

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

Структурирование затрат на рекламную кампанию с учетом асимметрии интереса пользователей

А.Н. Кисляков E-mail: ankislyakov@mail.ru

Российская академия народного хозяйства и государственной службы при Президенте Российской Федерации. Владимирский филиал

Адрес: 600017, г. Владимир, ул. Горького, д. 59а

Аннотация

Работа посвящена актуальной проблеме структурирования затрат на контекстную и таргетированную рекламу в сети интернет. Выбор структуры финансирования рекламной кампании рассматривается с позиции нарушения принципа симметрии интереса пользователей к рекламным объявлениям. Целью работы является разработка методики структурирования затрат на рекламную кампанию на основе кластерного анализа с учетом асимметрии интереса пользователей к рекламе. Ключевой особенностью проводимого исследования является описание возможности использования асимметрии интереса пользователей в прикладных решениях, таких как интернет-реклама. В качестве показателя, характеризующего степень дисбаланса проявления какого-либо признака, при кластеризации используется коэффициент Джини. Также рассматриваются особенности применения коэффициента подъема и кривой Лоренца для оценки эффективности контекстной и таргетированной рекламы для различных групп клиентов. Использование индекса Джини и кластерного анализа позволяет проанализировать возможности повышения дохода от рекламы и сравнить его с показателями, достигаемыми при отсутствии какой-либо политики структурирования рекламных затрат. Выявление подобных закономерностей в группах потребителей позволяет наметить основные направления развития продукта и интерес покупателей к нему. Описанную методику следует применять в целях повышения эффективности использования баннерной рекламы и алгоритмов кластеризации. Указанный подход не позволяет улучшить показатель кликабельности баннеров, но позволяет реализовать индивидуальный подход к рекламе продуктов с текущим количеством кликов, а также более эффективно структурировать затраты на различные виды рекламы.

Ключевые слова: информационная асимметрия; индекс Джини; кластерный анализ; баннерная реклама; иерархическая кластеризация.

Цитирование: Кисляков А.Н. Структурирование затрат на рекламную кампанию с учетом асимметрии интереса пользователей // Бизнес-информатика. 2020. Т. 14. № 4. С. 7–18. DOI: 10.17323/2587-814X.2020.4.7.18

Введение

На сегодняшний день одним из наиболее динамично развивающихся сегментов рекламной деятельности является реклама в сети интернет. Например, баннерная реклама позволяет более точно и эффективно донести рекламное объявление до заинтересованного клиента. При этом существует достаточно много вариантов монетизации показа рекламных объявлений в зависимости от назначения и возможностей.

Такой вид рекламы, разумеется, требует определенных вложений, которые не всегда бывают оправданы, поскольку распространение информации и технологии подачи рекламных объявлений, а также их стоимость зависят от большого количества факторов и носят сложный случайный характер [1]. Зачастую трудно понять, какая именно категория пользователей окажется заинтересована в рекламном объявлении, еще труднее спрогнозировать рост продаж в зависимости от вложений в рекламу.

В этой связи существует два принципиально разных вида рекламы: контекстная и таргетированная. Контекстная реклама позволяет выполнять автоматизированный показ рекламного объявления в соответствии с предметной областью поиска клиентом товаров. Таргетированная реклама, напротив выполняет поиск аудитории под предложение по признакам, что является более сложной задачей. Несмотря на то, что и тот и другой вид рекламы имеет тонкие настройки показа, эффективность контекстной рекламы зачастую оказывается выше, ввиду работы с заинтересованной аудиторией.

Однако компании также заинтересованы в расширении своей целевой аудитории. Поэтому возникает задача определения структуры финансирования рекламной кампании по видам в зависимости от видов бизнеса и особенностей поведенческой активности клиентов, как заинтересованных, так и потенциальных.

Зачастую целесообразно использовать оба вида рекламы, вместо того, чтобы отдавать предпочтение одному виду рекламы за счет другого. Поэтому цель исследования – разработка методики структурирования затрат на рекламную кампанию, учитывающей нарушение принципа симметрии интереса пользователей к рекламе.

Выбор структуры финансирования рекламной кампании не столь очевиден, как может показаться на первый взгляд. Он обусловлен явлением информационной асимметрии [2, 3] на рынке онлайн-про-

даж товаров и услуг. В этом случае продавцы ведут бизнес, не владея полной информацией о конкурентной среде, а также намерениях покупателей [1, 4]. В свою очередь, покупатели формируют свое мнение о товаре или услуге на основе различного набора факторов и источников, постоянно прислушиваются к мнению сетевых сообществ, читая статьи, отзывы в сети интернет, следуя за лидерами мнений и т.п. Таким образом, взаимодействие между рекламодателем и пользователем продукта усложняется с учетом информационной асимметрии рынка. В качестве основной гипотезы принимается связь явления рыночной асимметрии с дисбалансом в поведенческой активности групп клиентов [5, 6].

Существуют различные методики структурирования и планирования затрат на рекламную кампанию с учетом предпочтений целевой аудитории [7]. При этом используются различные офлайн- и онлайн-инструменты, например, анкетирование покупателей, либо выбор одного из видов рекламы (контекстной или таргетированной) в зависимости от цели организации (вывод на рынок нового продукта, наращивание целевой аудитории и т.д.). С одной стороны, эти подходы позволяют значительно упростить процесс планирования, но с другой стороны не позволяют гибко настраивать и эффективно управлять рекламной кампанией. Предлагаемый подход является одной из модификаций алгоритма структурирования затрат на рекламную кампанию на основе оценки экономического эффекта кликабельности и использования методов классификации [8]. В частности, предлагается использование методов кластерного анализа для создания более адекватной модели структурирования затрат на рекламу.

