

*Розглядається задача формування підгруп за ступенем ризику інфікування вірусом імунодефіциту у групі трудових мігрантів, як задача кластеризації. Побудовано математичну модель задачі кластеризації та виконано адаптацію методу еволюційної кластеризації об'єктів до визначення груп трудових мігрантів. Здійснено розбиття групи трудових мігрантів на кластери відповідно до ступеня ризику інфікування вірусом імунодефіциту*

*Ключові слова: еволюційна кластеризація, група трудових мігрантів, ризик інфікування вірусом імунодефіциту людини*

*Рассматривается задача формирования подгрупп по степени риска инфицирования вирусом иммунодефицита в группе трудовых мигрантов, как задача кластеризации. Построена математическая модель задачи кластеризации и выполнена адаптация метода эволюционной кластеризации объектов к определению групп трудовых мигрантов. Осуществлено разбиение группы трудовых мигрантов на кластеры в соответствии со степенью риска инфицирования вирусом иммунодефицита*

*Ключевые слова: эволюционная кластеризация, группа трудовых мигрантов, риск инфицирования вирусом иммунодефицита человека*

UDC 004.9:614.1

DOI: 10.15587/1729-4061.2016.71203

# FORMING THE CLUSTERS OF LABOUR MIGRANTS BY THE DEGREE OF RISK OF HIV INFECTION

**O. Mulesa**

PhD, Associate Professor

Department of cybernetics and applied mathematics\*

E-mail: mulesa.oksana@gmail.com

**V. Snytyuk**

Doctor of technical sciences,

Professor, Head of Department

Department of intellectual information systems

Taras Shevchenko National University of Kyiv

Volodymyrska str., 81, Kyiv, Ukraine, 03022

E-mail: snytyuk@gmail.com

**I. Myronyuk**

Candidate of Medical Sciences,

Associate Professor\*

Chief doctor

Trans-Carpathian center of AIDS prevention and fighting

Druhetiv str., 72, Uzhhorod, Ukraine, 88000

E-mail: ism69@ukrpost.ua

\*Uzhhorod National University

Narodna sq., 3, Uzhhorod, Ukraine, 88000

## 1. Introduction

Under conditions of concentrated stage of HIV-infection/AIDS epidemic that is taking place in Ukraine now [1], when the epidemic spreads mainly among specific groups of population, vulnerable to the HIV infection [2], the main measures against epidemic are focused exactly on the representatives of these groups. Given the fact that the commonly-recognized groups with high risk (GHR) of HIV infection are very diverse in their structure [3], and also considering the limited resource base, it is argued that there is a need to develop targeted preventive intervention and influences on their separate subgroups, maximally homogeneous by socio-demographic characteristics, based on a client-centered approach. Since it is based on the implementation of measures that are attractive and comfortable for representatives of the target groups [4], an important condition for its implementation is the study of the main socio-demographic characteristics of such subgroups of GHR and determination of the level of dissemination of risky behavior in each subgroup in terms of HIV infection. This approach will allow determining the priority of the need to implement prevention programs for different subgroups of GHR and designing action program that is the most convenient and close to the representatives of GHR.

One of the leading GHR of HIV infection in Zakarpattia Oblast is external and internal labor migrants [5]. This group is quite numerous and diverse in its composition, therefore, the formation of separate subgroups, homogeneous by socio-demographic characteristics and defined degree of risk of HIV infection for each of these subgroups in this GHR, is necessary and relevant.

The task of dividing the GHR in subgroups according to socio-demographic portrait of their representatives can be represented mathematically as the problem of clustering [6].

## 2. Analysis of scientific literature and the problem statement

The task of clustering is to divide the main set on the subsets of objects, similar to each other by certain criteria [7, 8]. There are several methods of solving clustering problems. Analysis of modern scientific publications devoted to the models and methods of clustering objects showed how actual the development of methods of clustering to solve real-world problems is. Thus, the papers [9, 10] describe statistical methods of clustering, based on the theory of probabilistic distributions and are used, for example, in the adaptive systems that learn. The paper [11] presents neural network

method of solving the problems of clustering and proves its efficiency. Theoretical bases of neural network methods are contained in [12]. Such methods require considerable experience from researchers when setting up the values of the main parameters, which makes it difficult to apply when solving practical tasks. The methods of group of tree clustering [13] consist in the construction of clustering tree. The methods of fuzzy clustering, among which the main method is a method of fuzzy c-means [14] that is based on application of the apparatus of fuzzy sets, allow simultaneous assigning of the same item to different clusters with different degrees of membership. Adaptation of the method of fuzzy c-means, which is provided in [15], allows dividing objects into clusters with regard to degrees of the impact of the characteristics on the object belonging in a cluster.

