

УДК 519.237.8

---

## ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ПОТРЕБИТЕЛЬСКОГО СПРОСА В УСЛОВИЯХ ИНФОРМАЦИОННОЙ АСИММЕТРИИ<sup>1</sup>

---

**Кисляков Алексей Николаевич**, канд. техн. наук

Российская академия народного хозяйства и государственной службы при Президенте Российской Федерации, Владимирский филиал, ул. Горького, 59а, Владимир, Россия, 600017; e-mail: ankislyakov@mail.ru

*Цель:* работа направлена на решение актуальной проблемы сегментации клиентской базы, выявление намерений потребителей и исследование потребительского спроса на выбранные категории товаров и услуг в течение определенного периода времени с целью оптимизации плана продаж. *Обсуждение:* предлагаемый в работе метод основан на применении алгоритма кластеризации на основе метода k-средних относительно динамических признаков клиентов о совершенных ими покупках, когда нет необходимости вести персонифицированный учет клиентов. Реализация алгоритма k-средних и его модификаций является достаточно простой и гибкой, показана в работе с использованием инструментария электронных таблиц. *Результаты:* рассмотрены вопросы выбора количества кластеров, мер близости объектов для их разделения между кластерами, а также вопросы интерпретации результатов кластеризации и дальнейшего анализа и применения полученного разбиения. Реализованная методика позволяет, располагая данным о портрете целевой аудитории, обращаться к различным сегментам клиентской базы с определёнными предложениями, тем самым увеличивая лояльность клиентов. С помощью описанных методов и подходов возможно эффективно исследовать модели потребительского спроса на отдельные наборы товаров в зависимости от намерений различных групп покупателей.

**Ключевые слова:** информационная асимметрия, кластерный анализ, машинное обучение, метод k-средних.

**DOI:** 10.17308/meps.2019.10/2221

---

<sup>1</sup> Исследование выполнено при поддержке гранта РФФИ № 18.07.00170

## 1. Введение

Несомненно, основной задачей специалистов в области маркетинга является привлечение и удержание клиентов путем выявления их намерений и потребностей для того, чтобы сделать наиболее выгодные предложения или откорректировать план продаж, т.е. оптимизировать деятельность компании в постоянно меняющихся внешних условиях. Для этого необходимо исследовать и построить модели потребительского спроса на различные группы товаров и услуг. Сложность решения этих задач обусловлена изменчивостью таких важных внешних факторов, как конкурентная среда, а также намерения покупателей, которые трудно спрогнозировать, потому как постоянно возникает дисбаланс в информированности участников рынка (продавцов и покупателей) об условиях сделки – качестве товара, его цене, характеристиках и т.д. Такое явление информационной асимметрии [5, 6] представляет собой неравномерное распределение между участниками рынка информации об условиях осуществления рыночной сделки и намерениях друг друга.

В этой связи целесообразным является описание «портрета» целевой аудитории, т.е. признаков клиентов [11]. Для этого необходимо учесть большое количество скрытых факторов их поведения, не имеющих ярко выраженных причинно-следственных связей. Конечно, необходимо собирать как можно больше статических данных о клиентах (пол, возраст, географические данные), но в большинстве случаев сбор, хранение и использование таких данных ограничены действием Федерального закона о персональных данных № 152-ФЗ. Поэтому проще всего вести обезличенный учет клиентов, например, по выданным на руки дисконтным (бонусным) картам с уникальным идентификатором. Разумеется, к данной карте можно привязать личный кабинет на портале компании, где с согласия клиента возможно получить его дополнительные статические признаки [10].

Еще сложнее обстоят дела с динамическими показателями, которые формируются на основании поведения отдельного клиента относительно других клиентов. В этом случае нужно не только вести строгий учет покупок отдельного клиента, но и выполнять обобщенный анализ клиентской базы по известным методикам и показателям эффективности: RFM-классификация, определение размера среднего чека, ценности клиента и т.д. [1, 9]. Так или иначе необходимо сопоставить признаки покупателей с признаками сделок и товаров, которые они потребляют в течение определенного периода времени.

