

DOI: [10.17323/2587-814X.2020.4.7.18](https://doi.org/10.17323/2587-814X.2020.4.7.18)

# Structuring advertising campaign costs considering the asymmetry of users' interests

**Alexey N. Kislyakov** 

E-mail: [ankislyakov@mail.ru](mailto:ankislyakov@mail.ru)

Russian Academy of National Economy and Public Administration under the President of the Russian Federation. Vladimir Branch

Address: 59a, Gorky Street, Vladimir 600017, Russia

## Abstract

This work is devoted to the highly topical problem of structuring costs for contextual and targeted advertising on the Internet. The choice of the ad campaign financing structure is considered from the point of view of violating the principle of symmetry of user interest in ads. The purpose of this work is to develop a methodology for structuring advertising campaign costs based on cluster analysis, taking into account the asymmetry of user interest in advertising. The key feature of the research is the description of the possibility of using the asymmetry of user interest in application solutions, such as online advertising. The Gini coefficient is used as an indicator of the degree of imbalance in the manifestation of a feature in clustering, and the features of using the lift coefficient and the Lorentz curve to evaluate the effectiveness of contextual and targeted advertising for various groups of customers are also considered. Using the Gini index and cluster analysis, you can analyze the possibilities of increasing ad revenue and compare it with the absence of any policy for structuring advertising costs. Identifying such patterns in consumer groups allows you to identify the main directions of product development and customer interest in it. The method described here should be used to improve the effectiveness of banner advertising and clustering algorithms. This approach does not improve banner clickability, but allows you to implement an individual approach to advertising products with the current number of clicks and more effectively structure the cost of various types of advertising.

**Key words:** information asymmetry; Gini index; cluster analysis; banner advertising; hierarchical clustering.

**Citation:** Kislyakov A.N. (2020) Structuring advertising campaign costs considering the asymmetry of users' interests. *Business Informatics*, vol. 14, no 4, pp. 7–18.

DOI: [10.17323/2587-814X.2020.4.7.18](https://doi.org/10.17323/2587-814X.2020.4.7.18)

## Introduction

Today, one of the most dynamically developing segments of advertising activity is advertising on the Internet. For example, banner advertising allows you to more accurately and effectively deliver an ad to an interested customer. However, there are quite a few options for monetizing the display of ads depending on the purpose and capabilities.

This type of advertising, of course, requires certain investments, which are not always justified, since the distribution of information and ad delivery technologies, as well as their cost, depend on a large number of factors and are complex and random [1]. It is often difficult to understand which category of users will be interested in an ad, and it is even more difficult to predict sales growth depending on the investment in advertising.

In this regard, there are two fundamentally different types of advertising: contextual and targeted. Contextual advertising allows you to perform automated display of an ad in accordance with the subject area of the customer's search for products. Targeted advertising, on the other hand, searches for the audience for the offer by features, which is a more complex task. Despite the fact that both types of advertising have fine-tuned display settings, the effectiveness of contextual advertising is often higher due to working with an interested audience.

However, companies are also interested in expanding their target audience. Therefore, the task arises of determining the structure of advertising campaign financing by type, depending on the types of business and the behavioral activity of customers, both interested and potential.

It is often advisable to use both types of advertising, instead of giving preference to one type of advertising at the expense of the other. Therefore, the purpose of this study is

to develop a methodology for structuring the costs of an advertising campaign that takes into account the violation of the principle of symmetry of user interest in advertising.

The choice of ad campaign financing structure is not as obvious as it might seem at first glance. It is created by the phenomenon of information asymmetry [2, 3] in the market of online sales of goods and services. In this case, sellers conduct business without having full information about the competitive environment, as well as the intentions of buyers [1, 4]. In turn, customers form their opinion about a product or service based on a different set of factors and sources, constantly listen to the opinion of online communities, reading articles, reviews on the Internet, following opinion leaders, etc. Thus, the interaction between the advertiser and the product user becomes more complicated due to the information asymmetry of the market. The main hypothesis is that the phenomenon of market asymmetry is associated with an imbalance in the behavioral activity of customer groups [5, 6].

There are various methods for structuring and planning advertising campaign costs, taking into account the preferences of the target audience [7]. In this case, various offline and online tools are used, such as customer surveys, or selecting one of the types of advertising (contextual or targeted), depending on the organization's goal (launching a new product on the market, increasing the target audience, etc.). On the one hand, these approaches can significantly simplify the planning process, but on the other hand, they do not allow you to flexibly configure and effectively manage your advertising campaign. The approach proposed is one of the modifications of the algorithm for structuring advertising campaign costs based on the assessment of the economic effect of clickability and the use of classification methods [8]. In particular, we propose using cluster analysis methods to create a more suitable model for structuring advertising costs.

### 1. The proposed approach to structuring advertising costs

The contextual advertising mechanism is an automated transaction for the implementation of advertising, and the usefulness of displaying an advertising banner is measured using the clickability indicator CTR (Click Through Rate) [1, 9]:

$$CTR = \frac{\text{number of clicks}}{\text{number of impressions}} \cdot 100\%. \quad (1)$$

When a company knows about the preferences of some of its customers, it can contact one of the ad providers that implements the contextual advertising mechanism for the corresponding users. Such ads are less effective in finding and expanding the target audience, since they are only shown to those customers who are already interested in purchasing this product.

