Converting Feature Types in Analysis of Different Types of Data

D.Z. Narzullaev, K.K. Shadmanov, Sh.T. Ilyasov

***Abstract:** Today there are the large number of methods of data analysis for solving problems of pattern recognition of regression, correlation and factor analysis, which are not applicable in the case of different types of features in the source information. In this paper we propose an approach to solving this problem, named the conversion of feature types. The conversion of feature types is considered as an independent task that allows you to make the transition from non-quantitative features to quantitative ones and in further processing to apply the full range of classical methods of data analysis. The proposed algorithm is implemented in Delphi 10 Seattle the integrated software development sphere. The result of the study was tested when solving the task of recognition of several sets of known data.*

***Keywords:** data analysis, conversion of feature type, pattern recognition, imperative scale, experimental data table.*

I. INTRODUCTION

One of the current problems is the analysis of pattern recognition problems in the data analysis sphere and introduction of methods and algorithms to solve such problems-the results received in all production areas, including the pharmaceutical industry. Experts in this field are scientists who conduct research in applied mathematics, computer science, and information and communication technologies. A review of the results on pattern recognition and in-depth study were shown in the following papers [1-6]. Models were studied based on estimation, mathematical statistics, probability theory, potentials, and pattern recognition based on the results obtained. In works [7-13], new and modified methods of pattern recognition based on the evaluation of informative features are proposed. In the article [14] questions of classification of cardiovascular diseases of the heart and information content of the data are solved on the basis of algorithms for calculating estimates [2]. This work is a further development of the approach proposed in [15].

A necessary part of data analysis is the conversion of feature types, i.e. the transition from the various types of features available in the experimental data table to one type: quantitative, qualitative or classification type. Each scale provides calculation of only a certain set of statistical characteristics of the studied phenomenon. In this regard, many classical methods of data processing, such as factor, discriminant, regression analysis, pattern recognition, are not applicable to the entire volume of different types of information obtained in research activities in different areas and are used only for processing quantitative data.

Revised Manuscript Received on February 06, 2020.

***** Correspondence Author

D.Z. Narzullaev*, Tashkent pharmaceutical Institute, Tashkent, Uzbekistan.

K.K. Shadmanov, Tashkent pharmaceutical Institute, Tashkent, Uzbekistan.

Sh.T. Ilyasov, Tashkent pharmaceutical Institute, Tashkent, Uzbekistan.

The main purpose of this article is to develop methods and algorithms for digitizing of non-quantitative features in solving the problem of pattern recognition, as well as its implementation in solving the problem of recognizing several sets of known data.

II. PROPOSED METHODOLOGY

In the framework of this study, the following methods were used: 1) theoretical: analysis of the literature on the subject of studying approaches to the study of non-quantitative features in solving problems of recognizing the educated; 2) empirical: comparison of the results of developed algorithms. As a result, we developed a new algorithm for converting feature types.

III. RESULT ANALYSIS

Let the initial information be represented as the table of experimental data (TED) «object-feature (feature)» size N*p, where N – number of objects -lines, and p – number of features –columns in TED:

$$\mathbf{X}=(X_1, X_2, \ldots, X_N),$$

wherein $X_i=(x_i^{(1)}, x_i^{(2)}, \ldots, x_i^{(p)})$ – vector of analyzed features $x^{(1)}, x^{(2)}, \ldots, x^{(p)}$, registered at i researched object. Splitting the initial set of N objects into disjoint classes in the presence of a training sample (a set of objects that belong to the images specified) is the main goal of solving the problem of image recognition. In this case, it is necessary to use some decisive rule that would classify the newly emerging observations with a minimum number of errors. Different approaches to solving the problem of pattern recognition are calculated for the presence of one type of characteristics in the original TED – quantitative. Therefore, in the process of pre-processing data, it may be necessary to convert the types of features. This is due to the fact that very often the original TED contains features that can be measured in different scales. The most common in practice are the following types of scales: names, order, quantitative. When solving real problems, in many cases, the results of research are presented by a table of experimental data, which includes simultaneously the values of various types of features. Such initial data is called heterogeneous, and the space under consideration is called the space of heterogeneous features. The scale values of each feature type allow only certain operations and transformations. For example, for quantitative features, all mathematical operations and transformations are acceptable, for qualitative ones, relations of order and comparison, and for classification ones, only equivalence relations are acceptable. From this it follows that a certain scale allows the calculation of a certain set of statistical characteristics.

