MATHEMATICAL MODELING OF PROCESSES AND SYSTEMS

UDC 004.896 (045)
DOI:10.18372/1990-5548.54.12333

[1]V. Y. Snytyuk,
[2]O. O. Suprun

EVOLUTIONARY CLUSTERING AS TECHNIQUE OF ECONOMICS PROBLEMS SOLVING

[1]Intellectual and Information Systems Department,
Taras Shevchenko National University of Kyiv, Kyiv, Ukraine
[2]Institute of Mathematical Machines and Systems Problems, NASU, Kyiv, Ukraine
E-mails: [1]snytyuk@gmail.com, [2]oleh.o.suprun@gmail.com

***Abstract*—*The article presents the method, developed to use evolutionary technologies for clustering large amount of objects that are specified by their characteristics values. The need to analyze big data and to extract the necessary data from multidimensional databases makes the classic methods ineffective, or they require a lot of recourses or time to give an appropriate solution to the stated practical problem. Such problems very often appear in economical and financial spheres, where an expert has to make right decisions, based on various information from different sources, this information may have noise effects, or even be unreliable. Solving these problems requires gathering and formalization of available information that can take a lot of time. The presented method allows to use evolutionary technologies, such as genetic algorithms and evolution strategies elements to solve clustering problems with minimal constraints on the initial data – the situation that represents real practical problems. The experimental results of using the method are given, which proof the effectiveness of the proposed methods.*

Index Terms—Clustering problem; complex objects; evolution technologies; genetic algorithm; evolution strategies.

I. INTRODUCTION

The process of progressive movement towards the information society creation is related to problems associated with big data storage and processing. Their solution is related to the intellectual analysis of data, the technologies of which are formed at the intersection of artificial intelligence, statistics, the databases theory. These include KDD (knowledge discovery in databases), data mining, OLAP (On-line analysis processing) – extracting information from multidimensional databases and others. Elements of these technologies are an integral part of electronic data warehouses. A significant part of information is represented by data that are socio-economic indicators of the complex systems functioning.

Large amounts of information are characterized by the presence of noise effects, their processing leads to the accumulation of a cumulative error. To overcome this problem, it is necessary to determine the important factors and carry out their analysis. Information entropy reduction can also be achieved by grouping objects and extracting knowledge in smaller and functionally related assemblies. Such procedures are aimed at successively overcoming uncertainty. The first step is the clustering problem solving.

The clustering problem is to define groups of objects (processes) that are closest to each other by some criterion. However, usually no assumptions about their structure are made [1], [2]. Most clustering methods are based on the analysis of the matrix of similarity coefficients, such as distance, conjugacy, correlation, etc. If the distance is the criterion or metric, then the cluster is a group of points Ω, such that the average square of the intra-group distance to the center of the group is less than the average distance to the common center of the original set of objects, i.e. $\overline{d}_\Omega^2 < \sigma^2$, where

$$\overline{d}_\Omega^2 = \frac{1}{n}\sum_{X_i \in \Omega}\left(X_i - \overline{X}_\Omega\right)^2,\ \overline{X}_\Omega = \frac{1}{n}\sum_{X_i \in \Omega} X_i.$$

Solving the problem of minimizing the distance between objects is equivalent to solving the problem of minimizing the distance to an object having average characteristics, since, for example, for Hamming distance:

$$\sum_{\substack{j=1\\k<l}}^{n}\left|X_{kj} - X_{lj}\right| = \sum_{\substack{j=1\\k<l}}^{n}\left|X_{kj} - \overline{X} + \overline{X} + X_{lj}\right|$$

$$\le \sum_{\substack{j=1\\k<l}}^{n}\left|X_{kj} - \overline{X}\right| + \sum_{\substack{j=1\\k<l}}^{n}\left|X_{lj} - \overline{X}\right| \le \sum_{j=1}^{n}\left|X_{kj} - \overline{X}\right|$$

$$+\sum_{j=1}^{n}\left|X_{lj} - \overline{X}\right| = 2\sum_{j=1}^{n}\left|X_{kj} - \overline{X}\right|.$$

