Zatsarynin Serhii
Postgraduate Student of the Department of Marketing
National Aviation University
(Kyiv, Ukraine)

MARKET SEGMENTATION OF INNOVATIVE PRODUCTS USING GENETIC ALGORITHMS

One way to increase the company's competitiveness is to find new market niches. The market niche is the result of innovations that stimulate hidden, potential demand, as a result of which the company, developing a new market, avoids intense competition and receives a higher rate of return. It is proved that the growing number and complexity of tasks in the field of marketing research, working with a large amount of information, leads to the need to group data. The aim of the study is to develop a universal approach to solving the problem of market segmentation of innovative products based on a combination of genetic algorithm with traditional clustering methods. An ideal market niche can be defined as a compact and isolated series of points, which in some space of characteristics are objects or data elements. The selection of a market niche in the medical equipment market is carried out using a top-down approach. This approach implies the traditional segmentation of customers, which is carried out in the following order: segmentation, segment selection, positioning. It is believed that segmentation is the starting point for the formation of a market niche. To segment the medical equipment market, it is proposed to use cluster analysis methods. According to the results of the analysis, it can be seen that the market segments of potential consumers of medical equipment and consumables of Siemens in Ukraine are characterized by a fairly dense grouping of images of consumers around the center of its cluster in the space of features. The presented genetic clustering algorithm is flexible in relation to the decision-making process, as it allows to perform clustering based on various criteria, such as maximum mutual removal of clusters, proximity of geometric images of objects to the center of the cluster, other criteria. This is achieved by changing the calculation formula of the fitness function, which takes into account the necessary combination of clustering criteria without changing the structure of the algorithm. The algorithm is insensitive to initialization, as in the process of evolution of chromosomes through the use of genetic operators, the algorithm completely covers the whole set of acceptable solutions, which, in turn, provides high quality market segmentation.

Keywords: market segmentation, niche market, innovative products, clustering, genetic algorithm

Problem statement in general and its connection with important scientific or practical tasks. The current market situation is characterized by saturation and fragmentation of markets, a high degree of competition in each segment and the ever-increasing demands of consumers for product quality. Under such conditions, any company must change, improve, constantly trying to offer the buyer new products, better quality at a better price than competitors. And as practice shows, the generally accepted approaches to marketing activities are becoming less effective and lead to a decrease in the return of marketing tools.

New ideas are needed that would bring higher returns compared to traditional methods of competition. One way to stay one step ahead of the competition is to find new market niches that are not occupied by competitors. At the same time, the niche is not just a market segment in which companies are more comfortable to work. It is the result of innovations that stimulate hidden, potential demand, as a result of which the company, developing a new market, avoids intense competition and receives a higher rate of return. In addition, the market for innovative products needs segmentation under the following conditions: changes in sales of specific innovative products or services of an innovative nature; a significant decline (increase) in potential demand for the relevant industrial product; significant increase in the level of competition; significant differences in consumer requirements for a particular product.

Analysis of the latest research and publications, which initiated the solution of this problem. The issue of market segmentation is given attention in the vast majority of marketing educational literature, from foreign specialists-classics of marketing, to domestic scientists: Zozulova O.V. [1], Ilyashenko S.M. [2], Beloved T.V. [3], Kudenko O.V. [4], Lotish O.Ya. [5], Aucklander M.A. [6], Smerichevsky S.F. [7] and others.

Highlighting previously unsolved parts of the general problem to which the article is devoted. However, in most of these works, only a general analysis is performed for educational purposes. In the context of globalization and digital transformation of the economy, the amount of information about the consumer has increased tenfold, the number of consumers has increased. All this has led to the need to use data mining techniques to market segmentation.

Formulation of the purpose of the article (problem statement). The aim of this study is to develop a universal approach to solving the practical problem of segmenting the market of innovative products based on a combination of genetic algorithm with traditional clustering methods.

Statement of the main material of the research with full justification of the scientific results obtained. The processes of all stages of the life cycle of knowledge-intensive innovative products (aviation, electronics, medical equipment, etc.) are accompanied by design, management, marketing and other decisions in conditions of uncertainty caused by ignorance or inadequacy of data, their incompleteness and inaccuracy (error) measurement or observation , subjective fuzzy assessment. To reveal this uncertainty, information technologies to support artificial intelligence are used, which implement "soft computing" using neural networks, genetic algorithms, fuzzy sets, clustering [8]. The greatest effect is achieved when using hybrid technologies that use a combination of the above models and methods [9].