1. Предлагаемый подход к структурированию затрат на рекламу

Механизм контекстной рекламы представляет собой автоматизированную сделку по реализации рекламы, а полезность показа рекламного баннера измеряется с помощью показателя кликабельности CTR (Click Through Rate) [1, 9]:

$$CTR = \frac{\text{количество кликов}}{\text{количество показов}} \cdot 100\%. \quad (1)$$

Когда компания знает о предпочтениях части своих клиентов, она может обратиться к одному из провайдеров объявлений, который реализует механизм контекстной рекламы для соответствующих пользователей. Такие объявления менее эффективны при

поиске и расширении целевой аудитории, поскольку показываются только тем клиентам, которые уже заинтересованы в покупке данного товара.

В случае таргетированной рекламы владельцы баннеров получают оплату по количеству кликов и по показам, когда пользователь видит баннер, но не нажимает на него. Способ монетизации баннеров заключается в их продаже на аукционе объявлений, где рекламодатели делают ставки на эти баннеры при различном количестве участников аукциона и условиях размещения. Таргетированная реклама позволяет привлечь больше целевой аудитории, однако является менее точным, хотя и более дешевым инструментом интернет-маркетолога.

Если бюджет рекламной кампании ограничен, то эти средства могут быть распределены между поставщиками рекламы, к примеру, следующим образом: 30% средств – провайдеру контекстной рекламы, а 70% – поставщикам таргетированной рекламы. В итоге компании необходима модель, которая позволит установить пропорции финансирования указанных видов рекламы в соответствии с интересами пользователей.

Эффективность проведения рекламной политики с учетом построенной модели может быть оценена с помощью коэффициента подъема [9]:

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

где $P(A)$ и $P(B)$ – вероятности интереса к контекстной и таргетированной рекламе соответственно;

совместная вероятность $P(A \cap B)$ – вероятность интереса обоим видам рекламных объявлений со стороны пользователей.

Коэффициент подъема является показателем эффективности таргетированной рекламы и используется при прогнозировании или классификации групп пользователей [10], проявляющих повышенный интерес к рекламе. Модель хорошо работает, если реакция в рамках целевого сегмента аудитории намного лучше, чем в среднем для общего количества пользователей, которым была показана реклама.

Отметим, что если учитывать только 30% возможного распределения для контекстной рекламы, то это означает, что нас интересуют только клиенты децилей 1, 2 и 3. Однако может сложиться ситуация, когда значение CTR выше среднего также и для дециля 4 (рисунки 1). Данный подход основан на классификации пользователей и аналогичен идее ABC-анализа [11].



Рис. 1. Вероятность клика для каждого дециля клиентов

Один из подходов к расчету коэффициента подъема заключается в разделении пользователей на квантили и ранжировании квантилей по степени подъема. Далее необходимо рассмотреть каждый квантиль и, взвесив прогнозируемую вероятность отклика (и связанную с ней финансовую выгоду) по отношению к затратам на рекламу, принять решение о финансировании рекламной кампании. Описание данного принципа на одном из примеров показано на рисунке 2. Кривая, обозначенная как «случайная модель», характеризует ситуацию абсолютно равномерного распределения интереса к рекламе у всех пользователей. Эта кривая называется кривой абсолютной симметрии и означает отсутствие экономического эффекта от рекламы, поскольку пользователи проявляют интерес к рекламным объявлениям в случайном порядке, вне зависимости от действий продавца, что в реальных условиях практически недостижимо [12, 13]. В этой связи существует возможность оценить экономический эффект, который получен от пользователей, проявивших интерес к рекламе. Кривые, обозначенные на рисунке 2 как «идеальная модель» и «нормальная модель», характеризуют эффективность рекламы для различных сегментов пользователей. Такая модель может рассматриваться как разновидность кривой рабочей характеристики приемника (ROC) [14, 15], которая также известна под названием кривой Лоренца [5, 14].