Однако в большинстве случаев известные методики анализа и сегментирования клиентской базы [8, 9] либо ориентированы на получение статических признаков клиентов, что не позволяет получить сбалансированный статистически значимый портрет целевой аудитории, либо являются классификационными, т.е. разделяют клиентскую базу по одному из заранее определенных признаков, например, давность покупки, средний

размер покупки и т.д. [4], при этом не учитываются комплексные факторы, обобщающие и разделяющие клиентов на отдельные кластеры. Кластерный анализ выполняется для объединения клиентов по отдельным признакам или группам признаков. Причем эти признаки могут относиться как к статическим персонифицированным признакам клиента, так и к динамическим признакам, характеризующим его действия – заказы, покупки, т.е. признаки сделок.

Целью исследования является разработка методики сегментации клиентской базы с использованием методов кластерного анализа данных о продажах. В работе рассматривается выполнение кластерного анализа с использованием методов машинного обучения относительно динамических признаков клиентов, когда нет необходимости вести персонифицированный учет клиентов.

## **2. Методология исследования**

Для того чтобы исследовать намерения покупателей, необходимо в первую очередь идентифицировать каждого покупателя, а затем сопоставить признаки товаров, которые он приобретал в течение определенного периода времени. Однако в реальности при большом количестве клиентов-покупателей высока трудоёмкость построения модели поведения для каждого клиента в отдельности, к тому же это будет малополезным и информативным по отношению к моделям поведения остальных клиентов. Очевидно, что необходимо выполнить сегментирование клиентской базы и для каждого сегмента построить собственную модель, усредняющую модели предпочтений всех клиентов, попавших в данный сегмент по различным характеристикам сделок [3]. Располагая данным о портрете целевой аудитории, продавцы без труда смогут обращаться к различным сегментам клиентской базы с определёнными предложениями, тем самым увеличивая лояльность клиентов [2].

Наиболее популярным классическим алгоритмом кластеризации является метод  $k$ -средних [7]. Реализация такого алгоритма и его разновидностей относительно проста и показана в работе с использованием инструментария электронных таблиц [12].

Первым шагом и одновременно недостатком данного алгоритма является необходимость задания количества кластеров  $k$ , на которое будет делиться пространство признаков. Существуют отдельные методы вычисления оптимального количества кластеров. В остальных случаях удобнее задать значение  $k$  вручную и проанализировать результат разбиения.

Следует также отметить, что разбиение может осуществляться и в двумерном пространстве признаков, например, ID клиента/№ сделки, а подробности сделки смогут дополнительно отражать характеристики купленного товара.

На следующем шаге  $k$  произвольных точек назначаются центрами кластеров, и вычисляется расстояние между назначенными кластерными

центрами и всеми остальными точками внутри пространства кластеризации. Принадлежность точки к кластеру определяется вычислением наименьшего расстояния до одного из k-центров.

В качестве мер может быть использовано евклидово расстояние между  $m$  объектами  $x_i = (x_{i1}, \dots, x_{im}) \in X_Q, i = \overline{1, Q}$  и центрами кластеров

$$d_{2k}(x_i, x_j) = \sqrt{\sum_{t=1}^m (x_{it} - x_{jt})^2}. \quad (1)$$

Алгоритм является итеративным и на следующем шаге изменяется положение кластерных центров и вновь проводится распределение точек по k-кластерам, и операция повторяется до тех пор, пока значения расстояний всех объектов до каждого кластера не минимизируются, это означает, что достигнуто оптимальное разбиение.

Целевой функцией является минимизация расстояний до каждого  $i$ -го объекта внутри  $j$ -го кластера является:

$$I(k, m) = \sum_{i=1}^m d_{2i \min}. \quad (2)$$

Завершающим этапом работы алгоритма являются анализ и интерпретация сформированных кластеров с целью определения обобщающих признаков, по которым точки в кластерах являются похожими, то есть какие из особенностей поведения пользователей являются систематическими.

### 3. Обсуждение результатов

Рассмотрим пример сегментации клиентской базы с использованием электронных таблиц. Для начала необходима выгрузка данных из базы по продажам товаров. На рис. 1 показан пример структуры базы данных по учету продаж по дисконтным картам клиентов.

