Practice of applying models of clustering of incoming semi-structured data in management systems

O. U. Askaraliyev*, *G. X.* Muxtarova, *N. T.* Malikova, and *R. H.* Yuldashev

Tashkent University of Information Technologies named after Muhammad al-Khwarizmi, Tashkent, Uzbekistan

Abstract. A scheme for solving the problem of binary clustering of incoming data in the control system is proposed. Different ways of presenting the initial data of the clustering problem are considered. Methods of sequential shortening and sequential merging of clusters and the model of initial placement of clusters are considered. Estimating the number of clusters is given to solve the clustering problem. A binary clustering method has been proposed for points in a circle. The clustering practice was performed based on the research results, and the developed model and algorithms were implemented. The optimal variant of the processing process based on clustering methods for semi-formed information sources is proposed based on the proposed structure.

1 Introduction

Many theoretical issues of geometric clustering and classification of weakly structured information remain unresolved to date [1-6]. Semi-structured data is data for which the exact structure is not known in advance and can change in the stream. Such data usually includes texts and pictures, which can be either separately or simultaneously present in documents presented in various formats.

Mathematical models of information representation and algorithms for transformation, feature extraction, and processing are needed to analyze such data. In this case, the difficulties are overcome by a set of different methods.

Fig. 1 presents weakly structured information's main stages, models, and cluster analysis methods. Much attention is paid to metrics, choosing the number, and initial placement of clusters. Formal formulations of several problems are common in the clustering of texts and images. Therefore, an urgent task is the integration of various models in a single processing system.

2 Methods

1. Models for representing semi-structured information. Models of data representation and knowledge significantly impact the choice of cluster analysis method. In the case of

*Corresponding author: oasqaraliyev77@gmail.com

multimodal information, characterized by heterogeneous sources of information (texts, images, sounds) and in the presence of many features, increased requirements are imposed on the form of data presentation. We can distinguish, in our opinion, the four most significant systems for presenting initial data:

1) a tabular method, which is widely used in solving classification problems by the algebraic method [1], neural networks, decision trees, and in constructing a reference set of clusters;

2) a method of representing data by multisets [3], the effectiveness of which has been shown in economic problems and is being studied concerning the problems of recognition of graphic images, for example, block letters;

3) the method of phase trajectories, proposed for processing multimodal information [4], and so far insufficiently studied in practice. The problem of forming trajectories in the phase space and their intelligent control based on the rules is considered in [5];

4) procedural representation, when knowledge about a specific problem area is specified in a set of rules, and declarative representation of knowledge, when information is stored in the form of a database and/or knowledge base [6].

Let us consider some data representation models that can be applied to cluster analysis problems [7-9].

2. Tabular way of setting initial data. The tabular method of assignment considered the main one, was applied and studied in [7] concerning the problem of binary classification. Semantic models are relatively rarely used in clustering and classification problems [6]. The representations 2) and 3) indicated above are considered an addition to the tabular method of data presentation.

Let the initial information for the objects $\omega_i, i = 1, \dots, m$ be given as training samples. The corresponding data structure is presented in Table 1. Classes $\Omega_1 \, and \, \Omega_2$ are represented by matrices of feature values $X_1 \, and \, X_1$ of dimensions $(m_1 \times p) \, and \, (m_2 \times p)$, respectively, with $(m_1 + m_1 = m)$.

Thus, Table 1 contains data on m objects $\{\omega_1, \dots, \omega_m\}$, each of which is represented by a vector of values of information features $x = \{x_1, \dots, x_m\}$ and assigned by the expert to one of two classes $\{\Omega_1, \Omega_2\}$.

Table 1. Representation of the training sample

Objects	Signs and their meanings			Class
	x_1	…	x_p	
ω_1	x_{11}	…	x_{1p}	Ω_1
…	…	…	…	…
ω_{m_1}	$x_{m_1 1}$	…	$\omega_{m_1 p}$	Ω_1
ω_{m_1+1}	$\omega_{(m_1+1)1}$	…	$\omega_{(m_1+1)p}$	Ω_2
…	…	…	…	…
ω_m	x_{m1}	…	x_{mp}	Ω_2

The tabular method conveniently represents information for constructing decision rules for classification problems. In particular, it is well suited for solving classification problems by the algebraic method. Since the dimension of the attribute space changes in document processing tasks, the tabular method should work in conjunction with the variation method of the attribute space [7].