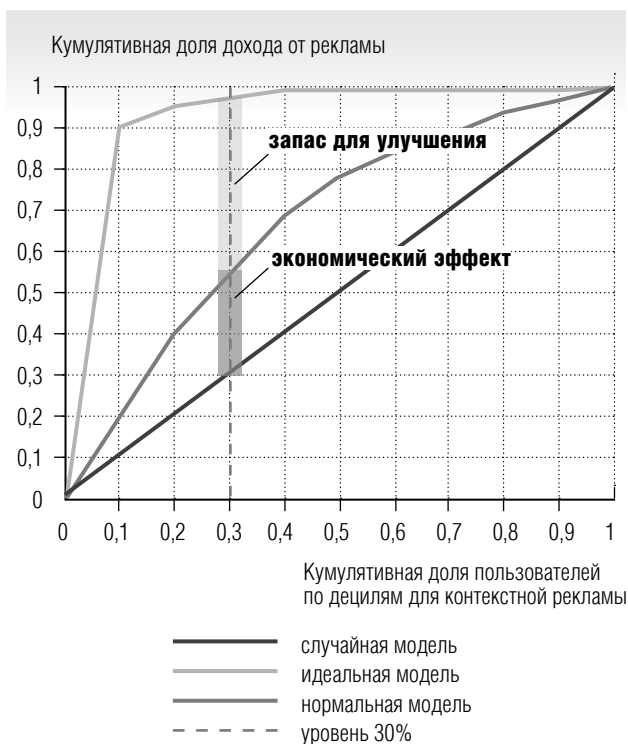


Рис. 2. Оптимизации затрат на контекстную рекламу

Кривая Лоренца для идеальной модели характеризует случай, когда реклама эффективна строго для одного небольшого сегмента пользователей, на которых приходится около 90% прибыли от рекламных объявлений, а остальные пользователи, которым была показана реклама, не проявляют к ней интереса. В этом случае затраты на контекстную рекламу могут быть структурированы в меньшей пропорции в целях поиска новых заинтересованных клиентов с помощью таргетированной рекламы.

Если максимальный экономический эффект по сравнению со случайной моделью достигается на четвертом дециле, то максимум заинтересованных пользователей составляет около 40% от их общего количества. Это случай характеризует нормальную модель, которая наиболее часто встречается на практике. Таким образом, поиск максимального экономического эффекта для различного количества заинтересованных пользователей позволяет предвзительно структурировать затраты на рекламу.

2. Методика исследования

Разделение пользователей на децили целесообразно выполнять тогда, когда отсутствует дополнительная информация о пользователях, не позво-

ляющая выявить закономерности в их поведении. Следует отметить, что в случае контекстной рекламы маркетологи обладают достаточно обширным массивом сведений, характеризующих поведенческую активность пользователей продукта. Поэтому второй подход к расчету коэффициента подъема состоит в использовании кластерного анализа [16, 17] для построения кривой Лоренца. Для того чтобы оценить однородность интереса пользователей к рекламному объявлению, а также сопоставить объявления с их интересами, необходимо перейти к кластерному анализу реакции пользователей на рекламные объявления [18].

Существующие подходы [19] используют в качестве показателя асимметрии дисперсию различий между тестовой и обучающей выборками относительно среднего уровня. Однако методы кластерного анализа, являясь методами машинного обучения без учителя, требуют использования иных показателей нарушения симметрии классов, которые используются в настоящей работе.

Разработанная методика включает в себя следующие этапы:

Этап 1. Для оценки качества кластеризации покупателей необходимо оценить количество кластеров (групп разбиения), а также однородность кластеров с точки зрения количества входящих в них клиентов.

Следует отметить, что количество групп разбиения клиентов заранее неизвестно, поэтому использовать кластеризацию на основе метода k -средних не представляется возможным. Кластеризация с использованием алгоритмов на основе деревьев решений требует тестовой обучающей выборки. Однако в случае рекламной кампании эта выборка может динамически изменяться. Поэтому в качестве наиболее подходящего метода в работе используются иерархические методы кластеризации [17], не требующие обучающей выборки и позволяющие осуществлять разбиение клиентов на группы по признакам.

Иерархические методы разбиения на кластеры позволяют выбрать один из двух вариантов объединения:

1) Агломеративная кластеризация начинается с n кластеров, где n — число наблюдений (предполагается, что каждое из них представляет собой отдельный кластер). Затем алгоритм пытается найти и сгруппировать наиболее схожие между собой точки данных;

2) Дивизионная кластеризация выполняется противоположным образом: изначально полагается, что все n точек данных представляют собой один большой кластер, после чего наименее схожие из них разделяются на отдельные группы.

При этом агломеративная кластеризация лучше подходит для выявления небольших кластеров, а применение дивизионной кластеризации целесообразно для выявления крупных кластеров. Поскольку предполагаемые признаки клиентов описываются категориальными переменными, в качестве метрики разделения кластеров используется расстояние Гоуэра [18].

Этап 2. На этом этапе необходимо построить кривую Лоренца для оценки дисбаланса интереса пользователей к рекламным объявлениям. В качестве показателя, характеризующего степень дисбаланса проявления какого-либо признака, часто используется коэффициент Джини [20, 21]. На *рисунке 3* представлен пример, показывающий зависимость доли точек в i -м кластере (от общего количества точек в выборке) от кумулятивной доли количества кластеров.

Например, для четырех кластеров доля первого кластера будет составлять 0,25 (25%). Этот кластер будет содержать в себе 25 точек из 100, поэтому на графике будет отображаться точка (0,25; 0,25). Если все кластеры имеют одинаковое количество точек, то наблюдается абсолютная симметрия в группах разбиения и коэффициент Джини равен нулю. Соответственно,

