарлық және нақты ғылымдар, Хорезм, Өзбекстан
[4]Қаракалпақ мемлекеттік университеті, Нукус, Өзбекстан
∗e-mail: tursunovr484za@gmail.com

Белгілердің жұптасқан өзара тәуелділіктерін зерттеу мәселелеріне арналған белгі түрлерін түрлендіру алгоритмдері

Ғалымдар символдарды цифрлық өңдеу, ақпараттық символдарды іріктеу, кескіндерді алдын ала өңдеу және символдарды тану әдістерін әзірлеп, жетілдіріп, оларды тәжірибеде қолдану мәселелерін зерттеген. Алайда, деректерді алдын ала өңдеу, символдарды тану және ақпараттық символдарды іріктеу әдістері мен алгоритмдерін әзірлеу, сондай-ақ осылардың негізінде деректерді өңдеуге арналған бағдарламалық жасақтама құру әлі де жеткілікті зерттелмеген. Мақалада символдардың жұптық өзара байланыстарын зерттеуде қолданылатын эмпирикалық реттілік шкаласын анықтау үшін пайдаланылатын g орын ауыстырулар санының бүтін санды градация деңгейлерін оңтайлы ауыстыруды табудың есептеу алгоритмдеріне шолу жасалған.

Түйін сөздер: Белгі түрінің өзгерісі, көп айнымалы деректерді талдау, эксперименттік деректер кестесі, корреляциялық талдау.

А. Турсунов[1*], Д. Нарзуллаев[1], Х. Илхамов[1], О. Аноров[2], Б. Абдурахманов[1], Ш. Тажибаев[2]
[1]Ташкентский фармацевтический институт, Ташкент, Узбекистан
[2]Ташкентский государственный транспортный университет, Ташкент, Узбекистан
[3]Гуманитарные и точные науки Университета имени Мамуна, Хорезм, Узбекистан
[4]Каракалпакский государственный университет, Нукус, Узбекистан
∗ e-mail: tursunovr484za@gmail.com

Алгоритмы преобразования типов признаков для задач исследования парных взаимозависимостей признаков

Разработка и совершенствование методов цифровой обработки символов, выбора информативных символов, предварительной обработки изображений и распознавания символов, а также их внедрение в практику были изучены учеными, но разработка методов и алгоритмов предварительной обработки данных, распознавания символов, выбора информативных символов и создание программного обеспечения для обработки данных на их основе изучены недостаточно. В статье рассматриваются вычислительные алгоритмы нахождения оптимальной замены целочисленных уровней градации числа перестановок, используемых при определении эмпирической порядковой шкалы в исследовании парных взаимосвязей символов.
Ключевые слова: Изменение типа признака, многомерный анализ данных, таблица экспериментальных данных, корреляционный анализ

1 Introduction

Information analysis, intellectual systems and modern statistician modeling in the fields research deepening to go with, various typical from signs organization found experimental information tables again work current scientific to the point is spinning. In fact, in practice common almost every one big information complex (Big Data) is one of time in itself quantitative, orderly, nominal and classification to the features owner was indicators own inside takes. Such information heterogeneous signs table because is called and them analysis in doing classic statistician methods limited to the opportunity owner will be. of information heterogeneity from that that is, every one sign certain scale type owner Quantity in signs arithmetic actions complete execution possible if so, then in signs only coloring and comparison possible, nominal in signs and only equivalence relationship exists. This because of, deep modeling or correlational analysis take on the way the information complete cover to take opportunity It won't happen. and research of the results one side by side interpretation to be done, sometimes and of information to the loss take comes.

Classic methods regression, discriminant analysis, factor analyses only quantitative to the signs justification due to, orderly and face value in signs common association, similarity, compatibility like complicated relationships to find out can't as a result, research object about general information fragmented in case will remain and information between internal connections complete open not given.

Scientific in literature scientists by signs types scaling and digital to appear to bring according to important theoretical basics created [1–7]. That's it with together, next ten in years research to practice directed in case developing information mechanical teaching to the models suitable to the form bringing, correlative and clustering in his duties the signs transformation to do algorithms improvement main from directions one as remains.

Especially medical information analysis, genomics, agricultural technology forecasting, digital economy, industry IoT (Internet of Things) information) sensors through analysis take to go like in the fields heterogeneous the information again work important importance profession This in the fields of signs nature sharp difference For example, if a patient's age or blood pressure is quantitative, clinical cases are ordinal, diagnostic types are nominal, and biomarkers are classificatory. In order to analyze such data as a whole, it is necessary to bring the types of features into a single type and determine the required order scales. This process is central to all stages of the research: correlation analysis, clustering, classification, forecasting, and selection of informative features.