A separate group includes heuristic clustering methods. Thus, the paper [16] presented a heuristic method of clustering, based on the analysis of matrices of differences. The paper [17] is devoted to the design of hybrid heuristics for solving the problems of clustering. The problem of clustering of time series and methods of its solution, based on the evaluation of degree of uncertainty of parameters of autoregressive models, are given in [18]. Methods of automatic clustering, based on the method of supporting vectors, are the subject of the papers [19, 20].

The accuracy of solution of the clustering problem depends on the distribution of the objects in the field of research, the number of necessary iterations for the formation and clarification of clusters, other problems associated with the peculiarities of the application of different methods.

Given the importance and urgency of the problem of determining the risks of HIV infection by certain groups of the population, it is expedient to design such models and methods of clustering, which would allow its efficient solving.

### 3. The purpose and objectives of the study

The purpose of the research is to increase the efficiency of decision-making processes when determining clusters of migrant workers by the feature of infection with the human immunodeficiency virus through the development of models and methods of evolutionary clustering.

According to the stated goal, the following tasks were set:

- to build a mathematical model of the problem of clustering and perform the adaptation of the method of evolutionary clustering of objects to define the groups of migrant workers;
- to perform experimental validation of the designed method for the task of defining the groups of people by socio-demographic characteristics in GHR – migrant workers;
- to carry out a comparative analysis of the results of solving the problem of clustering of the given GHR by different methods of clustering.

### 4. Adaptation of method of evolutionary clustering to the formation of groups of migrant workers

Let us consider the task of clustering objects in this formulation. Let there is a set of objects  $O = \{O_1, O_2, \dots, O_n\}$ , each of which is characterized by a set of features  $X = \{X_1, X_2, \dots, X_m\}$ . Let for each object  $O_i \in O$  ( $i = \overline{1, n}$ ) the vector of values of characteristics is set  $x^{(i)} = (x_{i1}, x_{i2}, \dots, x_{im})$ ,

where  $x_{ij}$  is the value of the feature  $X_j \in X$  for the corresponding object. It is necessary to build the rule  $s = D(x^{(i)})$  that fulfills the reflection of the set  $X^{(M)}$  of possible values of characteristics  $\{X_1, X_2, \dots, X_m\}$  on the set  $\{1, 2, \dots, K\}$  of numbers of clusters, i. e.

$$s : X^M \rightarrow \{1, 2, \dots, K\}. \tag{1}$$

In other words, the problem of clustering is to define the indicator variable  $s$  in the following way:

$$s = \begin{cases} 1, & \text{if } x^{(i)} \in \Omega_1, \\ \dots & \\ K, & \text{if } x^{(i)} \in \Omega_K, \end{cases} \tag{2}$$

where  $\Omega_k, (k = \overline{1, K})$  are the regions of space  $X^{(M)}$  that match the clusters.

For the solution of the problem we will apply the method of evolutionary clustering EvoClast [8], based on the genetic algorithm, which classically includes the following stages:

1. Determining the population of individuals  $\Theta$ , who are potential solutions to the problem of optimization of objective function.
2. The implementation of the previous steps of the algorithm, which consist in determining the number of elements  $l$  of the initial population  $\Theta$ , where  $l \ll |\Theta|$ ; selection of rule of normalization of the input data; the choice of ways of recombination, mutation and inversion and the corresponding probabilities.
3. For each element  $\theta_i \in \Theta, i = \overline{1, l}$  we compute the values of the objective function  $f_i = F(\theta_i)$ .
4. With probabilities  $p_i^l$ , proportionate to the values  $f_i$ , we select two individuals and carry out recombination to receive two new individuals.
5. With probability  $\frac{1}{2}$  we choose one of the received individuals and with probability  $P^m$  we implement the mutation.
6. Repeat steps 3–5  $\left\lceil \frac{1}{2} \right\rceil$  times.
7. We form a new composition of the population  $\Theta$ , delete individuals with worse values of the objective function.