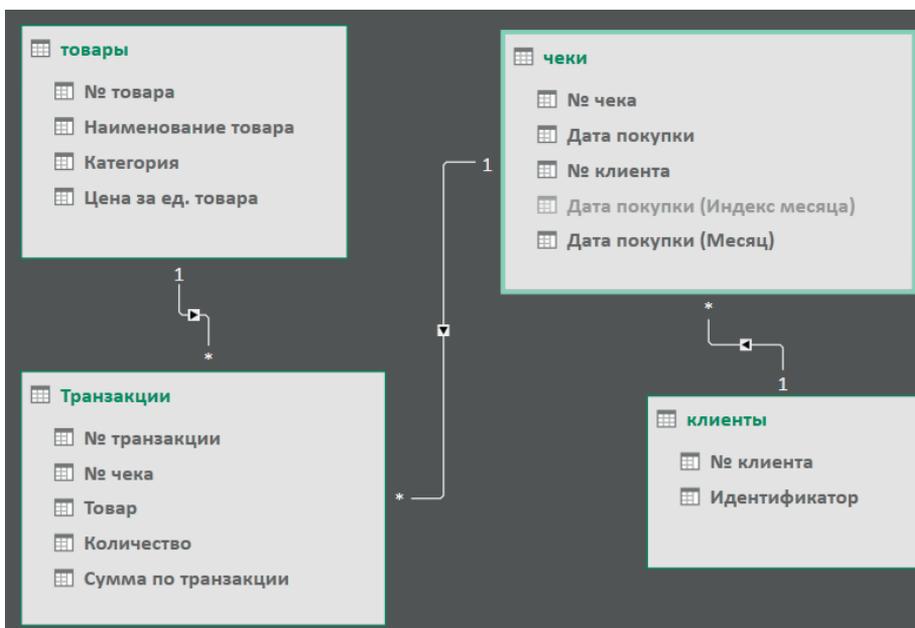


Рис. 1. Структура базы данных по учету продаж по дисконтным картам

Из этой базы нас будут интересовать в первую очередь наименование товаров и временной интервал, в течение которого эти товары приобретались, поэтому на основе базы для начала необходимо построить сводную таблицу по сделкам, клиентам и периодам продаж (рис. 2). Тем самым определяются ключевые признаки сделок, по которым будет выполняться кластеризация клиентов.

	A	B	C	D	E	F
2						
3		Идентификатор клиента				
4	Наименование товара	1003500010001	1003500010002	1003500010004	1003500010006	1003500010007
5	Апельсины	2,56			9,84	
6	Баклажаны			5,12		
7	Бананы	3,09		8,71		
8	Виноград	6,61				
9	Груши	3,41				
10	Капуста	14,9		6,28		
11	Картофель	11,03			3,15	
12	Огурцы	8,55	2,97		5,52	
13	Томаты	9,25	3,67			
14	Яблоки			7,33		1,12
15	Дата покупки (Месяц)					
16	<input type="button" value="январь"/> <input type="button" value="февраль"/> <input type="button" value="март"/> <input type="button" value="апрель"/>					

Рис. 2. Сводная таблица по товарам, клиентам и периодам продаж

В качестве примера рассматривается анализ на основе данных о продажах за 4 месяца и выявляются особенности поведения всего массива покупателей-держателей дисконтных карт за этот период времени. Следует отметить, что в данном случае используется ограниченный набор товаров из 10 товарных позиций для удобства иллюстрации работы алгоритма. В качестве значений элементов данной кросс-таблицы используется количество единиц (в данном случае кг) продукции. На самом деле, на данном этапе важен сам факт покупки, поэтому в дальнейшем следует преобразовать данную таблицу так, чтобы на пересечении строк и столбцов были показаны единицы только в случае совершения покупки (как показано на рис. 3).