As a result, many classical mathematical and statistical methods of data processing, such as discriminant, factor, and regression analysis, are not applicable to the entire volume of information of different types and are used only for processing experimental data measured on a quantitative scale. Methods based on the use of ordinal statistics and various measures of communication of categorized variables are used to analyze the initial data measured in the name and order scales. The results of the analysis of different types of data are interpreted independently of each other and this prevents the understanding of the studied phenomenon as a whole. The regularities and properties of the studied objects of the original TED, which are reflected in the relationships between different types of indicators, are not revealed. Thus, the final results of processing different types of data do not reflect the diversity of internal relationships and distort the essence of the phenomenon under study.

In the work [16] we propose an approach called digitization, according to which qualitative and classification features are transformed into quantitative ones for the purpose of their further, most effective use along with the existing quantitative features in classical statistical models. In the works [17,18] digitization is already considered as an independent research task aimed at identifying and meaningful interpretation of hidden, unobservable relations between the objects of the studied set, which are not reflected in the measurement scales of the original feature system. These methods are considered as methods for converting feature types that allow you to move from less strong scales to stronger ones.

In the work [18], the procedure for pass from less “strong” scales to more “strong " ones in the framework of solving a certain problem of data analysis is based on the following main provisions:

- consistent strengthening of the scale;
- when enriching the scale, a priori assumptions of the researcher of a particular subject area about hidden relationships that are not reflected in the original system of features are taken into account.

In the work [18] the imperative scale is understood as an isomorphism of an empirical system of relations into a numerical system of relations. In this paper, we consider the problem of finding the optimal imperative scale of order, when each i gradation is awarded the meaning of an integer label d_i of $\mathbf{K_g}$ ={1,2,...,g}, called range of gradation.

In the work [18] we propose generalized formula for choose imperative ordinal scales предлагается обобщённая формула для выбора императивных порядковых as:

$$\frac{\overline{\mathbf{d}}^{\mathbf{T}}\mathbf{V}\overline{\mathbf{d}}}{\overline{\mathbf{d}}^{\mathbf{T}}\mathbf{Z}\overline{\mathbf{d}}} \to \min, \qquad (1)$$

where $d \in P^g$, P^g – set of moves g of real digits. Concrete expressions for matrix V and Z depend on the type of solving task.

Let’s go to solve the optimizing task (1).

Let in table of experimental data of size N*p (N – number of objects, p – number of features) be l number of features

$\mathbf{x}^1$, $\mathbf{x}^2$, ... , $\mathbf{x}^l$ and l-p classification features $\mathbf{x}^{l+1}$, $\mathbf{x}^{l+2}$, ... , $\mathbf{x}^p$.

Let’s consider the procedure of numeralization for feature x^{l+1}.

In case if the number g of gradation of classified feature x^{l+1} does not exceed six, i.e. $g \le 6$, optimal value category (1) is proposed to search by method of full search of all conversions.

Let us consider the case when g>6. Under this condition, the total number of possible permutations from g integer ranks is quite large (g!) and the application of the full search method to obtain an optimal imperative scale that satisfies the criterion (1) becomes inefficient due to the huge expenditure of machine time. Therefore, the optimal permutation of ranks in [18] is proposed to be found using the method of successive approximations. Thus initial transposition $\overline{\mathbf{d}}_0$ is chosen by at random and the final result depends just on $\overline{\mathbf{d}}_0$. Thus different local minimums are received for different initial values (1).