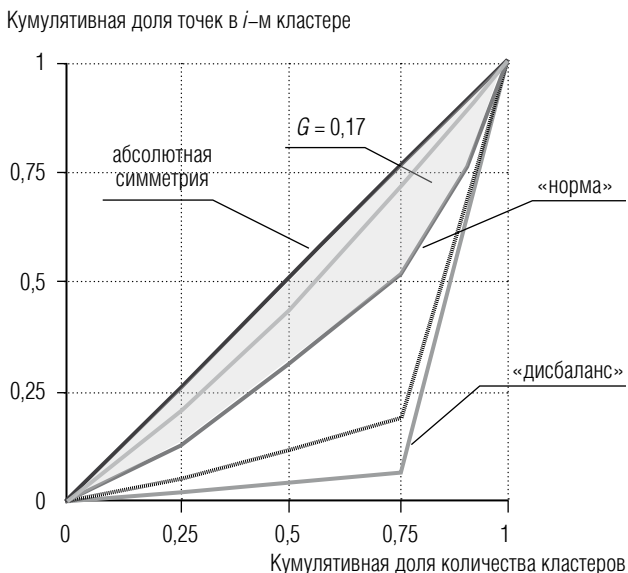


Рис. 3. Интерпретация коэффициента Джини в задачах кластеризации

дисбаланс описывается площадью ограниченной ломанной кривой Лоренца и кривой абсолютной симметрии и рассчитывается по формуле:

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

где n — количество кластеров;

X_k — кумулятивная доля количества кластеров;

Y_k — кумулятивная доля количества точек в кластере.

Чем больше значение коэффициента Джини отклоняется от нуля, тем в большей степени проявляется асимметрия в характеристиках кластеров [21–23]. Расчет коэффициента Джини дает возможность найти наилучшее соответствие между вариантами кластеризации продуктов и покупателей, что способствует повышению клиентоориентированности продуктов и повышению эффективности рекламы. В итоге необходим переход от методов классификации пользователей к их кластеризации, что одновременно позволит выполнить более точные настройки как контекстной, так и таргетированной рекламы. Однако для этого сначала необходимо оценить качество разбиения групп на кластеры с помощью коэффициента Джини.

На *рисунке 4* представлен пример, в котором рассматривается сравнение нескольких кривых Лоренца для четырех, пяти и десяти кластеров. Показано, что в данном случае при пяти кластерах не только наблюдается наилучшее качество разбиения пользователей на группы, но и имеется возможность сделать вывод о том, что пользователи первого кластера обеспечивают максимальный прирост прибыли от рекламы.

Этап 3. На завершающем этапе необходимо определить, при каком значении кумулятивной доли кластеров наблюдается максимальный коэффициент подъема, что позволяет сделать вывод о выделении той части пользователей, для которых более эффективна контекстная реклама, в соответствии с чем устанавливается доля контекстной рекламы в общем объеме затрат.

3. Пример применения предлагаемого подхода

С помощью индекса Джини и кластерного подхода можно рассчитать, насколько возможно улучшить доход от рекламы в одних и тех же условиях с помощью использования описанной в работе методики и

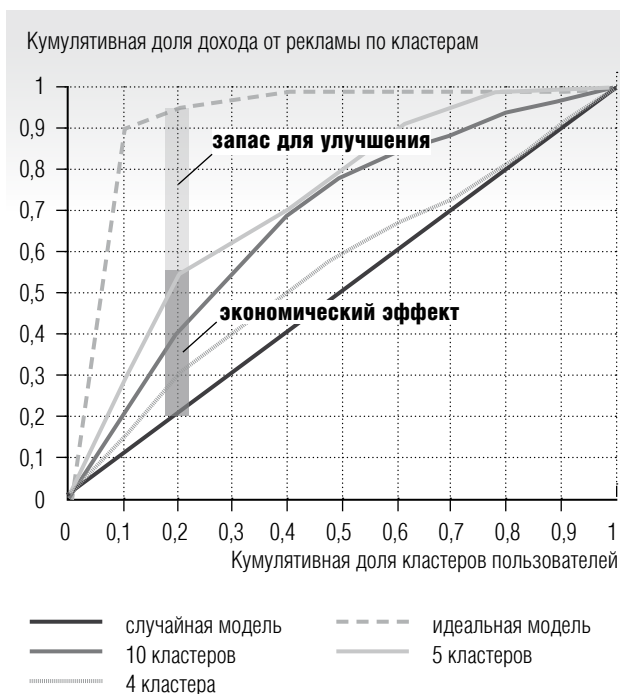


Рис. 4. Пример структурирования затрат на рекламу на основе кластерного анализа

сравнить его с результатами, получаемыми при отсутствии какой-либо политики структурирования рекламных затрат. Следует отметить, что указанная методика не позволяет улучшить *CTR* в целом, но позволяет оптимизировать затраты на рекламу продукта с текущим количеством кликов и покупок.

Рассмотрим на конкретном примере, как работает данный подход к оценке эффективности рекламы. В первую очередь, были смоделированы исходные данные средствами языка *R* с помощью пакетных функций *dunif* и *dbinom*. Моделирование выполнялось на основе различных функций распределения, характеризующих появление того или иного признака. Синтезированная тестовая выборка состояла из 10 тыс. точек, каждая из которых описывает действие пользователя в соответствии со следующими признаками:

- ◆ уникальный идентификатор действия, тип «string» — последовательная нумерация;
- ◆ дата и время, тип «object» — дискретное равномерное распределение в пределах от начальной до конечной даты;
- ◆ операционная система пользователя, тип «string» — дискретное равномерное распределение по четырем типам операционных систем;
- ◆ браузер пользователя, тип «string» — дискретное

равномерное распределение по шести наименованиям браузеров;

- ◆ страна, тип «string» — дискретное равномерное распределение по девяти странам;

- ◆ тип реферальной ссылки, тип «string» — дискретное равномерное распределение по пяти типам ссылок, соответствующих различным частям сайта, где размещаются баннеры;

- ◆ название баннера, тип «string» — дискретное равномерное распределение;

- ◆ действие (заинтересованность или отсутствие интереса), тип «binary int» — биномиальное распределение с вероятностью покупки 0,05;

- ◆ покупка (покупка или отсутствие покупки), тип «binary int» — биномиальное распределение с вероятностью покупки 0,02.