	A	B	C	D	E	F	G	H	I	J
1	Месяц	Товары	Цена	категория	Кластер 1	Кластер 2	Кластер 3	Кластер 4	1003500010001	1003500010002
2	Январь	Апельсины	82	Фрукты					1	
3	Январь	Баклажаны	142	Овощи						
4	Январь	Бананы	59	Фрукты					1	
5	Январь	Виноград	249	Фрукты					1	
6	Январь	Груши	129	Фрукты					1	
7	Январь	Капуста	45	Овощи					1	
8	Январь	Картофель	35	Овощи					1	
9	Январь	Огурцы	106	Овощи					1	1
10	Январь	Томаты	132	Овощи					1	1
11	Январь	Яблоки	89	Фрукты						
12	Февраль	Апельсины	82	Фрукты						
13	Февраль	Баклажаны	142	Овощи						
14	Февраль	Бананы	59	Фрукты						
15	Февраль	Виноград	249	Фрукты						
16	Февраль	Груши	129	Фрукты						

Рис. 3. Подготовка данных под кластеризацию (часть данных скрыта)

На рис. 3 в столбцах A-D показаны признаки сделок по каждой товарной позиции по месяцам. Всего получилось 40 строк, которые и являются

обобщающими признаками всех сделок за 4 месяца. Столбцы Е-Н отведены под массивы весовых коэффициентов принадлежности каждой группы сделок к определенному кластеру. Далее начиная со столбца I и далее расположены индикаторы совершения того или иного типа сделки каждым из учтенных по дисконтным картам клиентов. В примере используется выборка более чем из 700 клиентов.

На рис. 4 показано начальное состояние кластеризации, когда все объекты по умолчанию принадлежат к кластеру №1. Для каждого клиента рассчитывается расстояние до каждого кластера с использованием метрики по формуле 1, затем по минимальному расстоянию определяется принадлежность данного клиента к определенному кластеру. В качестве целевой функции используется минимум суммы расстояний каждого клиента до кластера, а в качестве переменных – весовые коэффициенты в столбцах Е-Н, лежащие в пределах от 0 до 1. Эти весовые коэффициенты показывают, насколько данный тип сделки по всем клиентам близок к центру кластера. Также после определения принадлежности, к какому кластеру относятся все сделки по каждому клиенту, нетрудно вести подсчет доли клиентов, относящихся к тому или иному кластеру, а также выполнить интерпретацию результатов (рис. 5).

	Е	Г	Н	И	К					
1	Кластер 1	Кластер 2	Кластер 3	Кластер 4	1003500010001	1003500010002	1003500010004			
2					1					
3										1
4								1		1
5								1		
6								1		
7								1		1
8								1		
9								1	1	
10								1	1	
11										1
12										1
13										
14										
42						Доля клиентов в кластере	100	Расст. до кластера 1	3,16227766	2
43		Доля клиентов в кластере	0	Расст. до кластера 2	3,16227766	2	2,645751311			
44		Доля клиентов в кластере	0	Расст. до кластера 3	3,16227766	2	2,645751311			
45		Доля клиентов в кластере	0	Расст. до кластера 4	3,16227766	2	2,645751311			
46	Целевая функция	<b>1727,039929</b>		Мин. Расст.	3,16227766	2	2,645751311			
47				Принадлежность	1	1	1			

Рис. 4. Расчет мер близости объектов к кластерам

Минимизация суммы расстояний сделок каждого клиента до кластера выполняется за счет подбора весовых коэффициентов в столбцах Е-Н с помощью эволюционного поиска решения. Разумеется, чем точнее сходимость алгоритма поиска решения и чем больше время поиска, тем точнее разбиения. На рисунке 5 показаны результаты разбиения пар клиент/сделка по кластерам. При этом доля клиентов, отнесенных к кластеру 1, составила 22%, к кластеру 2 – 17%, к кластеру 3 – 28% и к кластеру 4 – 33%.

На примере указанных признаков попытаемся интерпретировать полученное разбиение по кластерам. Так, например, кластер №1, для которого наблюдается пиковый весовой коэффициент, равный 0,98, для сделки, которая характеризуется покупкой товара «капуста» в апреле месяце, можно



остаётся неизменным, а изменяются лишь номера кластеров, что является принципиальным при интерпретации результатов.