For adopt this shortage we propose to do as follows. Let’s define first of all, the way of calculation of initial approach $\overline{\mathbf{d}}_0$, let’s solve accessorial task.

Supposing minimum (1) it is necessary to find on set of actual numbers, i.e. expand many allowed solutions to $\mathbf{R}^{\mathbf{g}}$ and the task of search of optimal vector will have the view like:

$$\frac{\overline{\mathbf{a}}^{T}\mathbf{V}\overline{\mathbf{a}}}{\overline{\mathbf{a}}^{T}\mathbf{Z}\overline{\mathbf{a}}} \to \min, \qquad (2)$$

where $\overline{\mathbf{a}}^{\mathbf{T}} \in \mathbf{R}^{\mathbf{g}}$.

Ratio (2) called generalized ratio of Reley. You know [19], that for ordinary ratio of Relay of kind

$$\frac{(\mathbf{A}x, x)}{(x, x)} \qquad (3)$$

inequality

$$v_1 \le \frac{(\mathbf{A}x, x)}{(x, x)} \le v_n$$, are fair, also ratio

$$v_1 = \min_{x \ne 0} \frac{(\mathbf{A}x, x)}{(x, x)} \qquad (4)$$

where x – any nonzero vector,

$v_1, \ldots, v_n$ - eigenvalues of matrix A, a v_1 - minimum of these numbers.

Lets give provision of centering of vector $\overline{\mathbf{a}}$, i.e.

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

In this case the matrix **Z** for tasks of analysis of mutual dependence of features, factor and regressive analysis put on diagonal view with elements of n_i/N and therefore it will be non-degenerate. So ratio (2) we may bring to view (3), as

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

where $\overline{\mathbf{b}} = \mathbf{Z}^{\mathbf{T}^{-1/2}}\overline{\mathbf{a}}$.

Point B= $\mathbf{Z}^{\mathbf{T}^{-1/2}}\ \mathbf{V}^{\mathbf{T}}\ \mathbf{Z}^{\mathbf{T}^{-1/2}}$. So we have

Retrieval Number: D1441029420/2020©BEIESP
DOI: 10.35940/ijitee.D1441.029420

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

$$\frac{(\mathbf{V}^{\mathbf{T}}\bar{\mathbf{a}},\bar{\mathbf{a}})}{(\mathbf{Z}^{\mathbf{T}}\bar{\mathbf{a}},\bar{\mathbf{a}})}=\frac{(\mathbf{B}\bar{\mathbf{b}},\bar{\mathbf{b}})}{(\bar{\mathbf{b}},\bar{\mathbf{b}})}.$$

Using for this congruence correlation (4), we may find own vector $\bar{\mathbf{b}}_{\mathbf{0}}$, meeting minimal own number of matrix B. Further Далее, vector looked for is calculated as follows:

$$\bar{\mathbf{a}}=\mathbf{B}^{-1/2}\bar{\mathbf{b}}.$$

But the process of find of matrix $\mathbf{B}^{-1/2}$ is rather bulky, therefore let's enter one more additional limitation:

$$\bar{\mathbf{a}}^{\mathbf{T}}\mathbf{Z}\bar{\mathbf{a}}=\mathbf{1}. \quad (5)$$

So the task of minimizing (2) will reduced to search conditional minimum of quadratic form

$$\bar{\mathbf{a}}^{\mathbf{T}}\mathbf{V}\bar{\mathbf{a}} \quad \to\min,$$
(6)

Proving for (5).

For solve the task (6) let's use method of Lagrange multiplier, on which optimal vector $\bar{\mathbf{a}}$ transfers to zero the first derivate of Lagrange function

$$\mathbf{F}(\bar{\mathbf{a}})=\bar{\mathbf{a}}^{\mathbf{T}}\mathbf{V}\bar{\mathbf{a}}-\nu\,\bar{\mathbf{a}}^{\mathbf{T}}\mathbf{Z}\bar{\mathbf{a}},$$

where ν - input Lagrangian coefficients.

