An Algorithm for Classification, Localization and Selection of Informative Features in the Space of Politypic Data

Akhram Khasanovich Nishanov
DSc, Professor, Department of Software of Information Technologies, Tashkent University of Information Technologies named after Muhammad Al-Khwarizimi, Tashkent, Uzbekistan.

Bakhtiyorjon Bakirovich Akbaraliev
PhD, Docent, Dean of Software Engineering Faculty, Tashkent University of Information Technologies named after Muhammad Al-Khwarizimi, Tashkent, Uzbekistan.

Samandarov Batirbek Satimovich
PhD, Department of Applied Mathematics, Karakalpak State University named after Berdakh, Nukus, Uzbekistan.

Akhmedov Oybek Kamarbekovich
Assistant, Department of Systematic and Practical Programming, Tashkent University of Information Technology named after Muhammad Al-Khwarizimi, Tashkent, Uzbekistan.

Tajibaev Shukhrat Khudaybergenovich
Assistant, Department of Applied Mathematics, Karakalpak State University named after Berdakh, Nukus, Uzbekistan.

Received April 04, 2020; Accepted May 22, 2020
ISSN: 1735-188X
DOI: 10.14704/WEB/V17I1/WEB17009

Abstract

Dimensionality reduction and feature subset selection are very important and challenging issues in preliminary processing of the large amount of data for its intellectual analysis, pattern recognition and clustering. In particular, the relevance of these issues will only grow if the preliminary data is derived from real-life and defined by qualitative indictors. Similarly, when problem is related to the selection of complex of important features for classification and localization of agricultural crops, the results are immediately reflected in practice. Thus, the problem would appear unclear should the initial data be polytypic i.e. defined by quality indicators and quantitative features. To address the problem, algorithms and programs have been developed for determining the degree of similarity of the objects based on analysis of the existing literature, and then textual, nominal and quantitative definition of the features in the form of qualitative indicators, and mathematical interpretation of the problem.

Keywords

Classification and Localization of Objects, Qualitative and Quantitative Features, Evaluation of Objects and Features, Feature Selection.

Introduction

The development of technical equipment and tools of communication, algorithmic and software program development for processing large amount of data have become a pressing challenge. The only reasonable way to solve the problem is to have a scientifically sound approach. The human activity has become automated as a result of introduction of the scientific approach to the solution of vital problems. These approaches have become a basis for introducing information and communication technologies to all sectors of the society.

The application of information and communication technologies to solve industry problems is related to the initial projection, in-depth analysis and study of these problems and processes.

In the present article, the initial data presents real-life varieties of crops, and this study is aimed at addressing important issues such as textual, nominal and quantitative definition of the objects, classification and localization as well as selection of a set of features for its adequate definition.

Scientific approaches and researches of many scientists have been studied and analyzed in addressing these problems. Herewith we bring to your attention some of the following methods, algorithm as well as scientific approaches for software programs that are used in classification, clustering, and selection of a set of features.

F-evaluation method based on Fisher’s function is a simple method of selecting features [4]. Usually, two classes work perfectly for study option, so the result is based on the order of the features selected. Methods developed by these authors [4] are based on Fischer's discriminant analysis, whereby a set of features [2,3,4] is selected based on spacing the distances between classes and compacting internal distances between the class objects. The achieved results can be applied for several classes too.

Selection of a set of features is but one of the main issue in pattern recognition, in addressing the problems of classification and clustering. A perfect way to solve the problem and to build an ensemble of features has been studied by the authors in their researches [5,6,7,8,9,10]. It presents the initial data in this way. In the proposed methods, efforts have been made to solve key problems by maximizing the percentage of object classification and minimizing the set of features.

This approach is best suited for pre-defined baseline data and minor study samples. Similarly, features taken from real-life situations may not always accurately represent objects, for such cases the issue of classification and ensuring object reliability is essential. Usually, this type of information is considered vague and vague cases of study objects are researched and the classification problem is resolved on the bases of these data in the scientific research [8].

For purposes of classification, clustering, pattern recognition and the initial data analysis it is quite important to diminish the set of patterns in cases where data features are given in textual, nominal and quantitative forms and to solve proposed it is quite important to diminish the set of patterns and to solve problems on their basis. The researchers [9, 10] have proposed an algorithm to reduce the set of considered features for cases where the data features consist of multiple textual, nominal and quantitative forms.

We believe that the suggested research in this direction [11-15], the proposed method, algorithm and program researches to address these issues will not leave the reader indifferent, and we are confident that these studies will be applied in the increase of the well-being and ease of our lives. The article describes how to solve a problem, which includes the following items:

Materials and Methods

Textual, Nominal and Quantitative Expression of Features in Preliminary Data Processing

The classification of cotton-plant varieties included in the State Register of agricultural crops recommended for planting in the territory of the Republic of Uzbekistan, identification of their closely related varieties, and their allocation to the areas suitable for planting have been developed [1]. Initially, the varieties of cotton are sorted into tables. Information about the varieties of cotton, their species, height, shape, stem, stem, leaf, flower, chest, weight, is presented in textual, nominal and quantitative forms. The following table #1 is developed based on these data.

Table 1. The initial data about cotton-plant varieties

№	Cotton-plant	Type G. hirsutum	Bush height (sm)	Shape	Cotton trunk (hairy/color/bending)	Branch (type)	Leaf (size/lobes/color)	Flower (size/color)	Capsule	Weight of 1000 seeds (gr)
1	Bukhara-8	+	90-120	columnar	less/ dark green/no	2	medium/ 3-5/ green	large/light yellow at the top, orange in the main	large roundish, underdeveloped, nosed, dark green, it blooms well, fruits do not fall	138
2	Chimboy-5018	+	110	cone	less/ dark green/no	1,5	medium/ 5/ dark green	medium/ yellowish	avarage, round, small nose, dark green, it blooms well, fruits do not fall	131
3	Andijan-36	+	110-120	cylindrical	medium/ green/ no	1	medium/ 3-5/ green	medium/ yellowish	avarage round, oval-shaped, green, it blooms well and fruits do not fall	108
4	Namangan -34	+	110-120	cone	weak/ green/ no	1-1,5	Small/ 3-5/ green	medium/ yellow	average elongate, green, it blooms well and fruits do not fall	117
5	C-6541	+	100-120	cone	medium/ green/ no	1-1,5	medium/ 3-5/ green	medium/ yellow	average round, elongate, green, it blooms well and fruits do not fall	116
6	Khorezm 150	+	90-110	pyramid	medium/ dark blue/ no	1,5-2	Large/ 3-5/ dark blue	medium/ yellowish	large, round, smooth, blue-colored, it blooms well and fruits do not fall	128
7	Dustlik-2	+	90-100	cone	medium/ green/no	1-1,5	medium/ 5/ green	medium/ yellow	average, round oval-shaped, small nosed, green, it blooms well and fruits do not fall	126
8	Sultan	+	130-140	cone	medium/ green/no	1,5	medium/ 3-5/ green	medium/ yellowish	average, elliptic, green, it blooms well and fruits do not fall	125,6
9	Besh Qakhramon	+	110-120	cone	medium/ green/no	1,5	medium/ 3-5/ green	medium/ yellowish	average sized, egg-shaped, green, it blooms well and fruits do not fall	118,2