#### 4. Заключение

Указанные методы позволяют эффективно выстраивать маркетинговую стратегию, опираясь на модели потребительского спроса отдельных групп покупателей на основе их намерений относительно выбора подходящего им товара. На основе предложенной методики удобно выполнять кластеризацию, не требующую использования сложного инструментария и программного обеспечения. Особенно полезным подход окажется для исследования моделей потребительского спроса как с использованием данных о состоянии намерений покупателей в условиях информационной асимметрии, так и с использованием подходов без персонифицированного учета клиентов.

#### Список источников

1. Андреева А.В. Оптимальное управление клиентской базой омпании на основе показателя долгосрочной стоимости клиента // *Бизнес-информатика*, 2012, no. 4 (22), с. 61-68
2. Белоцерковская М.Г. Кластеризация клиентской базы участников программы лояльности // *Московский экономический журнал*, 2017, no. 2, с. 112-11911.
3. Галямова А.Ф., Тархов С.В. Управление взаимодействием с клиентами коммерческой организации на основе методов сегментации и кластеризации клиентской базы // *Вестник УГАТУ*, 2014, т. 18, no. 4 (65), с. 149-156.
4. Кисляков А.Н. Модель поведения участников розничного рынка топлива в условиях информационной асимметрии // *Экономические отношения*, 2019, т. 9, no. 1, с. 219-234.
5. Рау В.Г., Кисляков А.Н., Тихонюк Н.Е., Рау Т.Ф. Принцип нарушения асимметрии в моделях развития экономических систем // *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.
6. Тихонюк Н.Е., Кисляков А.Н. Экономические модели работы с асимметрией информации: эволюция подходов // *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. 236-244.
7. Форман Джон. *Много цифр: анализ больших данных при помощи Excel*. Альпина Паблишер, 2016.
8. Цой М.Е., Залешин П.А. Сегментация потребителей на основе изучения стилей покупательского поведения // *Российское предпринимательство*, 2017, т. 18, no. 21, с. 3313-3326.
9. Цой М.Е., Щеколдин В.Ю., Лежнина М.Н. Построение сегментации на основе модифицированного RFM-анализа для повышения лояльности потребителей // *Российское предпринимательство*, 2017, т. 18, no. 21, с. 3113-3134.
10. Anderson Carl. *Creating a Data-Driven Organization: Practical Advice from the Trenches*. O'Reilly Media, 2015.
11. Hughes A.M. *The Customer Loyalty Solution*. McGraw-Hill Professional, 2008.
12. Winston W. *Microsoft Excel 2016 Data Analysis and Business Modeling*. Microsoft Press, 2016.

---

# INTELLECTUAL ANALYSIS OF CONSUMER DEMAND IN TERMS OF INFORMATION ASYMMETRY

---

**Kislyakov Aleksej Nikolaevich**, Cand. Sc. (Tech.)

Vladimir Branch of RANEPa, Gor'kogo, 59a, Vladimir, Russia, 600017; e-mail: ankislyakov@mail.ru

*Purpose:* the work is aimed at solving the relevant problem of segmentation of the customer base, identifying the intentions of consumers and studying consumer demand for selected categories of goods and services for a certain period of time in order to optimize the sales plan. *Discussion:* the proposed method is based on the application of clustering algorithm based on the method of k-means relative to the dynamic characteristics of customers about their purchases, when there is no need to maintain personalized customer accounting. The implementation of the k-means algorithm and its modifications, that is quite simple and flexible, is shown in the work with the use of spreadsheet tools. *Results:* the questions of choosing the number of clusters, measures of proximity of objects for their separation between clusters, as well as the interpretation of the results of clustering and further analysis and application of the obtained partition are considered. The implemented technique allows, having data about the portrait of the target audience, to address different segments of the customer base with certain proposals, thereby increasing customer loyalty. With the help of the described methods and approaches it is possible to effectively investigate the models of consumer demand for individual sets of goods depending on the intentions of different groups of buyers.

