

РАЗРАБОТКА СИСТЕМЫ РЕКОМЕНДАЦИЙ

М. А. Терёхин

*Белорусский государственный университет, г. Минск;
max.teryokhin@gmail.com; науч. рук. – В. Ю. Сакович*

В настоящее время широкое распространение и применение на большинстве веб-сайтов как коммерческой, так и социальной направленности получили системы рекомендаций. В зависимости от вида бизнеса эти системы могут составлять его основу или быть дополнительным сервисом, задача которого – сократить время поиска покупателем товаров (услуг) и улучшить уровень обслуживания. В статье кратко изложены основные подходы и алгоритмы, используемые в системах рекомендаций, выявлены сильные и слабые стороны каждого из них. Рассмотрены подходы к разработке алгоритма собственной системы рекомендаций и создания на его основе сервиса рекомендаций для интернет-магазинов.

Ключевые слова: персональные рекомендации; коллаборативная фильтрация; история транзакций; оценка точности; алгоритмы подбора; сервис рекомендаций.

ВВЕДЕНИЕ

Выбор – это то, с чем сталкивается каждый человек ежедневно. При наличии различных возможностей обязательно возникает проблема выбора одной из альтернатив: какой товар купить, фильм посмотреть, музыкальное произведение послушать, статью прочитать и т.д. Системы рекомендаций – это набор программных средств и методов, призванных помочь конкретному пользователю в осуществлении отбора среди альтернативных объектов.

Область применения рекомендательных систем разнообразна. Наиболее значимую роль они играют для интернет-магазинов. Персонализация онлайн-маркетинга является заметным трендом последних десятилетий. По оценкам McKinsey, 35% выручки Amazon и 75% Netflix приходится именно на рекомендованные товары и этот процент, скорее всего, будет расти [1]. Широко используются рекомендательные системы и на сайтах, производящих или публикующих контент.

В электронной торговле системы рекомендации нацелены на оказание людям помощи в осуществлении покупок; на сайтах контент-проектов – призваны увеличить время пребывания посетителя на сайте, разнообразить просмотр и повысить посещаемость сайтов.

КЛАССИФИКАЦИЯ

При построении рекомендательных систем в настоящее время используется три основных типа алгоритмов:

- алгоритмы коллаборативной фильтрации;
- алгоритмы, основанные на содержании;
- гибридные алгоритмы.

Основная идея алгоритмов коллаборативной фильтрации заключается в построении рекомендаций для конкретного потребителя, имея определенные данные о его профиле (предпочтениях, истории просмотров, оценок) или мнения (оценки) группы единомышленников. «Похожестью» пользователей в данном случае считается корреляция векторов их оценок, которая может рассчитываться множеством способов [2]. Возможность использования методов коллаборативной фильтрации предполагает, что интересы пользователей будут представлены оценками, которые они выставят объектам после их приобретения, просмотра и т.д. Никакой информации о самих пользователях и оцениваемых объектах не требуется.

Алгоритмы, основанные на содержании, наоборот, для создания рекомендаций предполагают наличие информации о характеристиках объектов. При этом для работы подобных алгоритмов необходимо, чтобы характеристики были представлены в виде структурированных текстов.

Каждый из названных методов имеет свои достоинства и недостатки. Например, метод коллаборативной фильтрации способен рекомендовать объекты, не имея никакого представления о том, что они собой представляют. Данные алгоритмы опираются на историю действий или оценок, проставляемых пользователями, и не учитывают содержание самих объектов. Основные проблемы возникают при появлении новых объектов (известные как проблемы холодного старта). Для их решения существуют различные подходы [3–4].

Главным ограничением в работе алгоритмов, основанных на содержании, является отсутствие обширных характеристических данных об объектах. При недостаточном количестве доступных данных точную систему рекомендаций разработать невозможно.

Минимизировать отрицательные стороны отдельных алгоритмов зачастую позволяет использование гибридного (смешанного) подхода, основанного на сочетании методов коллаборативной и контентной фильтрации. Применение нескольких алгоритмов повышает возможности получения безошибочной, ценной, а, главное, более точной рекомендации.

Именно точность рекомендательной системы является одним из наиболее важных признаков, с помощью которых оценивается ее каче-

ство. В свою очередь, оценка точности помогает в настройке самой системы рекомендаций, показывая, с какими значениями настраиваемых параметров она работает лучше всего. Конечный результат состоит в прогнозе количества потребителей, внимание которых привлечет хотя бы одна из предложенных рекомендаций.

