

МАТЕМАТИЧЕСКИЕ И ИНСТРУМЕНТАЛЬНЫЕ МЕТОДЫ В ЭКОНОМИКЕ

УДК 366.615

ВЫЯВЛЕНИЕ УСЛОВИЙ УДЕРЖАНИЯ КЛИЕНТОВ КОМПАНИИ

Казаков Олег Леонидович, д-р техн. наук, проф.

Царькова Наталия Ивановна, канд. пед. наук, доц.

Пекова Екатерина Александровна, маг.

Московский политехнический университет, ул. Большая Семеновская, 38, Москва, Россия, 107023; e-mail: olekazakov@gmail.com; tsarkovani@mail.ru; pekkonen97@gmail.com

Цель: статья посвящена разработке способа выявления из больших накопленных данных об индивидуальных характеристиках, используемых услугах и текущих статусах клиентов условий, позволяющих удерживать их в статусе «клиент компании». *Обсуждение:* выявление необходимых знаний требует применения математических и инструментальных методов для представления базы данных о клиентах большой размерности в виде матрицы и осуществления кластеризации ее строк. Поэтому актуальным является поиск в области Big Data новых способов решения и их сочетаний. *Результаты:* предлагаемое представление исходных данных позволяет сопоставлять пары подматриц, содержащих множество Парето строк, для определения условий удержания клиентов. Разработанный в статье способ решения неформальной задачи большой размерности отличается оригинальностью. Он позволит автоматизировать составление рекомендаций и повысить их объективность за счет возможности объяснения предоставляемых референций. Кроме того, использование технологии Big Data обеспечит самоадаптацию предлагаемого способа к изменяющимся интересам компаний.

Ключевые слова: атрибуты клиентов, надежные и ненадежные клиенты, множество Парето векторов атрибутов клиентов, выявление и установление приоритетности атрибутов.

DOI: 10.17308/meps.2022.3/9190

Введение

В настоящее время эффективность маркетинговой политики компании в значительной степени определяется анализом управления взаимоотношениями с клиентами. Убедительно показано, что возможность сбора и

использования огромного объема данных о клиентах позволяет проводить эффективный мониторинг [3]. В то же время это требует немалых затрат на техническое и программное обеспечение, а также на обучение персонала [10]. Все это в итоге приводит к получению прогнозируемых вероятностных оценок [11]. Для повышения достоверности таких оценок предлагается интегрировать модели [14]. Однако опыт показывает, что при интеграции может быть и уменьшение, и увеличение ошибок. Более продуктивным оказывается применение наиболее приемлемых методов, таких как «Случайный лес» и «Метод опорных векторов» [4]. Но и не отрицается, что необходим дальнейший экспериментальный поиск других методов анализа больших данных о клиентах.

Одним из методов такого анализа является выявление взаимозависимости между классификационными признаками клиентов. Определение закономерностей не посредством средней статистической оценки, а извлечение знаний об условиях поведения клиентов позволяет построить модель, опирающуюся на сопоставление характеристик тех, кто перестал быть клиентами компании, и тех, кто ими остался.

Таким образом, актуальность предлагаемой темы обусловлена тем, что выявление и анализ таких тенденций обеспечит оперативную корректировку маркетинговой политики и своевременный вывод клиентов из «группы риска». Помимо этого, прогнозирование предполагаемого набора признаков нового клиента позволит на первых этапах сотрудничества подобрать оптимальный формат информационной политики на рынке.

Постановка задачи

В качестве исходных данных используются атрибуты, взятые из клиентской базы компании Публичное акционерное общество «МегаФон» (ПАО «МегаФон») [6]. Набор данных представляет собой обезличенную информацию о перечне индивидуальных характеристик клиентов, используемых ими услуг и текущих статусов в компании (табл. 1).

Представим имеющиеся данные в компании матрицей (множеством векторов) $Q = \{a_{i(27)} | i = \overline{1, n}\}$, в которой a_{ij} – значение j -го ($j = \overline{1, 27}$) атрибута (компонента вектора) i -го ($i = \overline{1, n}$) клиента (вектора).