In the case of tagged ads, banner owners are paid based on the number of clicks and impressions when the user sees the banner but doesn't click on it. The way to monetize banners is to sell them at an ad auction, where advertisers bid on these banners with different numbers of auction participants and placement conditions. Targeted advertising allows you to attract more of the target audience, but it is less accurate, although it is a cheaper tool for an Internet marketer.

If the ad campaign budget is limited, these funds can be distributed among ad providers, for example, as follows: 30% of the funds are allocated to the contextual advertising provider, and 70% to the targeted advertising providers. As a result, the company needs a model that will allow you to set the proportions of funding for these types of advertising in accordance with the interests of users.

The effectiveness of advertising policy based on the model built can be estimated using the lift coefficient [9]:

$$\text{lift} = \frac{P(A \cap B)}{P(A) \cdot P(B)}, \quad (2)$$

where  $P(A)$  and  $P(B)$  – probabilities of interest in contextual and targeted ads, respectively;

joint probability  $P(A \cap B)$  – probability of interest in both types of ads from users.

The lift coefficient is an indicator of the effectiveness of targeted advertising and is used when predicting or classifying groups of users [10] who show increased interest in advertising. The model works well if the response within the target audience segment is much better than the average for the total number of users who were shown the ad.

Note that if we consider only 30% of the possible distribution for contextual advertising, this means that we are only interested in customers of deciles 1, 2, and 3. However, there may be a situation where the CTR value is also higher than the average for decile 4 (*Figure 1*). This approach is based on user classification and is similar to the idea of ABC analysis [11].

One approach to calculating the lift coefficient is to divide users into quantiles and rank the quantiles by the degree of lift. Next, you need to consider each quantile and, after weighing the predicted probability of response

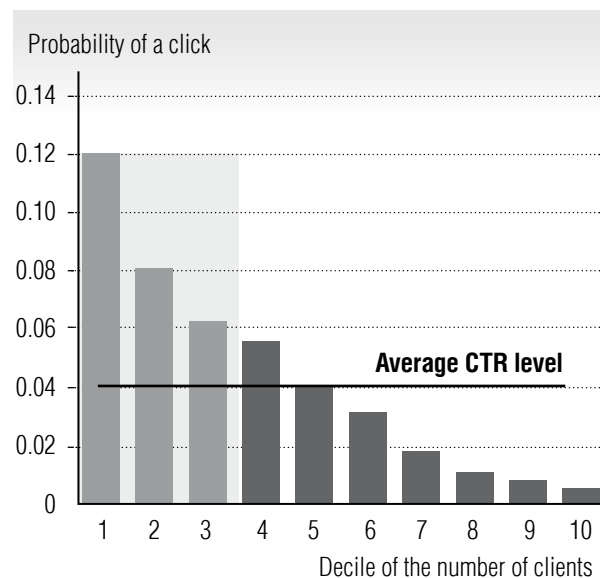


Fig. 1. The probability of a click for each decile of customers

(and the associated financial benefit) in relation to advertising costs, make a decision on the financing of the advertising campaign. This principle is described in one of the examples shown in *Figure 2*. The curve, designated as a “random model”, characterizes the situation of absolutely uniform distribution of interest in advertising among all users. This curve is called the absolute symmetry curve and means that there is no economic effect from advertising, since users show interest in ads in a random order, regardless of the seller’s actions, which is practically unattainable in real conditions [12, 13]. In this regard, it is possible to evaluate the economic effect that is received from users who have shown interest in advertising. The curves shown in *Figure 2* as “ideal model” and “normal model” characterize the effectiveness of advertising for different user segments. This model can be considered as a kind of receiver performance curve (ROC) [14, 15], which is also known as the Lorentz curve [5, 14].

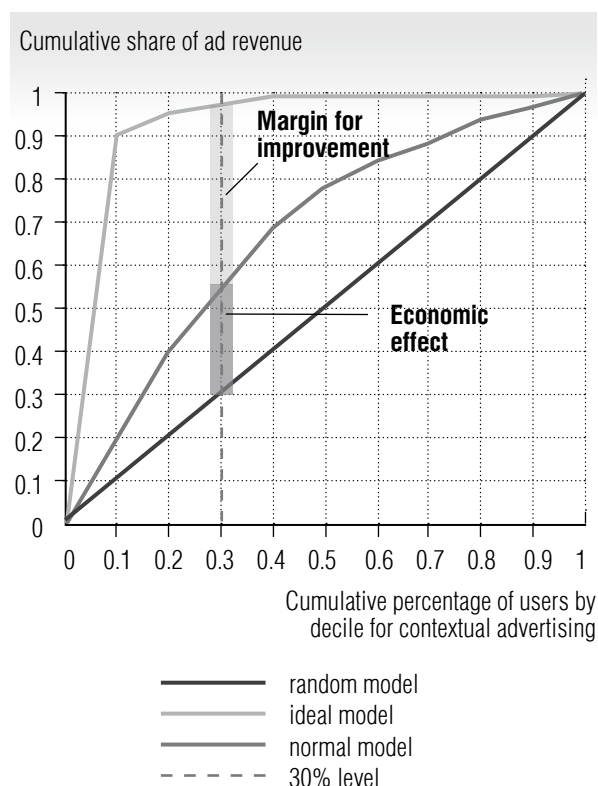


Fig. 2. Optimization of contextual advertising costs

The Lorentz curve for an ideal model describes the case when advertising is effective only for one small segment of users, who account for about 90% of the profit from ads, and the rest of the users who were shown the ad do not show interest in it. In this case, the cost of contextual advertising can be structured in a smaller proportion in order to find new interested customers using targeted advertising.