В качестве одного из упрощений принимается, что кумулятивный доход от покупки измеряется в относительных единицах — вероятностях совершения покупки одной единицы одинаковых товаров.

Следующим шагом является выбор пользователей, которые отреагировали на рекламные объявления. Таких в данном случае оказалось 453 человека (*CTR* = 4,53%), из которых совершили покупки 50 человек. Затем необходимо оценить однородность по кластерам заинтересованных клиентов, совершивших покупки.

На следующем шаге необходимо выполнить кластеризацию клиентов иерархическими методами [22], с помощью двух алгоритмов: на основе дивизионной и агломеративной кластеризации.

Показатели суммы квадратов расстояний между точками внутри кластера и средней ширины силуэта [24, 25] позволяют оценить качество кластеризации. Для суммы квадратов расстояний используется метод «локтя» [22, 26] для определения оптимального количества кластеров, а локальный максимум значения ширины силуэта позволяет выбрать количество кластеров с наилучшим разделением. Таким образом, оптимальное количество групп разбиения — кластеров равно пяти для агломеративного, и восьми — для дивизионного алгоритма кластеризации (рисунок 5). Кроме того, можно оценить внутрикластерное разнообразие действий пользователей относительно покупок. Для этого необходимо сравнить, как могут быть структурированы затраты на рекламу с учетом коэффициента Джини. Показатели, характеризующие кластерное разнообразие для разбиения на 5 и 8 кластеров, приведены в таблице 1.

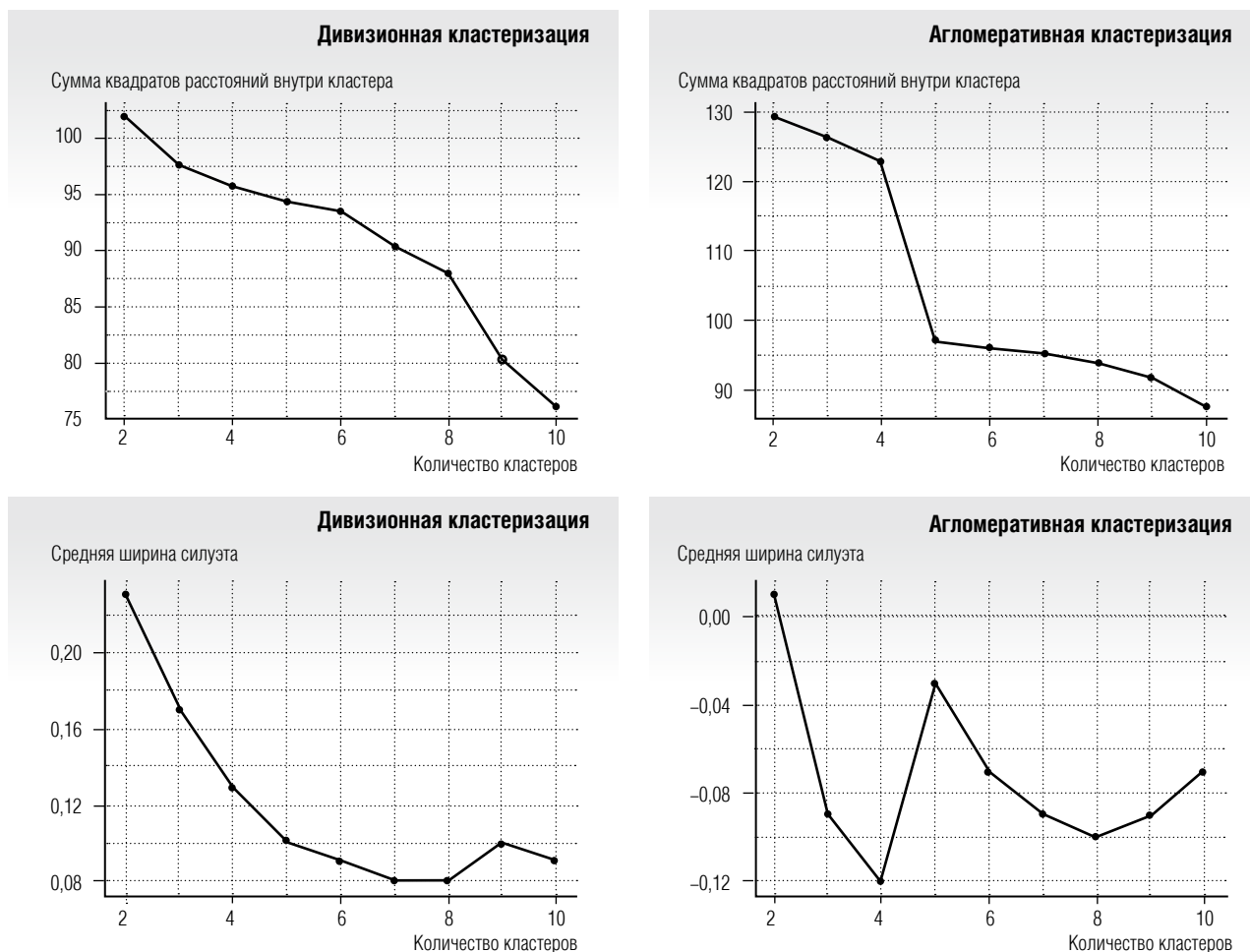


Рис. 5. Показатели оценки качества кластеризации

Результаты иерархической кластеризации заинтересованных пользователей

Таблица 1.

