ПРИМЕНЕНИЕ АЛГОРИТМА ПОИСКА АССОЦИАТИВНЫХ ПРАВИЛ APRIORI ДЛЯ ФОРМИРОВАНИЯ РЕКОМЕНДУЕМЫХ ТОВАРНЫХ КОМПЛЕКТОВ

THE APPLICATION OF APRIORI ASSOCIATION RULES LEARNING FOR GENERATION OF RECOMMENDED PRODUCTS SETS

УДК 004.62

Брусенцев Александр Григорьевич, профессор кафедры программного обеспечения вычислительной техники и автоматизированных систем, Белгородский государственный технологический университет им. В.Г. Шухова, г. Белгород, Россия

Кательников Александр Александрович, магистрант, Белгородский государственный технологический университет им. В.Г. Шухова, г. Белгород, Россия

Brusentcev Aleksandr Grigorievich, Professor of the Department of software of computer technology and automated systems, Belgorod State Technological University

Katelnikov Aleksandr Aleksandrovich, Graduate Student, Belgorod State Technological University

Аннотация

В статье рассматривается применение одного из методов анализа данных – алгоритма Apriori - для стимулирования покупок в интернет-магазине продукции повседневного спроса (FMCG). В рассматриваемом случае алгоритм анализа исторических данных покупок и формирования продуктовых корзин реализован

на языке PHP (как основном серверном языке проектов, с которыми будет связана библиотека с кодом алгоритма) с применением функций фреймворка Laravel с ORM Eloquent для обращений в базу данных на СУБД PostgreSQL. После обработки заказов алгоритмом Apriori получены как ожидаемые наборы пищевых продуктов (картофель, хлеб, овощи, фрукты), так и довольно нестандартные, не учитываемые в аналитических отчетах.

Annotation

The article discusses the use of one of the methods of data analysis – the Apriori algorithm-to stimulate purchases in the online store of consumer goods (FMCG). In this case, the algorithm for analyzing historical purchase data and forming grocery baskets is implemented in PHP (as the main server language of projects that will be associated with the library with the algorithm code) using the functions of the Laravel framework with ORM Eloquent for accessing the database on the PostgreSQL DBMS. After processing orders by the Apriori algorithm, both expected sets of food products (potatoes, bread, vegetables, fruits) and rather non-standard ones are obtained, which are not taken into account in analytical reports.

Ключевые слова: анализ данных, алгоритм Apriori, корзина покупателей, PHP

Keywords: data analysis, Apriori algorithm, grocery basket, PHP

Масштабное развитие компьютерных технологий, вызванное в первую очередь удобством хранения и обработки больших объемов данных, ведет к еще большему увеличению этих объемов. Это, в свою очередь, заставляет человека придумывать новые способы анализировать эти данные, поскольку очевидно, что в них может содержаться информация, которая может иметь достаточно высокую важность в процессе принятия решений. При этом эти методы для поддержания своей «жизнеспособности» должны отвечать определенным требованиям к количеству затрачиваемых временных и технологических ресурсов.

Одним из таких методов стал алгоритм Apriori – техника сбора данных (Data Mining) для исследования связей и корреляций переменных в базах данных,

содержащих большое количество транзакций [1]. При этом корреляционная связь не обязательно является причинной и может формироваться на основе факторов, вызывающих появление этой связи по иным причинам [2]. Данный алгоритм применяется в рекомендательных системах и, в частности, в системах анализа рыночной корзины как в онлайн, так и в оффлайн-торговле [3]. Такие системы уже широко применяются в мире, но в российских торговых сетях только сейчас осознают важность данного инструмента как одного из способов увеличения продаж.

В данной статье рассматривается использование алгоритма для стимулирования покупок в интернет-магазине продукции повседневного спроса (FMCG). В текущей версии уже существует похожий механизм – комплекты товаров, создаваемые вручную по указанию лица, принимающего решения (менеджмент компании или B2B-клиенты в случае рекламных кампаний). Задача алгоритма – расширить данный функционал за счет рекомендаций покупателям магазина комбинаций товаров, пользующихся наибольшим спросом.