The research of A. Nishanov and co-authors has provided important scientific results in the development of similarity functions for various types of data, digitization of signs, automatic search of ordinal scales, comparison of objects based on information measurements, structural clustering and classification algorithms[9–14]. For example, successful work has been carried out in such areas as modification of the "ball Apolonia"rules, transformation of signs in medical data, and integration of heterogeneous data for early detection of oncological diseases.

Also, the research results are developed in the form of practical software, expanding the practical application of transformation algorithms, allowing them to be effectively implemented in areas such as education, medicine, economic analysis, and intelligent monitoring in industry.

Another important aspect of feature type transformation is that it unifies the general TED (table of experimental data) structure, reveals hidden connections between data, and forms a complete map of correlation relationships. Models (clustering, classifying, predicting) built on the basis of a unified structure achieve high accuracy, stability, and reliability.

2 Materials and methods

The following methods were used in this study:

1. Theoretical: analysis of literature on the subject of studying approaches to transforming feature types when solving the problem of studying paired interdependencies of features;

2. Empirical: comparison of the results of the developed algorithms for transforming feature types using the example of solving the problem of studying paired interdependencies of features. As a result, new algorithms for transforming feature types were developed when solving the problem of studying paired interdependencies of features.

3 Results

Suppose we can represent the matrix X in the set of symbols in the following form:

$$X = [X_1,\ X_2, X_3] \tag{1}$$

of dimension m_p, where X_1 is a submatrix having l quantitative features $x^{(1)},\ x^{(2)}, \ldots, x^{(l)}, X_2$ – has only qualitative features with indices $i = \overline{l+1, m}$, and X_3 – a submatrix of that X has only classification features with indices $i = \overline{m+1, p}$. We will consider the problem of finding imperative order scales as a problem of minimizing a function of many variables, represented by the following formulas:

$$F(X_1, X_2, D_{m+1}, \ldots, D_p) \rightarrow \ min \tag{2}$$

The numerical set B^1 can be any linearly ordered set, such as a portion of the natural series of integers $K_g = \{1, 2, \ldots, g\}$. In this instance, the mapping ξ will allocate the value of the integer label d_i from K_g to each i-th gradation, which is known as the gradation's rank. Permutations of the ranks $D \in \ P^g$, where P^g is the set of permutations g of natural numbers, will then provide the mapping ξ.

Let us apply the generalized formula for the criterion for searching imperative scales to the problem of studying paired interdependencies of features.

The analysis of the interdependencies of features involves computing and interpreting the matrix $Z = (r_y)$ of dimension $p \times \ p$, which elements are sample estimates of the indicators of the paired relationship of features of a specific type. Let the initial TED be presented as a matrix X. The formation of Z in the presence of various TED characteristics is depicted in Fig.1. The elements of the submatrices $Z_{12}\,(Z_{21})\,, Z_{13}\,(Z_{31})\,, Z_{23}(Z_{32})$ describe the relationship between features of various types, while the submatrices Z_{11}, Z_{22}, Z_{33} comprise measures of connection for quantitative, qualitative, and classification features, respectively.

It should be mentioned that each kind of characteristic has a variety of correlation indices. Various studies conclude that most correlation measurement methods are based on either the covariance principle or the mutual conjugacy principle.