10	Ibrat	+	110-120	cone	weak/green/ no	1,5-2	medium/ 3-5/ green	medium/ yellowish	average, egg-shaped, green, it blooms well and fruits do not fall	110
11	Kopaysin	+	100-120	cone	medium/ green/no	1	medium/ 3-5/ green	large/ yellow	average sized, egg-shaped, green, it blooms well and fruits do not fall	103
12	Gulbakhor-2	+	110-120	cone	weak/green/ no	1-1,5	medium/ 3-5/ green	large/ yellowish	large, elliptic, green, it blooms well and fruits do not fall	113,7
13	Andijan-37	+	100-110	pyramid	weak/ green/no	2	medium/ 3-5/ green	medium/ yellow	average, rounded oval-shaped, green, it blooms well and fruits do not fall	105
14	C-6775	+	100-120	cone	medium/ green/no	1	medium/ 3-5/ green	medium/ yellowish	average, egg-shaped, green, it blooms well and fruits do not fall	120,7
15	Navbakhor-2	+	110-120	cylindrical	weak/ green/no	0,5-1	medium/ 3-5/ green	medium/ yellowish	large, round, green, it blooms well and fruits do not fall	126
16	C-8284	+	95-105	cone	weak/green/ no	1,5-2	medium/ 3-5/ green	medium/ yellowish	large, elliptic, green, it blooms well and fruits do not fall	136
17	Porloq-1	+	110-120	cone	medium/ green/no	1-2	medium/ 3-5/ green	medium/ yellowish	large, oval-shaped, green, it blooms well and fruits do not fall	135
18	Porloq-2	+	100-110	cylindrical	less/green/ no	2-3	medium/ 3-5/ green	medium/ yellowish	large, elliptic, green, it blooms well and fruits do not fall	145
19	Porloq-4	+	120-130	cone	weak/green/ no	1-2	medium/ 3-5/ green	medium/ yellowish	large, oval-shaped, green, it blooms well and fruits do not fall	115
20	C-6550	+	120-130	cone	weak/ green/ no	1-2	medium/ 3-5/ green	medium/ yellow	average, round, elongate, green, it blooms well and fruits do not fall	110-115

21	Namangan-102	+	120-130	cone	weak/ green/ no	1-2	Medium/ 3-5/ green	medium/ yellow	average, round, elongate, green, it blooms well and fruits do not fall	110-115
22	Pakhtakor-1	+	120-125	cone	weak/ green/ no	1-1,5	medium/ 3-5/ green	medium/ yellow	average, round, elongate, green, it blooms well and fruits do not fall	110
23	Istiqlol -14	+	120-130	cone	weak/ green/ no	1-2	medium/ 3-5/ green	medium/ yellow	average, round, elongate, green, it blooms well and fruits do not fall	111
24	Kelazhak	+	120-130	cone	weak/ green/ no	1-2	medium/ 3-5/ green	medium/ yellow	Average, round, elongate, green, it blooms well and fruits do not fall	122

Based on the available initial data, the cotton varieties have been classified and localized using the practical experience. As a result of localization, it was decided where and which cotton variety should be planted. The only problem is to justify this in theory. The construction of a mathematical process to solve this problem caused the following problems.

Definition of the Problem and Formation of the Space of Textual, Nominal and Quantitative Features

In the light of the above, the following problems need to be solved:

1. The problem of representing cotton characteristics in numbers, based on Table № 1;
2. The problem of determining the degree of similarity of objects;
3. Classification and localization of cotton varieties in agriculture;
4. Development of an algorithm for selection of the set of important features based on classification and localization.

Let's take the first problem. Defining cotton varieties with textual, nominal and quantitative features, i.e. the issue of numbering. To address all of the above mentioned the following designations are suggested:

x_i – variety of cotton-plant, $i=\overline{1,24}$. This means that there are 24 varieties of cotton and each variety is represented as follows: x_1 – Bukhara-8; x_2 – Chimboy-5018; x_3 – Andijan-36; x_4 – Namangan-34; x_5 – C-6541; x_6 – Khorezm-150; x_7 – Dustlik-2; x_8 – Sulton; x_9 – Besh Qakhramon; x_{10} – Ibrat; x_{11} – Kopaysin; x_{12} – Gulbahor-2; x_{13} – Andijan-37;

x_{14} – C-6775; x_{15} – Navbakhor-2; x_{16} – C-8284; x_{17} – Porloq-1; x_{18} – Porluk-2; x_{19} – Porluk-4; x_{20} – C-6550; x_{21} – Namangan-102; x_{22} – Pakhtakor-1; x_{23} – Istiqlol-14; x_{24} – Kelajak.

Each variety is considered a study object and expressed as follows: $x_i = \left(x_i^1, x_i^2, ..., x_i^N\right),\ i = \overline{1,24}$.

In this, x_i^1 – the length of the cotton bush, when expressed on a scale, receives five different measures: if $x_i^1 = 1$, then the maximum height of cotton-plant is 105 sm; if $x_i^1 = 2$, then the maximum height of cotton reaches 110 sm; if $x_i^1 = 3$, then the maximum height of cotton increased to 120 sm; if $x_i^1 = 4$, then the maximum height of the cotton is 125 sm, and if $x_i^1 = 5$, then the maximum cotton height is 130 sm. Thus, x_i^1 accepts five different values.

x_i^2 – shape of cotton-plant, represented in four different definitions: columnar, cone, cylindrical and pyramidal. Therefore, when expressed on scale, it accepts the following values: $x_i^2 = 1$; $x_i^2 = 2$; $x_i^2 = 3$, $x_i^2 = 4$.

x_i^3 – the hairiness of cotton plant, which is expressed in three types: less hairy, medium hairy and weak hairy. That is, in scale they are represented as $x_i^3 = 1$; $x_i^3 = 2$; $x_i^3 = 3$.

x_i^4 – color of the cotton-plant stalks, and are represented by three types: dark green, green and dark blue. That is, they represent the following values by scale: $x_i^4 = 1$; $x_i^4 = 2$; $x_i^4 = 3$.

x_i^5 – branch of cotton–plant and can be expressed in eight different types: second, one-half, one, one-and-one-a-half, one-and-a-half-to-two type, one-a-half-to-two, one-to-two, and two-three.

That is, they represent the following values by scale: $x_i^5 = 1$; $x_i^5 = 2$; $x_i^5 = 3$; $x_i^5 = 4$; $x_i^5 = 5$; $x_i^5 = 6$; $x_i^5 = 7$; $x_i^5 = 8$.

x_i^6 – the shape of a cotton-plant leaf, which is expressed in three forms: medium, small and large, that is, when expressed on a scale, represents the following values: $x_i^6 = 1$; $x_i^6 = 2$; $x_i^6 = 3$.

x_i^7 – the color of the cotton leaf and is expressed in two different ways: green, dark green, when expressed on a scale, represents the following values: $x_i^7 = 1$; $x_i^7 = 2$;

x_i^8 – the form of the cotton-plant flower and is expressed in three different ways: large, medium and big, when expressed on a scale, represents the following values: $x_i^8 = 1$; $x_i^8 = 2$; $x_i^8 = 3$.

x_i^9 – the size of cotton capsule and is expressed in three different ways: big, medium and large, when expressed on a scale, represents the following values: $x_i^9 = 1$; $x_i^9 = 2$; $x_i^9 = 3$;

x_i^{10} – the shape of the cotton capsule and is expressed in ten different ways: underdeveloped nosed, round small nosed, round oval-shape, elongate, rounded elongate, rounded smooth, elliptic, egg-shaped, round, oval-shaped, when expressed on scale, represents the following values: $x_i^{10} = 1$, $x_i^{10} = 2$, $x_i^{10} = 3$, $x_i^{10} = 4$, $x_i^{10} = 5$, $x_i^{10} = 6$, $x_i^{10} = 7$, $x_i^{10} = 8$, $x_i^{10} = 9$, $x_i^{10} = 10$;

x_i^{11} – the color of the cotton capsule and is expressed in three different ways: dark green, green and blue, and, when expressed on a scale, represents the following value: $x_i^{11} = 1$, $x_i^{11} = 2$, $x_i^{11} = 3$;

x_i^{12} – the weight of 1000 cotton seeds, which is expressed in four different ways. If the weight of seeds is between 103-113 gr, then $x_i^{12} = 1$, If the weight of seeds is between 115-122 gr, then $x_i^{12} = 2$, If the weight of seeds is between 123-131 gr, then $x_i^{12} = 3$ and If the weight of seeds is between 132-145 gr, then $x_i^{12} = 4$, in other words, when expressed on a scale the weight of 1000 represents the following value: $x_i^{12} = 1$, $x_i^{12} = 2$, $x_i^{12} = 4$.