If the maximum economic effect compared to the random model is achieved at the fourth decile, then the maximum of interested users is about 40% of their total number. This case characterizes the normal model, which is most often found in practice. Thus, the search for the maximum economic effect for a different number of interested users allows you to pre-structure advertising costs.

## 2. Research methods

It is advisable to divide users into deciles when there is no additional information about users that does not allow us to identify patterns in their behavior. It should be noted that in the case of contextual advertising, marketers have a fairly extensive array of information that characterizes the behavioral activity of product users. Therefore, the second approach to calculating the lift coefficient is to use cluster analysis [16, 17] to construct the Lorentz curve. In order to assess the uniformity of user interest in an ad, as well as to compare ads with their interests, it is necessary to proceed to cluster analysis of user reaction to ads [18].

Existing approaches [19] use the variance of differences between the test and training samples relative to the average level as an indicator of asymmetry. However, cluster analysis methods, being methods of machine learning without a teacher, require the use of other indicators of class symmetry breaking that are used in this paper.

The method developed includes the following steps:

**Step 1.** To assess the quality of customer clustering, you need to evaluate the number of clusters (groups of splitting) as well as the uniformity of clusters in terms of the number of customers included in them.

It should be noted that the number of client partitioning groups is unknown in advance, so it is not possible to use clustering based on the  $k$ -means method. Clustering using algorithms based on decision trees requires a test training sample. However, in the case of an ad campaign, this selection may change dynamically. Therefore, hierarchical clustering methods are used as the most appropriate method [17], which do not require a training sample and allow for dividing clients into groups based on characteristics.

Hierarchical clustering methods allow you to choose one of two options for joining:

1) Agglomerative clustering starts with  $n$  clusters, where  $n$  is the number of observations (each of them is assumed to be a separate cluster). The algorithm then tries to find and group the most similar data points;

2) Divisional clustering is performed in the opposite way: initially, it is assumed that all  $n$  data points are one large cluster, after which the least similar ones are divided into separate groups.

At the same time, agglomerative clustering is better suited for identifying small clusters, while the use of divisional clustering is advisable for identifying large clusters. Since the assumed characteristics of clients are described by categorical variables, the Gower distance is used as the cluster separation metric [18].

**Step 2.** At this stage, building a Lorentz curve to assess the imbalance of user interest in ads. The Gini coefficient is often used as an indicator of the degree of imbalance in the manifestation of a feature [20, 21]. *Figure 3* shows an example showing the dependence of the share of points in the  $i$ -th cluster (on the total number of points in the sample) on the cumulative share of the number of clusters.

For example, for four clusters, the share of the first cluster will be 0.25 (25%). This cluster will contain 25 points out of 100, so the graph will display a point (0.25; 0.25). If all clusters have the same number of points, then there is absolute symmetry in the partition groups and the Gini coefficient is zero. Accordingly, the imbalance is described by the area of the bounded Lorentz polyline and the absolute symmetry curve and is calculated by the formula:

$$G = 1 - \sum_{k=1}^n (X_k - X_{k-1}) \cdot (Y_k + Y_{k-1}), G \in [0; 1], \quad (3)$$

where  $n$  – number of clusters;

$X_k$  – cumulative percentage of the number of clusters;

$Y_k$  – cumulative percentage of the number of points in the cluster.

The greater the value of the Gini coefficient deviates from zero, the greater the asymmetry in the characteristics of clusters [21–23].