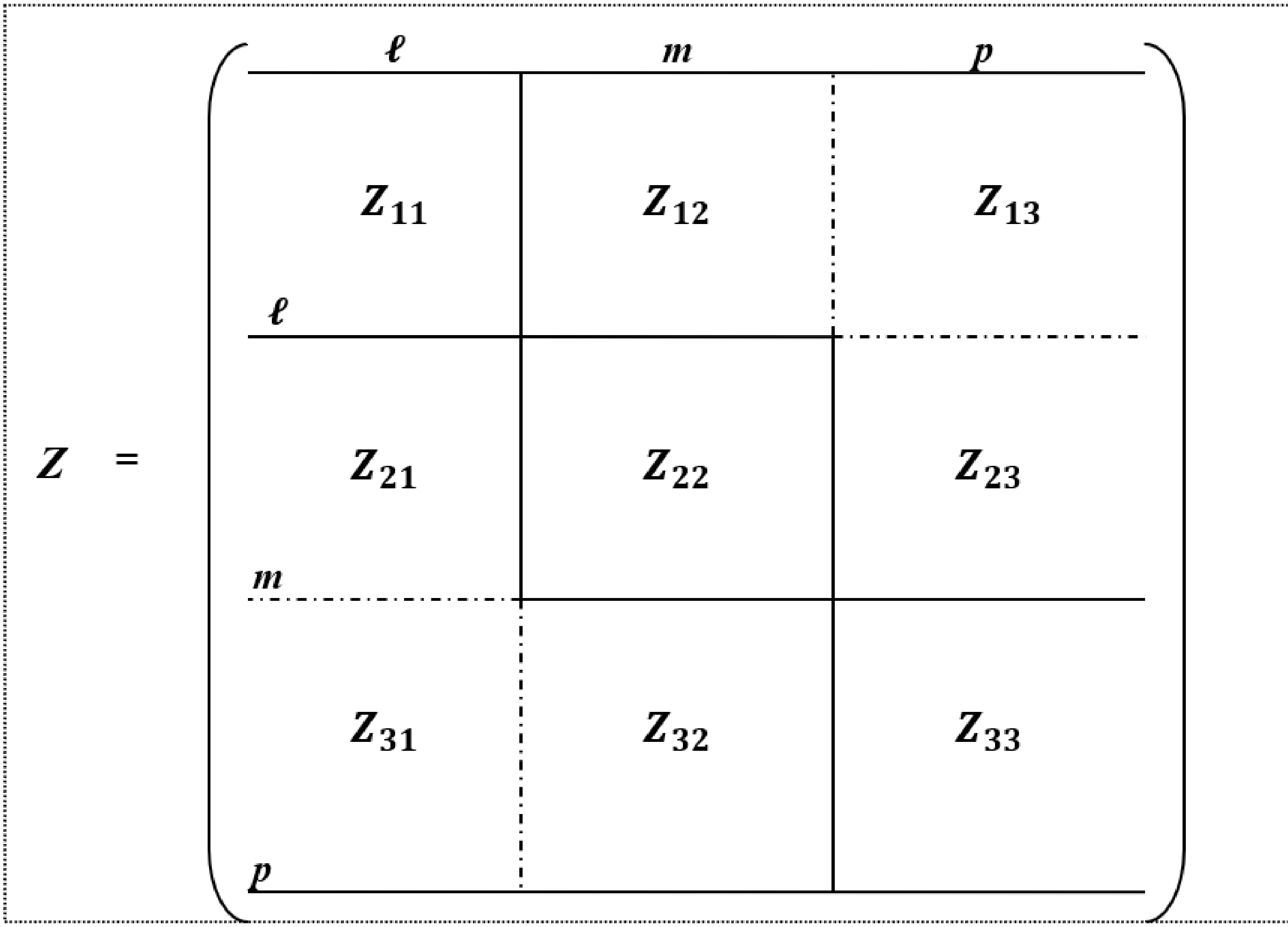

Figure 1. Matrix of indicators of paired relationships of different types of features

In the context of applying the covariance principle, a determination regarding the existence of a relationship between features is established when an escalation in the numerical values of one feature is paired with a consistent increase or decrease in another feature. Mathematically, this issue is reduced to the computation of the covariance value, the corresponding variations in the numerical values of the features, and the subsequent normalization of this value. In this particular instance, submatrices Z_{11} and Z_{22} are utilized to represent quantitative and qualitative features, respectively. As a metric of association, one may consider both the conventional correlation coefficient and the rank correlation coefficients (ρ -coefficients) established by Kendal, Spearman, and Stewart [2]. The elements within the submatrix Z_{22} are derived through the application of the formula for the conventional correlation coefficient based on rank values.

The measures of connection of classification features that form the elements of the submatrix Z_{33} are built on the principle of mutual conjugacy. These include Pearson, Cramer, Chuprov coefficients.

To calculate the measures of connection between different types of features, we will also use the usual correlation coefficient. This is done in order to strengthen the weak scale and prevent loss of information[3].

Let's assume that the feature x_1 – quantitative, and the feature x_2 – classification. It is necessary to identify relationships not reflected by the non-quantitative scale that would allow us to talk about a coordinated change in the values of the features under consideration. In other words, it is necessary to determine the imperative scale generated by the original non-quantitative scale and allowing us to transform the non-quantitative feature x_2 in quantitative x_2^{ξ} [4-7]. To solve this problem, we consider the set $\{\xi\}$ of imperative scales generated by the original non-quantitative scale. Then the value of the correlation coefficient r_{12}, calculated for x_1 and x_2^{ξ} and measured in some scale $\{\xi\}$, will characterize the degree of tightness of the linear relationship between x_1 and x_2, which is provided by this imperative scale. The correlation coefficient is a measure of the accuracy of the prediction of a quantitative feature and therefore r_{12} evaluates the quality of the prediction of the values of x_1, produced by numerical labels ξ using regression

$$x_1 = ax_2^{\xi} + b \tag{3}$$

he imperative scale ξ^*, at which the maximum value of the correlation coefficient r_{12} is achieved, ensures the most accurate forecast of feature values along the regression line (3). Therefore, when analyzing the paired interdependence of different types of features, the criterion for selecting a quantitative imperative scale is the maximization of r_{12}:

$$1 - r_{12} \to \ min \tag{4}$$

Formula (4) as applied to the optimization problem (2) can be written as:

$$1 - r_{12}^2(x_1, D) \to \ min, \quad D \in \ P^g \tag{5}$$

where x_1 are the values of the feature measured on a quantitative or rank scale.