3. Multisets for describing data. The work [3] proposes a method for classifying a set of multi-attribute objects based on their representation as points in a metric space of multisets.

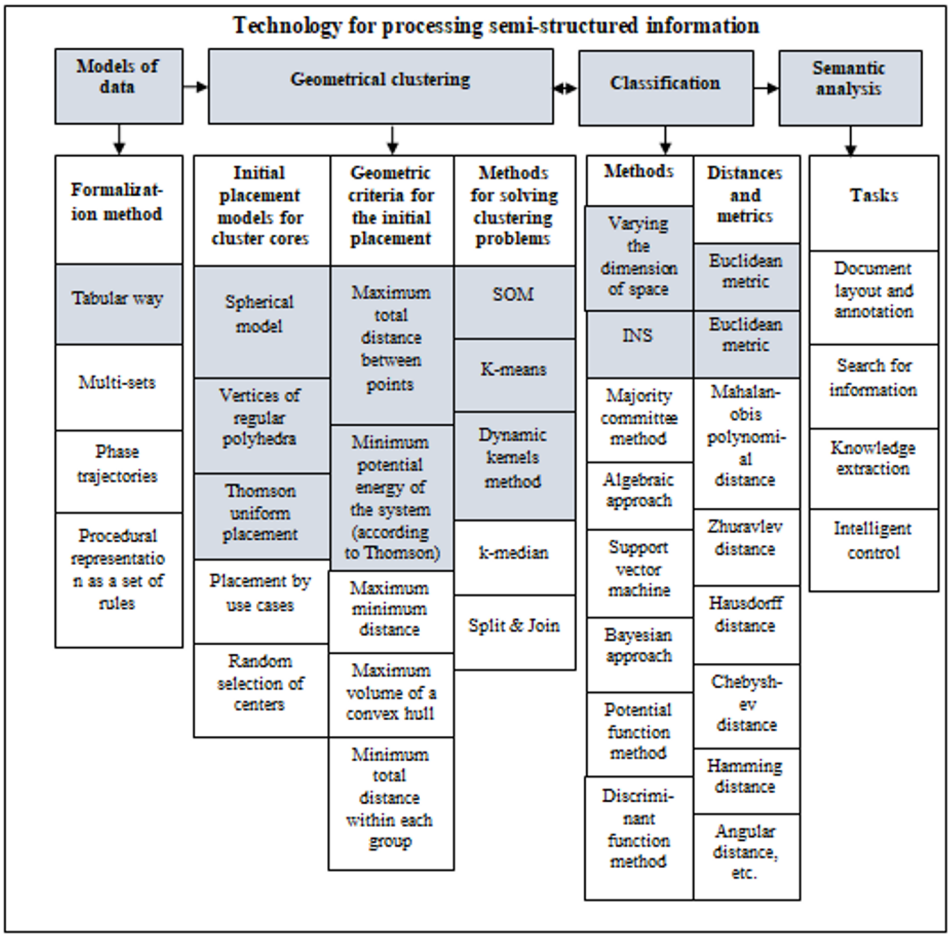


Fig. 1. Components of semi-structured information processing technology

A multiset or a set with repeating elements serves as a convenient mathematical model for describing objects that are characterized by many heterogeneous (quantitative and qualitative) features and can exist in several instances with different, in particular, contradictory values of features, the convolution of which is either impossible or mathematically incorrect. A multiset A generated by an ordinary set $U = \{x_1, x_2, \dots\}$, all elements of which are different, is a set of groups of elements of the form $A = \{k_A(x) \bullet x \,|\, x \in U,\ k_A(x) \in Z+\}$. Here $k_A: U \to Z+ = \{0,1,2,\dots\}$ is called the function of the number of instances of the multiset, which determines the multiplicity of occurrence of the element $x_i \in U$ in the multiset A, which is denoted by the symbol $\bullet$. If $k_A(x) = \chi_A(x)$ where $\chi_A(x) = 1$ for $x \in A$ and $\chi_A(x) = 0$ for $x \notin A$, then the multiset A becomes an ordinary many. If all multisets of the family $A = \{A_1, A_2, \dots\}$, are formed from elements of the set G, then G is called a domain for the family A, and the set $SuppA = \{x | x \in G, \chi_{SuppA}(x) = \chi_A(x)\}$ – support set or carrier of the multiset A.