Разработка алгоритма

При разработке автором собственного алгоритма персональных и предметных рекомендаций использовался метод коллаборативной фильтрации. Схема работы алгоритма изображена на рисунке.

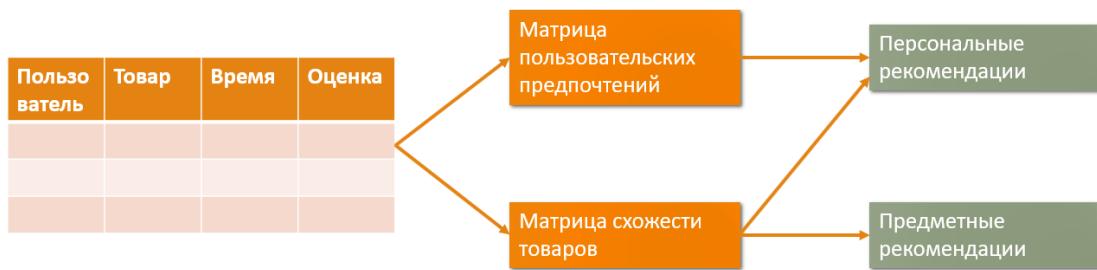


Рис. Схема работы алгоритма

На вход алгоритма поступает таблица с историей транзакций и оценкой приобретенных товаров пользователями интернет-магазина. Так как каждый покупатель оценивает товары по-разному, перед дальнейшими расчетом данные проходят предварительную нормализацию. В последующем производится расчет матриц схожести товаров и пользовательских предпочтений.

Для каждой пары объектов, матрица схожести товаров содержит числовое значение их похожести, а матрица пользовательских предпочтений – значение привлекательности товара для данного пользователя.

В результате алгоритм способен выдавать два типа рекомендаций: персональные и предметные.

Персональные рекомендации для пользователя получаются путем умножения матрицы схожести товаров на вектор предпочтений пользователя, получаемый путем транспонирования строки матрицы предпочтений, соответствующей этому пользователю. Предметные рекомендации рекомендуют элементы, аналогичные заданному и основываются на строке матрицы схожести, соответствующей данному товару.

Для применения разработанного алгоритма рекомендаций в уже существующих интернет-магазинах был разработан веб-сервис.

Для интеграции системы в интернет-магазин фрагмент кода на языке JavaScript из панели управления необходимо вставить на те страницы сайта, где показывается блок рекомендаций. Дальнейшую работу по их получению скрипт берет на себя: при помощи HTTP-запросов к API сер-

виса рекомендаций, скрипт отправляет информацию о текущем просматриваемом товаре и идентификаторе пользователя магазина, в случае если он произвел вход. Полученные от алгоритма рекомендаций идентификаторы товаров сопоставляются с базой данных магазина для получения их наименований и изображений. Сформированный блок рекомендаций отображается на странице интернет-магазина.

Для тестирования полученного сервиса и алгоритма рекомендаций разработан демонстрационный интернет-магазин на основе реальных данных, предоставленный в исследовательских целях крупнейшим интернет-магазином Amazon[5]. С его помощью была произведена оценка точности рекомендаций, генерируемых разработанным алгоритмом. На используемом наборе данных при показе блока из 5 рекомендаций, она составила 16,4%. Данное значение означает, что при использовании разработанного алгоритма приблизительно каждый шестой покупатель приобретет хотя бы один товар из предоставленных ему рекомендаций.

По данным [1], система рекомендаций Amazon достигает точности в 35 %. С одной стороны, меньшая точность разработанного алгоритма связана с гораздо более ограниченным набором данных, чем тот, который использует система Amazon. С другой стороны, это свидетельствует о необходимости повышения качества самого алгоритма, доработка которого будет продолжена.

Библиографические ссылки

1. *MacKenzie I., Meyer C., Noble S.* How retailers can keep up with consumers // McKinsey & Company. URL: <https://www.mckinsey.com/industries/retail/our-insights/how-retailers-can-keep-up-with-consumers> (date of access: 11.04.2019).
2. Котоцигов К. Анатомия рекомендательных систем // Хабр. URL: <https://habr.com/company/lanit/blog/420499> (дата обращения: 04.05.2019).
3. Ricci F., Rokach L., Shapira B. Recommender Systems Handbook. New York, 2015.
4. Kula M. Metadata Embeddings for User and Item Cold-start Recommendations // E-print archive. URL: <https://arxiv.org/pdf/1507.08439.pdf> (date of access: 04.03.2019).
5. Amazon product data / The University of California. URL: <http://jmcauley.ucsd.edu/data/amazon/> (date of access: 02.04.2019).