Generalizing the above to the multidimensional case, we can conclude that the use of the method of transforming feature types for the data analysis task under consideration will allow us to apply conventional correlation coefficients to calculate the relationship indicators in the Z_{12} submatrices. $Z_{12}\left(Z_{21}\right), Z_{13}\left(Z_{31}\right), Z_{23}\left(Z_{32}\right)$. Furthermore, the correlation coefficients derived from Z, computed for features of varying types using imperative scales, can be utilized in the development of multidimensional models.

Let us now proceed to the analytical reformulation of criterion (4), aimed at designing an algorithm for converting feature types. To this end, we examine a scenario in which the empirical data table (EDT), of dimension $N \times (\ell + 1)$, includes ℓ quantitative features $x_l, x_i, \ i = \overline{1, \ell}$, along with a non-quantitative feature $x_{\ell+1}$ comprising g distinct gradations.

Let be $\overline{a}^T = (a_1, \ldots, a_g)$ some set of numerical labels for feature $x_{\ell+1}$; n_i be the number of objects possessing the i - th gradation of feature $x_{\ell+1}$; m_k be the sample mean of the k -th feature, $k = \overline{1, \ell}; m_k^i$ be the partial sample mean of the k -th feature in the i - th gradation of the $\ell + 1$ - th feature. Let us calculate the elements of the $\ell + 1$- th row and the $\ell + 1$- th column of the covariance matrix Λ:

$$\lambda_{k,\ell+1} = \ \frac{1}{N}\sum_{i=1}^{g} a_i n_i \left(m_k^i - m_k\right).k = \overline{1, \ell};$$

$$\lambda_{\ell+1,\ell+1} = \frac{1}{N}\sum_{i=1}^{g} a_i n_i (a_i - \frac{1}{N}\sum_{i=1}^{g} a_i n_i)$$

Let $\mu_k^i = \frac{n_i}{N}(m_k^i - m_k)$. then we get

$$\lambda_{k,\ell+1} = \lambda_{\ell+1,k} = \sum_{i=0}^{n} a_i \mu_k^i \tag{6}$$

Let us denote by $T = [t_{ij}]\,, i,j = \overline{1,g}$,
where

$$t_{ij} = \begin{cases} -\dfrac{n_i n_j}{N}, & i = j, \\ \dfrac{n_i}{N^2}(N - n_i), & i \neq j. \end{cases}$$

Then in matrix form $\lambda_{\ell+1,\ell+1}$ the equality holds:

$$\lambda_{\ell+1,\ell+1} = \overline{a}^T T \overline{a} \tag{7}$$

Let us denote by $\underline{\Lambda}_{ij}$- the submatrix of the matrix obtained $\underline{\Lambda}$ by removing from $\underline{\Lambda}_i$ - th rows and j - th column; $\Lambda_{ij,pq}$ - by deleting the i - th and p - th rows, i - th and q - th columns; Λ_{ij} - the additional minor of the element λ_{ij} in the determinant Λ; $\Lambda_{ij,pq}$- the determinant of the matrix $\Lambda_{ij,pq}$. Assuming that rang $(\Lambda_{\ell+1,\ell+1}) = \ell$, we define the matrices:
$M = [\mu_k^i]$ of dimension $g \times \ell$,

$$C = M \underline{\Lambda}_{\ell+1,\ell+1}^{-1} M^T, \tag{8}$$

$$S = T - C \tag{9}$$

It is possible to prove the equality:

$$\Lambda = \Lambda_{\ell+1,\ell+1} \bar{a}^T S \overline{a} \tag{10}$$

Which is the desired representation of the determinant in the form of an explicit dependence on the numerical labels. Here, the determinant of the matrix Λ, which includes the digitized variable, is computed based on the covariance matrix $\Lambda_{\ell+1,\ell+1}$, considering only the quantitative variables. This approach enables the formulation of criteria for identifying imperative scales in which the influence of quantitative features is inherently preserved, eliminating the need for their prior conversion to nominal form.