We have:

$$\frac{\partial\mathbf{F}(\bar{\mathbf{a}})}{\partial\bar{\mathbf{a}}}=2\,\mathbf{V}\bar{\mathbf{a}}-2\,\nu\,\mathbf{Z}\bar{\mathbf{a}}=0,$$

Or we receive the system of rectilinear homogeneous equation

$$\mathbf{V}\bar{\mathbf{a}}-\nu\,\mathbf{Z}\bar{\mathbf{a}}=0. \quad (7)$$

Multiply this formula from left to $\bar{\mathbf{a}}$ and, accounting (5), we receive

$$\bar{\mathbf{a}}^{\mathbf{T}}\mathbf{V}\bar{\mathbf{a}}=\nu, \quad (8)$$

i.e. value (2) is defined by entered Lagrange multiplier providing for (5). Consequently, by choosing as ν minimal own number and solving system of formula (7), we will have $\bar{\mathbf{a}}$, minimizing (2).

By multiply (7) from left to $\mathbf{Z}^{-1}$, we will receive

$$\mathbf{Z}^{-1}\,\mathbf{V}\bar{\mathbf{a}}=\nu\,\bar{\mathbf{a}},$$

i.e. in this case sought optimal solution (2) is defined by own vector of matrix $\mathbf{Z}^{-1}\,\mathbf{V}$, meeting minimum own digit ν.

Mark as $\bar{\mathbf{a}}_{\mathbf{0}}$ of found optimal vector, minimizing (2).

Regularize components of vector $\bar{\mathbf{a}}_{\mathbf{0}}$ on increase, i.e.

$$a_{i_1}\le a_{i_2}\le\ldots a_{i_g},$$

where i_j - index showing the place of component a_{i_j} of vector $\bar{\mathbf{a}}_{\mathbf{0}}$, takes the values of sets {1,2, . . . ,g}.

let's award to components of vector $\bar{\mathbf{d}}_{\mathbf{0}}$ integer-valued values on the following scheme:

$$d_{i_1}=a_{i_1}=1;\ d_{i_2}=a_{i_2}=2;\ \ldots\ d_{i_{1g}}=a_{i_g}=g;$$

As a result we receive some vector

$$\bar{\mathbf{d}}_{\mathbf{0}}=(d_1,\ldots,d_g),$$

which we will consider as the initial shift of ranks used when determining an imperative scale of an order by method of consecutive approximations which essence is stated below.

Let's mark numerator and a denominator of ratio (1) via M_1 and M_2 accordingly. At transposition *k*- and *l*-gradation increment of numerator and denominator is calculated on formulas [18]:

$$\Delta M_1=\bar{c}\,\bar{d},\quad \bar{c}=(c_1,\ldots,c_g),$$

where

$$c_i=\begin{cases}2(d_k-d_l)(h_{il}-h_{ik}), & i\ne k,l;\\ (d_k-d_l)(h_{ll}-h_{kk}), & i=k,l;\end{cases}$$

$$\Delta M_2=\bar{f}\,\bar{d},\quad \bar{f}=(f_1,\ldots,f_g),$$

where

$$f_i=\begin{cases}2(d_k-d_l)(t_{il}-t_{ik}), & i\ne k,l;\\ (d_k-d_l)(t_{ll}-t_{kk}), & i=k,l;\end{cases}$$

It is obvious that subject to

$$\frac{1+\Delta M_1/M_1}{1+\Delta M_2/M_2}<1 \quad (9)$$

transposition of ranks of *k*- and *l*- gradation will decrease value (1).

Now we shall describe computational algorithm of search of optimal transposition of integer-valued ranks of gradation, minimizing correlation (1) for case g>6:

Step 1. Definition of initial shift of ranks of gradation $\bar{\mathbf{d}}_{\mathbf{0}}$ on method described above.

