

МЕТОД АВТОМАТИЧЕСКОЙ КЛАССИФИКАЦИИ КЛИЕНТОВ РОЗНИЧНЫХ ТОРГОВЫХ СЕТЕЙ НА ОСНОВЕ КЛАСТЕРИЗАЦИИ

Н.В. Москвичев

kolmosckvich@yandex.ru

SPIN-код: 3192-7660

Д.Е. Бекасов

bekasov@bmstu.ru

SPIN-код: 6513-4597

МГТУ им. Н.Э. Баумана, Москва, Российская Федерация

Аннотация

Персонализированный подход к клиентам, так необходимый в современной розничной торговле, может быть обеспечен с помощью анализа клиентской базы и предоставления каждой выделенной среди клиентов группе уникальных акционных предложений. В статье описан метод автоматической классификации клиентов на основе кластеризации клиентской базы с использованием нейронной сети Кохонена. Данный метод дает возможность выделить группы клиентов по схожести их характеристик. Использование нейронной сети позволяет распределять клиентов по группам без априорно известного точного количества групп, а также учитывать социально-поведенческую структуру покупателей для каждого магазина или сети. В качестве анализируемых параметров выделены характеристики, используемые в стандартном для маркетинга методе RFM-анализа. Осуществляется классификация выделенных кластеров на основе средних значений характеристик клиентов каждого кластера. Выполнено сравнение с методом RFM-анализа и описаны результаты применения разработанного метода на реальных данных. Представленный метод позволяет сократить количество рассматриваемых групп клиентов более чем в 2 раза. Его применение может позволить розничным сетям более точно корректировать предложение в соответствии с характером спроса.

Ключевые слова

Розничная торговля, программирование, нейронная сеть, сеть Кохонена, маркетинг, сегментация, классификация, кластеризация

Поступила в редакцию 22.12.2020

© МГТУ им. Н.Э. Баумана, 2021

Введение. Современные сети розничной торговли привлекают все большее количество покупателей. При этом в крупных городах практически все магазины розничной торговли принадлежат крупным сетям [1]. Количество покупателей в магазинах данных сетей также велико, и обеспечить какой-либо персонализирован-

ный подход к каждому покупателю практически невозможно. Персонализированный подход к клиентам позволяет получать информацию о спросе покупателей на определенные группы товаров и увеличивать объем продаж товаров или услуг посредством персональных скидочных акций [2]. Возникает необходимость выделить основные группы покупателей для дальнейшего анализа характера спроса и других покупательских характеристик (например, частоты совершения покупок) среди данных клиентов [3]. Также полезно разработать алгоритмы формирования акционных предложений для каждой группы клиентов [4].

В маркетинге для выделения групп покупателей обычно применяют методы линейной сегментации, такие как методы RFM-анализа [5]. Данные методы дают возможность распределить клиентов по кластерам на основе характеристик, интересующих маркетологов, но не позволяют определять реально существующие группы клиентов, что в совокупности с большим количеством получаемых кластеров не позволяет обеспечивать гибкие подходы к различным клиентам. В результате для определения основных групп пользователей с целью их дальнейшей классификации необходимо проводить кластеризацию клиентской базы с последующим анализом полученных кластеров и определением подхода к их клиентам.

В данной статье описан метод автоматической классификации клиентов, основанный на кластеризации клиентской базы с дальнейшим анализом полученных кластеров и определения общего характера покупательского спроса среди клиентов данного кластера. В этом методе кластеризация используется для определения групп клиентов со схожими покупательскими характеристиками, после чего клиенты классифицируются аналогично методу RFM-анализа. В качестве анализируемых параметров выделены характеристики, используемые в стандартном для маркетинга методе RFM-анализа: частота совершения покупок, время с момента совершения последней покупки, средняя суммарная стоимость покупок (средняя сумма чека) [6], а также определены самые популярные у клиента группы товаров и количество совершенных покупок, в которых присутствовали товары из данных групп, что позволило определить группы клиентов со схожими покупками.