Требуется определить условия, т.е. значения атрибутов a_j ($j = \overline{2, 23}$), при которых клиент удерживается в течение максимального периода времени a_{26} (в месяцах). Это позволит судить о степени надежности клиента и предпринимать возможные действия по его удержанию, создавая ему соответствующие условия [12].

Таблица 1

Набор исходных данных

Обозначение	Атрибут	Значение	
		да	нет
a_1	id клиента	№	-

Окончание табл. 1

Обозначение	Атрибут	Значение	
		да	нет
a_2	пол муж	1	0
a_3	пенсионер	1	0
a_4	в браке	1	0
a_5	есть иждивенцы	1	0
a_6	телефонная связь	1	0
a_7	несколько телефонных линий	1	0
a_8	онлайн-безопасность	1	0
a_9	online backup	1	0
a_{10}	страховка оборудования	1	0
a_{11}	техническая поддержка	1	0
a_{12}	стриминговое телевидение	1	0
a_{13}	стриминговой кинотеатр	1	0
a_{14}	безбумажный биллинг	1	0
a_{15}	интернет-провайдер DSL	1	0
a_{16}	интернет-провайдер Fiber optic	1	0
a_{17}	тип контракта месячный	1	0
a_{18}	тип контракта годовой	1	0
a_{19}	тип контракта на два года	1	0
a_{20}	метод оплаты Electronic check	1	0
a_{21}	метод оплаты Mailed check	1	0
a_{22}	метод оплаты Bank transfer	1	0
a_{23}	метод оплаты Credit card	1	0
a_{24}	месячный размер оплаты	руб.	-
a_{25}	общая заплаченная сумма за все время	руб.	-
a_{26}	продолжительность статуса «клиент»	месяц	-
a_{27}	отток	1	0

Проведем упорядочение строк матрицы по величине атрибута a_{26} , начиная с максимальной. Полученную матрицу разделим на две все также упорядоченные матрицы, в первой $Q_1 = \{a_{i(27)} | i \in I_1\}$, из которых атрибуты $a_{27} = 1$, а во второй $Q_0 = \{a_{i(27)} | i \in I_0\}$, из которых атрибуты $a_{27} = 0$. Другими словами, матрица Q_1 соответствует ненадежным клиентам, а матрица Q_0 – надежным, у которых наибольшая продолжительность статуса «клиент». Сравнивая множества I_1 и I_0 строк этих матриц, можно выявить признаки оттока клиентов. По существу требуется осуществить кластеризацию атри-

бутов клиентов по соответствию их надежности и ненадежности в совокупности. Для этого могут применяться разные методы.

Например, в [1] использован метод самоорганизующейся карты Кохонена, позволяющий получать автоматическую кластеризацию потребителей на рынке по различным признакам. Его недостатком является необходимость знания взаимосвязи между такими признаками.

Аналогичными недостатками обладают предлагаемые в [5] методы ассоциативных правил, деревьев решений, нечеткой логики и генетических алгоритмов. С подобными трудностями сталкиваются, когда пробуют построить математические модели с помощью системы линейных алгебраических уравнений, чтобы группировать наблюдения или процессы и находить в них закономерности [9].

Для классификации неформализованных значений атрибутов применяются массивы ключевых слов [2] или описательные модели [7]. В нашем случае бинарное представление достаточно.

Имеющиеся недостатки известных методов заставляют искать дополнительные специфические способы решения поставленной задачи.

Способ решения поставленной задачи

Представление исходных данных

Для того чтобы выявить признаки оттока клиентов, предлагается сравнивать клиентов из разных полученных матриц, но имеющих одинаковые продолжительности статуса «клиент» (атрибут a_{26}). Тогда, наверно, и обнаружатся условия, приводящие к оттоку одних из клиентов, в то время как отсутствие этих условий не будет способствовать оттоку других клиентов [8, 13].