Based on the data given in **Table 1** above and the featured data, the following **Table 2** can be formed. The table rows represent objects (varieties), the column represent the features. For example,

$$x_1 = \left(x_1^1, x_1^2, ..., x_1^{12}\right) = \left(3,1,1,1,1,1,1,1,1,1,1,4\right)$$

Hence, in the **first issue of numbering** mentioned above, cotton symbols are expressed in numbers using the scale of cotton symbols in **Table №1**, and the first problem is resolved in **Table 2** below.

Table 2. Textual, nominal and quantitative features of cotton-plant varieties

Variety	Features											
	x_i^1	x_i^2	x_i^3	x_i^4	x_i^5	x_i^6	x_i^7	x_i^8	x_i^9	x_i^{10}	x_i^{11}	x_i^{12}
x_1	3	1	1	1	1	1	1	1	1	1	1	4
x_2	2	2	1	1	2	1	2	2	2	2	1	3
x_3	3	3	2	2	3	1	1	2	2	3	2	1
x_4	3	2	3	2	4	2	1	2	2	4	2	2
x_5	3	2	2	2	4	1	1	2	2	5	2	2
x_6	2	4	2	3	5	3	2	2	3	6	3	3
x_7	1	2	2	2	4	1	1	2	2	3	2	3
x_8	6	2	2	2	4	1	1	2	2	3	2	3
x_9	3	2	2	2	2	1	1	2	2	8	2	2
x_{10}	3	2	2	3	2	3	1	1	2	2	8	1
x_{11}	3	2	2	2	3	1	1	3	2	8	2	1
x_{12}	3	2	3	2	4	1	1	3	1	7	2	1
x_{13}	2	4	3	2	1	1	1	2	2	3	2	1
x_{14}	3	2	2	2	3	1	1	2	2	8	2	2
x_{15}	3	3	3	2	6	1	1	2	1	9	2	2
x_{16}	1	2	3	2	5	1	1	2	1	7	2	3
x_{17}	3	2	2	2	7	1	1	2	1	10	2	4
x_{18}	2	3	1	2	8	1	1	2	1	7	2	4
x_{19}	5	2	2	2	7	1	1	2	1	11	2	4
x_{20}	5	2	2	2	7	1	2	2	2	5	2	2
x_{21}	5	2	3	2	7	1	1	2	2	5	2	1
x_{22}	4	2	3	2	4	1	1	2	2	5	2	1
x_{23}	5	2	3	2	7	1	1	2	2	5	2	1
x_{24}	5	2	3	2	7	1	1	2	2	5	2	2

The table contains textual, nominal and quantitative features based on numbers. Identical textual, nominal and quantitative features in the table column have identical meaning. However, identical textual, nominal and quantitative features in the table rows are different – they have different meaning. Their appearance is the same, but their meaning does not match. The following is the rules for determining the degree of similarity of research objects using this table.

Determination of the Degree of Similarity of Objects

This section addresses the issue of determining the degree of similarity of cotton varieties defined by textual, nominal and quantitative features.

Second problem. The problem of determining the degree of similarity of objects in the context of textual, nominal and quantitative features.

Statement of the problem. As described in the literature [11-15], let us assume that the set of study options is defined as follows $x_{p1}, x_{p2}, ..., x_{pm_p} \in X_p, \ p = \overline{1, r}$. In this x_{pi} is a $N-$dimensional vector of the space of textual, nominal and quantitative features, and each object $x_{pi} = \left(x_{pi}^1, x_{pi}^2, ..., x_{pi}^N\right), \ i = \overline{1, m_p}$, if considered in the $N-$dimensional space, mean a set of $X_p, \ \ p = \overline{1, r}$ classes and consists of $x_{p1}, x_{p2}, ..., x_{pm_p}$ objects in the amount of m_p.

The proximity function is introduced to determine the degree of similarity of objects. Each object belonging to the class has a one-to-one comparison with objects in its class, and the inter-object proximity function is described below.

The Function of Proximity in the Space of Textual, Nominal and Quantitative Features

Let's assume, the space of textual, nominal and quantitative features of the class X_p is given two x_{p1}, x_{p2} objects. The function of proximity between the objects $\rho_j\left(x_{p1}, x_{p2}\right)$ is entered into the space of textual, nominal and quantitative features as follows:

$$\rho_j\left(x_{p1}, x_{p2}\right) = \begin{cases} 1 \ \ if \ \ \left(x_{p1}^j - x_{p2}^j\right) = 0, \ j = \overline{1, N} \\ 0 \ \ \ \ \ otherwise \end{cases} \tag{1}$$

If the first condition indicates the degree of similarity between the two objects, then the second condition indicates that there is a big difference between them, meaning that the components are not identical.

The following formula is used to calculate the magnitude demonstrating the similarity degree of $k-$object with $t-$object in an arbitrary $p-$class in the space of textual, nominal and quantitative features:

$$\Gamma_t\left(x_{pk}, x_{pt}\right) = \sum_{j=1}^{N} \rho_j\left(x_{pk}, x_{pt}\right), \ \ k = \overline{1, \mathrm{m}_p}; t = \overline{1, \mathrm{m}_p} \ \ k \neq t \tag{2}$$

The following formula is used to calculate the magnitude demonstrating the similarity degree of $k-$object with all other objects in any $p-$class in the space of textual, nominal and quantitative features:

$$\Gamma_k\left(x_{pk},x_{pt}\right)=\sum_{t=1}^{m_p}\Gamma_t\left(x_{pk},x_{pt}\right),\ k,t=\overline{1,m_p};\ k\neq t \quad (3)$$

A Decisive Rule in the Space of Informative Features [2-5]

Let us assume that a new unknown diagnostic object $w=\left(w^1,w^2,...,w^N\right)$ is given. The similarity of this object to any of the given diagnostic objects is determined based on the following formula:

$$\Gamma_w=\left(w,x_{pk}\right)=\sum_{k=1}^{m_p}\sum_{j=1}^{N}\rho_j\left(w,x_{pk}\right),\ k=\overline{1,m_p} \quad (4)$$

If $\Gamma_w\left(w,x_{pi}\right)>\Gamma\left(w,\mathrm{x}_{pk}\right)$ inequality (4) is observed then, the degree of affiliation of the object $w=\left(w^1,w^2,...,w^N\right)$ to the $i-$diagnosis object is considered to be higher compared to others.

Algorithm for Solving the Given Problems

This paragraph describes an algorithm for finding solutions to the issues outlined in the article. The algorithm consists of six items, and it is expedient to apply the pattern identification issues only to specific class objects.

Step 1. Selective study objects are entered into database. The Primary Database is formed by all $X_p,\quad p=\overline{1,r}$ class objects;

Step 2. The proximity function of the space of informative features used in determining the value of contribution of X_p class objects to the formation of their own class is computed the convergence function based on the formula [1].