We propose an algorithm for constructing the objective function to solve the problem, formulated above.

Let us assume that all of the components of vectors of characteristics of objects  $x^{(i)} = (x_{i1}, x_{i2}, \dots, x_{im})$  are numeric. Then, on the initial stage we perform normalization of the data by the formula

$$x_{ij} := \frac{x_{ij} - x_{j\min}}{x_{j\max} - x_{j\min}}. \tag{3}$$

After the conversion, the values of all vectors of the data will be allocated to the single hypercube  $[0; 1]^m$ .

Set the value  $K$  – the number of clusters. Define  $T = m \cdot K$  – the number of elements in the vector-individual of the population. Then, the objective function  $F$  will depend on  $T$  variables  $F = F(y_{11}, y_{12}, \dots, y_{1m}, y_{21}, y_{22}, \dots, y_{2m}, \dots, y_{Km})$ , where  $(y_{k1}, y_{k2}, \dots, y_{km})$  is the center of the  $k$ -th cluster on the corresponding iteration ( $k = \overline{1, K}$ ).

Then we successively perform the following steps:

*Step 1.* Set the initial value of the fitness function  $F := 0$ .

*Step 2.* For each element of the sample set  $O$  we successively perform the following steps. Let  $i = 1$ .

Step 3. We calculate distances from the object  $O_i$  to each of the  $K$  clusters by the formula:

$$d_k = \sqrt{\sum_{j=1}^m (x_{ij} - y_{kj})^2} \quad (4)$$

Step 4. Attribute the  $i$ -th object to the  $q$ -th cluster, where

$$q = \arg \min_{k=1, K} d_k.$$

Step 5. Adjust the value of the objective function by the rule:  $F := F + d_q$ . Proceed to the next object:  $i := i + 1$ .

Step 6. If all the items from the set  $O$  have been considered, that is  $i = n + 1$ , then the process of calculating the value of the function  $F$  is completed, otherwise, proceed to step 3.

The expression to calculate the objective function can be written down in the following way:

$$F(y_{11}, \dots, y_{Km}) = \sum_{i=1}^n \left( \min_{k=1, K} \left\{ \sqrt{\sum_{j=1}^m (x_{ij} - y_{kj})^2} \right\} \right) \quad (5)$$

Then the problem of clustering is to find the minimum value of the objective function (5).

### 5. Experimental verification of the method of evolutionary clustering for the problem of defining subgroups of GHR – migrant workers

To perform the experimental verification of the modified method of evolutionary clustering, the problem of defining subgroups of GHR – migrant workers was studied. Initial data in the set task are the results of a previously conducted sociological survey among labour migrants – residents of Zakarpatska Region of Ukraine, in which 561 representatives of the target group were polled.

Analysis of the submitted questionnaires was conducted by the following areas: socio-demographic characteristics of survey respondents (gender, age, residence location, education and marital status); the direction of labour migration (external labour migration – European countries and, separately, Russian Federation; internal labour migration – other regions of Ukraine); behavioral peculiarities (the practice of HIV infection risky behavior).

Mathematically, the problem of selecting sub-groups of GHR was presented as a problem of clustering. To perform clustering, we used  $n=561$  questionnaires of the persons – migrant workers. The following socio-demographic characteristics were selected as the basis: gender of a person, age category, residence location, education, marital status and direction of labour migration ( $m=6$ ). A snippet of the data is in Table 1.

Each value of a qualitative characteristic was assigned with a numeric value according to Table 2.

Table 2

Table of accordance of characteristics of socio-demographic portrait to their numeric equivalents

Socio-demographic characteristic	Value of a characteristic	Numeric equivalent
Gender	female	0
	male	1
Age category	1 (18–24)	0
	2 (25–34)	1
	3 (35–44)	2
	4 (45 and older)	3
Place of residence	rural	0
	urban	1
Education	incomplete secondary	0
	full secondary	1
	secondary vocational	
	incomplete higher higher	2
Marital status	never married	0
	married/civil marriage	1
	divorced/widow	2
Direction of migration	West	0
	Ukraine	1
	Russia	2

After the transfer of quality characteristics into numerical equivalents and fulfillment of normalization by the formula (3), we carried out clustering by the algorithm of clusterization of  $k$ -means and by evolutionary algorithm for varying numbers of clusters ( $K=3,8$ ), described in the study. For each break-down, we calculated the value of the objective function  $F$  by the formula (5). Results of the calculations are listed in Table 3 and in Fig. 1.