The cardinality of the multiset $|A|$ is defined as the total number of instances of all its elements; the dimension of the multiset $|SuppA|$ is defined as the total number of distinct elements.

The concept of a measure of a multiset, which has the properties of monotonicity, symmetry, continuity, and elasticity, as well as the concept of a metric space with three

types of distances, is introduced. The considered mathematical apparatus provides ample opportunities for describing data and performing operations on them [8, 9]. The scheme of the classification method using the proposed description is to build decision rules over a small number of the most significant (selected as a result of selection) features.

All initial multisets are grouped (summed up) into two multisets representing two classes of objects. Multisets-sums, in turn, are divided into several multisets-summands according to the number of features that characterize objects. For each group of features, each pair of summands of multisets generates a pair of new multisets that are maximally distant from each other in the metric space. The boundary between new terms in each pair is determined by some value of the corresponding feature. Various combinations of such "boundary" feature values provide generalized decision rules for classifying objects.

Classification is carried out with the help of generalized decision rules, composed of different "boundary" values of features, which resemble the method of decision trees, providing the desired level of accuracy of object classification.

4. Phase trajectories for describing multimodal data. In [4], a method for classifying multimodal data for processing speech, images, and texts is proposed. The associativity of accessing information allows you to quickly obtain the necessary information, regardless of the sample size, and the structural approach to information processing allows you to automatically restore the structure and compactly store the information received.

Any keyword can be represented as a sequence of symbols and spoken aloud - as a sequence of phonemes. The image is presented for processing as a sequence of codes of its pixels or a sequence of numbers characterizing the selected features (Fourier coefficients, wavelet coefficients, invariants, descriptions of hierarchies, etc.). Thus, regardless of modality, information about a semantic unit of information (word, image) is represented as a sequence of binary numbers. The transformation F of a binary sequence into the space R^n is considered in such a way that each n - member fragment of the sequence (i.e., a vector of dimension n) corresponds to a point $a(t)$ в R^n with the corresponding coordinates, and the entire sequence A corresponds to a sequence of points - a trajectory: $\breve{A} = F(A)$. Here F is a mapping into the signal space, the basis for structural information processing. The transformation F has the associativity property of addressing the points of the trajectory $\breve{A}$: any n symbols address the corresponding trajectory point. In the general case, among the n-membered fragments of the information sequence, there may be a fragment already stored in memory (possibly a reference one), and the trajectories, in this case, will intersect or coincide. Identical fragments of the sequence are transformed into the same trajectory. Here, recognition is understood as the process of deciding on the degree of coincidence of the input information with the previously stored information. The recognition mechanism is based on comparing the input sequence $\breve{A}$ and the stored sequence A closest to it with the calculation of the proximity measure (according to Hamming).

The decision on a match with a given degree of accuracy is compared with the recognition threshold. In a simpler case, when learning, the sequence corresponding to the event to be remembered is used as the carrier sequence A, and the sequence of symbols of the code corresponding to this event is used as the information sequence J. In this case, recognition is understood as the reproduction of the information sequence of the J-code of the event that initiates the input sequence A. It is shown [4] that the described information processing processes: learning, reproduction, formation of a dictionary, and syntactic sequence, are effective both within the framework of one associative process and in multilevel systems. The use of all the properties of the associative process is possible only if it is included in a hierarchical structure that performs a structural analysis of information. Unfortunately, the author of this interesting work failed to fully implement practice the proposed method of presenting multimodal information within a single information system.

The considered information representation models complement each other and, in our opinion, can be used jointly in clustering and classification systems. The task of cluster analysis of text and graphic information is:

- in extracting informative parameters, for example, extracting the text index and informative parameters of images (integral or/and invariant parameters). This stage can be called information preprocessing;
- in clustering, i.e., automatic formation of headings and clusters of images;
- in the classification of text or images, the creation of an abstract;
- in the semantic processing of the received information for the purpose of generalization.

Models should be applied in the order in which they are considered in this work, together with the corresponding information processing algorithms.

3 Results and discussion

Scheme for solving the problem of binary clustering. The proposed technology is based on a network analysis model that allows solving problems in a complex: traveling salesman based on constructing a minimum length contour, clustering based on the method of dynamic kernels, and classification with fixed kernels using a set of metrics [10-13].