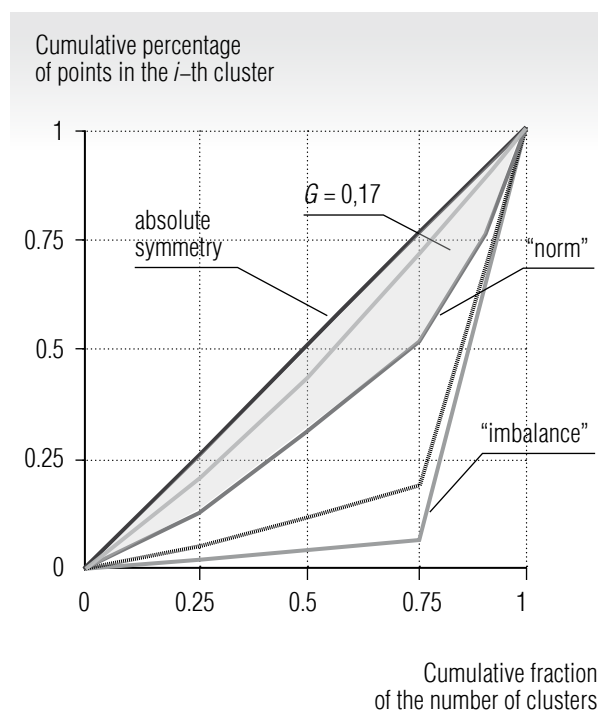


Fig. 3. Interpretation of the Gini coefficient in clustering problems

Calculating the Gini coefficient makes it possible to find the best match between the clustering options for products and customers, which helps to increase the customer orientation of products and increase the effectiveness of advertising. As a result, switch from user classification methods to clustering, which will simultaneously allow you to make more precise settings for both contextual and targeted advertising. However, to do this, it is necessary to evaluate the quality of grouping into clusters using the Gini coefficient.

Figure 4 shows an example that compares several Lorentz curves for four, five, and ten clusters. It is shown that in this case, with five clusters, not only the best quality of splitting users into groups is observed, but it is also possible to conclude that users of the first cluster provide the maximum increase in advertising revenue.

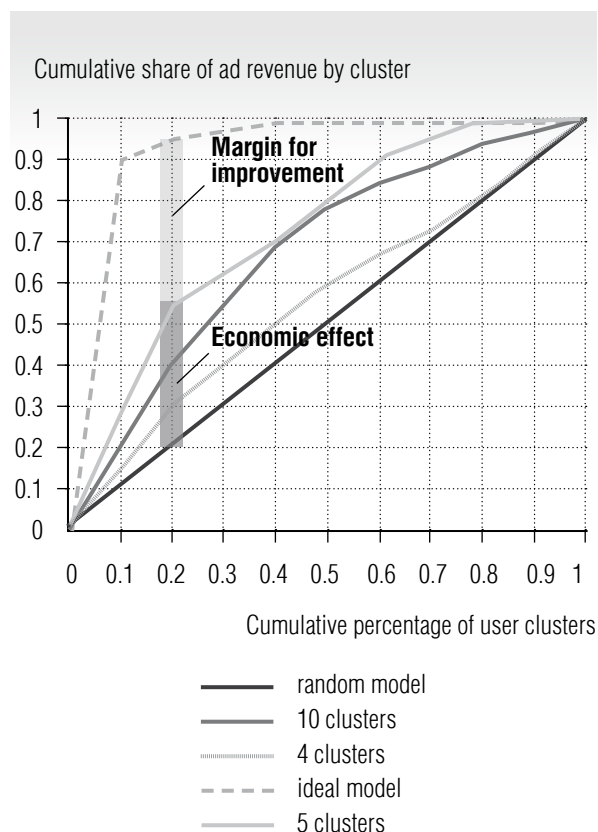


Fig. 4. Example of structuring advertising costs based on cluster analysis

**Step 3.** At the final stage, it is necessary to determine at what value of the cumulative share of clusters the maximum lift coefficient is observed, which allows us to conclude that the part of users for whom contextual advertising is more effective is allocated, according to which the share of contextual advertising in the total cost is set.

### 3. Example of application of the proposed approach

Using the Gini index and the cluster approach, you can calculate how much it is possible to improve advertising revenue in the same conditions by using the methodology described in this paper and compare it with the results obtained in the absence of any policy for structuring advertising costs. It should be noted that this method does not allow you to improve *CTR* in general, but it allows you to optimize the cost of advertising a product with the current number of clicks and purchases.

Let's take a concrete example of how this approach to evaluating the effectiveness of advertising works. First of all, the source data was modeled using the *R* language using the "dunif" and "dbinom" batch functions. Modeling was performed on the basis of various distribution functions that characterize the appearance of a particular trait. The synthesized test sample consisted of 10 thousand points, each of which describes the user's action in accordance with the following criteria:

- ◆ unique identifier of the action, type "string" – sequential numbering;
- ◆ date and time, type "object" – discrete uniform distribution within the range from the start date to the end date;
- ◆ user's operating system, type "string" – discrete uniform distribution across four types of operating systems;
- ◆ the user's browser, type "string" – the discrete uniform distribution on the six kinds of browsers;
- ◆ country, type "string" – discrete uniform distribution across nine countries;

♦ referral link type, type “string” – discrete uniform distribution across five types of links corresponding to different parts of the site where banners are placed;

♦ banner name, type “string” – discrete uniform distribution;

♦ action (interest or lack of interest), type “binary int” – binomial distribution with a purchase probability of 0.05;

♦ buy (or no buy), type “binary int” – binomial distribution with a purchase probability of 0.02.

As one of the simplifications, it is assumed that the cumulative income from the purchase is measured in relative units – the probability of making a purchase of one unit of identical goods.

The next step is to select the users who responded to the ads. In this case, there were 453 people ( $CTR = 4.53\%$ ), of whom 50 people made purchases. Then you need to evaluate

the uniformity across clusters of interested customers who have made purchases.

The next step is to perform client clustering using hierarchical methods [22], using two algorithms: based on divisional and agglomerative clustering.