Каждую из двух упорядоченных матриц разделим на подматрицы с одинаковыми величинами атрибута a_{26} . Тогда возможно, что не окажется подматриц разных матриц с одинаковыми величинами атрибута a_{26} или не у всех подматриц разных матриц совпадают величины атрибута a_{26} .

В этом случае возьмем за основу матрицу с наименьшим количеством подматриц и в другой матрице уменьшим количество подматриц, чтобы их число было одинаковым в каждой матрице. Уменьшение количества подматриц нужно производить путем объединения соседних подматриц, у которых разность величин атрибута a_{26} минимальна.

Сравнивая между собой подматрицы разных матриц, нужно убедиться в их примерном совпадении по атрибутам a_{26} . Для этого необходимо, чтобы атрибуты a_{26} у них совпадали или интервалы атрибутов a_{26} у них пересекались. Может оказаться, что существуют пары подматриц разных матриц, у которых такие условия не выполняются. В этом случае эти пары, т.к. они несравнимы, могут быть удалены, но только если такое удаление не приведет к обнулению количества всех подматриц.

Окончательно упорядоченные две рассматриваемые матрицы с мно-

жеством строк I_1 и I_0 будут теперь иметь одинаковое число m подматриц. Множества подматриц для этих двух матриц обозначим $Q_1 = \{Q_{1k} | k = \overline{1, m}\}$ и $Q_0 = \{Q_{0k} | k = \overline{1, m}\}$, а множества строк в каждой подматрице соответственно I_{1k} и I_{0k} , ($k = \overline{1, m}$). Выполняется:

$$I_1 = \bigcup_{k=1}^m I_{1k}, I_0 = \bigcup_{k=1}^m I_{0k}.$$

Рассмотрим каждую из $2m$ подматриц и оставим в них только одну из повторяющихся строк с атрибутами от a_2 по a_{23} . Будем обозначать возможно изменившиеся множества строк I_1^* , I_0^* , I_{1k}^* и I_{0k}^* , ($k = \overline{1, m}$).

Может оказаться, что после этого окажутся одинаковыми подматрицы в парах (Q_{1k}, Q_{0k}) ($k = \overline{1, m}$) с теми же атрибутами строк от a_2 по a_{23} . Тогда такие пары можно удалить как бесполезные для решения поставленной задачи, что приведет к уменьшению числа подматриц, которое будем обозначать m^* . Если в результате окажется $m^* = 0$, то поставленная задача не имеет решения при привлекаемых для этого исходных данных.

Будем считать, что после всех произведенных преобразований имеется две матрицы Q_1^* и Q_0^* , каждая из которых включает m^* подматриц $Q_{1k}^* = \{Q_{1k}^* | k = \overline{1, m^*}\}$ и $Q_{0k}^* = \{Q_{0k}^* | k = \overline{1, m^*}\}$, образующих пары (Q_{1k}^*, Q_{0k}^*) и содержащих соответствующие множества строк I_{1k}^* , I_{0k}^* , I_{1k}^* и I_{0k}^* ($k = \overline{1, m^*}$). Можно утверждать, что в каждой полученной подматрице будет множество Парето строк.

Определение условий удержания клиентов

Рассмотрим пару подматриц (Q_{11}^*, Q_{01}^*) с максимальной величиной атрибута a_{26} . Наибольшее возможное число несовпадающих, т.е. несравнимых между собой, строк в них не будет превышать 222, т.к. в предельном случае будут сравниваться 22-атрибутные строки с бинарными значениями этих атрибутов, а точнее число несовпадающих строк не будет превышать значения $|I_{11}^*| + |I_{01}^*|$. Таким образом, для решения задачи необходимо сопоставить два множества Парето строк.

При таком сопоставлении некоторые строки разных подматриц могут оказаться сравнимыми между собой и такие пары строк не представляют интереса для решения задачи и должны быть удалены. После этого может остаться только одна непустая подматрица или обе будут непустыми.