Step 3. In the space of textual, nominal and quantitative features, the magnitude indicating the degree of similarity of any x_{pk} object through other x_{pt} object by textual, nominal and quantitative features is calculated based on the formula (2);

Step 4. In the space of textual, nominal and quantitative features, evaluation of the magnitude indicating the degrees of similarity of any x_{pk} object with all other $x_{pt},\ t=\overline{1,m_p};\ k\neq t$ objects is calculated based on the formula (2);

Step 5. The decision rule for determining a new object, i.e. to which objects in which classes of cotton-plant variety does it belong to, is constructed on the basis of the paragraph five, based on the formula (4).

Based on the proposed theoretical studies and algorithm we solve the above problems. The similarity degrees of cotton varieties according to the function of similarity are shown in the fourth table below. Based on the formula above (3), **Table № 3** has been created. The intersection of rows and columns in the table indicates the value of these objects in respect to each other. For example, at the intersection of row 15 and 17 column is the number 7 (in red). Thus, the 15 variety of cotton-plant and variety of cotton-plant in the 17^{th} column indicate that the degree of their similarity is 7 out of 12 possibilities. As can be seen from the table, if an object evaluates itself, it gets 12 out of 12 possibilities. Certainly, self-evaluation is not relevant.

Results

Table 3. Degrees of similarity of cotton-plant varieties by the function of proximity

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}
X_1	12	4	3	2	3	0	2	2	2	3	3	4
X_2	4	12	3	3	4	4	5	5	5	4	3	2
X_3	3	3	12	6	8	2	8	8	8	5	9	2
X_4	2	3	6	12	9	1	7	7	8	4	9	7
X_5	3	4	8	9	12	2	9	9	8	5	8	7
X_6	0	4	2	1	2	12	3	3	2	2	1	0
X_7	2	5	8	7	9	3	12	11	8	4	7	6
X_8	2	5	8	7	9	3	11	12	8	4	7	6
X_9	2	5	8	8	8	2	8	8	12	6	9	6
X_{10}	3	5	5	4	5	2	4	4	6	12	6	4
X_{11}	3	3	9	9	8	1	7	7	9	6	12	8
X_{12}	4	2	2	7	7	0	6	6	6	4	8	12
X_{13}	3	4	8	6	6	3	7	7	6	4	6	6
X_{14}	3	4	9	8	10	2	8	8	11	6	10	6
X_{15}	4	2	7	7	7	4	5	5	7	2	5	7
X_{16}	3	4	5	6	6	3	8	7	6	2	5	8
X_{17}	5	3	7	6	8	2	7	7	8	4	7	7
X_{18}	5	4	6	4	5	2	5	5	5	1	4	6
X_{19}	4	3	6	5	7	2	7	7	7	6	6	6
X_{20}	1	5	6	6	9	3	7	7	8	3	6	4
X_{21}	2	4	6	8	9	1	7	7	8	4	6	8
X_{22}	2	4	7	8	9	1	8	8	7	5	7	8
X_{23}	2	4	7	7	8	1	7	7	7	4	7	7
X_{24}	2	4	6	8	9	1	7	7	8	3	6	6

	X_{13}	X_{14}	X_{15}	X_{16}	X_{17}	X_{18}	X_{19}	X_{20}	X_{21}	X_{22}	X_{23}	X_{24}
X_1	3	3	4	3	5	5	4	1	2	2	2	2
X_2	4	4	2	4	3	4	3	5	4	4	4	4
X_3	8	9	7	5	7	6	6	6	6	7	7	6
X_4	6	8	7	6	6	4	5	6	8	8	7	8
X_5	6	10	7	6	8	5	7	9	9	9	8	9
X_6	3	2	4	3	2	2	2	3	1	1	1	1
X_7	7	8	5	8	7	5	7	7	7	8	7	7
X_8	7	8	5	7	7	5	7	7	7	8	7	7
X_9	6	11	7	6	8	5	7	8	8	7	7	8
X_{10}	4	6	2	2	4	1	6	3	4	5	4	3
X_{11}	6	10	5	5	7	4	6	6	6	7	7	6
X_{12}	6	6	7	8	7	6	6	4	8	8	7	6
X_{13}	12	6	6	6	5	6	5	5	7	8	8	7
X_{14}	6	12	7	6	8	5	7	8	8	7	7	8
X_{15}	6	7	12	7	7	7	6	5	7	5	6	7
X_{16}	6	6	7	12	7	7	7	5	7	7	7	7
X_{17}	5	8	7	7	12	7	10	7	7	6	7	7
X_{18}	6	5	7	7	7	12	7	5	5	5	5	5
X_{19}	5	7	6	7	10	7	12	8	8	6	8	8
X_{20}	5	8	5	5	7	5	8	12	10	7	9	10
X_{21}	7	8	7	7	7	5	8	10	12	9	11	12
X_{22}	8	7	5	7	6	5	6	7	9	12	10	9
X_{23}	8	7	6	7	7	5	8	9	11	10	12	11
X_{24}	7	8	7	7	7	5	8	10	12	9	11	12

Now let's conduct the following classification using the fourth table:

The objects that voted most for object x_1 are objects x_{17} and x_{18}. The degree of similarity of these objects is 5. That is, 5 out of 12 probabilities;

The objects that voted most for object x_2 are objects x_7, x_8, x_9 and x_{19}. The degree of similarity of these objects is 5. That is, 5 out of 12 probabilities;

The objects that voted most for object x_3 are objects x_{11} and x_{14}. The degree of similarity of these objects is 9. That is, 9 out of 12 probabilities;

The objects that voted most for object x_4 are objects x_5 and x_{11}. The degree of similarity of these objects is 9. That is, 9 out of 12 probabilities;

The object that voted most for object x_5 is object x_{14}. The degree of similarity of these objects is 10. That is, 10 out of 12 probabilities;

The objects that voted most for object x_6 are objects x_2 and x_{15}. The degree of similarity of these objects is 4. That is, 4 out of 12 probabilities;

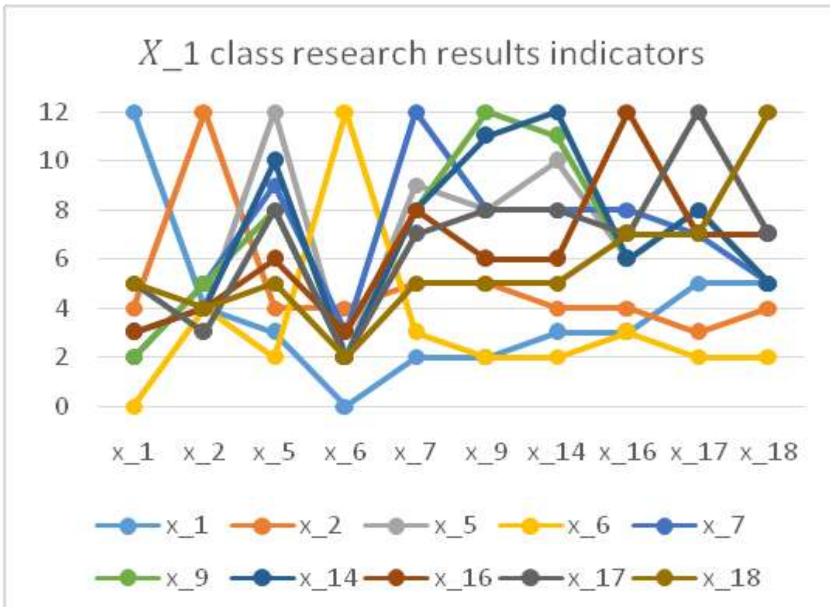
The object that voted most for object x_7 is object x_8. The degree of similarity of these objects is 11. That is, 11 out of 12 probabilities;