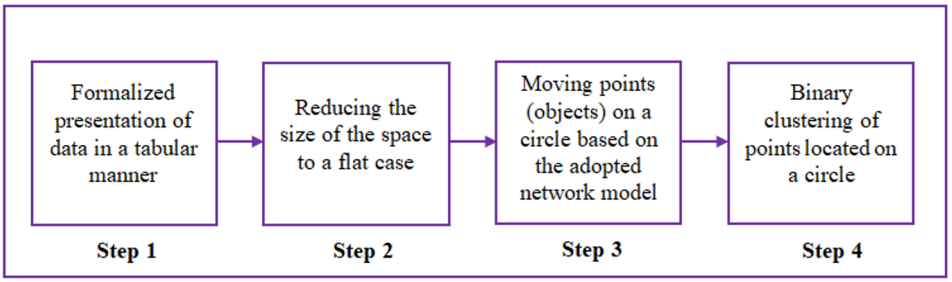


Fig. 2. Scheme for solving problem of binary clustering

In Fig. 2, a scheme for the joint application of various cluster analysis methods for solving the binary clustering problem is considered, which contains the following steps:

1) a formalized presentation of data in a tabular way is given, in which objects are described as vectors of a feature space;

2) there is a reduction in the dimension of the feature space to a flat case.

For this, the method of feature space variation is used [7];

3) the traveling salesman problem is solved to find the minimum length of the round of points.

An iterative procedure based on the Kohonen neural network is proposed, with the help of which the points move on a circle, which can reduce the time and improve the quality of further clustering;

4) binary clustering is performed based on the method proposed in the article.

Statement of clustering problems. Clustering aims to combine objects into non-overlapping groups (clusters, classes, sets) of "close" objects. The problem of geometric clustering is formulated as follows: split the set of points $X \subset R^p$, $|X| = n$, into k $(k > 1)$ − subsets (clusters, classes) $C_1, C_2, \dots, C_k$, $C_i \cap C_j = \emptyset \ \forall i, j, \ i \neq j \quad X = \bigcup_{i=1}^{k} C_i$ so that the following condition is satisfied:

$$H = \sum_{i=1}^{k} \sum_{x \in C_i} d(a_i, x) \rightarrow \min, \qquad (1)$$

where $a_i \in R^p$ is the core of the cluster C_i, $d(a_i, x)$ is the distance between the core a_i and any point x from the set *X*.

One of the well-known methods for solving the problem — the method of dynamic kernels uses the squared Euclidean distance as a measure

$$D_E(a_i, x) = \|x - a_i\|^2 \quad (2)$$

where $\sum_{x \in C_i} \|x - a_i\|^2$ defines the intraclass distance for the cluster C_i.

The method is as follows. First, the initial values of the kernels are set. Then perform the following steps:

- division into classes C_i with fixed values of kernels:

$$C_i = \{x: D_E(a_i, x) \leq D_E(a_j, x)\} \quad (3)$$

-optimization of kernel values for a fixed division into classes:

$$\sum_{x \in C_i} d(a_i, x) \rightarrow \min \quad (4)$$

Procedure (3)-(4) stops if, after the next execution of the division into classes, the composition of any class has not changed. Characteristics of clusters are shown in Fig. 3.

To study the resulting division of objects into homogeneous groups, the following mathematical characteristics of clusters are used:

cluster center – the mean locus of points in the space of variables;

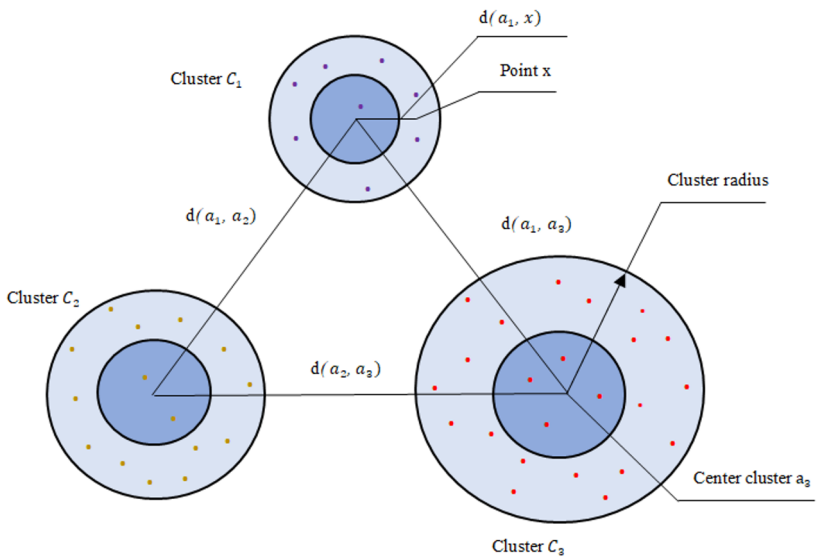


Fig. 3. Geometric interpretation of clusters

-cluster dispersion - a measure of the scattering of points in space relative to the center of the cluster;

-standard deviation (SD) of objects relative to the center of the cluster;

-cluster radius - the maximum distance of points from the cluster's center.