В случае одной непустой подматрицы Q_{11}^{**} решение сводится к отысканию таких атрибутов, которые во всех строках этой подматрицы равны 0 (нулю). Тогда равенство 1 (единице) этих атрибутов у клиента свидетельствует о том, что он расположен к продолжительному удержанию статуса «клиент», т.е. надежен. Наоборот, если существуют атрибуты, равные 1 (единице) во всех строках рассматриваемой подматрицы, то их равенство 1 (единице) у клиента говорит о его ненадежности. Таким образом, могут найдены атрибуты, свидетельствующие о надежности клиента.

Если отыскание таких атрибутов, которые во всех строках этой подма-

трицы равны 0 (нулю) или 1 (единице), не дает результатов, то поставленная задача не имеет решения при привлекаемых для этого исходных данных.

В случае одной непустой подматрицы Q_{01}^{**} решение сводится к отысканию таких атрибутов, которые во всех строках этой подматрицы равны 0 (нулю). Тогда равенство 1 (единице) этих атрибутов у клиента свидетельствует о том, что он не расположен к продолжительному удержанию статуса «клиент». Наоборот, если существуют атрибуты, равные 1 (единице) во всех строках рассматриваемой подматрицы, то их равенство 1 (единице) у клиента говорит о его надежности. Таким образом, могут найдены атрибуты, свидетельствующие о надежности клиента.

Если отыскание таких атрибутов, которые во всех строках этой подматрицы равны 0 (нулю) или 1 (единице), не дает результатов, то поставленная задача не имеет решения при привлекаемых для этого исходных данных.

В случае двух непустых подматриц Q_{01}^{**} и Q_{11}^{**} описанные выше действия нужно произвести отдельно с каждой из них, но окончательные выводы о признаках надежности клиентов можно сделать только при совпадении частных выводов по каждой из них. В противном случае, когда таких совпадений не наблюдается, поставленная задача не имеет решения при привлекаемых для этого исходных данных.

Аналогичные действия производятся для остальных пар подматриц (Q_{1k}^*, Q_{0k}^*) ($k = \overline{1, m^*}$). В результате могут найдены дополнительные атрибуты, свидетельствующие о надежности клиента, или сделан вывод о том, что поставленная задача не имеет решения при привлекаемых для этого исходных данных.

Если все же искомые атрибуты будут найдены у нескольких из указанных пар подмножеств, то приоритетность заслуживают те атрибуты, которые соответствуют большей величине атрибута a_{26} этих пар, т.к. подтверждаются в течение более продолжительного периода времени.

Описанный выше способ учитывает все варианты и является завершающим для выявления условий удержания клиентов компании.

Заключение

Удержание клиентов компании способствует повышению эффективности ее деятельности. Для этого необходимо выявить соответствующие тенденции и учитывать их при проведении маркетинговой политики. Поэтому задача состоит в установлении закономерностей между классификационными признаками клиентов, одним из которых является их статус.

Исходные данные о классификационных признаках каждого клиента представляются в виде матрицы. Производится упорядочение ее строк и разделение на подматрицы, соответствующие надежным и ненадежным клиентам. Предполагается, что выявить знания о признаках надежных клиентов можно путем сравнения этих подматриц.

Сравнению подматриц предшествуют их преобразования до вида

множеств Парето строк, позволяющих осуществить такое сравнение. Приводится описание алгоритма выявления и установления приоритетности атрибутов, свидетельствующих о надежности клиента, или получения вывода о том, что поставленная задача не имеет решения при привлекаемых для этого исходных данных.

Разработанный способ решения поставленной задачи учитывает специфику исходных данных и основан на матричном их представлении в виде множества векторов атрибутов клиентов. Это позволило использовать понятие множества Парето для агрегирования исходной информации до уровня отражения закономерностей. Полученная возможность автоматизированного извлечения знаний о рассматриваемых намерениях клиентов компании является главным результатом проведенного исследования.