The clustering problem is accompanied by two problems: determining the clusters optimal number and obtaining their centers. The initial data for the clustering problem are the parameters values of the objects of the study. Obviously, determining the clusters optimal number is the prerogative of the researcher. Suppose that the clusters number m in known and $m << n$, where n is the number of objects. The problem is obtained:

$$\sum_{i=1}^{m}\sum_{j=1}^{n_i}\left\| X_j - \overline{X}_i \right\| \rightarrow \min,$$

where $\overline{X}_i$, $i=\overline{1,m}$ is the average value in the cluster, $\left\| X_j - \overline{X}_i \right\|$ is the distance between objects. The clustering problem solution are the cluster centers $\overline{X}_i$ that can be contained among the objects being considered, which is a fairly strict condition, and can be represented by any points in the field of study.

Traditional cluster analysis methods include tree clustering, two-input integration, the K-medium method, the dendrite method, the correlation pleiad method, and the ball method. The advantages of these methods are their simplicity, the invariance of their technique relative to the nature of the initial data and the metrics used. The disadvantages include a weak formalization, which hinders the computer technology use, as well as low accuracy, resulting in preliminary estimating of the factors space structure and their informativeness. Another method for solving the clustering problem is to use the self-organized Kohonen map [3]. The problem of using such a card (neural network) is the choice of initial weighting coefficients, the continuous operation nature, and efficiency, the evaluation of which remains a problem for today.

The complex objects and systems clustering problem is not new and has been considered before. The most famous analogue is the Mendeleev's periodic table. Nowadays it's difficult to find a field in which the clustering problem solving is not required. Suffice it to recall the need of recognition of satellite images or telemetric information, data from video sources about stationary or moving objects. Such recognition is based on a clustering, that uses an "object-property" type table.

Numerous papers have been devoted to clustering problems and methods; their review is given in [4]. First of all, the whole family of clustering heuristic algorithms that were obtained by the Novosibirsk School of Data Analysis [5] should be noted. An important contribution to the theory and practice of clustering was made in [6], the results, obtained in it, allowed to separate linearly inseparable objects by moving into higher-dimensional spaces. Without considering classical clustering methods, it should be noted, that research and development of new methods continues. In particular, the automatic determination of clusters number [7], application of clustering for objects, whose characteristics are specified by their membership functions [8], designing of algorithms, that consider the clustering criteria weight coefficients [9], development of algorithm for analyzing large data, based on the clustering development, which includes the analysis of probability function [10], improvement of satellite image clustering algorithms [11] development of algorithms for mixed types data [12], are investigated.

Analysis of the results, obtained in modern sources, shows the existence of two tendencies: first, the development of clustering methods, based on new scientific paradigms, and second, the clustering results are not used for solving the problems of identifying unknown dependencies and their optimization.

At the same time, it should be noted that several problems still exist, in particular:

– problem of clustering algorithms convergence accuracy and speed;

– problem of clustering objects in images that were subjected to linear or other transformations;

– problem of clustering algorithms convergence accuracy and speed.

II. Fitness-Function and Evolutionary Methods for Clustering Problem Solving

The problem is the following: to split the set Ω, that consists of n objects into m clusters $\Omega = \{S_1, S_2, ..., S_m\}$. Every object S_i has the set of characteristics, $i=\overline{1,n}$. The objects data are presented in the "object-property" type Table I.