На текущий момент существует несколько реализаций алгоритма Apriori на различных языках программирования, но в определенных случаях их использование считается нецелесообразным или неприменимым (например, требуется быстрое внедрение на текущем стеке технологий или невозможно применить те или иные реализации по причине отсутствия компетенций у специалистов компании).

В рассматриваемом случае требовалось реализовать алгоритм на языке PHP (как основном серверном языке проектов, с которыми будет связана библиотека с кодом алгоритма). После обзора существующих реализаций и выявления их недостатков («сырости» текущей версии и ориентации на обработку данных в текстовых файлах, что является относительно медленной операцией) было принято решение разработать собственную библиотеку, не зависящую от используемого в проекте фреймворка (на случай изменения технологического стека), но позволяющую через интерфейсы [4] добавлять различные варианты

вызовов методов алгоритма (например, для разных фреймворков, ORM или СУБД).

В данном случае в дополнение к библиотеке используется вариант вызова методов с применением функций фреймворка Laravel с ORM Eloquent для обращений в базу данных на СУБД PostgreSQL [5…6]. Использование данной СУБД обосновано по двум причинам:

1. Это основная СУБД, которая применяется на проекте. Получать из нее данные проще всего напрямую без посредников.
2. PostgreSQL является одной из самых быстрых реляционных БД с открытым исходным кодом, что позволит ускорить процесс обработки по сравнению с MySQL или посредством текстовых файлов.

Поскольку скорость обработки данных зависит от количества заказов, количества позиций в заказе и требуемого количества позиций в комплектах на выходе, совместно в менеджментом компании было принято компромиссное решение ограничить эти параметры (заказы за последние 2 месяца, не больше 30 товаров в заказе и не больше 3 товаров в комплекте), с возможностью их изменения в будущем.

Помимо непосредственно библиотеки и реализации ее методов под технологический стек создаются дополнительные инструменты и функции:

1. Сценарий импорта данных. Исходные данные (Data Set) для работы алгоритма могут быть получены не только из таблиц, относящихся к актуальным заказам, но и загружены извне. Например, можно дополнить их данными от партнеров компании или загрузить какие-либо старые, но все еще актуальные данные.
2. Сценарий обработки данных. Из-за большого объема данные не могут быть обработаны в реальном времени под контролем специалиста, поэтому предполагается периодическое выполнение алгоритма раз в 2-4 недели (или реже, если этого будет достаточно для успешного использования библиотеки). В этом сценарии выполняются шаги создания таблиц с записями рассчитанных порогов (threshold), поддержки (support) и доверия/достоверности

(confidence) [3], а также записываются готовые для дальнейшей обработки комплекты.

3. Страница управления комплектами в панели администрирования интернет-магазина (Back-office) – дает возможность контент-менеджеру дополнительно настраивать свойства комплекта, включать или выключать отображение на сайте, управлять скидками.

4. Карточка комплекта и товара в комплекте. На рисунке 1 показан дизайн новых компонентов в карточке комплекта. В нем отображены товары, которые будут добавлены в заказ при покупке данного комплекта. Изменения в карточке товара выполнены схожим образом, но вместо товаров отображаются комплекты, в которые входит данный товар, в виде слайдера-карусели. Новые компоненты выполнены на основе JavaScript-библиотеки React [7].