Table 1

Snippet of the socio-demographic characteristics of persons – migrant workers

№	Socio-demographic characteristics					
	Gender of person	Age category of person	Place of residence of person	Education of person	Marital status of person	The direction of migration of person
...	...	...	...	...	...	...
10	male	3 (35–44)	rural	secondary vocational	married/civil marriage	Ukraine
11	male	2 (25–34)	rural	incomplete secondary	never married	Ukraine
12	male	2 (25–34)	rural	full secondary	married/civil marriage	West
13	female	4 (45 and older)	rural	secondary vocational	married/civil marriage	West
14	female	2 (25–34)	rural	secondary vocational	never married	Ukraine
15	male	4 (45 and older)	urban	full secondary	divorced / widow	West
16	male	2 (25–34)	rural	incomplete secondary	married/civil marriage	West
17	male	2 (25–34)	rural	full secondary	married/civil marriage	Ukraine
18	male	2 (25–34)	rural	secondary vocational	married/civil marriage	Ukraine
...	...	...	...	...	...	...

Table 3

Values of the objective function

The number of clusters (K)	Method of k-means (F <sub>1</sub> )	Method of evolutionary clustering (F <sub>2</sub> )
3	380,6709	361,5065
4	322,4295	321,5448
5	311,1273	305,5683
6	295,9276	294,6649
7	290,8294	282,5672
8	276,0774	275,7684

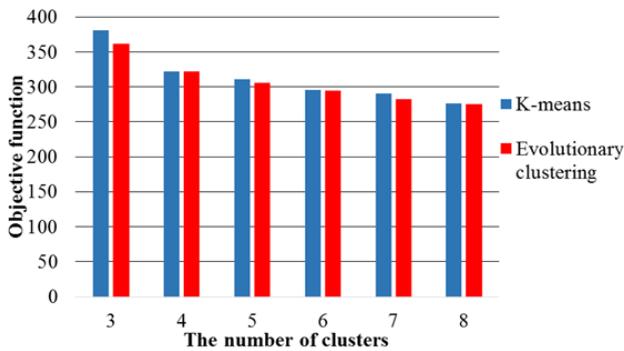


Fig. 1. Diagram of values of the objective function for the various methods of clustering

According to the data from Table 3 and Fig. 1, the values of the objective function obtained by evolutionary clustering are lower than those when applying the method of k-means, indicating the expedience of solving a clustering task of the group of persons – labour migrants, by the method of evolutionary clustering.

With the number of clusters K=6, the following results were obtained by the method of evolutionary clustering, which are in Table 4.

Table 4

Results of clustering (centres of clusters)

№ of cluster	Y <sub>k1</sub>	Y <sub>k2</sub>	Y <sub>k3</sub>	Y <sub>k4</sub>	Y <sub>k5</sub>	Y <sub>k6</sub>
1	0,975	0,481	0,99994	0,773	0,499	0,266
2	0,0013	0,438	0,027	0,546	0,537	0,2219
3	0,99997	0,507	0,0053	0,494	0,473	0,2004
4	0,99984	0,755	0,00703	0,4708	0,5297	0,965
5	0,998	0,254	0,00378	0,478	0,236	0,902
6	0,959	0,452	0,9998	0,556	0,491	0,976

Clusters by the socio-demographic characteristics

Cluster (the number of representatives)	Socio-demographic characteristics					
	Gender	Age category	Place of residence	Education	Marital status	Direction of migration
1 (n=63)	male	25–44	urban	secondary/higher	married/civil marriage	West/Ukraine
2 (n=83)	female	24–34	rural	secondary	married/civil marriage	West/Ukraine
3 (n=168)	male	25–44	rural	secondary	married/civil marriage	West/Ukraine
4 (n=22)	male	35 and older	rural	secondary	married/civil marriage	Russia
5 (n=190)	male	25–34	rural	secondary	never married/ married/civil marriage	Russia
6 (n=35)	male	25–34	urban	secondary	married/civil marriage	Russia

Based on the obtained results of the centres of clusters, according to Table 2, it is possible to construct a socio-demographic portrait of the typical representative-cluster by its characteristics. Division into clusters by socio-demographic characteristics of the representatives of the target group of the study are presented in Table 5.