TABLE I. OBJECTS DATA

S \ X	X_1	X_2	…	X_3
S_1	x_{11}	x_{12}	…	x_{1k}
S_2	x_{21}	x_{22}	…	x_{2k}
…	…	…	…	…
S_n	x_{n1}	x_{n2}	…	x_{nk}

Let the set $C^* = \{C_1^*, C_2^*, ..., C_m^*\}$ be the solution of the clustering problem. Considering that the objects are points located inside a k-dimensional hyperparallelepiped, the following can be obtained:

$$(x_{i1}, x_{i2}, ..., x_{ik}) \in \Theta = \left\{ \underset{i=1}{\overset{k}{X}} [a_i, b_i] / a_i, b_i \in R^+ \right\}$$

$$.(x_{i1}, x_{i2}, ..., x_{ik}) \in \Theta = \left\{ \underset{i=1}{\overset{k}{X}} [a_i, b_i] / a_i, b_i \in R^+ \right\}.$$

Thus the following equality holds:

$$\sum_{i=1}^{n} \sum_{j=1}^{m} \chi\{S_i \in R_j\} d(S_i, C_j^*)$$
$$= \min_{(C_1, C_2, ..., C_m)} \sum_{i=1}^{n} \sum_{j=1}^{m} \chi\{S_i \in R_j\} d(S_i, C_j).$$

where R_j is the j-cluster and

$$\chi\{S_i \in R_j\} = \begin{cases} 1, \text{ if } d(S_i, C_j) = \min_{l=\overline{1,m}} d(S_i, C_l), \\ 0, \text{ otherwise.} \end{cases}$$

The original problem reduces to the following: find

$$C^* = \arg\min_{C \in \Theta} F(c)$$
$$= \arg\min_{C \in \Theta} \min_{(C_1, C_2, ... C_m)} \sum_{i=1}^{n} \sum_{j=1}^{m} \chi\{S_i \in R_j\} \cdot d(S_i, C_j),$$

with restriction that all potential cluster centers lie inside the hyperparallelepiped, $C = (C_1, C_2, ..., C_m)$.

To find the problem solution, evolutionary methods can be used. The solution search can be presented as a sequence of the following steps:

Step 1. Normalize the elements of Table I according to the formula $x'_{ij} = \frac{x_{ij} - x_{\min j}}{x_{\max j} - x_{\min j}}$, where $x_{\min j}$ and $x_{\max j}$ are the minimum and maximum values in j-column.

Step 2. Generate q sets, that consist of m elements $C^i = (C_1^i, C_2^i, ..., C_m^i), i = \overline{1,q}$ and, as usual, q is a number from the set $\{20, 21, ..., 50\}$. The values C_j^i are uniformly distributed in the hypercube, $i = \overline{1,q}, j = \overline{1,m}$,

Step 3. Calculate the distance from each object to each cluster center:

$$d_{jp}^i = d(S_j, C_p^j) = \left(\sum_{l=1}^{k} (x_{jl} - c_{pl}^j)^2 \right)^{1/2}.$$

where $i = \overline{1,q}, j = \overline{1,n}, p = \overline{1,m}$.

In order to determine which cluster the objects belong to, the search problem must be solved: for $\forall i = \overline{1,q}$, $\forall j = \overline{1,n}$ find $\arg\min_p d(S_j, C_p^i)$.

Table II is obtained, where p_{ij} is a cluster number, which the object S_j Belongs to, for ith potential problem solution.

TABLE II. OBJECT-CLASTER ATTACHMENT

S	S_1	S_2	…	S_k
1	p_{11}	p_{12}	…	p_{1n}
2	p_{21}	p_{22}	…	p_{2n}
…	…	…	…	…
q	p_{q1}	p_{q1}	…	p_{qn}

Step 4. Calculate the distance form every object to the center of appropriate cluster that is the potential problem solution:

$$d_i = \sum_{j=1}^{n} d(S_j, C_{p_{ij}}^i), \forall i = \overline{1,q}.$$

Step 5. If the Genetic Algorithm is chosen to solve the problem, then, considering the value d_i, the following operations are made with the set C^i: crossover operation, mutation, and elite selection into the new population of potential solutions is made.