Indicators of the sum of squares of distances between points within the cluster and the average width of the silhouette [24, 25] allow us to assess the quality of clustering. For the sum of squared distances, the “elbow bend” method is used [22, 26] to determine the optimal number of clusters, and the local maximum of the silhouette width value allows you to select the number of clusters with the best separation. Thus, the optimal number of partitioning groups – clusters is five for the agglomerative and eight for the divisional clustering algorithm (Figure 5). In addition, it is possible to evaluate

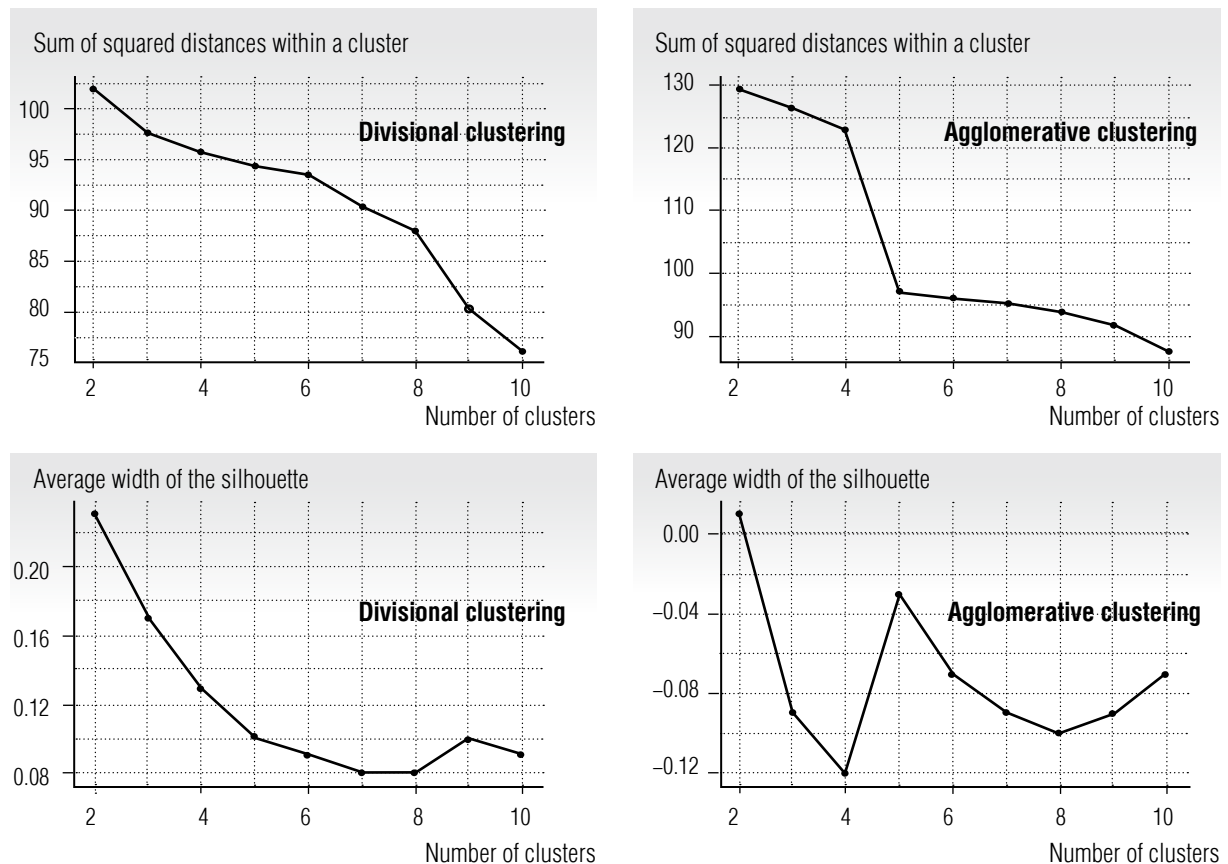


Fig. 5. Indicators for evaluating the quality of clustering

the intra-cluster variety of user actions regarding purchases. To do this, you need to compare how advertising costs can be structured based on the Gini coefficient. The indicators that characterize cluster diversity for splitting into 5 and 8 clusters are shown in *Table 1*.

The threshold value of the buyers share (*Figure 6*), which characterizes the maximum economic effect (lift coefficient) from advertising, for five clusters varies from 0.66 to 0.71 for agglomerative and divisional clustering algorithms, respectively. This means that for optimal structuring of advertising costs, it should be taken into account that the majority of users (about 80%) belonging to cluster 2 (*Table 1*) do not have clear intentions and signs of actions related to the purchase of a product, i.e. they most likely bought it spontaneously [2, 27],

searching by needs. Therefore, in this example, about 70% of advertising costs should be given to contextual advertising, which will allow you to target only interested users, while the remaining 30% should be given to targeted advertising to attract new customers.