Описание метода. Кластеризация. Большинство методов кластеризации требуют априорного знания о количестве кластеров, но при кластеризации покупателей розничных сетей точное количество групп покупателей заранее неизвестно [7]. Данную проблему частично позволяет устранить метод кластеризации, основанный на нейронной сети Кохонена, с помощью которой можно решить задачу кластеризации, зная лишь ориентировочное количество кластеров [8].

Сеть Кохонена является однослойной сетью, построенной из нейронов типа “Winner Takes All”, и позволяет выявлять кластеры (группы) входных векторов, обладающих некоторыми общими свойствами. Для решения текущей задачи кластеризации клиентов сеть создается из 100 выходных нейронов. Данное количество нейронов было выбрано из соображений разбиения клиентов на приемлемое для анализа человеком число кластеров.

На вход сети Кохонена подается вектор характеристик клиентов. Данный вектор необходимо нормализовать. При нормализации исходные значения приводят к значению, находящемуся в пределах от -1 до 1 . Это преобразование необходимо для смещения центра масс исходных данных в центр координат и уравнивания влияния изменения различных характеристик на изменение расстояния между кластеризуемыми объектами [9]. После нормализации характеристик необходимо нормализовать весь входной вектор, т. е. привести его к единичной длине.

После первичной инициализации в процессе обучения осуществляется корректировка векторов сети в соответствии с поступающими обучающими данными. Во время обучения для получения более равномерного распределения весовых коэффициентов нейронов применяется метод выпуклой комбинации [10]. В данном методе обучение проводится на скорректированных векторах:

$$x_i = \alpha x_i + \frac{1 - \alpha}{\sqrt{n}},$$

где x — входной вектор; α — распределяющий коэффициент; n — длина вектора.

Коэффициент α линейно возрастает на каждой эпохе обучения от $0,01$ до 1 . Обучение сети проходило на протяжении 100 эпох с линейным изменением коэффициента скорости обучения от $0,5$ до $0,0001$.

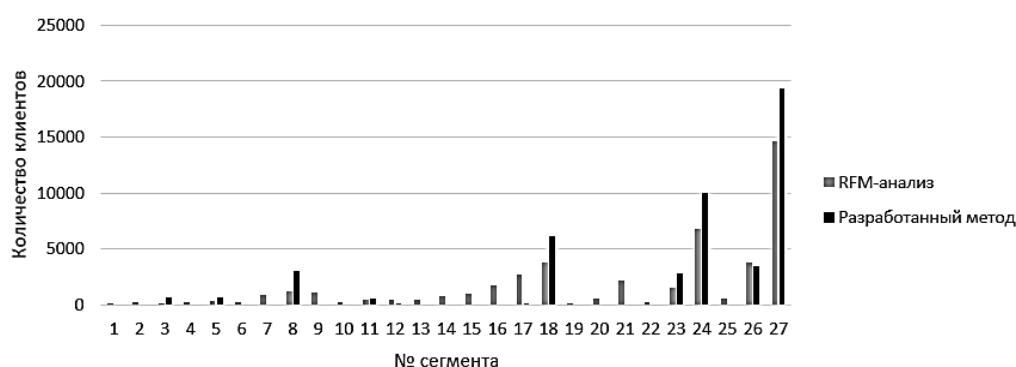
В результате работы нейронной сети на выходе получают кластеры, которые содержат клиентов со схожими входными векторами (со схожими характеристиками). Полученные кластеры в дальнейшем необходимо проанализировать и классифицировать.

Классификация. После кластеризации необходимо определить, какие кластеры были выделены и какие характерные особенности присущи покупателям каждого кластера. В данном методе определяются средние показатели характеристик среди клиентов по каждому кластеру, а также выделяются три самые популярные группы товаров (аналогично тому, как они ранее определялись у каждого клиента). На основе полученных данных формируются соответствующие текстовые метки для дальнейшего анализа результатов. Каждый кластер классифицируется с помощью полученных меток. В дальнейшем, используя размеченные классы, можно обеспечивать различный подход к клиентам каждого класса.