It can be shown that the matrices T, C and S are degenerate. Let us calculate the value of the quadratic form $\overline{a}^T S \overline{a}$:

$$\Lambda = \sigma_1^2 \left(1 - r_{12}^2\right) \ldots \sigma_\ell^2 \left(1 - r_{\ell,1..\ell-1}^2\right) \sigma_{\ell+1}^2 \left(1 - r_{\ell+1,1...\ell}^2\right) = \Lambda_{\ell+1,\ell+1} \sigma_{\ell+1}^2 (1 - r_{\ell+1,1..\ell}^2)$$

where

$$\overline{a}^T S \overline{a} = \sigma_{\ell+1}^2 (1 - r_{\ell+1,1...\ell}^2) \tag{11}$$

Let us proceed to obtaining an expression for the functions of the criteria for selecting imperative scales for the problem of analyzing the interdependencies of features. In this problem of data analysis for transforming types of features, it is proposed to use as a criterion for selecting imperative scales the minimization of the value $1-r_{12}^2$, where r_{12}^2 is the correlation coefficient between the quantitative feature x_1 and a digitized non-quantitative feature x_2^{ξ}. Using (6) and (7) for r_{12}, we can obtain the following relationship:

$$r_{12}^2 = \frac{\overline{a}^T C \overline{a}}{\overline{a}^T T \overline{a}}$$

where $C = \overline{\mu}\lambda_{11}^{-1}\mu^{-T}, \mu^T = [\mu_1^1,\ \mu_1^2, \ldots, \mu_1^g]$ from these relations we obtain:

$$1 - r_{12}^2 = \frac{\overline{a}^T H \overline{a}}{\overline{a}^T \overline{a}} \tag{12}$$

where $H = T - C$. Thus, the criterion for searching for numerical labels when studying paired dependencies has the form

$$\frac{\overline{a}^T H \overline{a}}{\overline{a}^T T \overline{a}} \rightarrow\ min,\quad \overline{a} \in\ B^g \tag{13}$$

Let us proceed to solving the optimization problem (13) under the condition that it is necessary to find the optimal integer numerical labels. In this case, problem (13) will take the form:

$$\frac{\overline{D}^T H \overline{D}}{\overline{D}^T T \overline{D}} \rightarrow\ min \tag{14}$$

where $D \in\ P^g$.

First, consider the case where the number of gradations g of the classification feature $x_{\ell+1}$ does not exceed six, i.e., $g \leq 6$. In this scenario, each rank-based imperative scale corresponds to a permutation of g integer ranks, denoted as $D^T = (d_1, \ldots, d_g)$, where $d_i \in K_g$. Given that the total number of permutations, $g!$, remains relatively small for this range, we propose evaluating criterion (14) through exhaustive enumeration of all possible permutations. Accordingly, a procedure is defined for generating all permutations corresponding to the g gradations of the classification feature $x_{\ell+1}$.

Let us begin by setting the initial value $k_1 = 1$. By sequentially adding the second number $k_2 = 2$—first to the beginning, then to the end—we obtain 2! permutations: $\{2, 1\}$ and $\{1, 2\}$. At the third step, the process yields 3! permutations, and by step g, all $g!$ permutations are generated.

The computational algorithm for determining the optimal permutation of integer ranks corresponding to gradations (for $g \leq 6$), used in constructing the imperative order scale, proceeds as follows [8-10]:

Step 1. Determine all $g!$ permutations using the scheme proposed above.

Step 2. Select the initial permutation of gradations $\overline{D}_0 = (1, 2, \ldots, g)$ and calculate the value $F\left(\overline{D}_0\right) = \frac{\overline{D}_0^T H \overline{D}_0}{\overline{D}_0^T T \overline{D}_0}; F_{min} = F(\overline{D}_0)$

Step 3. Move to the next permutation $\overline{D}_1$. Calculating the value of $F(\overline{D}_1)$. If $F(\overline{D}_1) < F_{min}$, then $F_{min} = F(\overline{D}_1)$

Step 4. If all $g!$ permutations have been viewed, then go to step 5. Otherwise, go to step 3.

Step 5. The vector $\overline{D}_k$ corresponding to the minimum value of F_{min} is the desired imperative order scale.

4 Discussion

Let us now consider the case when $g > 6$. Under this condition, the total number of possible permutations of g integer ranks becomes prohibitively large ($g!$), making the exhaustive search method impractical due to excessive computational time. To address this, reference [4] proposes using the method of successive approximations to identify the optimal permutation that satisfies criterion (14). In this approach, the initial permutation $\overline{D}_0$ is selected arbitrarily, and the resulting solution depends heavily on this initial choice. Consequently, different initializations may lead to different local minima of (14).

To mitigate this limitation, we propose an alternative strategy. First, we define a method for computing the initial approximation $\overline{D}_0$ by solving an auxiliary optimization problem. Assuming that the minimum of (14) is to be found over the set of real numbers, we extend the admissible solution space to B^g. The resulting problem takes the form of expression (13), which is known as the generalized Rayleigh relation. It is well established that, for the standard Rayleigh relation expressed as

$$\frac{(Ax,x)}{(x,x)} \tag{15}$$

the inequalities are valid

$$v_1 \leq \frac{(Ax,x)}{(x,x)} \leq \ v_n$$

and also the ratio

$$v_1 = min_{x\neq 0}\frac{(Ax,x)}{(x,x)} \tag{16}$$

where x is any nonzero vector, and let $v_1, \ldots, v_n$ denote the eigenvalues of the matrix A, where v_1 is the smallest among them. We now impose a condition for the vector $\overline{a}$ to be considered centralized, that is,