However, in the case of five clusters, the area under the curve is significantly smaller than in the case of eight clusters, when there is a more detailed partition. It should be noted that the agglomerative clustering algorithm provides a large area, which in this case shows the best results, despite the increase in the number of clusters. The threshold values for the share of buyers for both algorithms have not changed much, which indicates that the results are balanced and reliable. However, the very value of the lift coefficient has grown significantly and

*Table 1.*

### Results of hierarchical clustering of interested users

Cluster No	1	2	3	4	5	6	7	8
Divisional clustering								
Percentage of the total number of interested customers	0.071	0.717	0.082	0.029	0.040	0.035	0.015	0.011
Probability of buying	0.04	0.9	0.04	0	0.02	0	0	0
Percentage of the total number of interested customers	0.071	0.717	0.168	0.029	0.015	–	–	–
Probability of buying	0.04	0.82	0.1	0	0.04	–	–	–
Agglomerative clustering								
Percentage of the total number of interested customers	0.2649	0.6225	0.0442	0.0265	0.0155	0.0110	0.0110	0.0044
Probability of buying	0.1	0.86	0.02	0	0	0.02	0	0
Percentage of the total number of interested customers	0.265	0.660	0.044	0.026	0.004	–	–	–
Probability of buying	0.16	0.76	0.04	0.04	0	–	–	–

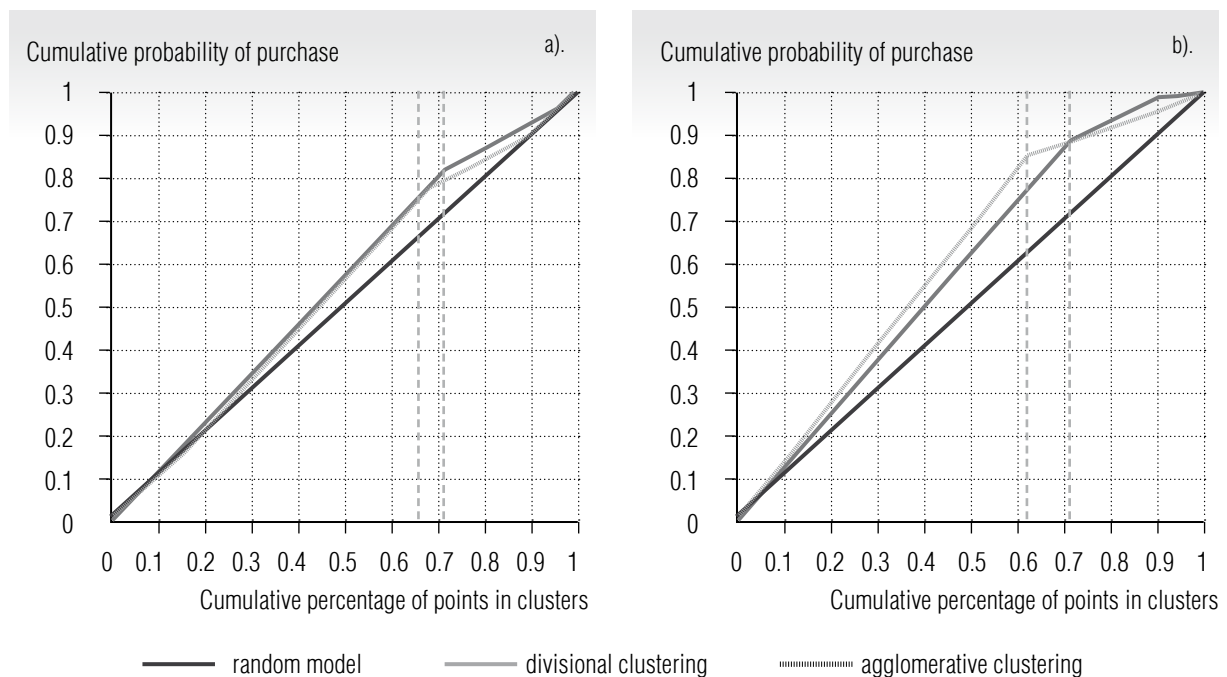


Fig. 6. Example of structuring advertising costs based on hierarchical clustering for five (a) and eight (b) clusters

it can be judged that the maximum economic effect of advertising is achieved by financing contextual advertising in the amount of 62% of total costs. This clarifies the characteristics of users who are most likely to buy the product, and the probability of purchase increases from 82% to 90%.

#### 4. Discussion of the proposed approach

In order to evaluate the possibilities of the proposed approach, let's look at the diagram that shows the results of structuring the costs of an advertising campaign based on the ranking of customers by deciles (Figure 7).

Figure 7 shows that the cost structure for contextual and tagged advertising by ranking by decile differs markedly from the cluster approach. In the case of ranking clients by deciles, the maximum lift coefficient was obtained for 20% of clients, which indicates that contextual advertising is financed in the