Исследование применимости метода. Работу метода проверяли на данных о 500 тыс. транзакций, совершенных 47 тыс. клиентами в течение года в магазинах одной из розничных сетей крупного российского города. На основе этих данных также было проведено сравнение работы предложенного метода с методом RFM-анализа.

В разработанном методе кластеризация основывается на характеристиках, используемых в RFM-анализе, также присутствует информация о трех самых популярных у клиентов группах товаров. Для получения достоверных результа-

тов при сравнении данных из входного вектора характеристик была исключена информация о количестве транзакций по каждой из рассматриваемых групп товаров. При данном сравнении в качестве относительных границ сегментов (в процентах общего количества клиентов) были взяты используемые в сфере маркетинга значения: 10 % клиентов — верхний сегмент, следующие 25 % — средний сегмент, остальные — низший сегмент. В результате использования метода RFM-анализа было получено распределение клиентов по сегментам, отображенное на рисунке.



Распределение клиентов по сегментам

Видно, что пиковое количество клиентов расположено в одних и тех же сегментах. Также было определено, что 69,2 % клиентов в случае применения двух рассматриваемых методов оказывались в одних и тех же сегментах. При использовании разработанного метода было получено меньшее (10 вместо 27) по сравнению с методом RFM-анализа количество классов. На основе результатов можно сделать вывод, что описываемый метод позволяет объединить клиентов со схожим характером совершения покупок в меньшее количество классов, что, в свою очередь, позволяет обеспечить в сфере маркетинга более уникальный подход к клиентам каждого из получившихся классов.

В результате работы данного метода при учете информации о самых популярных у каждого покупателя группах товаров было выделено 77 кластеров покупателей. Размер кластеров варьируется от 16 тыс. клиентов до одного клиента в кластере. Кластеры с малым объемом клиентов характеризуют клиентов с уникальными характеристиками. К примеру, был определен клиент, совершивший всего одну покупку на крупную сумму в конце рассматриваемого периода. Также был выделен кластер из трех клиентов с очень крупным средним чеком и самой популярной группой товаров (колготки), купленных в большом количестве. Данный факт показывает, что при кластеризации выбросы выделяются в отдельные кластеры, что позволяет легко находить их при анализе результатов работы программы. Дальнейший анализ полученных данных может позволить определять сильно отклоняющееся поведение клиентов, например, может быть обнаружена массовая скупка какого-либо товара с целью его дальнейшей перепродажи за большую цену.

Кластеры с большим объемом характеризуют основную массу покупателей, характеристики которых слабо различаются. Но даже в таких кластерах можно обнаружить некоторые закономерности. Например, среди клиентов кластера с объемом в 10 тыс. человек, которые часто совершают покупки на небольшую сумму, самой популярной группой товаров является питьевая вода.

Также были обнаружены несколько кластеров с клиентами, имеющими некоторое характерное покупательское поведение. Был определен кластер клиентов, часто покупающих алкогольные напитки на крупную сумму. Также были обнаружены кластеры «любителей слабоалкогольных напитков». В кластере, в котором было популярно разливное пиво, средний показатель частоты совершения покупок ниже, чем соответствующий показатель среди клиентов кластера, в котором самым популярным товаром является обычное светлое пиво, а средний показатель чека — выше. Подобным образом можно выявить и другие закономерности между типами покупаемых товаров и характером совершаемых покупок (по сумме чека, частоте покупок, и т. п.).

Заключение. В статье предложен метод автоматической классификации клиентов на основе кластеризации. Проведен анализ результатов испытания разработанного метода на реальных данных о покупках клиентов розничной сети и выполнено сравнение предложенного метода с со стандартным в сфере маркетинга методом RFM-анализа. Показана возможность сокращения количества групп клиентов более чем в 2 раза. Отметим, что разработанный метод может быть применен не только в масштабах одного города, но и в масштабах отдельного района или даже магазина.