Список источников

1. Брызгалов А.А., Ярошенко Е.В. Применение методов Data Mining при проектировании и создании новой продукции и услуг // *Открытое образование*, 2020, no. 6.
2. Ерисов В.Д., Суворов В.С., Царькова Н.И. Автоматическая классификация обращений на основе анализа больших данных // *Российский экономический интернет-журнал*, 2020 [Электронный ресурс], no. 4.
3. Криштопова Е.А., Терех И.С. Большие данные операторов сотовой подвижной электросвязи: содержание и возможности использования мониторинга поведения индивидов // *Современные средства связи. Материалы XXV Международной научно-технической конференции*. Минск: Учреждение образования «Белорусская государственная академия связи», 2020, с. 135-136.
4. Крюкова А.А., Пальмов С.В. Исследование применимости методов технологии data mining для анализа клиентской базы телекоммуникационной компании // *Прикладная информатика*, 2019, т. 14, no. 1 (79), с. 17-28.
5. Методы Data Mining: обзор и классификация. Доступно: <https://hsbi.hse.ru/articles/metody-data-mining-obzor-i-klassifikatsiya/> (дата обращения: 01.05.2021).
6. Официальный сайт ПАО «МегаФон». Доступно: <https://moscow.megafon.ru/> (дата обращения: 01.05.2021).
7. Суворов С.В., Царькова Н.И., Спиридонова А.К. Анализ больших данных компании Uber Technologies Inc с помощью технологии Data Mining // *Управление экономическими системами: электронный научный журнал*, 2019, no. 7.
8. Управление бизнесом в цифровой экономике: вызовы и решения / под ред. И.А. Аренкова, Т.А. Лезиной, М.К. Ценжарик, Е.Г. Черновой. Санкт-Петербург, Изд-во С.-Петерб. ун-та, 2019.
9. Царькова Н.И., Ерисов В.Д., Пекова Е.А. Технология Big Data как инструмент управления в межкультурной коммуникации // *Управление экономическими системами* [Электронный ресурс], 2019, no. 7.
10. Adel Ismail Al-Alawi. Customer Relationship Management: The Application of Data Mining Techniques in the Telecommunications Sector, 2020.
11. Customer Relationship Management: The Application of Data Mining Techniques in the Telecommunications Sector, Adel Ismail Al-Alawi, 2020 [Электронный ресурс].
12. D. Ćamilović. Data mining and CRM in telecommunications // *Serbian Journal of Management*, 2018, no. 3, pp. 61-72.
13. Data mining techniques: for marketing, sales, and customer, relationship management / Michael J.A. Berry, Gordon Linoff.
14. International Business & Economics Research Journal, Data Mining For Customer Relationship Management Savitha S. Kadiyala, Alok Srivastava, Georgia State University, 2018 [Электронный ресурс].

IDENTIFICATION OF THE COMPANY'S CUSTOMER RETENTION CONDITIONS

Kazakov Oleg Leonidovich, Dr. Sc. (Tech.), Prof.

Tsarkova Natalia Ivanovna, Cand. Sc. (Ped.), Assoc. Prof.

Pekova Ekaterina Alexandrovna, M.A. student

Moscow Polytechnic University, Bolshaya Semyonovskaya str., 38, Moscow, Russia, 107023; e-mail: olekazakov@gmail.com; tsarkovani@mail.ru; pekkonen97@gmail.com

Purpose: the article is devoted to the development of a method for identifying from the large accumulated data about individual characteristics, services used and current client statuses conditions that allow them to keep them in the status of «client of the company». *Discussion:* identifying the necessary knowledge requires the use of mathematical and instrumental methods to represent a database large-dimensional customer in the form of a matrix and the clustering of its rows. Therefore, it is relevant to search in the field of Big Data for new solutions and their combinations. *Results:* the proposed representation of the initial data allows you to map pairs of submatters containing a set of Pareto rows to determine the conditions of customer retention. The method of solving an informal problem of large developed in the article is original. It will automate the preparation of recommendations and increase their objectivity due to the possibility of explaining the references provided. In addition, the use of Big Data technology will ensure the self-adaptation of the proposed method to the changing interests of companies.