Step 2. Choice of some couple of gradation, other than previous iteration *k and l.*

Step 3. Definition ΔM_1 and ΔM_2 at transposition of ranges d_k and d_l. Calculation of value of relation (9).

Step 4. At fulfill of inequation (9) we receive new transposition $\bar{\mathbf{d}}_1$. Otherwise – transition to step 5.

Step 5. Check of a condition of a stop. An exit is carried out in case the transposition of any pair of ranks of the received sorting does not lead to reduction (9). Otherwise – transition to a step 2.

Thus, the algorithm described above allows to find optimum transposition of ranks $\bar{\mathbf{d}}$, minimizing correlation (1) for feature x^{l+1}. In case of existence of several not quantitative signs, the same operations are made with signs $x^{l+2}, \ldots, x^{p}$.

The described method allows to transform classification signs x^{l+1}, x^{l+2}, ... , x^{p} into quantitative and in further processing to use all range of methods and analysis algorithms of data.

Further we will consider criteria for selection of imperative scales in the models of recognition using the linear dividing (discriminant) functions.

For the task of recognition of patterns of imperative scales we suppose be set such numerical tags for which the following relation of dissipation would be minimized:

$$J=\frac{\Lambda_W}{\Lambda_T} ,$$

where

$$\Lambda_T = \Lambda = \Lambda_{l+1,l+1} \bar{a}^T S \bar{a},$$

Matrix S calculated likewise [18]. For calculation of Λ_W we mark $\Lambda_W = (\lambda_{ij}^W)$. Thus

$$\lambda_{ij}^W = \frac{1}{N}\sum_{p=1}^{Q} N_p \lambda_{ij}^{(p)} ,$$

where $i,\ j=\overline{1,l}$; Q – quantity of teaching selection (TS); N_p - number of objects in *p*- teaching selection;

$$\lambda_{k,l+1}^W = \lambda_{l+1,k}^W = \frac{1}{N}\sum_{i=1}^{g} a_i \sum_{p=1}^{Q} N_p \mu_k^{(p)i} ,$$

where $\mu_k^{(p)i}$ parameter μ_k^i, defined on *p*-teaching selection [10].

Mark as $m_k^{(p)i}$ and $m_k^{(p)}$ private and general average value of *k*- sign on *p*- TS; $n_i^{(p)}$ and $t_{ij}^{(p)}$ - number of objects in *i*- gradation and value of parameter t_{ij}, defined on *p*- TS, where

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

and n_i - the number of objects with the ith gradation of the attribute x_{l+1}.

Let's enter the following parameters:

$$\mu_k^{*i} = \frac{1}{N}\sum_{p=1}^{Q} N_p \mu_k^{(p)i} = \frac{1}{N}\sum_{p=1}^{Q} N_p \frac{n_i^{(p)}}{N_p}(m_k^{(p)i} - m_k^{(p)}) = \frac{1}{N}\sum_{p=1}^{Q}(n_i^{(p)} m_k^{(p)i} - n_i^{(p)} m_k^{(p)}) =$$

$$= \frac{n_i}{N}(m_k^i - \frac{1}{n_i}\sum_{p=1}^{Q} n_i^{(p)} m_k^{(p)});$$

$$t_{ij}^{*} = \frac{1}{N}\sum_{p=1}^{Q} N_p t_{ij}^{(p)} ,$$

and we may show that

$$\sum_{i=1}^{g} t_{ij}^{*} = 0; \qquad (10)$$

$$\sum_{i=1}^{g} \mu_k^{*i} = 0. \qquad (11)$$

Then we will receive:

$$\lambda_{k,l+1}^W = \sum_{i=1}^{g} a_i\ \mu_k^{*i},\ k=\overline{1,l};$$

$$\lambda_{l+1,l+1}^W = \sum_{i=1}^{g}\sum_{j=1}^{g} a_i a_j\ t_{ij}^{*}.$$

Having transformed determinant Λ_W, as Λ, we receive:

$$\Lambda_W = \Lambda_{Wl+1,l+1} \bar{a}^T W \bar{a}, \qquad (12)$$

where

$$W=(\omega_{ij})=T^{*} - M^{*} \Lambda_{Wl+1,l+1}^{-1} M^{*T},$$

$$T^{*} = (t_{ij}^{*}),\ M^{*} = (\mu_k^{*i}),\ i,j=\overline{1,g};\ k=\overline{1,l}.$$