№ кластера	1	2	3	4	5	6	7	8
Дивизионная кластеризация								
Процент от общего количества заинтересованных клиентов	0,071	0,717	0,082	0,029	0,040	0,035	0,015	0,011
Вероятность покупки	0,04	0,9	0,04	0	0,02	0	0	0
Процент от общего количества заинтересованных клиентов	0,071	0,717	0,168	0,029	0,015	–	–	–
Вероятность покупки	0,04	0,82	0,1	0	0,04	–	–	–
Агломеративная кластеризация								
Процент от общего количества заинтересованных клиентов	0,2649	0,6225	0,0442	0,0265	0,0155	0,0110	0,0110	0,0044
Вероятность покупки	0,1	0,86	0,02	0	0	0,02	0	0
Процент от общего количества заинтересованных клиентов	0,265	0,660	0,044	0,026	0,004	–	–	–
Вероятность покупки	0,16	0,76	0,04	0,04	0	–	–	–

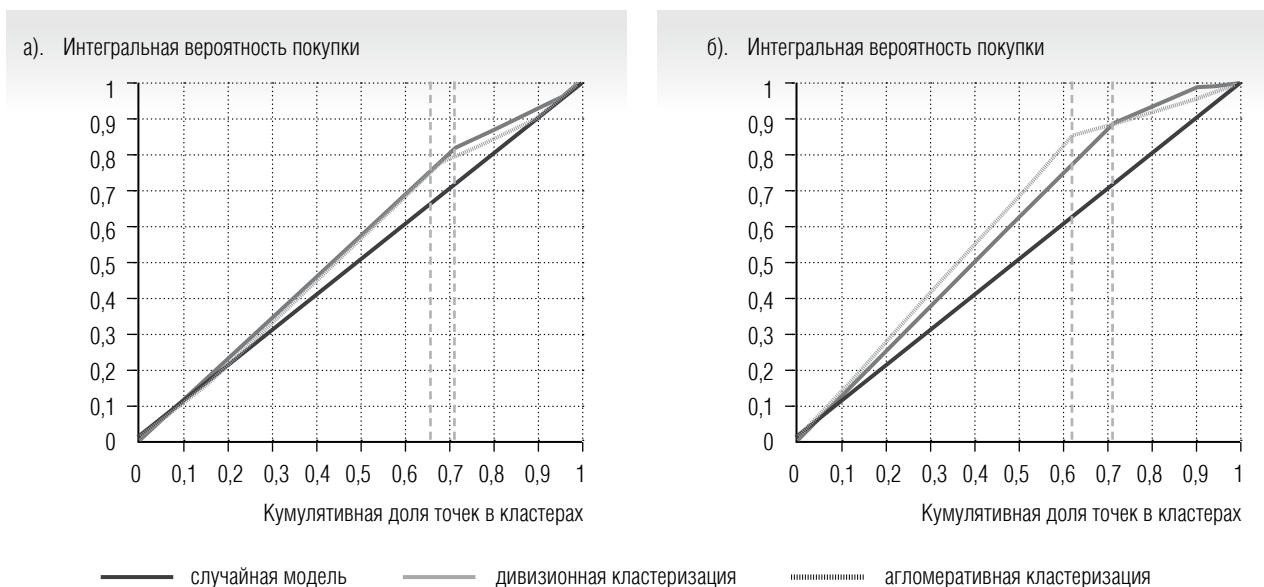


Рис. 6. Пример структурирования затрат на рекламу на основе иерархической кластеризации для пяти (а) и восьми (б) кластеров

Пороговое значение доли покупателей (*рисунок б*), характеризующее максимальный экономический эффект (коэффициент подъема) от рекламы, для пяти кластеров варьируются от 0,66 до 0,71 для алгоритмов агломеративной и дивизионной кластеризации соответственно. Это означает, что для оптимального структурирования затрат на рекламу следует учесть, что большинство пользователей (около 80%), относящихся к кластеру 2 (*таблица 1*), не имеют четких намерений и признаков действий, связанных с покупкой товара, т.е. купили его вероятнее всего спонтанно [2, 27], осуществляя поиск по потребностям. Поэтому в данном примере около 70% затрат на рекламу целесообразно отдать контекстной рекламе, которая позволит точно воздействовать лишь на заинтересованных пользователей, в то время как остальные 30% следует отдать таргетированной рекламе для привлечения новых клиентов.

Однако в случае пяти кластеров площадь под кривой существенно меньше, чем в случае восьми кластеров, когда имеется более детальное разбиение. Следует учесть, что большую площадь дает алгоритм агломеративной кластеризации, показывающий в этом случае наилучшие результаты, несмотря на увеличение количества кластеров. Пороговые значения доли покупателей для обоих алгоритмов практически не изменились, что говорит

о сбалансированности и достоверности полученных результатов. Однако само значение коэффициента подъема значительно выросло и по нему можно судить о том, что максимальный экономический эффект от рекламы достигается при финансировании контекстной рекламы в объеме 62% от общих затрат. При этом уточняются признаки пользователей, которые, вероятнее всего, купят товар, а вероятность покупки увеличивается с 82% до 90%.