If the Evolution Strategy method is used, the new potential population is generated, where every new solution is obtained from "parent" solution, by adding a normally distributed random displacement $X_{\text{new}} = X_{\text{parent}} + \xi(N(0, \delta^2))$. The amount of new potential solutions, as usual, is 7 times bigger, than the amount of "parent" solutions [13], [14]. The best solutions from the "parent" and intermediate populations are selected to the new population.

Step 6. Steps 3-5 are repeated until the criteria for iterative process stop are not achieved. Such criteria may be:

– the priori number of iterations;
– for a given ε:

$$\left| \max_i d_i^{(it)} - \max_i d_i^{(it+1)} \right| < \varepsilon.$$

or

$$\left| \operatorname{avg}_i d_i^{(it)} - \operatorname{avg} d_i^{(it+1)} \right| < \varepsilon.$$

or

$$\max_i d(C^{i(it)}, C^{i(it+1)}) < \varepsilon.$$

where it is iteration number.

Step 7. The clustering problem solutions, after the criteria for iterative process stop are achieved, are the following:

$$\arg\min_{C^{i(it)}} d_i^{(it)}$$

The proposed clustering method is a parametric method and it's efficiency depends on researcher's qualification and efficiency of parameters setting for each specific problem. These parameters are the following:

– the iteration process stop criteria;
– the crossover and mutation type;
– the type of parents selection and population of the next generation formation, if the Genetic Algorithm is chosen as optimization method;
– the parent-solutions and offspring-solution number;
– constant or variable value standard deviation;
– positive or negative dynamics of standard deviation for the Evolution Strategy.

It is known that the convergence in probability holds for a genetic algorithm with an elite selection to a new solutions population, and for $(\lambda+\mu)-$ Evolution Strategy, where λ is the parent-solution number, and μ is the offspring-solution number with iterations number tending to infinity.

III. EXPERIMENTS AND THEIR ANALYSIS

To test the method effectiveness, two experiments were performed. During the first experiment, Fisher's Iris data set was used. It is known, that this set consists of the data on 150 irises and contains the following characteristics: sepal length, sepal width, petal length and petal width. The irises classification is known, it consists of 3 types: Iris setosa, Iris virginica and Iris versicolor. The purpose of the experiment was to find out if the suggested method allows to find the clustering problem solution (the cluster centers), to compare it with already known traditional clustering methods, and to improve the objective function value.

During the second experiment, the problem of Ukrainian areas clustering was considered, based on their socio-economic indicators. As a result of preliminary analysis it was found out, that the most informative and important indicators for solving this problem are the following:

– added value (x_1);
– territory (x_2);
– investment in main capital per person (x_3);
– foreign direct investment per person (x_4);
– employment of the population per 10.000 people (x_5);
– average income per person (x_6);
– loans granted to business entities per person (x_7);
– number of received patents for 10.000 inventions (x_8).

As the clustering methods the following are chosen: *k*-means clustering as the most popular classic method, evolution clustering method based on Genetic Algorithm, evolution method based on Evolution Strategy. The main feature of the *k*-means clustering method is that cluster centers are defined a posteriori after the determination of the belonging of all objects to clusters. According to the proposed methods, the cluster centers are determined a priori and, in contrast to many other clustering methods, no pairwise comparisons are made, that is essential for the high dimensional problems because of calculations time saving. Based on the preliminary experiments results, the maximum number of iterations was chosen as criteria for the iteration process stop. For the Iris clustering problem the maximal iteration number is 180, for the clustering problem of 25 Ukraine areas – 200. Iris data were not standardized, since they are comparable. Ukraine areas data were standardized, since they were in different ranges, for example $x_1 \in [1000, 6000]$ and $x_2 \in [0, 2]$.

During the simulation, classic k-means method variations were used (KM), evolution method based on genetic Algorithm (EMGA) and evolution method based on Evolution Strategy (EMES), that means the absence in them of any procedures that optimize the method itself. The simulation results are the objective function values. Obviously, it is impossible to draw an opinion on the solution optimality, but it is possible to compare the obtained solutions and to choose the best among them.