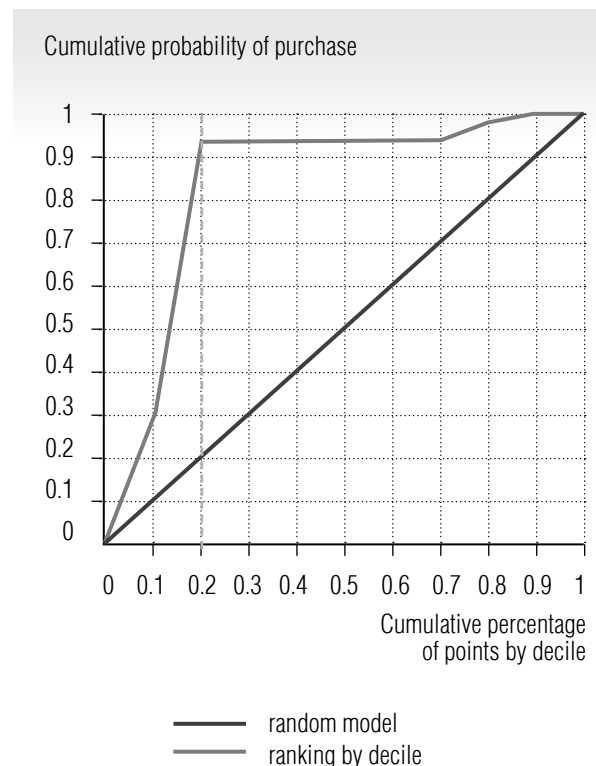


Fig. 7. Results of structuring advertising campaign costs by ranking by decile

amount of only 20% of the total cost. Despite the fact that this graph is closer to the ideal case, this model has a minimal margin for improvement. In addition, it is impossible to correctly compare the obtained Lorentz curves for the cumulative fraction of points by deciles and the cumulative fraction of clusters, since in the first case, ranking was performed, and in the second – feature clustering. For the same reason, the areas under the Lorentz curves obtained in *Figures 6 and 7* cannot be compared. However, the lift coefficient in both cases can be used to determine the proportions of advertising costs. The noticeable differences are explained by the fact that in the case of ranking by deciles, only the facts of the transaction are taken into account, but this approach does not take into account the characteristics of customers. The main advantage of the proposed method is the ability to more flexibly configure the structure of advertising costs depending on the signs of customer behavioral activity. The cluster approach allows you to build a more appropriate model and configure contextual advertising more precisely.

### Conclusion

Analysis of the results allows us to draw the following conclusions.

1. To expand the capabilities of approaches to structuring banner advertising costs, it is necessary to use clustering algorithms based on categorical characteristics of user actions.
2. Hierarchical clustering methods are well suited for estimating the required number of clusters, and also make it possible to identify hidden patterns in customer behavioral activity.
3. The Gini coefficient makes it possible to

evaluate the quality of clustering and determine the user groups that give the maximum purchase probability.

4. Using the cluster approach allows you not only to structure advertising costs, but also to determine which type of advertising should be applied to which users. This gives more opportunities to optimize costs and increase the effectiveness of your advertising campaign.

Identifying hidden patterns in consumer groups allows you to identify the main directions of product development and customer interest in the product as well as assess the stability of the market for products with similar characteristics and the stability of its development.

The results obtained reveal the applied possibilities of using the principle of symmetry breaking in business tasks, and, in contrast to existing works [6, 27], they reflect the possibilities of structuring the costs of an advertising campaign. This approach allows you not only to identify the popularity of products by characteristics, but also to determine the most effective ways to attract customers for a particular type of product. This is achieved by comparing the results of consumers' behavioral activity in relation to their performance of target actions, as well as the characteristics of the products for which they performed these actions. Also, one of the advantages is the ability to use the Gini uncertainty and the lift coefficient as indicators of user groups for which contextual or targeted advertising is more effective.■

### Acknowledgement

The study was supported by RFBR grant No 18.07.00170.

### References

1. Kislyakov A.N. (2019) Evaluation of the effectiveness of advertising campaigns in social networks using simulation methods. *Economics and Management: Problems, Solutions*, vol. 5, no 3, pp. 20–26 (in Russian).