For further analysis and interpretation of the results of clustering, a system of characteristics  $\{u_1, u_2, u_3, u_4, u_5, u_6, u_7\}$ , was formed from the behavioral questionnaires of persons – migrant workers, which define the practice of HIV infection risky behavior of these persons (Table 6).

Table 6

Descriptive characteristics of features that define the practice of HIV infection risky behavior of migrant workers

Feature	Description of a feature as the practice of HIV infection risky behavior
u <sub>1</sub>	Existence of casual sexual partners over the last 12 months
u <sub>2</sub>	Sexual relations with casual sex partners without the use of condoms in the region of work destination
u <sub>3</sub>	Presence of past infections that are transmitted primarily sexually
u <sub>4</sub>	Sexual relations with a casual sex partner without the use of condoms in the region of permanent residence
u <sub>5</sub>	Absence of casual sex partners
u <sub>6</sub>	Casual sexual relations with sex partners over the last 12 months, but sexual relations only with a condom
u <sub>7</sub>	No sexual relations or all sexual relations only with a condom

Availability of characteristics u<sub>1</sub> – u<sub>4</sub> display high behavioural risk of HIV infection for the person. Similarly, the same is expressed by the absence of features u<sub>5</sub> – u<sub>7</sub>. For each cluster we calculated percentage of its representatives (migrant workers), for whom these characteristics are inherent. The results are in Table 7.

Table 7

Percentage of representatives of clusters, for whom characteristics that define the practice of HIV infection risky behavior are inherent.

No. of cluster	u <sub>1</sub>	u <sub>2</sub>	u <sub>3</sub>	u <sub>4</sub>	u <sub>5</sub>	u <sub>6</sub>	u <sub>7</sub>
1	33	17,4	66,6	31	68,2	9	68
2	13,2	10,8	59	12	87,9	9,6	87,9
3	26,1	9,5	50,5	22	77,9	10,1	77,9
4	40,9	0	59	40,9	59	4,5	59
5	24,7	11,5	61	23,1	76,8	11,5	76,8
6	28	22	65,7	25,7	74,2	5,7	74,2

Table 5

As presented in Table 7, HIV infection risky behavior is practiced by the smallest share of the representatives of the second and third clusters. Thus, only 13.2 % of cluster 2 representatives noted the existence of casual sexual partners (feature u<sub>1</sub>) – this is the lowest indicator among all formed clusters. For the representatives of this cluster also characteristic is a low percentage of the representa-

tives, who display features of risky behavior ( $u_2, u_3, u_4$ ) and a relatively high proportion of representatives, for whom the features of the absence of the risk of HIV infection ( $u_5-u_7$ ) are characteristic. At the same time, the percentage of representatives of clusters 1 and 4, for whom the features of risky behavior ( $u_1-u_4$ ) are characteristic, is much higher.

Analysis of the results allowed dividing all persons – migrant workers, into three groups of clusters in ascending order by the degree of risk, according to the degree of riskiness of their behaviour regarding HIV infection: Group 1 – group of clusters with a relatively low risk of HIV infection, Group 2 – group of clusters with moderate risk, and Group 3 – group of clusters of high risk. As a result of the division, each group of clusters is homogeneous not only by socio-demographic portraits of its representatives, but also by the degree of prevalence of the practice of HIV infection risky behavior. Division of clusters by the socio-demographic characteristics into the groups according to the degree of risk of HIV infection is in Table 8.

Grouping of persons – migrant workers according to the degree of risk of HIV infection

Group of risk	Cluster	Socio-demographic characteristics					
		Gender	Age category	Place of residence	Education	Marital status	Direction of migration
1	2, 3	female	24–34	rural	secondary	married/civil marriage	West
		male	25–44				
2	5, 6	male	25–34	rural	secondary	never married/married/civil marriage	Russia
				urban		married/civil marriage	
3	4, 1	male	35 and older	rural	secondary	married/civil marriage	Russia
			25–44	urban	secondary/higher		Ukraine

As shown in Table 8, a group of clusters with high risk of infection includes representatives of 1 and 4 clusters, the socio-demographic portrait of whom is as follows: men of age 25–44 who have full secondary or higher education and are married. In this case the people with the highest risk among the rural residents of this group are labour migrants who travel to Russia, and among the urban population – to the countries of Western Europe and other regions of Ukraine. Representatives of clusters 2 and 3, which belong in the groups with relatively low risk of HIV infection, are residents of rural areas, married, with a full secondary education. These are young (25–34 years of age) and mature (up to 45 years of age) women and men, who travel to work to the countries of Western Europe.