It was assumed that in the first and second cases the number of clusters could be 2, 3 or 4. The simulation results are shown in Table III.

TABLE III. EXPERIMENTS RESULTS

	KM	**EMGA**	**EMES**
Iris	***Fitness-function values***		
2	154.31	144.20	142.76
3	132.21	121.14	130.16
4	126.42	112.15	109.42
Areas	***Fitness-function values***		
2	4.31	3.22	3.26
3	3.94	3.72	3.53
4	3.12	3.04	2.86

The results show that the fitness-function values, obtained for the developed methods, are less, which indicates a better clustering quality. The fitness-function dynamic and cluster centers for the clustering problem are shown in Fig. 1.

Both in the first and in the second cases, the initial population consisted of 20 potential solutions.

For EMGA the following settings were chosen: single-point crossover, the mutation probability 0.01, the offspring selection method – by tournament, the new population forming method – elite. For EMES the size of the parent population was also 20, all standard deviations were the same and equal 0.02. Such parameter values choice was made in accordance with the rule 3δ, known from information theory.

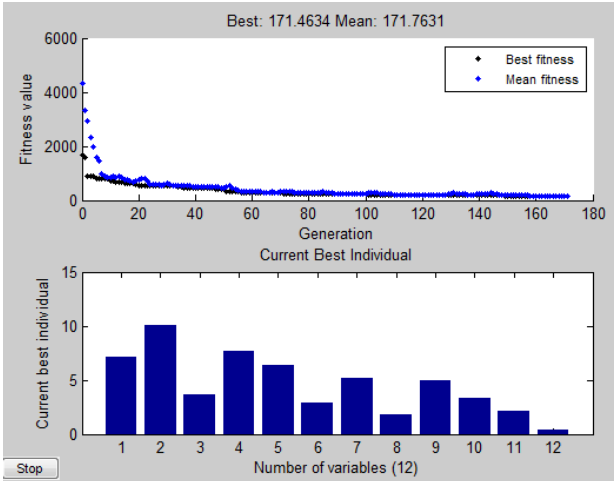

Fig. 1. Fitness-function dynamic and cluster centers

Data, averaged over 50 experiments, allow such conclusions to be made for each case:

For the Iris set clustering problem, the objective function value for the EMGA and EMES methods decreases, compared to KM, by:

- 6.6% and 5.8% accordingly, for two clusters;
- 8.3% and 8.5% accordingly, for three clusters;
- 11.2% and 13.4% accordingly, for four clusters.

For the Ukraine areas clustering problem, the objective function value for the EMGA and EMES methods decreases, compared to KM, by:

- 25.2% and 25.4% accordingly, for two clusters;
- 5.5% and 10.4% accordingly, for three clusters;
- 2.5% and 8.3% accordingly, for four clusters.

Note that, based on the experimental data and the meaning of the problem, clustering of a large number of clusters does not make sense. The obtained results proof an advantage of the developed methods EMGA and EMES. The objective function value, during an increasing of clusters number, should decrease to a certain point, which can be seen from the data in Table III. Of interest is the dynamics of comparing the objective function values, obtained by the method KM with methods EMGA and EMES. With increasing clusters number for more structured data, methods EMGA and EMES show a positive comparative accuracy dynamics. In the case of poorly structured Ukraine areas data, a reverse tendency occurs. However, both in the first and second cases, the cluster centers obtained by methods EMGA and EMES are more optimal.

IV. Conclusions

There is a huge number of clustering methods based on different paradigms. The presented methods do not limit the researcher to any requirements in this class of problems. Their use is advisable in large-dimensional spaces, for the recognition of satellite and other images. The application of preprocessing technologies for such data in combination with the power of evolutionary modeling will result in better quality solutions for the clustering problem.