Methods for choosing the initial number of clusters. The choice of methods for determining the initial number of clusters is very limited. The number of classes is often set

by experts based on expediency, considering the characteristics of the subject area. If no such considerations exist, then the number of classes is determined experimentally. The most common method is based on the sequential pairing of the closest objects of a limited sample with averaging their characteristics [2]. It allows you to automate the choice of the number of clusters, and the expert user makes the final decision.

The method of successive reduction in the number of clusters. The principle of constructing the method, also called the "annealing" method [14], is that, based on the class quality criterion, a decision is made to remove or merge this class with another. The algorithm is an analog of the DEL scheme designed to select the optimal set of features in recognition problems. Note that in some cases, on the contrary, it is expedient to split the class into two. First, there are m classes (clusters), which must be reduced to n. Then the number of tested systems of clusters or features will be:

$$N = m + (m+1) + (m-2) + \cdots + (n+1) = \sum_{i=1}^{m-n} (m-i).$$

The value of N, in this case, turns out to be significantly less than C_m^n. The process stops when a minimum-sized feature system with acceptable quality is obtained.

The following are used as class quality criteria:

1) Quantitative criterion. A class with fewer than N instances (points in space features) is considered empty and should be deleted. The expert chooses the threshold depending on the problem's meaning and the proximity measure type.

2) Criterion of spherical separability. Two classes are considered spherically separable if the sum of the radii of the two classes is less than the distance between the nuclei (centers) of these classes. If the classes are spherically inseparable, then they merge into one. The decision to split classes into two can serve as an extension of this method. In this case, the uniformity criterion is used.

3) The average measure of proximity of class points from

the nucleus must be at least half or one-third of the maximum measure of the proximity of points to the nucleus

(class radius). If not, the class is split into two (another kernel is generated near the original one).

Method of sequential increase in the number of clusters. An alternative is the successively increasing the number of clusters [8]. The idea of the method is to start with a small number of classes and increase it until a "good" classification is obtained.

The method is analogous to the general scheme of the ADD algorithm [12].

Such a concept as a frequently reproduced class is used as a quality criterion. A sufficiently large series of classifications are carried out with a different initial choice of classes. Classes that arise in various classifications are defined. The frequencies of occurrence of such classes are considered. The criterion for obtaining the "true" number of classes can be a reduction in the number of frequently repeated classes. Those. with the number of classes k, the number of frequently repeated classes is noticeably less than with the number of classes $(k-1)$ and $(k+1)$.

You should start with two classes. The complexity of the ADD method is approximately the same as that of the DEL algorithm. The method works well with a small number of classes. When enough a large number of classes and a large volume of instances need to be divided into classes, such a selection procedure becomes too slow. Both described algorithms are extremely difficult to implement due to the complexity of assessing the real quality of the solutions obtained. Therefore, in practice, it is better to use simpler and more accessible means of automated selection of the number of clusters.

The problem and models of the initial arrangement of clusters. A universal way to set the initial position of the cores is to set the initial partition of the representative part of objects into classes. In this case, not all objects can participate in the initial partition. Further, in solving problem (1), the initial values of the clusters are obtained. However, other ways or models of initial data representation may be interesting.

Consider some models for choosing the initial location of clusters. As a rule, the effectiveness of the location can only be verified experimentally as a result of solving specific problems and assessing them by an expert.

Random Clustering Model. With an automated approach, clusters can be set by the user based on the task's requirements. The main scheme in the automatic approach is the selection of random objects as clusters. In this case, the cores (centers, centroids) of the clusters are points in the same space as the objects. With a random assignment of kernels, the algorithm is sensitive to possible outliers, and the results obtained for the same sample of documents will differ.

Randomly selecting instances is not the best solution and can lead to wasted time in subsequent steps.