The use of traditional clustering algorithms, such as iterative ("k-means" [10], SOM [11]), hierarchical divisive (BIRCH [12], MST [13]) and hierarchical agglomerative (CURE, ROCK [14]) has changed to hybrid approach, which involves the use of clustering of the above technologies "soft computing". the possibilities of cluster analysis are expanded due to the use along with traditional new methods (the method of wave transformations in the Wave Cluster algorithm [15]), combining several methods into one, detection and accounting for clustering interdependence in data [16], and hybrid approach [17]. Many tasks involve multidimensional data analysis, and any of them can cause automatic grouping tasks. These may include, for example, tasks such as tasks of working with customer groups in CRM-systems, geographical grouping of customers of service points.

The growing number and complexity of tasks to be solved within the framework of market research, the need to work with a large amount of data, leads to the need to group data. Cluster analysis aims to identify in the initial multidimensional data such homogeneous subsets that the objects within the groups were in this case similar to each other, and the objects from different groups are not similar. Under "similarity" (degree of similarity) means the proximity of objects in a multidimensional space of features and the task is to highlight in this space of

natural clusters of objects that are considered homogeneous groups.

An ideal market niche can be defined as a compact and isolated series of points, which in some space of characteristics are objects or data elements. In fact, a market niche is a subjective concept, its definition may require knowledge in the relevant field. In the case of two- and three-dimensional data, a person is quite capable of identifying separate niches, but in practical tasks, the number of data measurements can be very large. For example, in the applied problem of clustering the medical equipment market, which is considered in this paper, the dimension of the data can vary from tens to hundreds of measurements.

The allocation of a market niche in the medical equipment market is traditionally carried out using a "top-down" approach. This approach involves finding your "place under the sun" in accordance with the traditional marketing segmentation of customers, which is carried out in the following order: segmentation, segment selection, positioning. It is believed that segmentation is the starting point for the formation of a market niche.

To segment potential consumers of medical equipment, it is proposed to use the method of k-means (k-means). The k-means method is the most popular algorithm for cluster analysis. The source data is a set of vectors in space with d features. Each vector belongs to a corresponding group of objects. The source data also contains the required number of clusters - k. The essence of the method is to find k points (centers) in the d-dimensional space, such that the sum of the squares of the distances from the sample vectors to the nearest of the centers reaches a minimum [18]:

$$\underset{X_1,\dots,X_k}{\mathrm{argmin}} F(X_1,\dots,X_k) = \sum_{i=1}^{N} \min_{j\in\{1,k\}} \left\| X_j - A_i \right\|^2,$$

where A_i - instance vectors (vector that characterizes the i-th client);
N - sample size (number of customers);
k - number of clusters (market segments);
$X_1, \dots, X_k$ - search points (centers of segments).

In our case, the data vectors are the vectors of the company's customer characteristics: one vector corresponds to one customer, and the dimension of the space corresponds to the number of parameters (characteristics) by which the customer is evaluated.

This clustering algorithm can also be classified as hybrid, as it uses an evolutionary approach specific to genetic algorithms. The basis for its development was the need to use in decision support systems a universal algorithm for the formation of clusters not one by one, but by a set of criteria. In the presented algorithm, this versatility is provided by changing the calculation of the fitness function (fitness function) used in the chromosome algorithm without changing the structure of the algorithm itself. In this case, clustering, in addition to the proximity of objects to the center of the cluster in the metric feature space (criterion for the formation of denser clusters), takes into account the distance between clusters (criterion of maximum difference between clusters).

In our case, the vectors in 5-dimensional space represent the following parameters of consumers of the market of medical equipment and consumables of Siemens in Ukraine:
1. Location (1 – Kyiv, 2 – regional centers, 3 – cities and district centers, 4 – rural areas).
2. Estimation of the consumer budget (1 – low, 2 – below average, 3 – average, 4 – above average, 5 – high).
3. Form of ownership of the consumer's company (1 – private, 2 – state, 3 – communal).
4. Frequency of purchases of equipment and consumables (1 – single, 2 – periodic, 3 – regular).

5. The cost of the order (from $ 10 thousand to $ 100 thousand).

It should be noted that the range of values of these parameters is different. Thus, the scatter of the values of the income level estimate (maximum distance between the extreme vectors for this parameter) is 4, for the order value – more than 100000. Thus, if the k-means method calculates the distance between the vectors as Euclidean distance, it is the cost of the order (the contribution of other parameters to the distance between the vectors will be insignificant in comparison with it). To level this effect when calculating the distances between the vectors, it is necessary to use normalized Euclidean distances:

$$\|X_j - A_i\| = \sqrt{\sum_{p=1}^{d}(x_{jp} - a_{ip})^2 / \sigma_p^2},$$

where σ_p is the standard deviation of the p-th parameter;
a_p is the average value of the p-th parameter;

The standard procedure for k-means includes the following steps:

1. From the whole set of vectors A_i, k vectors are randomly selected. Suppose that the centers of the segments $C_j=(c_{j1}, c_{j2}, ..., c_{jd})$ coincide with the selected vectors.