Литература

- [1] Доля оборота розничных сетей в общем объёме оборота розничной торговли. *fedstat.ru: веб-сайт*. URL: <https://www.fedstat.ru/indicator/40536> (дата обращения: 07.11.2020).
- [2] Сысоева С.В., Бузукова Е.А. Категорийный менеджмент. Курс управления ассортиментом в рознице. СПб., Питер, 2008.
- [3] Классификация B2B и типология клиентов в B2C. *salesportal.ru: веб-сайт*. URL: <https://salesportal.ru/statya-23-klassifikaciya-b2b-i-tipologiya-klientov-v-b2c> (дата обращения: 07.11.2020).
- [4] Логинова Е.Ю. Маркетинг. Конспект лекций. *marketing.wikireading.ru: веб-сайт*. URL: <https://marketing.wikireading.ru/17167> (дата обращения: 07.11.2020).
- [5] Сегментация клиентов по лояльности или RFM-анализ. *loginom.ru: веб-сайт*. URL: <https://loginom.ru/blog/rfm> (дата обращения: 07.11.2020).
- [6] Применение RFM-анализа в сегментировании клиентской базы. *habr.com: веб-сайт*. URL: <https://habr.com/ru/company/unisender/blog/131225> (дата обращения: 07.07.2020).
- [7] Тюрин, А.Г., Зуев И.О. Кластерный анализ, методы и алгоритмы кластеризации. *Вестник МГТУ МИРЭА*, 2014, № 2(3), с. 86–97.
- [8] Горбаченко В.И. Сети Кохонена. *masters.donntu.org: веб-сайт*. URL: <http://masters.donntu.org/2018/fknt/shumskiy/library/article8.pdf> (дата обращения: 07.11.2020).

- [9] Предобработка данных. *intuit.ru: веб-сайт*. URL: https://intuit.ru/studies/professional_skill_improvements/1609/courses/139/lecture/20569?page=3 (дата обращения: 07.11.2020).
- [10] Аксенов, С.В., Новосельцев В.Б. Организация и использование нейронных сетей. Томск, НТЛ, 2006.

Москвичев Николай Владимирович — студент кафедры «Программное обеспечение ЭВМ и информационные технологии», МГТУ им. Н.Э. Баумана, Москва, Российская Федерация.

Бекасов Денис Евгеньевич — старший преподаватель кафедры «Программное обеспечение ЭВМ и информационные технологии», МГТУ им. Н.Э. Баумана, Москва, Российская Федерация.

Научный руководитель — Барышникова Марина Юрьевна, кандидат педагогических наук, доцент кафедры «Программное обеспечение ЭВМ и информационные технологии», МГТУ им. Н.Э. Баумана, Москва, Российская Федерация.

Ссылку на эту статью просим оформлять следующим образом:

Москвичев Н.В., Бекасов Д.Е. Метод автоматической классификации клиентов розничных торговых сетей на основе кластеризации. *Политехнический молодежный журнал*, 2021, № 01(54). <http://dx.doi.org/10.18698/2541-8009-2021-01-665>

**METHOD FOR AUTOMATIC CLASSIFICATION
OF RETAIL CHAIN CLIENTS ON THE BASIS OF CLUSTERING**

N.V. Moskvichev

kolmosckvich@yandex.ru

SPIN-code: 3192-7660

D.E. Bekasov

bekasov@bmstu.ru

SPIN-code: 6513-4597

Bauman Moscow State Technical University, Moscow, Russian Federation

Abstract

The personalized approach to customers is necessary in modern retail and can be achieved by analyzing the customer base and providing each group of unique promotional offers. The article describes a method for automatic classification of clients based on clustering the client base using the Kohonen neural network. This method makes it possible to distinguish groups of customers by the similarity of their characteristics. The use of a neural network makes it possible to distribute customers into groups without a priori known exact number of groups, as well as to take into account the social-behavioral structure of customers for each store or network. The characteristics used in the RFM-analysis method, which are standard for marketing, are highlighted as the analyzed parameters. The allocated clusters are classified based on the average values of the characteristics of the clients of each cluster. Comparison with the RFM-analysis method is performed and the results of applying the developed method on real data are described. The presented method allows to reduce the number of considered groups of clients by more than 2 times. Its application can allow retail chains to adjust the supply more accurately in accordance with the nature of demand.