Model of placement of clusters on the boundary of the p-ball. The distribution of points on the m-sphere can serve as a geometric model for the placement of clusters. The following problem is considered. Arrange n points in a p -ball of radius r so that the sum of the distances between them is maximum [9]. The solution was obtained for the plane case and a particular three-dimensional case [10]. It is proved that the solution reduces the problem of placing n points on the surface of a ball. They should be located on a sphere to maximize the total distance between clusters [11]. This conclusion was used in the proposed scheme for solving the binary clustering problem.

Estimates of the required number of clusters in the clustering problem. Let's estimate the number of clusters based, for example, on the minimum time for solving the problem. Let n - be the number of points, m be the number of clusters ($n > m$), then $\left(\frac{n}{m}\right)$ is the average number of points per cluster.

The time complexity of the clustering algorithm is estimated in different sources as $O(m^2)$ or $O(m^3)$. We write the analytical dependence of the average duration of the solution of the problem with cubic complexity in the form [12]:

$$(m) = am^3 + am(\frac{n}{m} + 1)^3 = am^3 + a\frac{(n+m)^3}{m^2} \qquad (5)$$

Solving the optimization problem, we get:

$$3am^2 + a\frac{3(n+m)^2m^2 - 2m(n+m)^3}{m^4} = 0,$$

$$3m^5 + m^3 - 3n^2m - 2n^3 \qquad (6)$$

The results of the numerical solution of equation (6) are contained in Table. 2 (results rounded to whole numbers).

Table 2. Results of the numerical solution of the equation

n	10	50	100	200	300	400	500
m	4	10	15	22	29	34	39

Binary clustering. A solution to the problem of binary clustering of points located on a circle is proposed. Let there be n points (objects) ordered on the boundary of the circle L.

Without loss of generality, we divide the set of points located on a circle into two subsets (clusters) C_1 and C_2 so that the sum of intraclass distances is greater than the interclass distance. To do this, perform the following steps.

We number the points on the circle from 1 to n. Let a pair (l, k) characterize a tuple formed by a sequence of points, where $l(l = 1, \dots, n)$ – is the number of the initial point of the cluster, $k(2 \le k \le n - 2)$ –is the number of cluster points located on the circle.

For each l, we construct a discrete function $F(l, k)$, which characterizes the partitioning quality, depending on the number of elements in the cluster k. Best local solution:

$$F(l, k) = \frac{\text{distance between classes}}{\text{sum of intra-class distances}} \rightarrow_{k}^{\max} \quad (7)$$

The optimal solution corresponds to the global maximum among all $F(l, k) \rightarrow_{l,k}^{\max}$ a distance, we will use the well-established Euclid-Mahalanobis metric d_{E-M}, which takes into account the statistical properties of clusters.

Let us define the components of the quality function intra-class distance. Counting for each class

$C_1 = C_1(l, k)$ and $C_2 = C_2(l, k)$ covariance matrices $S_1 = S_2(l, k)$ and $S_2 = S_2(l, k)$, as well as intraclass distances as the sum of distances from each point of class $x \in C_1$ to the center of the corresponding class $\bar{x}_i$:

$$\sum_{x \in C_1} d_{E-M}(x, C_1) = \sum_{x \in C_1} \sqrt{(x - \bar{x}_1)^T S_1^{-1} (x - \bar{x}_1)},$$

$$\sum_{x \in C_2} d_{E-M}(x, C_2) = \sum_{x \in C_2} \sqrt{(x - \bar{x}_2)^T S_2^{-1} (x - \bar{x}_2)}.$$

Distance between classes. For each option (l, k), we calculate the interclass distance, i.e., the distance between classes C_1 and C_2 To do this, we build the combined covariance matrix according to the formula:

$$S_{1,2} = \frac{1}{n - 2}(S_1 + S_2).$$

The distance between classes is calculated as the distance between the centers of the classes with the merged matrix:

$$d_{E-M}(C_1, C_2) = \sqrt{(\bar{x}_1 - \bar{x}_2)^T S_{1,2}^{-1} (\bar{x}_1 - \bar{x}_2)}.$$

Quality control:

$$F(l, k) = \frac{d_{E-M}(C_1, C_2)}{\sum_{x \in C_1} d_{E-M}(x, C_1) + \sum_{x \in C_2} d_{E-M}(x, C_2)} \rightarrow_{l,k}^{\max}.$$

The procedure has time complexity (by the number of iterations $O(n(n - 3)/2)$.