$$\frac{1}{N}\sum_{i=1}^{g} a_i n_i = 0$$

In this case, the matrix T, used in analyzing the interdependence of features, assumes a diagonal form with elements $\frac{n_i}{N}$, and thus is non-degenerate. Consequently, relation (13) can be simplified to the form (15), since:

$$\frac{\overline{a}^T H\overline{a}}{\overline{a}^T T\overline{a}} = \frac{(H^T\overline{a,a})}{(T^T\overline{a,a})} = \frac{(T^{T^{-1/2}}H^T T^{T^{-1/2}}\overline{b},\overline{b})}{(\overline{b},\overline{b})}$$

Where $\overline{b} = T^{T^{-1/2}}\overline{a}$. Let us denote $B = T^{T^{-1/2}}H^T T^{T^{-1/2}}$ Then we have

$$\frac{\overline{a}^T H\overline{a}}{\overline{a}^T T\overline{a}} = \frac{(B\overline{b,b})}{(\overline{b,b})};$$

Using relation (16) for this equality, we can determine the eigenvector $\bar{b}_0$ corresponding to the smallest eigenvalue of the matrix B. The desired vector is then computed as follows:

$$\overline{a} = B^{-1/2}\overline{b}$$

However, the process of finding the matrix $B^{-1/2}$ is quite cumbersome, so we will introduce one more additional restriction:

$$\overline{a}^T T\overline{a} = 1 \tag{17}$$

Then the minimization problem (13) is reduced to finding the conditional minimum of the quadratic form

$$\overline{a}^T H\overline{a} \rightarrow \ min \tag{18}$$

provided.

To solve problem (18), we use the Lagrange multiplier method, according to which the optimal vector $\overline{a}$ makes the first derivative of the Lagrange function vanish.

$$F\left(\overline{a}\right) = \overline{a}^T H\overline{a} - va^T T\overline{a}$$

where v is the introduced Lagrange multiplier. We have:

$$\frac{\partial F(\overline{a})}{\partial\overline{a}} = 2H\overline{a} - 2vT\overline{a} = 0$$

or we obtain a system of linear homogeneous equations

$$H\overline{a} - vT\overline{a} = 0. \tag{19}$$

Let's multiply this equation on the left by $\overline{a}$ and, taking into account (17), we get

$$\overline{a}^T H\overline{a} = v \tag{20}$$

The value (13) is determined by the introduced Lagrange multiplier under condition (18). Therefore, by choosing as v the minimum eigenvalue and solving the system of equations (19), we obtain the vector $\overline{a}$, which minimizes (13). Multiplying (19) on the left by T^{-1}, we obtain

$$T^{-1}H\bar{a} = v\bar{a}$$

In this case, the optimal solution to (13) is given by the eigenvector of the matrix $T^{-1}H$ corresponding to its smallest eigenvalue v. Let us denote the resulting optimal vector that minimizes (13) by $\overline{a}_0$. We now sort the components of the vector $\overline{a} * 0$ in ascending order:

$$a_{i1} \leq \ a_{i2} \leq \ldots \ a_{ig}$$

where i_j denotes the index corresponding to the position of component a_{ij} in the vector $\overline{a} * 0$, taking values from the set $1, 2, \ldots, g$. Assign integer values to the vector components to obtain $\overline{D} * 0$ as follows:

$$d_{i1} = a_{i1} = 1; \ d_{i2} = a_{i2} = 2; \ldots \ d_{ig} = a_{ig} = g$$

This yields the vector

$$\overline{D}_0 = (d_1, \dots, d_g)$$

which we regard as the initial rank permutation used in determining the imperative order scale by the method of successive approximations, the essence of which is described below.

Let M_1 and M_2 denote the numerator and denominator, respectively, of the ratio in (14). When the k-th and l-th gradations are transposed, the corresponding increments of the numerator and denominator are computed using the formulas given in [4]:

$$\Delta M_1 = \overline{c}\,\overline{d}, \qquad \overline{c} = (c_1, \dots, c_g)$$

$$c_i = \begin{cases} 2(d_k - d_\ell)(h_{i\ell} - h_{ik}), & i \neq k, \ell, \\ (d_k - d_\ell)(h_{\ell\ell} - h_{kk}), & i = k, \ell. \end{cases}$$

$$\Delta M_2 = \overline{f}\,\overline{d}, \qquad \overline{f} = (f_1, \dots, f_g)$$

where

$$f_i = \begin{cases} 2(d_k - d_\ell)(t_{i\ell} - t_{ik}), & i \neq k, \ell, \\ (d_k - d_\ell)(t_{\ell\ell} - t_{kk}), & i = k, \ell. \end{cases}$$

Obviously, provided that

$$\frac{1 + \Delta M_1/M_1}{1 + \Delta M_2/M_2} < 1 \tag{21}$$

transposition of the ranks of the k -th and l -th gradations will decrease the value of (14).