Keywords: customer attributes, trusted and unreliable customers, many Pareto vectors of customer attributes, identification and prioritization of attributes.

References

1. Bryzgalov A.A., Yaroshenko E.V. *Primeneniye metodov Data Mining pri proyektirovaniy i sozdaniy novoy produktsii i uslug* [Application of Data Mining methods in the design and creation of new products and services]. *Open education*, 2020, no. 6. (In Russ.)
2. Yerisov V.D., Suvorov V.S., Tsarkova N.I. *Avtomaticeskaya klassifikatsiya obra-shcheniy na osnove analiza bol'shikh dannykh* [Automatic classification of appeals based on big data analysis]. *Russian Economic Online Journal*, 2020 [Electronic resource], no. 4. (In Russ.)
3. Krysh-topova E.A., Terekh I.S. *Bol'shiye dannyye operatorov sotovoy podvzhnoy elektrosvyazi: sodержaniye i vozmozhnosti ispol'zovaniya monitoringa povedeniya individov* [Big data of cellular mobile telecommunication operators: the content and possibilities of using individual behavior monitoring]. *Modern means of communication. Materials of the XV International Scientific and Technical Conference*. Minsk: Educational institution "Belarusian State Academy of Communications", 2020, pp. 135-136. (In Russ.)
4. Kryukova A.A., Palmov S.V. *Issledovaniye primenimosti metodov tekhnologii data mining dlya analiza kliyentskoy bazy telekommunikatsionnoy*

- kompanii [The study of the applicability of data mining technology methods for analyzing the customer base of a telecommunications company]. *Applied Informatics*, 2019, vol. 14, no. 1 (79), pp. 17-28. (In Russ.)
5. Metody Data Mining: obzor i klasifikatsiya [Data Mining methods: overview and classification: [Electronic resource]]. Available at: <https://hsbi.hse.ru/articles/metody-data-mining-obzor-i-klasifikatsiya/> (accessed: 01.05.2021). (In Russ.)
6. Ofitsial'nyy sayt PAO «MegaFon» [Official website of PJSC MegaFon [Electronic resource]]. Available at: <https://moscow.megafon.ru/> (accessed: 01.05.2021). (In Russ.)
7. Suvorov S.V., Tsarkova N.I., Spiridonova A.K. Analiz bol'shikh dannyykh kompanii Uber Technologies Inc s pomoshch'yu tekhnologii Data Mining [Big data analysis of Uber Technologies Inc using Data Mining technology]. *Management of economic systems: electronic scientific journal*, 2019, [Electronic resource], no. 7. (In Russ.)
8. *Upravleniye biznesom v tsifrovoy ekonomike: vyzovy i resheniya* [Business management in the digital economy: challenges and solutions] / ed. by I.A. Arenkov, T.A. Lezina, M.K. Tsenzharik, E.G. Chernova. St. Petersburg: Publishing House of St. Petersburg. Un-ta, 2019. (In Russ.)
9. Tsarkova N.I., Yerisov V.D., Pekova E.A. Big Data kak instrument upravleniya v mezhkul'turnoy kommunikatsii. *Upravleniye ekonomicheskimi sistemami* [Big Data technology as a management tool in intercultural communication. Management of economic systems], 2019, no. 7. [Electronic resource] (In Russ.)
10. Adel Ismail Al-Alawi. Customer Relationship Management: The Application of Data Mining Techniques in the Telecommunications Sector, 2020.
11. Customer Relationship Management: The Application of Data Mining Techniques in the Telecommunications Sector, Adel Ismail Al-Alawi, 2020. [Electronic resource]
12. D. Čamilović. Data mining and CRM in telecommunications. *Serbian Journal of Management*, 2018, no. 3, pp. 61-72.
13. Data mining techniques: for marketing, sales, and customer, relationship management / Michael J.A. Berry, Gordon Linoff.
14. International Business & Economics Research Journal, Data Mining For Customer Relationship Management Savitha S. Kadiyala, Alok Srivastava, Georgia State University, 2018. [Electronic resource]