2. For each of the centers C_j we find a subset S_j of vectors A_i for which this center is the closest.

3. For each of the subsets S_j its true center is determined. To do this, calculate the average value of each of the d parameters of this subset

$$c_{jp} = \sum_{Y=(y_1,...,y_d)\in S_j} y_p / S_j, p \in \{\overline{1, d}\},$$

where S_j is the power of the set.

4. If in the third step the coordinates of any of the centers have not changed, then the clustering process is complete. Otherwise, the next iteration is performed starting from step 2.

This procedure is an algorithm for local optimization, its result depends on the choice of initial centers. Various heuristic algorithms are used to find the global minimum, the simplest of which is to run the procedure described above repeatedly with different randomly selected starting centers. The quality criterion of clustering is determined taking into account the average size of the formed clusters and the assessment of the distance of the clusters from each other.

The target functions of market segmentation tasks usually have a large number of local lows. As a result, the search for a universal algorithm that guarantees an accurate solution (except for problems with a small amount of data) is hopeless. The literature offers many heuristic methods for setting the initial centers of the local search procedure. In the case of evolutionary algorithms, various methods of coding valid solutions that form a population are used. Hosage and Godchild [19] proposed the first genetic algorithm for the p-median problem. Quite an accurate algorithm with a special crossing operator called greedy heuristics [20]. If research on classical placement theory is aimed at obtaining accurate results, then research related to the development of methods of cluster analysis and automatic grouping, as a rule, focuses on improving the performance of appropriate algorithms. In some cases, fast and at the same time quite accurate methods are required, which the known methods do not allow.

In order to increase the accuracy of the result for the selection of initial centers, it is proposed to use a genetic algorithm with greedy heuristics. The choice of greedy heuristics as a method of local search is due to the fact that it allows to obtain high-precision results.

Algorithms of the method of greedy heuristics are an effective method of solving optimization problems of automatic grouping, placement and optimization with a large amount of input data depending on the conditions and parameters of the problem [20].

Genetic algorithm with greedy heuristics for the problem of k-means includes the following steps:

1. To form a "population" – to randomly select P sets of vectors S_j, in each of which k vectors. For each of the sets the standard procedure of the k-means method is implemented, the results (total square of the normalized Euclidean distance) are written in the variable f_j.
2. Randomly select two indices of the sets p and q. Create the set $S = S_q \cup S_p$.
3. If $| S | = k$, then the set S_j with the worst (greatest distance) value f_j is replaced by the set S.
4. For each $Y_i \in S$ we obtain the set S', from which the vector Y_i. is excluded. For the received set we start standard procedure, we write down result in a variable f_i'.
5. Exclude from the set S the vector Y_i. with the lowest value of f_i'. We return to step 3.

The result of this algorithm, which implements the method of k-means, is a set of points (centers) in a d-dimensional normalized Euclidean space. Each of the vectors A_i belongs to the cluster whose center is closest to it. This method is also convenient because the center is the average values of each of the cluster parameters.

According to the results of k-average calculations, consumers of medical equipment and consumables of Siemens in Ukraine were divided into 7 groups. The analysis of the selected segments showed that the bulk of the company's customers are distributed between three segments (Table).

Table – Results of segmentation of potential consumers of medical equipment and consumables of Siemens in Ukraine

№	Location	The size of the consumer's budget	Form of ownership of the consumer company	Frequency of purchases	The cost of the order	Segment characteristics
1	2,498	3,017	2,033	1,588	23,968	Regional centers, cities and district centers Average budget State form of ownership Single purchases Small batches of goods
2	1,921	3,229	2,061	2,061	84,655	Kyiv and regional centers Average and above budget State form of ownership Periodic purchases Large batches of goods
3	2,659	2,972	1,838	2,020	54,279	Regional centers, cities and district centers Average budget Private form of ownership Periodic purchases Medium batches of goods

The results of the study show that the market segments of potential consumers of medical equipment and consumables of Siemens in Ukraine are characterized by a fairly dense grouping of images of clustered objects around the center of its cluster in the space of features and remote location of these centers. The algorithm uses genetic operators of selection, crossover, mutation, as well as a new filtering operator, which excludes from consideration chromosomes that reflect unacceptable clustering, ie do not correspond to the number of clusters. The probability of such an event is quite high due to the random nature of the choice of the above values of the components in the formation of the initial population of chromosomes, as well as due to the work of other genetic operators.