The objects that voted most for object x_8 is object x_7. The degree of similarity of these objects is 11. That is, 11 out of 12 probabilities;

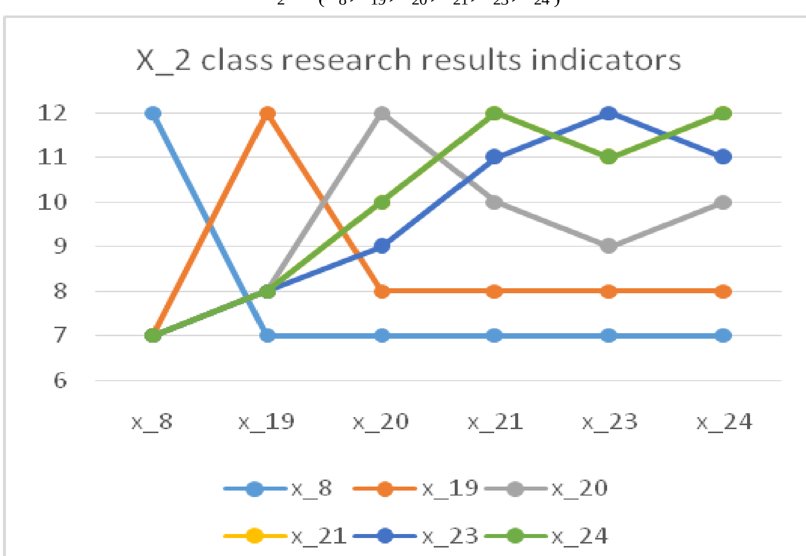
The object that voted most for object x_9 is object x_{14}. The degree of similarity of these objects is 11. That is, 11 out of 12 probabilities;

The object that voted most for object x_{10} is object x_{19}. The degree of similarity of these objects is 6. That is, 6 out of 12 probabilities. The following three classes are formed as a result of the research:

$$X_1 = \{x_1, x_2, x_5, x_6, x_7, x_9, x_{14}, x_{16}, x_{17}, x_{18}\}$$



$$X_2 = \{x_8, x_{19}, x_{20}, x_{21}, x_{23}, x_{24}\}$$



$$X_3 = \{x_3, x_4, x_{10}, x_{11}, x_{12}, x_{13}, x_{15}, x_{22}\}$$

The problem of localization is explored in the third problem by using these classes.

Third problem. The classification and localization of cotton varieties in the agricultural sector.

As a result of the initial data processing, the study choices are divided into 3 classes. As a result, there are 10 objects in the first class, which are as follows:

$$X_1 = \{x_1, x_2, x_5, x_6, x_7, x_9, x_{14}, x_{16}, x_{17}, x_{18}\} = \begin{Bmatrix} Bukhara-8,(1,3,5) \\ Chimboy-5018,(1) \\ C-6541,\ (11,2,4,10); \\ Khorazm-150,(13); \\ Dostlik-2,(1,13); \\ Beshqakhraman,(5,9); \\ C-6775,\ (12); \\ C-8284,\ (8,13); \\ Porloq-1,(4,8,9,10); \\ Porloq-2,(3,4,6,7,13); \end{Bmatrix}$$

As a result, the matrix view of the class, the objects of the class, and the representation object are as follows:

$$X_1 = \begin{pmatrix} 3 & 2 & 2 & 2 & 4 & 1 & 1 & 2 & 2 & 5 & 2 & 2 \\ 3 & 2 & 2 & 2 & 2 & 1 & 1 & 2 & 2 & 8 & 2 & 2 \\ 3 & 2 & 2 & 2 & 3 & 1 & 1 & 2 & 2 & 8 & 2 & 2 \\ 1 & 2 & 2 & 2 & 4 & 1 & 1 & 2 & 2 & 3 & 2 & 3 \\ 3 & 2 & 2 & 2 & 7 & 1 & 1 & 2 & 1 & 10 & 2 & 4 \\ 3 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 4 \\ 2 & 2 & 1 & 1 & 2 & 1 & 2 & 2 & 2 & 2 & 1 & 3 \\ 2 & 4 & 2 & 3 & 5 & 3 & 2 & 2 & 3 & 6 & 3 & 3 \\ 2 & 3 & 1 & 2 & 8 & 1 & 1 & 2 & 1 & 7 & 2 & 4 \\ 1 & 2 & 3 & 2 & 5 & 1 & 1 & 2 & 1 & 7 & 2 & 3 \end{pmatrix}$$

$x_{v1} = (3 \;\; 2 \;\; 2 \;\; 2 \;\; 4 \;\; 1 \;\; 1 \;\; 2 \;\; 2 \;\; 8 \;\; 2 \;\; 3)$ – this considered the representation object of X_1 class.

The second class consists 6 variety objects and they are as follows:

$$X_2 = \{x_8, x_{19}, x_{20}, x_{21}, x_{23}, x_{24}\} = \begin{Bmatrix} sulton \quad (1,2,4,11), \\ Porloq4 \quad (5,12), \\ C-6550 \quad (9), \\ Namangan-102 \; (5), \\ Istiqlol-14 \; (2,9), \\ Kelajak \quad (11) \end{Bmatrix}$$

The objects of the class and the representative object formed on the basis of these varieties are as follows:

$$X_2 = \begin{pmatrix} 5 & 2 & 3 & 2 & 7 & 1 & 1 & 2 & 2 & 5 & 2 & 2 \\ 5 & 2 & 3 & 2 & 7 & 1 & 1 & 2 & 2 & 5 & 2 & 1 \\ 5 & 2 & 3 & 2 & 7 & 1 & 1 & 2 & 2 & 5 & 2 & 2 \\ 6 & 2 & 2 & 2 & 2 & 1 & 1 & 2 & 2 & 7 & 2 & 3 \\ 5 & 2 & 2 & 2 & 7 & 1 & 1 & 2 & 1 & 11 & 2 & 4 \\ 5 & 2 & 2 & 2 & 7 & 1 & 2 & 2 & 2 & 5 & 2 & 2 \end{pmatrix}$$

$x_{v2} = (5 \;\; 2 \;\; 3 \;\; 2 \;\; 7 \;\; 1 \;\; 1 \;\; 2 \;\; 2 \;\; 5 \;\; 2 \;\; 2)$ – it is considered the representation object of X_2 class.

The third class consists 8 variety objects and they are as follows:

$$X_3 = \{x_3, x_4, x_{10}, x_{11}, x_{12}, x_{13}, x_{15}, x_{22}\} = \begin{Bmatrix} Andijon36-(2,7,11,12), \\ Namangan34-(6,7,8), \\ Ibrat-(13,1), \\ Kopay\sin-(4), \\ Gulbakhor2-(10), \\ Andijon37-(1), \\ Navbakhor-2-(10), \\ Paktakor-1-(4) \end{Bmatrix}$$

The class objects and representative objects, based on the proposed varieties, are as follows:

$$X_3 = \begin{pmatrix} 3 & 2 & 3 & 2 & 4 & 2 & 1 & 2 & 2 & 4 & 2 & 2 \\ 3 & 2 & 3 & 2 & 4 & 1 & 1 & 3 & 1 & 7 & 2 & 1 \\ 2 & 4 & 3 & 2 & 1 & 1 & 1 & 2 & 2 & 3 & 2 & 1 \\ 3 & 3 & 3 & 2 & 6 & 1 & 1 & 2 & 1 & 9 & 2 & 2 \\ 3 & 2 & 3 & 2 & 5 & 1 & 1 & 2 & 2 & 8 & 2 & 1 \\ 3 & 2 & 2 & 2 & 3 & 1 & 1 & 3 & 2 & 8 & 2 & 1 \\ 3 & 3 & 2 & 2 & 3 & 1 & 1 & 2 & 2 & 3 & 2 & 1 \\ 4 & 2 & 3 & 2 & 7 & 1 & 1 & 2 & 2 & 5 & 2 & 1 \end{pmatrix}$$

$x_{v3} = (3 \quad 2 \quad 3 \quad 2 \quad 3 \quad 1 \quad 1 \quad 2 \quad 2 \quad 8 \quad 2 \quad 1)$ – it is considered the representation object of X_3 class.