Experimental verification of the adapted method of evolutionary clustering for the solution of the problem of the formation of clusters of persons – migrant workers proved its efficiency. The advantage of the adapted method of evolutionary clustering for the solution of the set problem is its greater accuracy compared to the method of k-means.

Application of the proposed method allowed conducting more precise clustering of the representatives of a group with a higher risk of HIV infection – migrant workers, by the principal socio-demographic characteristics and forming the groups of clusters based on the degree of risk of HIV

infection of their representatives. This approach allows, when planning preventive measures and interventions in the programs against HIV epidemic, concentrating attention and allocating resources to the most vulnerable subgroups of representatives of GHR, which are homogeneous by their socio-demographic characteristics. This is especially relevant under conditions of limited both financial and human resources in the area of counteraction to HIV epidemic, which is currently observed in Ukraine.

It is advisable to apply the adapted method of evolutionary clustering to the analysis of GHR in other regions of Ukraine, as well as to social groups of different origin in general.

## 6. Conclusions

Solving the problem of determining the risks of HIV infection for certain groups of the population as the clustering problem was proposed. In the course of the study:

Table 8

1. Mathematical model of the problem of clustering as the problem of building up a rule was developed, by which reflection from the set of possible values of characteristics on a set of clusters is carried out and the method of evolutionary clustering of objects was modified, which allowed dividing the objects into clusters, using evolutionary paradigm. The modification of the method was performed by constructing a fitness function in the form of such an objective function, which provides assignment of an object to the cluster, the Euclidean distance from the center of which to the object is the smallest.

2. Experimental verification of the modified method of evolutionary clustering for the problem of defining subgroups according to socio-demographic characteristics in GHR – migrant workers was performed, allowing dividing the representatives of this GHR into clusters according to the degree of risk of HIV infection.

Three groups of clusters were formed in the ascending order by the degree of risk of HIV infection: a group of clusters with high risk, a group of clusters with moderate risk and a group with a relatively low risk of HIV infection. Therefore, the representatives of this GHR, who display socio-demographic features that match the characteristics of clusters 1 and 4 (group of clusters with relatively high risk of HIV infection) potentially have higher level of risk of infection and thus require focused prophylactic measures to prevent infection. The applied approach allows concentrating attention and resources on the most vulnerable category of persons of this GHR, which is the basis of improving the efficiency of combating HIV infection/AIDS among the representatives of this group of population.

3. Comparative analysis of the results of the problem solving of clustering of the set GHR by the method of k-means and by the method of evolutionary clustering was carried out. The comparison was performed by the values of the function F, which is the integral sum of the distances from objects to the centres of those clusters where they belong. Less value of this function corresponds to the best division into clusters. Therefore, according to the fulfilled calculations, the advantages of the evolutionary method have been proven.