Genetic algorithms have traditionally worked with populations of chromosomes that are modified in an iterative process of evolution. In our case, in a population of size w, each chromosome reflects one of the possible clustering options, which uses a vector (t is the chromosome number within the population). The number of chromosome vector components is fixed and equal to the number of objects to be clustered (k).

The presented genetic clustering algorithm is flexible in relation to the decision-making process, as it allows to perform clustering based on various criteria, such as maximum mutual removal of clusters, proximity of geometric images of objects to the center of the cluster, other criteria. This is achieved by changing the calculation formula of the fitness function, which takes into account the necessary combination of clustering criteria without changing the structure of the algorithm.

Like any genetic algorithm, this algorithm is simple in its software implementation, which makes it reliable to use. The high reliability of the algorithm was confirmed by computational experiments on data reflecting different cluster structures. Experiments have shown the rapid convergence of the algorithm, which is due to the small number of generations of evolution in the process of finding the optimal solution.

The algorithm is insensitive to initialization, as in the process of evolution of chromosomes through the use of genetic operators, the algorithm completely covers the whole set of acceptable solutions, which, in turn, provides high quality market segmentation.

Conclusions from this research and prospects for further developments in this area. Noting a certain deficit of compromise in time calculations and quality of results (accuracy - closeness of the value of the objective function to the global optimum, stability - closeness of the obtained values to each other at iterative starts of the algorithm), the aim of this study segmentation of the company's customers, which would meet the following criteria:

– tasks must be solved in a reasonable time;

– to carry out automatic segmentation of a large number of objects in the multidimensional space of characteristics (dimension - tens of measurements);

– solve problems using different clustering models;

– combine with local search methods used to solve specific tasks, provide the opportunity to use different global search strategies;

– give stable results with multiple launches (be deterministic), while the accuracy of the results should not be inferior to other known methods;

– must either estimate the number of groups (segments), or solve a series of problems with different numbers of groups.

The choice of genetic algorithm as a means of market segmentation is due to the fact that these methods allow to obtain high-precision results, characterized by deterministic procedures, which gives hope for more stable results when using such heuristics in various global search strategies, including marketing research.

1. Zozulov O.V. (2003) Segmentirovanie ryinka [Market segmentation]. Kiev, Studtsentr (in Russian)

2. Illiashenko S.M. (2005) Marketynhova tovarna polityka. [Marketing product policy]. Sumy: Universytetska knyha (in Ukrainian)

3. Kokhana T.V. (2013) Stratehichnyi pidkhid do sehmentatsii rynku. [Strategic approach to market segmentation]. *Agrarian economy*, vol. 6, No. 3-4, pp. 64−69.

4. Kudenko O.V. (2014) Krytychnyi analiz metodiv ta pidkhodiv do sehmentuvannia rynku. [Critical analysis of methods and approaches to market segmentation]. *Productivity of agro-industrial production. economic sciences*. No. 26, pp. 33−38.

5. Lotysh O. (2017) Sehmentatsiia rynku yak instrument stratehichnoho analizu haluzi. [Market segmentation as a tool for strategic analysis of the industry]. *Scientific Bulletin of the International Humanities University. Series: economics and management*. No. 24(1), pp. 120−124.

6. Oklander M.A. (2013) Suchasni transformatsii kulturnykh chynnykiv povedinky spozhyvacha. [Modern transformations of cultural factors of consumer behavior]. *Economic Bulletin of NTUU "KPI"*. No. 10, pp. 386−392.

7. Smerichevskyi S., Klimova O., Kniazieva T. (2020) Mechanisms to ensure the activation of the market of products of aircraft building based on clustering and outsourcing. *Cluster model of innovative development of the national economy: infrastructure and investment aspectsp*. Poznań: Wydawnictwo naukowe WSPIA. pp. 352−364.

8. Yudin V.N., Karpov L.E. (2017) Nepolnostyu opisannyie ob'ektyi v sistemah podderzhki prinyatiya resheniy. [Incompletely described objects in decision support systems]. *Programming*. No. 5, pp. 24−31.

9. Hu Z., Bodyanskiy Y., Tyshchenko O. (2019) Self-learning procedures for a kernel fuzzy clustering system. *Advances in Computing Science for Engineering and Education*. Vol. 754, pp. 487−497.