Based on the information in brackets, Pakhtakor-1 (4) is permitted to plant in Jizzakh region in accordance with the State Register of Agricultural Crops recommended for planting in the territory of the Republic of Uzbekistan, The following table provides information on localization.

Table 4. Numbering of regions according to the state register

Name	No.
Republic of Karakalpakstan	1
Andijan region	2
Bukhara region	3
Jizzakh region	4
Kashkadarya region	5
Navio region	6
Namangan region	7
Samarkand region	8
Surkhandarya region	9
Syrdarya region	10
Tashkent region	11
Fergana region	12
Khorezm region	13

Discussion

In the article, the issue of classification and localization of agricultural technical crops was fully analyzed in terms of varieties of cotton. According to the analysis, it is recommended to plant the first class of cotton varieties in the following regions:

$$X_1 = \{x_1, x_2, x_5, x_6, x_7, x_9, x_{14}, x_{16}, x_{17}, x_{18}\} => \begin{pmatrix} 1(3) \\ 2(1) \\ 3(2) \\ 4(2) \\ 5(2) \\ 6(1) \\ 7(1) \\ 8(2) \\ 9(2) \\ 10(2) \\ 11(1) \\ 12(1) \\ 13(4) \end{pmatrix} = (1,3,5,8,9,13)$$

$$X_2 = \{x_8, x_{19}, x_{20}, x_{21}, x_{23}, x_{24}\} => \begin{pmatrix} 1(1) \\ 2(2) \\ 4(1) \\ 5(1) \\ 9(1) \\ 11(2) \\ 12(1) \end{pmatrix} = (2,11,12)$$

$$X_3 = \{x_3, x_4, x_{10}, x_{11}, x_{12}, x_{13}, x_{15}, x_{22}\} => \begin{pmatrix} 1(2) \\ 2(1) \\ 4(2) \\ 6(1) \\ 7(2) \\ 8(1) \\ 10(2) \\ 11(1) \\ 12(1) \\ 13(1) \end{pmatrix} = (4,6,7,10).$$

The fourth issue. It is required to develop an algorithm for selecting the set of informative features in the space of textual, nominal and quantitative features.

Step 1. Entering the selected study objects

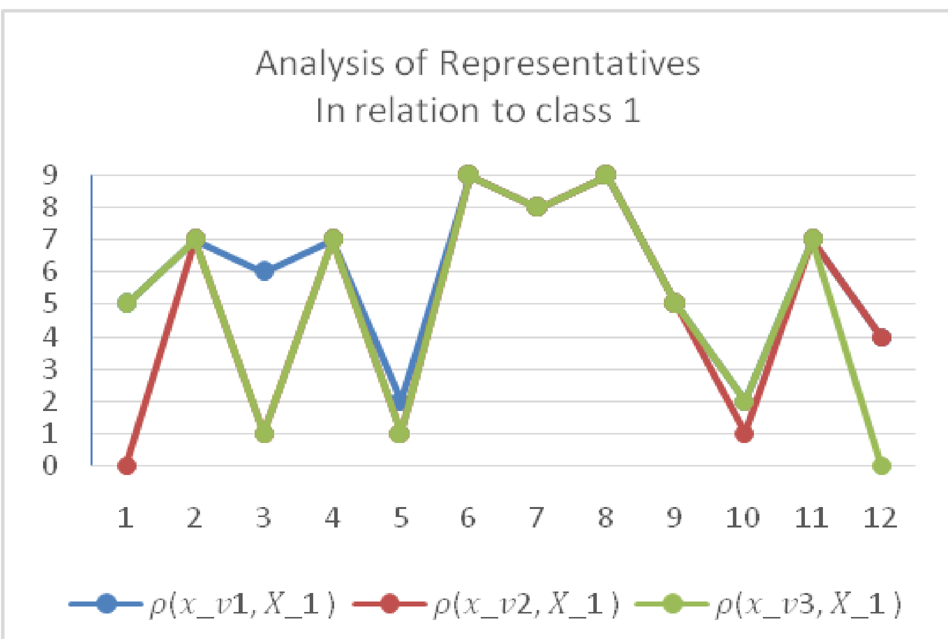
$$X_1=\{x_1,x_2,x_5,x_6,x_7,x_9,x_{14},x_{16},x_{17},x_{18}\};\ X_2=\{x_8,x_{19},x_{20},x_{21},x_{23},x_{24}\};$$
$$X_3=\{x_3,x_4,x_{10},x_{11},x_{12},x_{13},x_{15},x_{22}\};$$

Step 2. A representative $x_{v1}\in X_1$; $x_{v2}\in X_2$; $x_{v3}\in X_3$ is formed for each class. In this step, the most common textual, nominal and quantitative features is selected in the column of textual, nominal and quantitative features considered as a textual, nominal and quantitative feature of the class' representative,. For example, the representative objects for three classes above are:

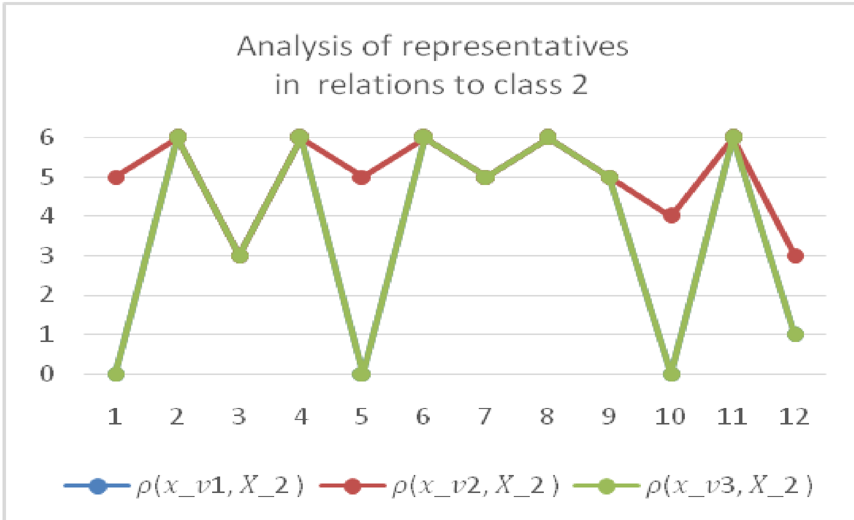
$$x_{v1}=(3\ \ 2\ \ 2\ \ 2\ \ 4\ \ 1\ \ 1\ \ 2\ \ 2\ \ 8\ \ 2\ \ 3)\in X_1;$$
$$x_{v2}=(5\ \ 2\ \ 3\ \ 2\ \ 7\ \ 1\ \ 1\ \ 2\ \ 2\ \ 5\ \ 2\ \ 2)\in X_2;$$
$$x_{v3}=(3\ \ 2\ \ 3\ \ 2\ \ 3\ \ 1\ \ 1\ \ 2\ \ 2\ \ 8\ \ 2\ \ 1)\in X_3.$$