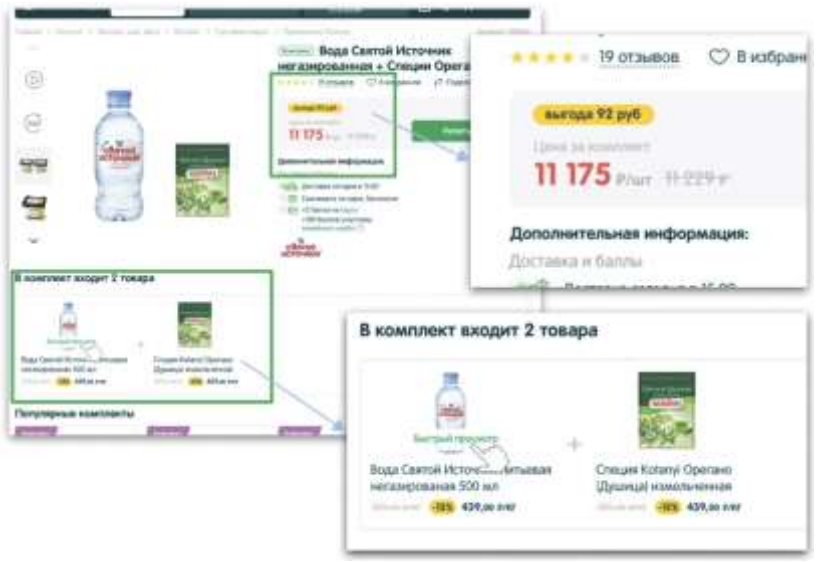

Рис. 1. Дизайн карточки комплекта

После обработки заказов алгоритмом Apriori получены как ожидаемые наборы пищевых продуктов (картофель, хлеб, овощи, фрукты), так и довольно нестандартные, не учитываемые в аналитических отчетах. В связи с чем ожидается более частое использование сущности комплекта как инструмента

продаж, а в последствии рост среднего чека и количества заказов, содержащих комплекты.

Литература

1. Ray Li. Top 10 data mining algorithms in plain English, 2015. – URL: https://hackerbits.com/data/top-10-data-mining-algorithms-in-plain-english/ (дата обращения: 21.05.2021)
2. Брусенцев А. Г. Анализ данных и процессов: учебное пособие для студентов вузов направлений подготовки 09.04.01 - Информатика и вычислительная техника и 09.04.04 - Программная инженерия. Ч.I. Методы статистического анализа данных. – Белгород: Издательство БГТУ им. В. Г. Шухова, 2017. 62 с.
3. Shreya Shetty. What is Apriori Algorithm? Apriori Algorithm Explained, 2020. – URL: https://www.mygreatlearning.com/blog/apriori-algorithm-explained/ (дата обращения: 21.05.2021)
4. Паттерн проектирования Interface (Интерфейс) на PHP, 2015. – URL: https://makedev.org/patterns/fundamental/interface.html (дата обращения: 21.05.2021)
5. Laravel – The PHP Framework for Web Artisans. – URL: https://laravel.com/ (дата обращения: 22.05.2021)
6. PostgreSQL: The World's Most Advanced Open Source Relational Database. – URL: https://www.postgresql.org/ (дата обращения: 22.05.2021)
7. React – A JavaScript library for building user interfaces. – URL: https://reactjs.org/ (дата обращения: 23.05.2021)

Literature

1. Ray Li. Top 10 data mining algorithms in plain English, 2015. – URL: https://hackerbits.com/data/top-10-data-mining-algorithms-in-plain-english/ (дата обращения: 21.05.2021)
2. Brusentcev A.G. Analysis of data and processes: a textbook for students of higher education institutions of training areas 09.04.01-Computer Science and Computer Engineering and 09.04.04-Software engineering. Part I. Methods of statistical

data analysis. - Belgorod: Publishing house of BSTU named after V. G. Shukhov, 2017. 62 p.

3. Shreya Shetty. What is Apriori Algorithm? Apriori Algorithm Explained, 2020. – URL: https://www.mygreatlearning.com/blog/apriori-algorithm-explained/ (дата обращения: 21.05.2021)
4. Design pattern Interface, 2015. – URL: https://makedev.org/patterns/fundamental/interface.html (date of request: 21.05.2021)
5. Laravel – The PHP Framework for Web Artisans. – URL: https://laravel.com/ (дата обращения: 22.05.2021)
6. PostgreSQL: The World's Most Advanced Open Source Relational Database. – URL: https://www.postgresql.org/ (дата обращения: 22.05.2021)
7. React – A JavaScript library for building user interfaces. – URL: https://reactjs.org/ (дата обращения: 23.05.2021)