10. Hartigan J.A., Wong M.A., Algorithm A.S. (1979) 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society, Series C (Applied Statistics)*. Vol. 28. № 1, pp. 100−108.

11. Kohonen T. (2001) Self-Organizing Maps: 3rd edition. Berlin-New York: Springer-Verlag. 521 p.

12. Zhang T., Ramakrishnan R., Livny M. (1996) BIRCH: an efficient data clustering method for very large databases. *Proceedings of the ACM SIGMOD international conference on Management of data (SIGMOD '96)*. pp. 103-114. Available at: https://doi.org/10.1145/235968.233324 (accessed 3 February 2021).

13. Päivinen N. (2005) Clustering with a minimum spanning tree of scale-free-like structure. *Pattern Recognition Letters*. Vol. 26. Iss. 7. pp. 921-930. Available at: https://doi.org/10.1016/j.patrec.2004.09.039 (accessed 3 February 2021).

14. Guha S., Rastogi R., Shim K. (2000). ROCK: a robust clustering algorithm for categorical attributes. *Information Systems*. Vol. 25. № 5. pp. 345-366. Available at: https://doi.org/10.1016/S0306-4379(00)00022-3 (accessed 3 February 2021).

15. Sheikholeslami G., Chatterjee S., Zhang A. (2000) WaveCluster: A Wavelet-Based Clustering Approach for Spatial Data. *VLDB Journal*. No. 8 (3-4). pp. 289−304. Available at: http://dx.doi.org/10.1007/s007780050009 (accessed 3 February 2021).

16. Wang C., She Z., Stantic B., Chi C.H., Cao L. (2018) Coupled Clustering Ensemble by Exploring Data Interdependence. *ACM Transactions on Knowledge Discovery from Data*. Vol. 12. No. 6. pp. 63:1-63:38.

17. Abasi A., Sajedi H. (2016) Fuzzy-clustering based data gathering in wireless sensor network. *International Journal on Soft Computing (IJSC)*. Vol. 7. No. 1. Pp.1-15. Available at: https://doi.org/10.5121/ijsc.2016.7101 (accessed 3 February 2021).

18. Antamoshkin A., Kazakovtsev L. (2013). Random Search Algorithm for the p-Median Problem. *Informatica (Ljubljana)*. Vol. 37. No. 3. pp. 267−278.

19. Hosage C.M., Goodchild M.F. (1986) Discrete Space Location Allocation Solutions from Genetic Algorithms. *Annals of Operations Research*. Vol. 6. pp. 35-46.

20. Alp O., Erkut E., Drezner Z. (2003) An Efficient Genetic Algorithm for the p-Median Problem. *Annals of Operations Research*. No.122 (1-4). pp. 21−42. Available at: https://doi.org/10.1023/A:1026130003508 (accessed 3 February 2021).

21. MacQueen J.B. (1967) Some Methods of Classification and Analysis of Multivariate Observations. *Proceedings of the 5th Berkley Symposium on Mathematical Statistics and Probability*. Vol. 1. pp.

281−297.
22. Mandel I.D. (1988) Klasternyiy analiz. [Cluster analysis]. Moscow: Finance and Statistics (in Russian)

Зацаринін С.А., *аспірант кафедри маркетингу, Національний авіаційний університет (Київ, Україна).*

Сегментування ринку інноваційної продукції за допомогою генетичних алгоритмів

Метою дослідження є розробка універсального підходу до вирішення задачі сегментування ринку інноваційної продукції на основі поєднання генетичного алгоритму з традиційними методами кластеризації. Доведено, що зростаюча кількість та складність завдань, що вирішуються в межах маркетингових досліджень ринку, необхідність роботи з великим обсягом інформації, призводить до необхідності групування даних. Для сегментування ринку медичного обладнання пропонується використовувати метод k-середніх – популярний алгоритм кластерного аналізу. Запропонований алгоритм кластеризації відноситься до категорії гібридних, так як він використовує еволюційний підхід, характерний для генетичних алгоритмів. Його перевагою є формування кластерів не по одному, а по сукупності критеріїв. У дослідженні розглядаються вектори в 5-вимірному просторі. За результатами дослідження виділено три основні ринкові ніші на ринку медичного обладнання та витратних матеріалів в Україні, які характеризуються достатньо щільним угрупованням споживачів навколо центру свого кластеру в просторі ознак.

Ключові слова: сегментування ринку, ринкова ніша, інноваційна продукція, кластеризація, генетичний алгоритм