4. Обсуждение предлагаемого подхода

Для того, чтобы оценить возможности предлагаемого подхода рассмотрим диаграмму, на которой показаны результаты структурирования затрат на рекламную кампанию на основе ранжирования клиентов по децилям (*рисунок 7*).

Из *рисунка 7* видно, что структура затрат на контекстную и таргетированную рекламу путем ранжирования по децилям заметно отличаются от кластерного подхода. В случае ранжирования клиентов по децилям получен максимальный коэффициент подъема для 20% клиентов, что говорит о финансировании контекстной рекламы в объеме лишь 20% от общего количества затрат. Несмотря на то, что данный график более близок к идеаль-



Рис. 7. Результаты структурирования затрат на рекламную кампанию путем ранжирования по децилям

ному случаю, данная модель имеет минимальный запас для улучшения. Кроме того, невозможно провести корректное сравнение полученных кривых Лоренца для кумулятивной доли точек по децилям и кумулятивной доли кластеров, поскольку в первом случае проводилось ранжирование, а во втором – признаковая кластеризация. По этой же причине нельзя сравнивать полученные на рисунках 6 и 7 площади под кривыми Лоренца. Однако по коэффициенту подъема и в том, и в другом случае можно судить о пропорциях затрат на рекламу. Заметные отличия объясняются тем, что в случае ранжирования по децилям учитываются лишь факты совершения сделки, но такой подход не учитывает признаки клиентов. Основным преимуществом предлагаемой методики является возможность более гибко настраивать структуру затрат на рекламу в зависимости от признаков поведенческой активности клиентов. Кластерный подход позволяет построить более адекватную модель и более точно настроить контекстную рекламу.

Заключение

Анализ результатов позволяет сделать следующие выводы.

1. Для расширения возможностей подходов к структурированию затрат на баннерную рекламу

необходимо использовать алгоритмы кластеризации на основе категориальных признаков действий пользователей.

2. Методы иерархической кластеризации хорошо подходят для оценки необходимого количества кластеров, а также дают возможность выявить скрытые закономерности в поведенческой активности клиентов.

3. Коэффициент Джини дает возможность оценить качество кластеризации и определить группы пользователей, дающие максимальное значение вероятности покупки.

4. Использование кластерного подхода позволяет не только структурировать затраты на рекламу, но и определить, какой тип рекламы следует применять по отношению к каким пользователям. Это дает более широкие возможности для оптимизации затрат и повышения эффективности рекламной кампании.

Выявление скрытых закономерностей в группах потребителей позволяет выявить основные направления развития продукта и интерес покупателей к нему, а также оценить устойчивость рынка продуктов с аналогичными характеристиками и стабильность его развития.

Полученные результаты раскрывают прикладные возможности использования принципа нарушения симметрии в бизнес-задачах, и, в отличие от существующих работ [6, 27], отражают возможности структурирования затрат на рекламную кампанию. Данный подход позволяет не только выявить популярность продуктов по признакам, но и определить наиболее эффективные способы привлечения клиентов для того или иного вида продукции. Это достигается путем сопоставления результатов поведенческой активности клиентов относительно совершения целевых действий, а также признаков товаров, относительно которых они совершили эти действия. Также одним из преимуществ является возможность использования неопределенности Джини и коэффициента подъема в качестве индикаторов групп пользователей, для которых более эффективна контекстная или таргетированная реклама. ■

Благодарности

Исследование выполнено при поддержке гранта РФФИ № 18.07.00170.