We now outline the computational algorithm used to identify the optimal permutation of integer ranks of gradations that minimizes the ratio (14) for the case when $g > 6$:

Step 1. Generate the initial permutation of gradation ranks $\overline{D}_0$ using the procedure previously described.

Step 2. Choose a pair of gradations k and l, where k was not selected in the previous iteration.

Step 3. Compute the increments ΔM_1 and ΔM_2 resulting from swapping the ranks d_k and d_l. Then calculate the value of ratio (21).

Step 4. If inequality (21) holds, update the current permutation to a new one, $\overline{D}_0$. If not, proceed to Step 5.

Step 5. Evaluate the stopping condition. If no transposition of any rank pair results in a further reduction of expression (14), terminate the algorithm. Otherwise, return to Step 2.

This algorithm provides a systematic method for obtaining the optimal rank permutation $\overline{D}$ that minimizes expression (14) for the feature $x^{\ell+1}$. If multiple non-quantitative features are present, the same procedure is applied to $x^{\ell+2}, \dots, x^p$.

5 Conclusion

In the research conducted, significant results were achieved aimed at increasing the efficiency of processing heterogeneous data and expanding the possibilities of correlation analysis under conditions of mixed types of traits. First and foremost, methods and algorithms have been developed for transforming different scales and types of features, ensuring a rigorous and systematic basis for subsequent statistical analysis.

Special emphasis was placed on non-quantitative characteristics, for which an algorithm was introduced to construct imperative order scales for their gradations. This approach enables the transformation of categorical features into ordered forms without compromising their semantic integrity. As a result, quantitative methods can be effectively applied to assess relationships, thereby overcoming the limitations inherent in classical techniques when handling qualitative data.

The developed algorithms are fully implemented in the integrated software development environment and included in the data analysis system, which confirms their practical significance and readiness for practical use. The integration of the proposed solutions into analytical tools opens up opportunities for a more in-depth study of relationships between features, improving the accuracy of correlation assessments, and expanding the range of tasks solved within the framework of data mining.

Thus, the presented methodological and software developments represent a comprehensive solution to the problem of processing heterogeneous features and form a scientifically grounded basis for further research and implementation in modern analytical systems.

References

[1] Pfanzagl, I. (1976). *Measurement theory.* Moscow: Mir.

[2] Did, E. (1985). *Data Analysis Methods.* Moscow: Finansi i statistika.

[3] Ayvazyan, S.A., Yenyukov, I.S., & Meshalkin, L.D. (1985). *Applied statistics: Dependency research.* Moscow: Finansi i statistika.

[4] Nikiforov, A.M., & Fazilov, Sh.Kh. (1988). *Methods and algorithms for converting feature types in data analysis tasks.* Tashkent: FAN.

[5] Enyukov, I.S. (1980). Methods for digitizing non-quantitative variables. In *Algorithms and software for applied statistical analysis* (pp. 309–315). Moscow: Nauka.

[6] Lbov, G.S. (1981). *Methods for processing different types of experimental data.* Novosibirsk: Nauka.

[7] Alexandrov, V.V., & Gorsky, N.D. (1983). *Algorithms and programs for the structural method of data processing.* Leningrad: Nauka.

[8] Narzullaev, D.Z., Abdurakhmanov, B.A., Baydullaev, A.S., Ilyasov, S.T., & Shadmanov, K.K. (1988). Transformation of types of signs for a task of regression analysis. IOP Publishing.

[9] Nishanov, A., Ruzibaev, O., & Tran, N. (2016). Modification of decision rules "Ball Apolonia" for the classification problem. In *Proceedings of the 2016 International Conference on Information Science and Communications Technologies (ICISCT)*.

[10] Nishanov, A., Saidrasulov, Sh., & Babadjanov, E. (2022). Analysis of methodology of rating evaluation of digital economy and e-government development in Uzbekistan. *International Journal of Early Childhood Special Education*, 14(2), 2447–2452.

[11] Nishanov, A., Ruzibaev, O., Chedjou, J.C., Kyamakya, K., Kolli, A., De Silva, D., Djurayev, G., & Khasanova, M. (2020). Algorithm for the selection of informative features in the classification of medical data. In *Developments of Artificial Intelligence Technologies in Computation and Robotics*, 12, 647–658.

[12] Nishanov, A., Akbarova, M., Tursunov, A., Ollamberganov, F., & Rashidova, D. (2024). Clustering algorithm based on object similarity. *Journal of Mathematics, Mechanics and Computer Science*, 123(3), 108–120.