An example of solving the problem of binary clustering. Let us consider an example of a joint application of the feature space variation method [7] and the method based on the network model. 1) Initial Table of precedents (Table 3).

Table 3. Source cases

	x_1	x_2	x_3	x_4	x_5	x_6	Class
W_1^1	1	2	0	1	-1	0	1
W_2^1	-1	0	-1	-2	-2	1	1
W_3^1	1	-1	2	-1	-1	0	1
W_1^2	0	2	1	-1	0	1	2
W_2^2	2	-1	0	-1	-1	0	2

2) Let's construct the covariance matrix C.

$$C = \begin{pmatrix} 1.3 & -0.55 & 0.45 & 0.60 & 0.25 & -0.55 \\ -0.55 & 2.30 & -0.20 & 0.90 & 0.50 & 0.30 \\ 0.45 & -0.20 & 1.30 & 0.15 & 0.50 & -0.20 \\ 0.60 & 0.90 & 0.15 & 1.20 & 0.25 & -0.35 \\ 0.25 & 0.50 & 0.50 & 0.25 & 0.50 & 0.00 \\ -0.55 & 0.30 & -0.20 & -0.35 & 0.00 & 0.30 \end{pmatrix} \quad (8)$$

3) Find the eigenvectors and eigenvalues for the covariance matrix (8) by solving the determinant (Table 4):

$$\begin{vmatrix} c_{11} & -\lambda_1 & c_{12} & c_{13} & c_{14} & c_{15} & c_{16} \\ c_{21} & c_{22} & -\lambda_2 & c_{23} & c_{24} & c_{25} & c_{26} \\ c_{31} & c_{32} & c_{33} & -\lambda_3 & c_{34} & c_{35} & c_{36} \\ c_{41} & c_{42} & c_{43} & c_{44} & -\lambda_4 & c_{45} & c_{46} \\ c_{51} & c_{52} & c_{53} & c_{54} & c_{55} & -\lambda_5 & c_{56} \\ c_{51} & c_{52} & c_{53} & c_{54} & c_{55} & c_{56} & -\lambda_6 \end{vmatrix} = 0 \quad (9)$$

Table 4. Vector of eigenvalues

λ_1	λ_2	λ_3	λ_4	λ_5	λ_6
0	2.96	1.18	2.45	0	0.29

4) The resulting Table,5 vectors in the feature system $Y = (y_1, \dots, y_n)$.

Table 5. Vectors in the new coordinate system

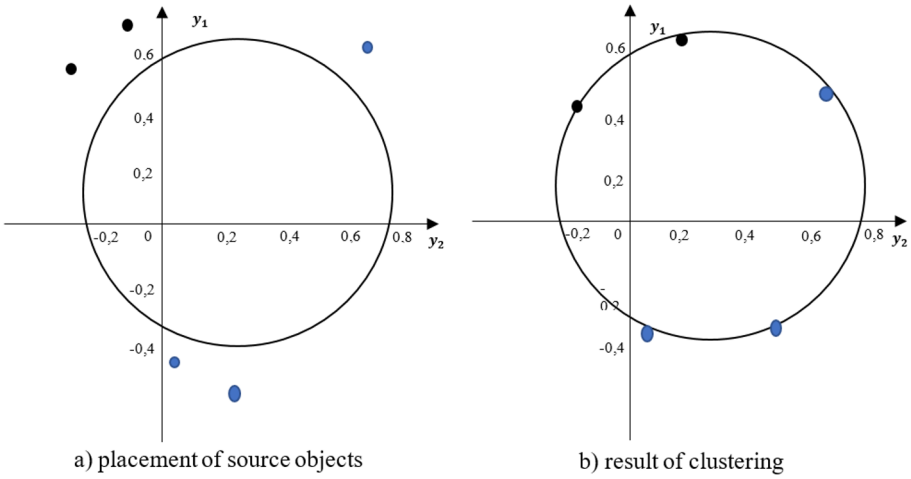
Vectors in the new system.	y_1	y_2	y_3	y_4	y_5	y_6	Class
V_1^1	-0.33	0.27	-0.01	-0.43	-0.02	-0.85	1
V_2^1	-0.16	0.87	-0.06	0.39	0.18	0.08	1
V_3^1	-0.32	0.07	0.75	-0.38	0.34	0.2	1
V_1^2	0.64	-0.06	0.47	0.45	0.24	-0.29	2
V_2^2	-0.45	-0.36	-0.22	0.31	0.66	-0.26	2

For the further course of the research, it is important for us to use the largest eigenvalues 2.96, $\lambda_4 = 2.45$ and the eigenvectors corresponding to them. After the dimension reduction, the original objects were placed on the plane, as shown in Fig. 4a).