Of (10) and (11) follows congruence:

$$\sum_{i=1}^{g} \omega_{ki} = 0, \qquad (13)$$

then the relation of dissipation we may write down as:

$$J=\frac{\Lambda_{Wl+1,l+1}}{\Lambda_{l+1,l+1}} \frac{\bar{a}^T W \bar{a}}{\bar{a}^T S \bar{a}}.$$

As $\Lambda_W \geq 0$ and (13) fulfilled, quadratic form $\bar{a}^T W \bar{a}$ is not negative. The first factor in the last equality does not depend on set $\bar{a}$ and the criterion of selection of imperative scales for a task of recognition of images can be written down as:

$$\frac{\bar{a}^T W \bar{a}}{\bar{a}^T S \bar{a}} \rightarrow \min, \qquad (14)$$

where $\bar{a} \in B^g$.

Provided that it is necessary to find optimum integer numerical tags, the task (14) will take a form:

$$\frac{\bar{D}^T W \bar{D}}{\bar{D}^T S \bar{D}} \rightarrow \min, \qquad 15)$$

where $D \in P^g$.

For optimization of category (15) approach stated in i. II. is used.

Thus, the possibility of use of not quantitative signs is considered $x^i,\ i=\overline{l+1,p}$ for increase in accuracy of linear model of recognition. The linear dividing functions can be constructed only for quantitative signs and use of signs $x^i,\ i=\overline{l+1,p}$ it is admissible only after their transformation to quantitative with use of imperative scales $\xi_i,\ i=\overline{l+1,p}$. For the purpose of reduction of level of an error of recognition at search of imperative scales, function of criterion of dispersion is optimized

$$J=\frac{\Lambda_W}{\Lambda_T},$$

calculated taking into account the digitized signs.

Further examines the effectiveness of the algorithms for converting feature types proposed in this paragraph of this article.

The efficiency of the algorithms for converting feature types was evaluated by the example of solving the tasks of pattern recognition. As initial information, well-studied data on three sorts of irises were taken [20,21]. Each variety of irises was represented by 50 flowers, which were initially measured by four quantitative characteristics that characterize the length and width of the petal, as well as the length and width of the sepals. In order to check the effectiveness of the algorithms for converting feature types, the fourth feature was converted to a classification with the number of gradations equal to 10. In this case, each variety was represented by at least two gradations, and the corresponding intervals of these gradations were assumed to be equal in magnitude. The assignment of numerical values to the classification feature was made for the purpose of mixing objects belonging to each of the three sorts.

In order to solve the task of pattern recognition for each of the three sorts of irises, a training sample was randomly compiled, including 25 objects. The remaining 25 objects of each class were a control sample.

The task of pattern recognition was solved for three data arrays:

A. the Initial data are taken directly from [20] and are represented by four quantitative features;

B. the Initial data are presented by three quantitative and one classification criteria obtained by the method described above. Quantitative features are also taken from [20];

C. the algorithm for finding integer marks for the pattern recognition task described in clause III was applied to the array B.

The task of recognition of unclassified irises with three specified variants of initial data was solved using the RECHKA algorithm. This algorithm is part of the SITO-PC data analysis system developed by the author's team. Recognition of unclassified irises occurs using three crucial rules: "far neighbor" (a), "medium connection" (b) and “near neighbor” (c) [21]. The research was conducted on a personal computer. The results of computer processing are shown in table 1. Here the number of incorrectly recognized objects is specified at the intersection of the array and the decisive rule.