## References

1. HIV infection in Ukraine: Information Bulletin number 44 [Text]. – Health of Ukraine, Eng. control center for the socially dangerous diseases. Kyiv, 2015. – 37 p.
2. Harmonised Ukraine report on progress in the implementation of the national response to the AIDS epidemic. Reporting period [Text]. – Ministry of Health of Ukraine, 2012. – 243 p.
3. Grushetsky, A. Monitoring of behavior and HIV prevalence among commercial sex workers as surveillance component of the second generation analytical report on the results of bio-behavioral studies 2011 [Text] / A. Grushetsky // International Alliance HIV/AIDS Alliance in Ukraine, 2012. – 120 p.
4. Bejan, A. B. Standards program correcting deviant behavior of adolescents aged 14–18 years who have experience of substance use [Text] / A. B. Bejan. – Kyiv, 2015. – 180 p.
5. Myronyuk, I. S. Regional high risk of HIV infection in Transcarpathia is workers [Text] / I. S. Myronyuk, G. A. Slabkiy // Ukraine. Health of the Nation. – 2012. – Vol. 3, Issue 23. – P. 201–205.
6. Mulesa, O. Yu. Information technologies of quantitative evaluation of risk groups of human immunodeficiency virus infection [Text] / O. Yu. Mulesa // Eastern-European Journal of Enterprise Technologies. – 2013. – Vol. 5, Issue 4 (65). – P. 10–15. Available at: <http://journals.urau.ua/eejet/article/view/18327/16087>
7. Gorban, A. N. Method of Elastic Maps and its Applications in Data Visualization and Data Modelling [Text] / A. N. Gorban, A. Yu. Zinovyev. // Int. Journal of Computing Anticipatory Systems, CHAOS. – 2002. – Vol. 12. – P. 353–369.
8. Snytyuk, V. Evolutionary clustering of complex objects and processes. Vol. 1 [Text] / V. Snytyuk // XI-th International Conference «Knowledge–Dialogue–Solution», 2005. – P. 232–237.
9. Baturkin, S. A. Statistical data clustering algorithms in adaptive learning systems [Text] / S. A. Baturkin, E. Yu. Baturkina, V. A. Zimenko, Y. V. Sihinov // Vestnyk RHRTU. – 2010. – Vol. 1, Issue 31. – P. 82–85.
10. Fukunaha, K. Introduction to statistical pattern recognition theory [Text] / K. Fukunaha. – Moscow: Nauka. Hlavnaia redaktsiya fizyko–matematycheskoi lyteratury, 1979. – 368 p.
11. Tsekouras, G. E. On training RBF neural networks using input–output fuzzy clustering and particle swarm optimization [Text] / G. E. Tsekouras, J. Tsimikas // Fuzzy Sets and Systems. – 2013. – Vol. 221. – P. 65–89. doi: 10.1016/j.fss.2012.10.004
12. Horbachenko, V. Y. Networks and Kohonen maps [Electronic resource] / V. Y. Horbachenko. – 2013. – Available at: [http://gorbachenko.self-organization.ru/articles/Self-organizing\\_map.pdf](http://gorbachenko.self-organization.ru/articles/Self-organizing_map.pdf)
13. Mundhada, M. R. Record Linkage in Various Types of Clustering Tree By using K–means Clustering Algorithm [Text] / M. R. Mundhada, A. R. Nimje // National Conference “CONVERGENCE”, 2015. – P. 28.
14. Smeshko, Yu. V. On a criterion to select the exponential weight classification algorithm of fuzzy C–means [Text] / Yu. V. Smeshko. – Molodezh y nauk, Krasnoyarsk, 2012. – P. 19–25.
15. Mulesa, O. Adaptation of fuzzy c-means method for determination the structure of social groups [Text] / O. Mulesa // Technology Audit and production reserves. – 2015. – Vol. 2, Issue 2 (22). – P. 73–76. doi: 10.15587/2312-8372.2015.41014
16. Santi, É. A model for clustering data from heterogeneous dissimilarities [Text] / É. Santi, D. Aloise, S. J. Blanchard // European Journal of Operational Research. – 2016. – Vol. 253, Issue 3. – P. 659–672. doi: 10.1016/j.ejor.2016.03.033
17. Oliveira, R. M. Hybrid heuristics based on column generation with path-relinking for clustering problems [Text] / R. M. Oliveira, L. A. Lorena, A. A. Chaves, G. R. Mauri // Expert Systems with Applications. – 2014. – Vol. 41, Issue 11. – P. 5277–5284. doi: 10.1016/j.eswa.2014.03.008
18. Bakoben, M. Improving clustering performance by incorporating uncertainty [Text] / M. Bakoben, A. Bellotti, N. Adams // Pattern Recognition Letters. – 2016. – Vol. 77. – P. 28–34. doi: 10.1016/j.patrec.2016.03.004
19. Wu, C.-H. Automated clustering by support vector machines with a local-search strategy and its application to image segmentation [Text] / C.-H. Wu, C.-C. Lai, C.-Y. Chen, Y.-H. Chen // Optik – International Journal for Light and Electron Optics. – 2015. – Vol. 126, Issue 24. – P. 4964–4970. doi: 10.1016/j.ijleo.2015.09.143
20. Carrizosa, E. Clustering categories in support vector machines [Text] / E. Carrizosa, A. Nogales-G mez, D. R. Morales. – Omega, 2016. doi: 10.1016/j.omega.2016.01.008