The proposed method of evolutionary modeling, based on the use of a genetic algorithm, effectively functions when processing large-scale arrays, since it optimally combines a purposeful search and elements of randomness, aimed at knocking the objective function out from local minima. No preconditions for its use are required. The main condition for optimizing calculations is the correct algorithmization of the objective function values calculation. The multi-vector process of algorithm speed improving (for genetic algorithms is especially relevant) and its accuracy (finding the fitness function global minimum), as well as its relevance, indicate the need to solve the optimization problem using the proposed method.

References

[1] I. Mandel, "Cluster analysis," *Finance and Statistics*, Moscow, 1988.

[2] A. N. Gorban and A. Yu. Zinovyev, "Method of Elastic Maps and its Applications in Data Visualization and Data Modeling," *Int. Journal of Computing Anticipatory Systems*, CHAOS, vol. 12, pp. 353–369, 2002.

[3] T. Kohonen, "Self-organization and associative memory," New-York, 2d. ed., Springer Verlag, 1988.

[4] A. Jain, M. Murty, and P. Flynn, "Data Clustering: A Review," *ACM Computing Surveys*, vol. 31, no. 3, 1999.

[5] N. G. Zagoruiko, *Applied methods of data and knowledge analysis*, Novosibirsk, MI SB RAS, 1999, (in Russian).

[6] M. Aizerman, E. Braverman, and A. Rozonoer, *Method of potential functions in learning theory*, Moskow, Nauka, 1970, (in Russian).

[7] A. Ryosuke, M. Miyamoto, E. Yasunori, and H. Yukihiro, "Hierarchical clustering algorithms with automatic estimation of the number of clusters," *9th International Conference on Soft Computing and Intelligent Systems (IFSA-SCIS)*, Otso, Japan, 2017.

[8] A. T. C. Francisco, "Fuzzy clustering algorithm with automatic variable selection and entropy regularization," *IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, Naples, Italy, 2017.

[9] M. Anusha and J. G. R. Sathiaseelan, "Evolutionary clustering algorithm using criterion-knowledge-ranking for multi-objective optimization," *World Conference on Futuristic Trends in Research and Innovation for Social Welfare* (Startup Conclave), Coimbatore, India, 2016.

[10] K. S. S. Reddy and C. S. Bindu, "A review on density-based clustering algorithms for big data analysis," *International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)* Palladam, Tamilnadu, India, 2017.

[11] S. Yadav and M. Biswas, "Improved color-based K-mean algorithm for clustering of satellite image," *4th International Conference on Signal Processing and Integrated Networks (SPIN) Noida*, Delhi-NCR, India, 2017.

[12] H. Yu, Z. Chang, and B. Zhou, "A novel three-way clustering algorithm for mixed-type data," *IEEE International Conference on Big Knowledge (ICBK)*, Hefei, China, 2017.

[13] E. Falkenauer, *Genetic algorithms and grouping problems*, Chichester, England: John Wiley & Sons Ltd., 1997.

[14] H.-P. Schwefel, *Evolution and optimum seeking*, New York: Wiley & Sons, 1995.

Received September 04, 2017.

Vitaliy Snytyuk. Doctor of Engineering Science. Professor.
Intellectual and Information Systems Department, Taras Shevchenko National University of Kyiv, Kyiv, Ukraine.
Education: Taras Shevchenko National University of Kyiv, Kyiv, Ukraine, (1991).
Research interests: computational intelligence, evolutionary modeling.
Publications: 255.
E-mail: snytyuk@gmail.com

Oleh Suprun. Post-graduate student.
Institute of Mathematical Machines and Systems Problems, NASU, Kyiv, Ukraine.
Education: Taras Shevchenko National University of Kyiv, Kyiv, Ukraine, (2014).
Research interests: artificial intelligence, evolutionary modeling.
Publications: 8.
E-mail: oleh.o.suprun@gmail.com

В. Є. Снитюк, О. О. Супрун. Еволюційна кластеризація як метод розв'язання економічних задач