The same figure shows neurons evenly spaced on a circle. Next, the clustering problem is solved based on the network model. At the same time, a network model (based on the Kohonen neural network) is necessary for the iterative "pulling" of objects to a circle. After that, the simplified problem of binary clustering (7) is solved. As can be seen from Fig.4, b), as a result, the objects are located on the circle as experts in the original Table classified them. Thus, the joint application of methods allows us to solve the problem of binary clustering.

4 Conclusion

It is expedient to take the tabular method as the basis for representing the semi-structured initial data of the clustering problem, which is widely used in solving classification problems by the algebraic method, neural networks, decision trees, and constructing the reference set of clusters. The considered methods for choosing the initial number of clusters and the model of their placement allow us to solve the problem of binary clustering. Moreover, the solution to the problem of placing points on a circle based on a network model is considered a step preceding clustering.



a) placement of source objects b) result of clustering

Fig. 4. An example of solving the clustering problem

The expediency of using different cluster analysis methods together is shown in the example of solving the problem of binary clustering. The binary cluster method was the most optimal for the study object. Based on the experiments performed, it is applied to the data array included in the management system.

References

1. Askaraliyev O.U., Zaynutdinova M.B. "Development of the structure of the intelligent data processing system (on the example of the Integrated Management System): Bulletin of TUIT: Management and Communication Technologies. 2020 year.
2. Askaraliyev O.U., Sharipov Sh.O., Akbarova N.R.: "Implementation of Decision Support Procedures using an Expert System in Integrated Management (On the Field of Tax Authorities)". Design Engineering: Y 2021 Issue 9, -P 4048-406,
3. Averkin A.N., Gaaze-Rapoport M.G., Pospelov D.A. Explanatory Dictionary of Artificial Intelligence. M .: Radio and communication, 1992.
4. Bolotova L.S. Artificial intelligence systems: models and technologies based on knowledge: textbook. / FGBOU VPO RGUITP; FGAU GNII ITT "Informika". Moscow: Finance and Statistics, 2012.
5. Osipov G. Strategies for Stabilization Behaviour of Intelligent Dynamic Systems. – Proc. of 20th European Meeting on Cybernetics and Systems, Vienna, 2010, pp.195-197.

6. Larichev OI, Mechitov A.I., Moshkovich E.M., Furems E.M. Systems for identifying expert knowledge in classification problems // Izv. USSR Academy of Sciences. - Ser. Technical cybernetics. - 2005. - No. 2.
7. Larichev OI, Moshkovich E.M. Qualitative decision making methods. - M .: Science. Fizmatlit. - 1996.
8. Simon H.A. The new science of managerment decision. Englewood Cliffs. N.J., Prentice-Hall Inc., 1998.
9. Levykin VM, Stopchenko GI, Aydarov AV Formation of models of weakly structured problems in decision-making systems // ACS and automation devices, 1998. No. 108. - P. 155–159.
10. J.B.Elov, U.R.Khamdamov, Dj.B.Sultanov, O.Q.Makhmanov, "Organizing functional processes of information system for the advanced training of medical personnel on the basis of IDEF methodology", International Journal of Advanced Research in Science, Engineering and Technology, vol. 6, Issue 12, December 2019. India. pp. 12085-12090.
11. Gavrilova T.A., Khoroshevsky V.F. Knowledge base of intelligent systems. SPb .: Peter, 2000.
12. Luger D. Artificial intelligence: strategies and methods for solving complex problems. 4th ed. M .: Williams, 2003.
13. Nechaev V.V., Koshkarev M.I. Intelligent Problem Solvers: Comparative Analysis and Architectural Model // Information and Telecommunication Technologies. 2014. No. 21, pp. 51–61.
14. Nechaev V.V. Trofimenko V.M. Analysis of methods of semantic search for information resources // Educational resources and technologies. 2014. No. 5. P. 127–135.
15. Parsaye K. A Characterization of Data Mining Technologies and Processes. The Journal of Data Warehousing, 1998, vol. 1, pp. 11-28.