2. Rau V.G., Kislyakov A.N., Tikhonyuk N.E. Rau T.F. (2018) The problem of asymmetry in the models of economic systems development. *Proceedings of the XI International Scientific and Practical Conference "Regional Economy: Experience and Challenges", 15 May 2018*. Eds. A.I. Novikov, A.E. Illarionov). Vladimir: RANEPa, Vladimir Branch, pp. 201–211 (in Russian).
3. Rau V.G., Polyakov S.V., Rau T.F., Firsov I.V., Togunov I.A. (2019) Some features of application of broken symmetry groups for "visualization" of processes in natural, "living" and socio-economic systems. *Proceedings of the XI International Scientific and Practical Conference "Regional Economy: Experience and Challenges", 15 May 2018*. Eds. A.I. Novikov, A.E. Illarionov). Vladimir: RANEPa, Vladimir Branch, pp. 111–119 (in Russian).
4. Kislyakov A.N., Tikhonyuk N.E. (2019) Model of price formation of a homogeneous market taking into account the asymmetry of information. *Innovative Development of Economy*, no 1, pp. 93–100 (in Russian).
5. Perskii Yu.K., Dmitriev D.V. (2009) Formation of the information-economic mechanism of information asymmetry level management at the regional branch market. *Bulletin of SUSU. Series: Economics and Management*, no 29, pp. 66–74 (in Russian).
6. Kislyakov A.N. (2020) Asymmetry of information in the analysis of socio-economic processes. *Vestnik NSUEM*, no 1. pp. 64–75 (in Russian). DOI: 10.34020/2073-6495-2020-1-064-075.
7. Baranovskaya T.P., Ivanova E.A., Khachak F.R. (2016) The automated subsystem for advertising budget planning. *Scientific Journal of KubSAU*, no 120, pp. 223–238 (in Russian).
8. Barnés M. (2020) *Calculate the economic impact of your click-through prediction*. Available at: <https://towardsdatascience.com/calculate-the-economic-impact-of-your-click-through-prediction-1fa6eee64494> (accessed 25 April 2020).
9. Galyamov A.F., Tarkhov S.V. (2014) Customer relationship management of a commercial organization based on methods of segmentation and clustering of customer database. *Vestnik USATU*, vol. 18, no 4, pp. 149–156 (in Russian).
10. Andreeva A.V. (2012) Optimal control of a company's customer base using the customer lifetime value parameter. *Business Informatics*, no 4, pp. 61–68 (in Russian).
11. Tsoy M.E., Zaleshin P.A. (2017) Consumer segmentation on the basis of the study of consumer behavior styles. *Rossiyskoe Predprinimatelstvo*, vol. 18, no 21, pp. 3313–3326 (in Russian). DOI: 10.18334/rp.18.21.38543.
12. Mishra B.K., Hazra D., Tarannum K., Kumar M. (2016) Business Intelligence using Data Mining techniques and Business Analytics. *Proceedings of the 5th International Conference on System Modeling & Advancement in Research Trends (SMART 2016), 25–27 November 2016, Moradabad, India*, pp. 84–89. DOI: 10.1109/SYSMART.2016.7894496.
13. James G., Witten D., Hastie T., Tibshirani R. (2013) *An introduction to statistical learning with applications in R*. New York: Springer.
14. Nielsen F. (2016) *Introduction to HPC with MPI for data science*. Springer.
15. Hastie T., Tibshirani R., Friedman J. (2017) *The elements of statistical learning: Data mining, inference, and prediction*. Second Edition. Springer.
16. Kassambara A. (2017) *Practical guide to cluster analysis in R: Unsupervised machine learning. Multivariate analysis I*. STHDA.
17. Tripathi S., Bhardwaj A., E P. (2018) Approaches to clustering in customer segmentation. *International Journal of Engineering & Technology*, vol. 7, no 3.12, pp. 802–807. DOI: 10.14419/ijet.v7i3.12.16505.
18. Gower J.C. (1971) A general coefficient of similarity and some of its properties. *Biometrics*, vol. 27, no 4, pp. 857–871.

19. Kislyakov A.N. (2020) Indicators of asymmetry in the tasks of studying the behavioral activity of product users. *Izvestia Sankt-Petersburgskogo Gosudarstvennogo Ekonomicheskogo Universiteta*, no 3, pp. 110–116 (in Russian).
20. Frunza M.-Ch. (2013) Computing a standard error for the Gini coefficient: An application to credit model validation. *Journal of Risk Model Validation*, vol. 7, no. 1, pp. 61–82. DOI: 10.21314/JRMV.2013.099.
21. Zorina A.A. (2014) The formation of fluctuating asymmetry during individual development of Beluta pendula. *Principles of the Ecology*, no 4, pp. 27–46 (in Russian).
22. Murtagh F., Contreras P. (2011) Methods of hierarchical clustering. *ArXiv*. Available at: <https://arxiv.org/pdf/1105.0121.pdf> (accessed 30 March 2020).
23. Korolev O.L., Coussy Yu.M., Segal A.V. (2013) *Application of entropy in modeling decision-making processes in the economy*. Ed. A.V. Segal. Simferopol: OJAK Publishing House (in Russian).
24. Pechenyi E.A., Nuriev N.K., Starygin S.D. (2019) Dynamic clustering of big data flow. Proceedings of the *International Scientific Conference “Mathematical Methods in Technics and Technologies”*, vol. 3, Ed. A.A. Bolshakov). Saint-Petersburg: Polytechnical Institute Publishing House, pp. 19–21 (in Russian).
25. Prokofyeva E.S., Zaytsev R.D. (2020) Clinical pathways analysis of patients in medical institutions based on hard and fuzzy clustering methods. *Business Informatics*, vol. 14, no 1, pp. 19–31. DOI: 10.17323/2587-814X.2020.1.19.31.
26. Yakimov V.N., Shurganova G.V., Cherepennikov V.V., Kudrin I.A., Il'in M.Yu. (2016) Methods for comparative assessment of the results of cluster analysis of hydrobiocenoses structure (by the example of zooplankton communities of the Linda River, Nizhny Novgorod Region). *Inland Water Biology*, vol. 9, no 2, pp. 200–208. DOI: 10.7868/S0320965216020169.
27. Baranov S.G., Burdakova N.E. (2015) *Assessment of development stability. Methodological approaches*. Vladimir: Vladimir State University Publishing House (in Russian).

### About the author

#### Alexey N. Kislyakov

Cand. Sci. (Tech.);

Associate Professor, Department of Information Technologies, Russian Academy of National Economy and Public Administration under the President of the Russian Federation, Vladimir Branch, 59a, Gorky Street, Vladimir 600017, Russia;

E-mail: [ankislyakov@mail.ru](mailto:ankislyakov@mail.ru)

ORCID: 0000-0001-8790-6961