Three selected representatives evaluate the objects of X_p class: the function of proximity of the space of informative features applied to compute the value of contribution to the formation of its own class is calculated on the basis of formula (1);

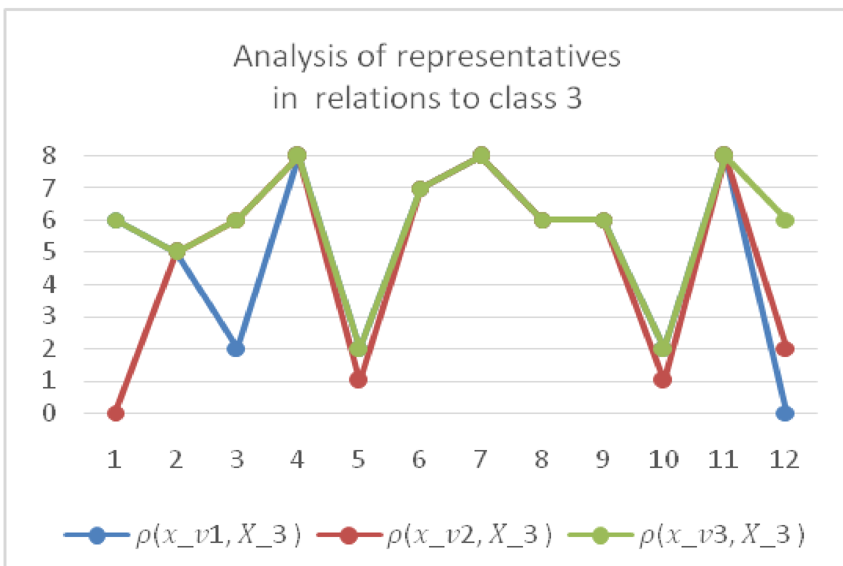
$$\rho(x_{v1},X_1)=(5\ \ 7\ \ 6\ \ 7\ \ 2\ \ 9\ \ 8\ \ 9\ \ 5\ \ 2\ \ 7\ \ 4);$$
$$\rho(x_{v2},X_1)=(0\ \ 7\ \ 1\ \ 7\ \ 1\ \ 9\ \ 8\ \ 9\ \ 5\ \ 1\ \ 7\ \ 4);$$
$$\rho(x_{v3},X_1)=(5\ \ 7\ \ 1\ \ 7\ \ 1\ \ 9\ \ 8\ \ 9\ \ 5\ \ 2\ \ 7\ \ 0);$$



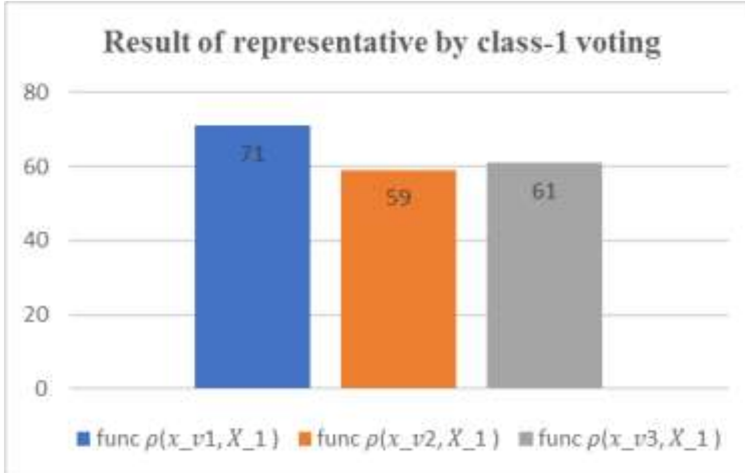
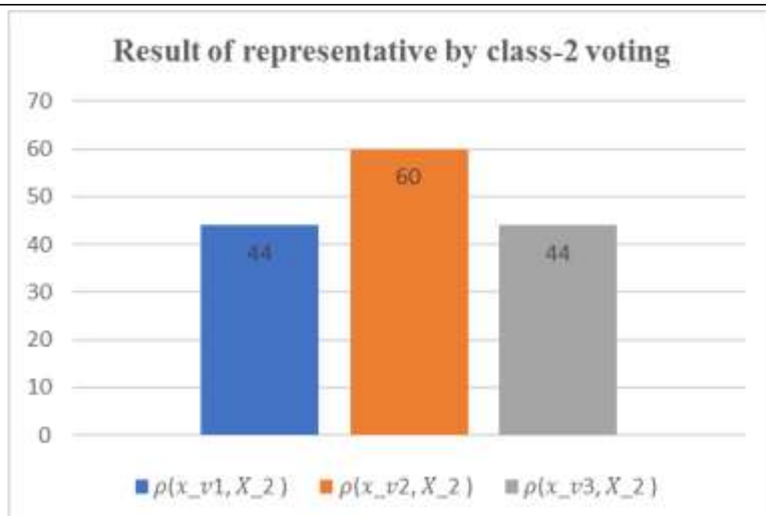
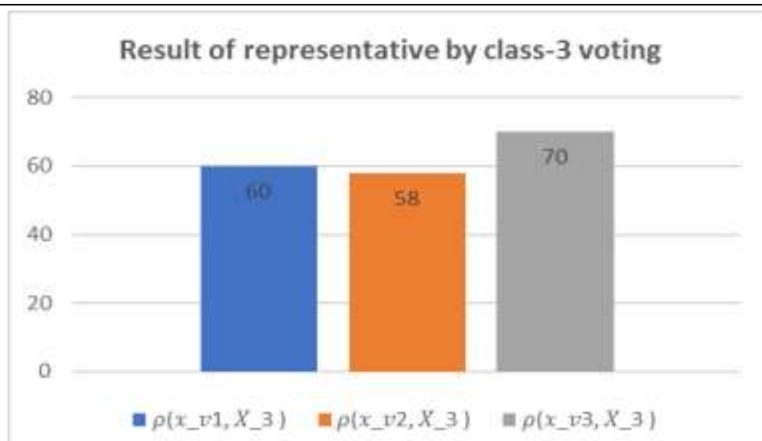
$$\rho(x_{v1}, X_2) = (0 \quad 6 \quad 3 \quad 6 \quad 0 \quad 6 \quad 5 \quad 6 \quad 5 \quad 0 \quad 6 \quad 1);$$
$$\rho(x_{v2}, X_2) = (5 \quad 6 \quad 3 \quad 6 \quad 5 \quad 6 \quad 5 \quad 6 \quad 5 \quad 4 \quad 6 \quad 3);$$
$$\rho(x_{v3}, X_2) = (0 \quad 6 \quad 3 \quad 6 \quad 0 \quad 6 \quad 5 \quad 6 \quad 5 \quad 0 \quad 6 \quad 1).$$



$$\rho(x_{v1}, X_3) = (6 \quad 5 \quad 2 \quad 8 \quad 2 \quad 7 \quad 8 \quad 6 \quad 6 \quad 2 \quad 8 \quad 0);$$
$$\rho(x_{v2}, X_3) = (0 \quad 5 \quad 6 \quad 8 \quad 1 \quad 7 \quad 8 \quad 6 \quad 6 \quad 1 \quad 8 \quad 2);$$
$$\rho(x_{v3}, X_3) = (6 \quad 5 \quad 6 \quad 8 \quad 2 \quad 7 \quad 8 \quad 6 \quad 6 \quad 2 \quad 8 \quad 6);$$