Table 1. Data Processing Results.

decisive rule / data array	a	b	c
A	2	6	8
B	14	19	21
C	3	8	10

Table 1 shows that the smallest number of errors are made when the “far neighbor” and “medium connection" algorithms work. This circumstance is explained by the fact that classification algorithms using the decisive rules of "far neighbor" and "medium connection" recognize mainly objects whose images are represented by groups of ellipsoid or spherical shape. In the case of use of array B, the number of incorrectly recognized objects increases significantly, which is explained by the strong distortion of the original data structure in the digitization method described above, since objects belonging to the same training samples are strongly mixed with objects from other training samples.

Analysis of table 1 shows that when considering the array C, the number of incorrectly recognized objects when using all three decision rules is significantly reduced compared to option B. And even in this case, the use of the “far neighbor” and “middle link " decision rules provides better recognition quality. This circumstance is explained by the fact that the criterion for selecting imperative scales for the recognition problem provides for the Union of objects of the training sample into compact groups of spherical or ellipsoid forms.

Thus, the use of the developed algorithms for converting feature types can significantly improve the adequacy of the data analysis model compared to the original data.

IV. CONCLUSION

The following main results are obtained:

1. Methods and algorithms of transformation of types of signs for a task of recognition of patterns are developed;
2. An algorithm for searching for imperative scales of the order of gradations of non-qualitative features for the task of pattern recognition is proposed;
3. The developed algorithms for searching for imperative scales of the order of gradations of non-qualitative features for the task of pattern recognition are implemented in the integrated Delphi 10 Seattle software development environment and are included in the data analysis system SITO-PC;
4. An experimental study of the developed algorithms is carried out on the example of solving the task of pattern recognition of several known data sets.

REFERENCES

1. ZhuravlevYu.I. Selected scientific works. – M:‘Magistr’ Publishing house, 1998. - 420 p.
2. Zhuravlev, Y., Kamilov, M.M., Tulyaganov, S.H.E. Algorithms for Calculating Estimates and Their Application. - Tashkent: Fan, 1974. – 119 p.
3. Flah P. Machine learning. Science and art of creation of algorithms which take knowledge from data. Translation from English by A. A. Slinkin. - M.: ДМК Press, 2015.-p.400: il. ISBN 978-5-97060-273-7, 600 dpi, OCR
4. Fazylov Sh.Kh., Nishanov A.Kh., Mamatov N.S. Methods and algorithms for selection of informative features based on heuristic criteria of informativeness. Tashkent: Fan va Tekhnologiya. 2017 - 132 p.
5. Braga-Neto UM, Dougherty ER. Error Estimation for Pattern Recognition. New York: Springer; 2016.
6. Ramakrishnan S. Pattern Recognition: Analysis and Applications. New York: lTexLi; 2016.
7. Nishanov, A.X., Beglerbekov, R.J., Axmedov, O.K. Hybrid Recognition Algorithm In The Space Of Informative Attributes, Vestnik TUIT, 2017, No.4. - pp. 62-69.
8. Kamilov, M.M., Nishanov, A.X., Beglerbekov, R.J. Modified stages of algorithms for computing estimates in the space of informative features. International Journal of Innovative Technology and Exploring Engineering,April 2019, Volume 8, Issue 6. – pp. 714-717.
9. Fazilov ShKh, Mirzaev ON, Radjabov SS. State of the art of the problems of pattern recognition. Problems of computational and applied mathematics 2015;**2**:99-112.