**Keywords:** information asymmetry, cluster analysis, machine learning, k-means method.

## References

1. Andreeva A.V. Optimal'noe upravlenie klientskoj bazoj kampanii na osnove pokazatelya dolgosročnoj stoimosti klienta [Optimal management of the company's client base based on the long-term value of the client]. *Biznes-informatika – Business Informatics*, 2012, no. 4, pp. 61-68. (In Russ.)
2. Belocerkovskaya M.G. Klasterizaciya klientskoj bazy uchastnikov programmy lojal'nosti [Clustering of customer base of loyalty programm participants]. *Moskovskij ehkonomicheskij zhurnal – Moscow economic journal*, 2017, no. 2, pp. 112-119. (In Russ.)
3. Galyamova A.F., Tarhov S.V. Upravlenie vzaimodejstviem s klientami kommercheskoj organizacii na osnove metodov segmentacii i klasterizacii klientskoj bazy [Management of interaction with clients of the commercial organization on the basis of methods of segmentation and clustering of client base] // *Vestnik UGATU*, 2014. v.18, no. 4, pp. 149-156. (In Russ.)
4. Kislyakov A.N. Model' povedeniya uchastnikov roznichnogo rynka topliva v usloviyah informacionnoj asimmetrii [Model of behavior of participants of the retail fuel market in the conditions of

information asymmetry]. *Ekonomicheskie otnosheniya – Economic attitude*, 2018, v. 9, no. 1, pp. 219-234. (In Russ.)

5. Rau V.G., Kislyakov A.N., Tihonyuk N.E., Rau T.F. Princip narusheniya asimmetrii v modelyah razvitiya ekonomicheskikh sistem opyt i problemy [The problem of asymmetry in the models of economic systems development] // *Regional'naya ekonomika: opyt i problemy. Materialy XI mezhdunarodnoj nauchno-prakticheskoy konferencii (Gutmanskije chteniya) 15 maya 2018 goda / pod obshch. red. A. I. Novikova i A. E. Illarionova. Regional economy: experience and challenges. Materials XI international scientific-practical conference may 15, 2018 under the General editorship of A. I. Novikov, and A. E. Illarionov.* Vladimir, Vladimir branch of RANEPa, 2018, pp. 201-211. (in Russ.)

6. Tihonyuk N.E., Kislyakov A.N. Ekonomicheskie modeli raboty s asimmetriiej informacii: evolyuciya podhodov [Economic models of work with information asymmetry: evolution of approaches Regional economy: experience and problems]. *Regional'naya ekonomika: opyt i problemy. Materialy XI mezhdunarodnoj nauchno-prakticheskoy konferencii (Gutmanskije chteniya) 15 maya 2018 goda / pod obshch. red. A.I. Novikova i A.E. Illarionova. Regional economy: experience and challenges. Materials XI international scientific-practical conference*

*may 15, 2018 under the General editorship of A. I. Novikov, and A. E. Illarionov.* Vladimir, Vladimir branch of RANEPa, 2018, pp. 236-244. (in Russ.)

7. Forman Dzh. *Lot numbers: big data Analytics using Excel*, 2016.

8. Tsoi M.E., Zaleshin P.A. Segmentaciya potrebitelej na osnove izucheniya stilej pokupatel'skogo povedeniya [Segmentation of consumers based on the study of styles of consumer behavior]. *Rossijskoe predprinimatel'stvo – Russian business*, 2017, v. 18, no. 21, pp. 3313-3326. (In Russ.)

9. Tsoi M.E., Shchekoldin V.YU., Lezhnina M.N. Postroenie segmentacii na osnove modificirovannogo RFM-analiza dlya povysheniya loyal'nosti potrebitelej [Construction of segmentation based on modified RFM-analysis to increase customer loyalty]. *Rossijskoe predprinimatel'stvo – Russian business*, 2017, v. 18, no. 21, pp. 3113-3134. (In Russ.)

10. Anderson Carl. *Creating a Data-Driven Organization: Practical Advice from the Trenches*. O'Reilly Media, 2015.

11. Hughes A.M. *The Customer Loyalty Solution*. McGraw-Hill Professional, 2008.

12. Winston W. *Microsoft Excel 2016 Data Analysis and Business Modeling*. Microsoft Press, 2016.