В статті представлено метод, розроблений для використання еволюційних технологій для кластеризації великої кількості об'єктів, що представлені за допомогою їх характеристичних значень. У зв'язку з необхідністю аналізувати великі дані та вилучати необхідні дані з багатовимірних баз, використання класичних методів аналізу не є ефективним, або ж потребує великої кількості ресурсів чи часу, щоб надати задовільне вирішення поставленої практичної задачі. Такі задачі дуже часто виникають в економічній та фінансовій сфері, де експерту потрібно приймати правильні рішення, згідно з інформацією, отриманою з різних джерел. Ця інформація може містити шуми, або навіть бути недостовірною. Для вирішення таких задач необхідні збір та формалізація доступної інформації, що потребує значних витрат часу. Запропонований метод дозволяє використовувати еволюційні технології, такі як генетичний алгоритм та елементі еволюційних стратегій, для вирішення задач кластеризації при мінімальних обмеженнях, що накладаються на початкові дані, що відповідає умовам реальних практичних задач. Наведено результати експериментів, при проведенні яких було використано такий метод, що доводить ефективність його використання.

Ключові слова: задача кластеризації; складні об'єкти; еволюційні технології; генетичний алгоритм; еволюційні стратегії.

Снитюк Віталій Євгенович. Доктор технічних наук. Професор.
Кафедра інтелектуальних та інформаційних систем, Київський національний університет імені Тараса Шевченка, Київ, Україна.
Освіта: Київський національний університет імені Тараса Шевченка, Київ, Україна (1991).
Направлення наукової діяльності: обчислювальний інтелект, еволюційне моделювання.
Кількість публікацій: 255.
E-mail: snytyuk@gmail.com

Супрун Олег Олексійович. Аспірант.
Інститут проблем математичних машин та систем при НАН України.
Освіта: Київський національний університет імені Тараса Шевченка, Київ, Україна, (2014).
Направлення наукової діяльності: штучний інтелект, еволюційне моделювання.
Кількість публікацій: 8.
E-mail: oleh.o.suprun@gmail.com

В. Е. Снитюк. О. О. Супрун. Эволюционная кластеризация как метод решения экономических задач
В статье представлен метод, разработанный для применения эволюционных технологий при кластеризации большого количества объектов, представленных с помощью их характеристических значений. В связи с необходимостью анализировать большие данные и изымать необходимые данные из многомерных баз, использование классических методов анализа не является эффективным, или же требует большого количества ресурсов или времени, чтобы предоставить удовлетворительное решение поставленной практической задачи. Такие задачи очень часто возникают в экономической и финансовой сферах, где эксперту необходимо принимать правильные решения, согласно информации, полученной из различных источников. Эта информация может содержать шумы, или даже быть недостоверной. Для решения таких задач необходимы сбор и формализация доступной информации, что требует значительных затрат времени. Предложенный метод позволяет использовать эволюционные технологии, в частности генетический алгоритм и элементы эволюционных стратегий, для решения задач кластеризации при минимальных ограничениях, налагаемых на начальные данные, что соответствует условиям реальных практических задач. Приведены результаты экспериментов, при проведении которых был использован такой метод, которые доказывают эффективность его практического применения.
Ключевые слова: задача кластеризации; сложные объекты; эволюционные технологии; генетический алгоритм; эволюционные стратегии.

Снитюк Виталий Евгеньевич. Доктор технических наук. Профессор.
Кафедра интеллектуальных и информационных систем, Киевский национальный университет им. Тараса Шевченка, Киев, Украина.
Образование: Киевский национальный университет им. Тараса Шевченка, Киев, Украина, (1991).
Направление научной деятельности: вычислительный интеллект, эволюционное моделирование.
Количество публикаций: 255.
E-mail: snytyuk@gmail.com

Супрун Олег Алексеевич. Аспирант.
Институт проблем математических машин и систем при НАН Украины.
Образование: Киевский национальный университет им. Тараса Шевченка, Киев, Украина, (2014).
Направление научной деятельности: искусственный интеллект, эволюционное моделирование.
Количество публикаций: 8.
E-mail: oleh.o.suprun@gmail.com