10. Sh.Kh.Fazilov, N.M.Mirzaev, G.R.Mirzaeva. Modified Recognition Algorithms Based on the Construction of Models of Elementary Transformations. Procedia Computer Science, Volume 150, 2019, Pages 671-678.
11. Fazilov, S.K., Mirzaev, N.M., Nurmukhamedov, T.R., Ibragimova, K.A., Ibragimova, S.N. Model of recognition operators based on the formation of representative objects. International Journal of Innovative Technology and Exploring Engineering, Volume 9, Issue 1, November 2019, Pages 4503-4508.
12. Fazilov Shavkat, Mamatov Narzillo, Niyozmatova Nilufar. Developing methods and algorithms for forming of informative features' space on the base K-types uniform criteria. International Journal of Recent Technology and Engineering, Volume 8, Issue 2 Special Issue 11, September 2019, Pages 3784-3786.
13. Fazilov Shavkat, Mamatov Narzillo, Samijonov Abdurashid. Selection of significant features of objects in the classification data processing. International Journal of Recent Technology and Engineering, Volume 8, Issue 2 Special Issue 11, September 2019, Pages 3790-3794.
14. Nishanov, A.K., Djurayev, G.P., Khasanova, M.A. Improved algorithms for calculating evaluations in processing medical data. Compusoft, 2019, Volume 8, Issue 6. – pp. 3158-3165.
15. Narzullaev D.Z. Conversion of feature types in the analysis of heterogeneous data in scientometrics. International Scientific and technical conference "Role of scientific - technical information in the development of innovation activity", Tashkent, 2012. - pp. 242-245.
16. Enyukov I.S. Methods of digitizations of not quantitative variables//Algorithms and software of applied statistical analysis. M.: Science. 1980. - pp. 309-315.
17. Lbov G.S. Methods of processings of polytypic experimental data. - Novosibirsk: Science, 1981. – p.160
18. Nikiforov A.M., Fazilovв Sh.H. Methods and algorithms of transformation of types of signs in tasks of the analysis of data. - Tashkent, Fan, 1988.- 131 p.
19. Rais J. Matrix calculations and software: translation from English. M.: Mir, 1984.- 320 p.
20. Kendall M., Stuart A. Multidimensional statistical analysis and temporary ranks.- M.: Science, 1976. – 736 p.
21. Aleksandrov V.V., Gorsky N.D. Algorithms and programs of a structural method of data processing. - L.: Science, 1983. - 208 p.

.

AUTHORS PROFILE

Narzullaev Davronbek is currently pursuing her D.Sc in the field of visual data mining, development of software on computer processing of visual data. He graduated from the National University of Uzbekistan, Faculty of Applied Mathematics, Department of Applied Mathematics. He holds a PhD in Technical Sciences. He currently works at Tashkent Pharmaceutical Institute, Republik of Uzbekistan. His professional interests include: developing intelligent computer processing systems in pharmaceuticals for application in the process of creating and manufacturing new medicines based on local raw materials and introducing advanced computer technologies into the process of training qualified pharmacists. He is the responsible executor of the national scientific grant and participant in international scientific conferences.

Ilyasov Shakir is currently pursuing his Ph.D. in the field of data processing in the organization of pedagogical activities. He graduated from Tashkent Polytechnic Institute, Engineering and Physics Faculty of Computer Engineering Department. He is qualified as an electrical engineer in computer engineering. He currently works at Tashkent Pharmaceutical Institute, Uzbekistan. His professional interests include: statistical methods in medicine and pharmaceuticals, introduction of computer technologies in pedagogical process and research of synergistics of process of training of specialists-engineers of industrial pharmaceuticals profile.It has publications in the Republic of Uzbekistan and Russia. He is a participant in scientific conferences, including international conferences.

Shadmanov Kamoliddin is currently is currently pursuing her D.Sc in data processing in research on the creation of new drugs. He graduated from the National University of Uzbekistan, Physical Faculty of Physical Electronics Department. He holds a PhD in Chemical Sciences and an associate professor.He currently works at Tashkent Pharmaceutical Institute, Uzbekistan. His professional interests include: the introduction of computer technologies in pharmaceuticals to apply CAD-CAM systems to the process of creating and producing new drugs based on local raw materials and the introduction of advanced pedagogical technologies in the process of training qualified pharmaceutical specialists. He has publications of his scientific papers in various well-known SCOPUS journals and is a participant in international scientific conferences.