Литература

1. Кисляков А.Н. Оценка эффективности рекламной кампании в социальных сетях с использованием методов имитационного моделирования // Экономика и управление: проблемы, решения. 2019. Т. 5. № 3 (87). С. 20–26.
2. Рау В.Г., Кисляков А.Н., Тихонюк Н.Е., Рау Т.Ф. Принцип нарушения асимметрии в моделях развития экономических систем: опыт и проблемы // Региональная экономика: опыт и проблемы. Материалы XI международной научно-практической конференции (Гутманские чтения) 15 мая 2018 года / под общ. ред. А.И. Новикова и А.Е. Илларионова. Владимир: Владимирский филиал РАНХиГС, 2018. С. 201–211.
3. Некоторые особенности применения групп нарушенной симметрии для «визуализации» процессов в природных, «живых» и социально-экономических системах / В.Г. Рау и [др.] // Региональная экономика: опыт и проблемы. Материалы XII международной научно-практической конференции (Гутманские чтения) 15 мая 2019 года / под общ. ред. А.И. Новикова и А.Е. Илларионова. Владимир: Владимирский филиал РАНХиГС, 2019. С. 111–119.
4. Кисляков А.Н., Тихонюк Н.Е. Модель ценообразования однородного рынка с учетом асимметричности информации // Инновационное развитие экономики. 2019. № 1. С. 93–100.
5. Перский Ю.К., Дмитриев Д.В. Формирование информационно-экономического механизма управления уровнем асимметрии информации на региональном отраслевом рынке // Вестник Южно-Уральского государственного университета. Серия: Экономика и менеджмент. 2009. № 29 (162). С. 66–74.
6. Кисляков А.Н. Асимметрия информации в задачах анализа социально-экономических процессов // Вестник НГУЭУ. 2020. № 1. С. 64–75. DOI: 10.34020/2073-6495-2020-1-064-075.
7. Барановская Т.П., Иванова Е.А., Хачак Ф.Р. Автоматизированная подсистема планирования бюджета рекламной кампании // Научный журнал КубГАУ. 2016. № 120. С. 223–238.
8. Barnés M. Calculate the economic impact of your click-through prediction. [Электронный ресурс]: <https://towardsdatascience.com/calculate-the-economic-impact-of-your-click-through-prediction-1fabeee64494> (дата обращения: 25.04.2020).
9. Галямов А.Ф., Тархов С.В. Управление взаимодействием с клиентами коммерческой организации на основе методов сегментации и кластеризации клиентской базы // Вестник УГАТУ. 2014. Т. 18. № 4 (65). С. 149–156.
10. Андреева А.В. Оптимальное управление клиентской базой компании на основе показателя долгосрочной стоимости клиента // Бизнес-информатика. 2012. № 4 (22) С. 61–68.
11. Цой М.Е., Залешин П.А. Сегментация потребителей на основе изучения стилей покупательского поведения // Российское предпринимательство. 2017. Т. 18. № 21. С. 3313–3326. DOI: 10.18334/гр.18.21.38543.
12. Mishra V.K., Hazra D., Tarannum K., Kumar M. Business Intelligence using Data Mining techniques and Business Analytics // 5th International Conference on System Modeling & Advancement in Research Trends (SMART 2016), 25–27 November 2016, Moradabad, India. P. 84–89. DOI: 10.1109/SYSMART.2016.7894496.
13. James G., Witten D., Hastie T., Tibshirani R. An introduction to statistical learning with applications in R. New York: Springer, 2013.
14. Nielsen F. Introduction to HPC with MPI for data science. Springer, 2016.
15. Hastie T., Tibshirani R., Friedman J. The elements of statistical learning: Data mining, inference, and prediction. Second Edition. Springer, 2017.
16. Kassambara A. Practical guide to cluster analysis in R: Unsupervised machine learning. Multivariate analysis I. STHDA, 2017.
17. Tripathi S., Bhardwaj A., E P. Approaches to clustering in customer segmentation // International Journal of Engineering & Technology. 2018. Vol. 7. No 3.12. P. 802–807. DOI: 10.14419/ijet.v7i3.12.16505.
18. Gower J.C. A general coefficient of similarity and some of its properties // Biometrics. 1971. Vol. 27. No 4. P. 857–871.
19. Кисляков А.Н. Показатели асимметрии в задачах изучения поведенческой активности пользователей продукта // Известия Санкт-Петербургского государственного экономического университета. 2020. № 3 (123). С. 110–116.
20. Frunza M.–Ch. Computing a standard error for the Gini coefficient: An application to credit model validation // Journal of Risk Model Validation. 2013. Vol. 7. No. 1. P. 61–82. DOI: 10.21314/JRMV.2013.099
21. Зорина А.А. Формирование флуктуирующей асимметрии в процессе индивидуального развития *Betula pendula* // Принципы экологии. 2014. № 4 (12). С. 27–46.
22. Murtagh F., Contreras P. Methods of hierarchical clustering // ArXiv. 2011. [Электронный ресурс]: <https://arxiv.org/pdf/1105.0121.pdf> (дата обращения: 30.03.2020).
23. Королев О.Л., Кусый М.Ю., Сигал А.В. Применение энтропии при моделировании процессов принятия решений в экономике / Под ред. А.В. Сигала. Симферополь: ОДЖАКЪ, 2013.
24. Печеный Е.А., Нуриев Н.К., Старыгина С.Д. Динамическая кластеризация потока больших данных // Математические методы в технике и технологиях: Сб. трудов международной научной конференции: в 12 т. Т. 3 / Под общ. ред. А.А. Большакова. СПб.: Изд. Политехнического университета, 2019. С. 19–21.
25. Prokofyeva E.S., Zaytsev R.D. Clinical pathways analysis of patients in medical institutions based on hard and fuzzy clustering methods // Business Informatics. 2020. Vol. 14. No 1. P. 19–31. DOI: 10.17323/2587-814X.2020.1.19.31.
26. Методы сравнительной оценки результатов кластерного анализа структуры гидробиоценозов (на примере зоопланктона реки Линда Нижегородской области) / В.Н. Якимов и [др.] // Биология внутренних вод. 2016. № 2. С. 94–103. DOI: 10.7868/S0320965216020169.
27. Баранов С.Г., Бурдакова Н.Е. Оценка стабильности развития: Методические подходы Владимир: Изд. ВлГУ, 2015.

Об авторе**Кисляков Алексей Николаевич**

кандидат технических наук;

доцент кафедры информационных технологий, Российская академия народного хозяйства и государственной службы при Президенте Российской Федерации, Владимирский филиал, 600017, г. Владимир, ул. Горького, д. 59а;

E-mail: ankislyakov@mail.ru

ORCID: 0000-0001-8790-6961

Structuring advertising campaign costs considering the asymmetry of users' interests**Alexey N. Kislyakov**

E-mail: 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**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: RANEPА, 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-Peterburgskogo 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

ORCID: 0000-0001-8790-6961