Keywords

Programming, neural network, Kohonen network, marketing, segmentation, classification, clustering

Received 22.12.2020

© Bauman Moscow State Technical
University, 2021

References

- [1] Dolya oborota roznichnykh setey v obshchem ob"eme oborota roznichnoy trgovli [Percentage of retail chains turnover in overall volume of retail turnover]. *fedstat.ru: website* (in Russ.). URL: <https://www.fedstat.ru/indicator/40536> (accessed: 07.11.2020).
- [2] Sysoeva S.V., Buzukova E.A. Kategoriynny menedzhment. Kurs upravleniya assortimentom v roznitse [Category management. Course on assortment management in retail]. Sankt-Petersburg, Piter Publ., 2008 (in Russ.).
- [3] Klassifikatsiya B2B i tipologiya klientov v B2C [B2B classification and B2C customer typology]. *salesportal.ru: website* (in Russ.). URL: <https://salesportal.ru/statya-23-klassifikaciya-b2b-i-tipologiya-klientov-v-b2c> (accessed: 07.11.2020).

- [4] Loginova E.Yu. Marketing. Konspekt lektsiy [Marketing. Lecture notes]. *marketing.wikireading.ru: website* (in Russ.). URL: <https://marketing.wikireading.ru/17167> (accessed: 07.11.2020).
- [5] Segmentatsiya klientov po loyalti ili RFM-analiz [Customer loyalty segmentation or RFM analysis]. *loginom.ru: website* (in Russ.). URL: <https://loginom.ru/blog/rfm> (accessed: 07.11.2020).
- [6] Primenenie RFM-analiza v segmentirovanii klientskoy bazy [Using RFM analysis for segmentation of customer base]. *habr.com: website* (in Russ.). URL: <https://habr.com/ru/company/unisender/blog/131225> (accessed: 07.07.2020).
- [7] Tyurin, A.G., Zuev I.O. Cluster analysis, methods and algorithms of the clustering. *Vestnik MGTU MIREA* [Herald of MSTU MIREA], 2014, no. 2(3), pp. 86–97 (in Russ.).
- [8] Gorbachenko V.I. Seti Kokhonena [Kohonen networks]. *masters.donntu.org: website* (in Russ.). URL: <http://masters.donntu.org/2018/fknt/shumskiy/library/article8.pdf> (accessed: 07.11.2020).
- [9] Predobrabotka dannykh [Data pre-processing]. *intuit.ru: website* (in Russ.). URL: https://intuit.ru/studies/professional_skill_improvements/1609/courses/139/lecture/20569?page=3 (accessed: 07.11.2020).
- [10] Aksenov, S.V., Novosel'tsev V.B. Organizatsiya i ispol'zovanie neyronnykh setey [Arrangement and application of neuron networks]. Tomsk, NTL Publ., 2006 (in Russ.).

Moskvichev N.V. — Student, Department of Computer Software and Information Technologies, Bauman Moscow State Technical University, Moscow, Russian Federation.

Bekasov D.E. — Senior Lecturer, Department of Computer Software and Information Technologies, Bauman Moscow State Technical University, Moscow, Russian Federation.

Scientific advisor — Baryshnikova M.Yu., Cand. Sc. (Pedagog.), Assoc. Professor, Department of Computer Software and Information Technologies, Bauman Moscow State Technical University, Moscow, Russian Federation.

Please cite this article in English as:

Moskvichev N.V., Bekasov D.E. Method for automatic classification of retail chain clients on the basis of clustering. *Politekhnichestkiy molodezhnyy zhurnal* [Politechnical student journal], 2021, no. 01(54). <http://dx.doi.org/10.18698/2541-8009-2021-01-665.html> (in Russ.).