[13] Nishanov, A., Zaripov, F., Akbaraliev, B., Babadjanov, E., & Geldibayev, B. (2025). Improved deep learning model for cattle identification using muzzle images. *Journal of Mathematics, Mechanics and Computer Science*, 125(1).

Information about authors:

Tursunov Alisher Tulkunovich – Doctor of philosophy in technical sciences, Associate Professor of the Department of Physics, Mathematics and Information Technology at the Tashkent Pharmaceutical Institute (Tashkent, Uzbekistan, e-mail: tursunovr484za@gmail.com).

Narzullaev Davronbek Zikrullaevich – Candidate of Technical Sciences, Associate Professor of the Department of Physics, Mathematics and Information Technology at the Tashkent Pharmaceutical Institute (Tashkent, Uzbekistan, e-mail: davr1960@mail.ru).

Ilkhamov Khisamiddin Shatursunovich – Candidate of Technical Sciences, Associate Professor of the Department of Physics, Mathematics and Information Technology of the Tashkent Pharmaceutical Institute (Tashkent, Uzbekistan, e-mail: khisamiddin@mai.ru).

Anorov Obidjon Umaralievich – Acting Associate Professor of the Department of Higher Mathematics, Tashkent State University of Transport (Tashkent, Uzbekistan, e-mail: anorovobidjon6@gmail.com).

Abdurakhmonov Bakhromjon Alisherovich – PhD, Associate Professor of the "Communication and Digital Technologies"Department of the University of Management and Future Technologies (Tashkent, Uzbekistan, email: b.abdurakhmonov75@gmail.com).

Tajibaev Shukhrat – PhD, Associate Professor of the Faculty of Social, Humanities and Exact Sciences of Mamun university (Khorezm, Uzbekistan, email: tajibaevsx@gmail.com).

Авторлар туралы мәлімет:

Турсунов Алишер Тулкунович – Ташкент фармацевтика институтының физика-математика және ақпараттық технологиялар кафедрасының доценті (Ташкент, Өзбекстан, электрондық пошта: tursunovr484za@gmail.com).

Нарзуллаев Давронбек Зикруллаевич – Ташкент фармацевтика институтының физика-математика және ақпараттық технологиялар кафедрасының доценті (Ташкент, Өзбекстан, электрондық пошта: davr1960@mail.ru).

Ильхамов Хисамиддин Шатурсунович - техникалық ғылымдар кандидаты, Ташкент фармацевтика институтының физика, математика және ақпараттық технологиялар кафедрасының доценті (Ташкент, Өзбекстан, электрондық пошта: khisamiddin@mai.ru).

Аноров Обиджон Умаралиевич – Ташкент мемлекеттік көлік университетінің жогары математика кафедрасының доценті міндетін атқарушы (Ташкент, Өзбекстан, электрондық пошта: anorovobidjon6@gmail.com).

Абдурахмонов Бахромжон Алишерович – физика-математика ғылымдар кандидаты, University of Management and Future Technologies «Communication and Digital Technologies» кафедрасының доценті (Ташкент, Өзбекстан, электрондық пошта: b.abduraxmonov75@gmail.com).

Тажибаев Шухрат – техникалық ғылымдар кандидаты, Мамун университетинин әлеуметтік, гуманитарлық және нақты ғылымдар факультеті доценті (Харезм, Өзбекстан, электрондық пошта: tajibaevsx@gmail.com).

Информация об авторах:

Турсунов Алишер Тулкунович – Доктор философии по техническим наукам, доцент кафедры физики, математики и информационных технологий ташкентского фармацевтического института (Ташкент, Узбекстан, электронная почта: tursunovr484za@gmail.com).

Нарзуллаев Давронбек Зикруллаевич – кандидат технических наук, доцент кафедры физики, математики и информационных технологий ташкентского фармацевтического института (Ташкент, Узбекистан, электронная почта: davr1960@mail.ru).

Илхамов Хисамиддин Шатурсунович - Кандидат технических наук, доцент кафедры физика, математика и информационной технологии Ташкентского фармацевтического института (Ташкент, Узбекистан, электронная почта: khisamiddin@mai.ru).

Аноров Обиджон Умаралиевич – и.о. доцента кафедры Высшей математики Ташкентского государственного университета транспорта (Ташкент, Узбекистан, электронная почта: anorovobidjon6@gmail.com).

Абдурахмонов Бахромжон Алишерович – кандидат физика-математических наук, доцент кафедры «Коммуникационные и цифровые технологии» Университета управления и технологий будущего (Ташкент, Узбекистан, электронная почта: b.abduraxmonov75@gmail.com).

Тажибаев Шухрат – кандидат технических наук, доцент факультета социальных, Гуманитарных и точных наук (Хорезм, Узбекистан, электронная почта: tajibaevsx@gmail.com).

Received: December 07, 2025
Accepted: June 10, 2026