Step 3. Representative objects in the space of textual, nominal and quantitative features evaluate their own class and other classes. The evaluation process is based on formula (2). The magnitude of the degree of similarity in the context of representative objects and class objects is computed with the formula (2) as follows:

$\begin{cases} \Gamma_{11}(x_{v1}, X_1) = \sum_{j=1}^{12} \rho(x_{v1}, X_1) = 71, \\ \Gamma_{21}(x_{v2}, X_1) = \sum_{j=1}^{12} \rho(x_{v2}, X_1) = 59, \\ \Gamma_{31}(x_{v3}, X_1) = \sum_{j=1}^{12} \rho_j(x_{v3}, X_1) = 61 \end{cases}$	
$\begin{cases} \Gamma_{12}(x_{v1}, X_2) = \sum_{j=1}^{12} \rho(x_{v1}, X_2) = 44, \\ \Gamma_{22}(x_{v2}, X_2) = \sum_{j=1}^{12} \rho(x_{v2}, X_2) = 60, \\ \Gamma_{32}(x_{v3}, X_2) = \sum_{j=1}^{12} \rho_j(x_{v3}, X_2) = 44 \end{cases}$	
$\begin{cases} \Gamma_{13}(x_{v1}, X_3) = \sum_{j=1}^{12} \rho(x_{v1}, X_3) = 60, \\ \Gamma_{23}(x_{v2}, X_3) = \sum_{j=1}^{12} \rho(x_{v2}, X_3) = 58, \\ \Gamma_{33}(x_{v3}, X_3) = \sum_{j=1}^{12} \rho_j(x_{v3}, X_3) = 70 \end{cases}$	

Conclusion

The application of Information Communication Technology to the solution of the problems of the sphere is associated with the initial design, in-depth analysis and study of these problems, processes.

In the article, the information obtained from the initial research was agricultural crop varieties derived from Real life, as a result of this research, it was required to address such important issues as the expression of objects by means of textual characters and then transfer them to nominal characters, the classification and documentation of objects and the selection of character sets in their perfect.

Scientific approaches, studies of many scientists in solving these issues positively were studied and analyzed. Let's take some of these, they are the following methods, algorithms, scientific approaches related to the development of software, which are used to solve the issues of classification, clustering and selection of character sets.

The F-evaluation method based on fisherionionality is a simple method of character selection [4]. Usually the training, consisting of two classes, works perfectly for the Chosen One, the result of which is obtained is based on the sorting of the characters. The method proposed by these authors [4] is based on the discriminant analysis of Fisher, in which a set of characters is selected on the basis of the distance from classlararo distances and the compactness of the internal distances of class objects [2,3,4]. The removable results can also be made for multiple classes.

In the recognition of logos, solving the issues of classification and clustering, initially the choice of a set of characters is one of the main issues. A great way to find a solution to the issue and build a character ensemble based on them has been studied by the authors [5,6,7,8,9,10] in research. In it, the initial data were expressed in abundance, and in the proposed methods it was attempted to determine the solution of the main issues, to maximize the percentage of classification of objects and to achieve the evasion of minimizing the set of characters. This approach is considered acceptable for non-educational selections, the size of the initial data, which is pre-determined. Just as well as from Real cases, characters from life can not at all times fully express objects, for these cases it is a matter of classification and ensuring the reliability of objects is extremely important. As a rule, the data in this view are called Noravshan, [8] in the scientific research, the Non-cases of research facilities were investigated and the question of classification based on this data was solved.

Classification, clustering, recognition of logos and data intellektual analysis in the initial processing for cases where data signs are given in the form of text, reducing the character set and solving the issues proposed on their basis is an extremely important study. By [9, 10], the researchers proposed an algorithm for reducing the size of the characters being looked at for cases where the data signs consist of multiple metrics.

We think that the proposed research on this subject [11-15], the proposed method, algorithm and software research in finding solutions to the problems posed will not leave the reader indifferent and we are confident that this research will find its place in making our lives comfortable, making our lives easy. In the article, the expression of finding a solution to the issue includes the following sources.

References

Recommended for cultivation on the territory of the Republic of Uzbekistan. Description of varieties and hybrids included in the State Register for 2007-2017.

Crops recommended for planting in the territory of the Republic of Uzbekistan. Description of varieties and hybrids included in the State Register for 2008-2018

Ahmedov OK, Khodjiev SS, Hasanov UA. Algorithm based on sorting standard tables for pattern recognition. Problems of Informatics and Energy 2016; 6: 34-38.

Akhmedov OK. Algorithm based on sorting standard tables for pattern recognition. International journal for innovative research in multidisciplinary field 2019; 5(8).

Nishanov AH, Beglerbekov RJ, Akhmedov OK. A hybrid algorithm for recognition of patterns in the space of informative features. TATU news 2017; 4 (44): 62-69.

Song Q, Jiang H, Liu J. Feature selection based on FDA and F-score for multi-class classification. Expert Systems with Applications 2017; 81: 22-27.

Bolón-Canedo V, Alonso-Betanzos A. Ensembles for feature selection: A review and future trends. Information Fusion 2019; 52: 1-12.

Arora S, Anand P. Binary butterfly optimization approaches for feature selection. Expert Systems with Applications 2019; 116: 147-160.

Hancer E, Xue B, Karaboga D, Zhang M. A binary ABC algorithm based on advanced similarity scheme for feature selection. Applied Soft Computing 2015; 36: 334-348.

Yong Z, Dun-wei G, Wan-qiu Z. Feature selection of unreliable data using an improved multi-objective PSO algorithm. Neurocomputing 2016; 171; 1281-1290.

Zhang Y, Gong DW, Sun XY, Guo YN. A PSO-based multi-objective multi-label feature selection method in classification. Scientific reports 2017; 7(1): 1-12.

Emary E, Zawbaa HM, Hassanien AE. Binary ant lion approaches for feature selection. Neurocomputing 2016; 213: 54-65.

Mafarja M, Mirjalili S. Hybrid binary ant lion optimizer with rough set and approximate entropy reducts for feature selection. Soft Computing 2019; 23 (15).

Nishanov A.Kh, Turakulov Kh.A, Turakhanov Kh.V. A decision rule for identification of eye pathologies. Biomedical Engineering 1999; 33(4): 178-179.

Nishanov A. Kh, Turakulov Kh.A,Turakhanov Kh.V. A decisive rule in classifying diseases of the visual system. Meditsinskaia tekhnika 1999; 4: 16-18.

Nishanov A.Kh, Djurayev GP, Kasanova M.Kh. Improved algorithms for calculating evaluations in processing medical data. National Institute of Science Communication and Information Resources (NISCAIR)-India 2019: 3158-3165.

Kamilov MM, Nishanov AX, Beglerbekov RJ. Modified stages of algorithms for computing estimates in the space of informative features. International Journal of Innovative Technology and Exploring Engineering 2019; 8(6): 714-717.

Nishanov AX, Ruzibaev OB, Tran NH. Modification of decision rules "ball Apolonia" the problem of classification. In International Conference on Information Science and Communications Technologies (ICISCT) 2016; 1-3.

Nishanov A, Avazov E, Akbaraliyev B. Partial selection method and algorithm for determining graph-based traffic routes in a real-time environment. International Journal of Innovative Technology and Exploring Engineering 2019; 8(6): 696-698.

Sarabi RE, Abdekhoda M, Dehnad A, Khajouei G. Environmental literacy and accountability of undergraduate students of medical sciences. Webology 2020; 17(1): 191-201.

Abdekhoda M, Sattari A, Mohammadi M, Salih KM. Presenting a conceptual model of adopting micro-blogging in learning. Webology 2020; 17(1): 98-108.