

УДК 330.133

© Л.В. Флакс, млад. науч. сотрудник

Институт телекоммуникаций и глобального информационного пространства, г. Киев

МЕТОД ПОДБОРА ТОВАРОВ-ЗАМЕНИТЕЛЕЙ НА ОСНОВАНИИ БЛИЗОСТИ ПОКУПАТЕЛЬСКИХ СВОЙСТВ

В данной статье рассматривалась задача подбора похожих товаров для крупных интернет-магазинов с большим и часто обновляемым ассортиментом. Целью была разработка автоматической системы поиска товаров-заменителей на основании схожести их характеристик. Предложенное решение позволяет оптимизировать существующие механизмы формирования групп похожих товаров за счет замены ручного труда автоматически генерируемыми рекомендациями.

Ключевые слова: товары-заменители, покупательские свойства, статус наличия, ключевой показатель эффективности, коэффициент конверсии, показатель отказов, интернет-магазин.

Введение

В электронной коммерции известна задача подбора похожих товаров. В ситуациях, когда товара, которым интересуется покупатель, нет в наличии, удачная рекомендация заменителя может удержать пользователя на сайте и в значительной мере способствовать совершению покупки. В условиях отсутствия живого консультанта, программное обеспечение интернет-магазина должно автоматически найти замену товару, недоступному для заказа. На данный момент разработано множество различных сценариев пользовательских рекомендаций, призванных повысить продажи и удовлетворенность покупателей [2, 3].

В рамках этой статьи будет рассматриваться метод, реализующий рекомендации похожих товаров на основании схожести их покупательских свойств.

Проблема

В крупных интернет-магазинах с большим ассортиментом товаров, который регулярно обновляется, часто возникает проблема несоответствия статусов наличия товаров в базах магазина и рекламных агентств. Специфика инфраструктуры обуславливает то, что зачастую статусы наличия обновляются быстрее соответствующей рекламы, ведущей на сайт.

В таком случае пользователь, который увидел рекламное объявление либо предложение на торговой площадке, может перейти на сайт и обнаружить, что на самом деле товар недоступен, и в данный момент оформить покупку невозможно.

Это известная проблема, и во многих интернет-магазинах она решается категориальными менеджерами, которые определяют группы похожих товаров вручную. Очевидно, что данный подход не является оптимальным в виду нерационального использования ресурсов, а также ошибок, обусловленных человеческим фактором. Стоит учесть, что составленные менеджерами группы носят субъективный характер и по определению не могут учитывать неявных связей между конкретными товарами, вследствие чего основанные на них рекомендации будут уступать в качестве своим аналогам, полученным автоматически.

Подбор похожих товаров будет полезен не только в случае, когда основного нет в наличии, – часто предложенные в рекомендациях товары могут удовлетворять потребностям пользователя даже лучше, чем тот, за которым он пришел изначально.

Отображение на карточке товара (страница сайта с подробным описанием товара) блока «Товары, похожие на этот» становится популярной тенденцией, которая позволяет существенно повысить эффективность интернет-магазина [4]. Ввиду актуальности задачи в данной статье будет рассмотрен алгоритм, позволяющий выделить группы похожих товаров автоматически, без участия человека, знакомого с ассортиментом.

Постановка задачи

Целью данной статьи является оптимизация подбора похожих товаров путем разработки и применения алгоритма автоматического поиска.

Основываясь на прикладной задаче, входными данными будем считать:

1. Конкретный товар, замену которому нужно найти;
2. Базу товаров со значениями их характеристик;
3. Базу с актуальными статусами наличия товаров;
4. Хранилище с историей активности пользователей на сайте.

Ниже приведено более подробное описание всех входных параметров.

1. Заинтересованность в конкретном товаре будем учитывать в случае, если пользователь:

- a. открыл карточку товара;
- b. добавил товар в лист ожидания;
- c. ввел модельный поисковый запрос (поиск на сайте конкретного товара);
- d. добавил товар в список желаний/избранное.

2. В базе хранятся значения для всех характеристик товаров в рамках одной категории.

Под характеристиками подразумевается перечень покупательских свойств, по которым пользователь производит свой выбор (например, разрешение экрана для мониторов).

3. В базе хранятся актуальные статусы наличия по каждой товарной позиции. Возможные варианты:

- a. «в наличии»;
- b. «заканчивается» – на складе находится ограниченное количество товара;
- c. «нет в наличии»;
- d. «архивный» – нет в наличии, и товар не будет закупаться, но он продавался ранее.

Список похожих строится на основании списка всех товаров за исключением архивных.

4. В хранилище находится вся информация об активности на сайте каждого авторизованного пользователя (например, история просмотренных и купленных товаров, факт добавления товара в лист ожидания или избранное, использование фильтров по характеристикам товара и пр.). Данные собираются с помощью инструмента веб-аналитики (в конкретном алгоритме, Google Analytics) и детальнее будут рассмотрены в следующем разделе.

Очевидно, что подбор похожих необходимо осуществлять среди товаров той же категории, а не всего имеющегося множества товаров. Таким образом, задача сводится к расчету показателя близости покупательских свойств между конкретным товаром и множеством товаров той же категории.

Описание решения

Предлагаемый метод основан на представлении пользовательских свойств товара в виде вектора в многомерном пространстве. Следовательно, искомым показателем сходства будет расстояние между точками (товарами) в этом пространстве [1].

Проиллюстрируем это на примере, в рамках которого сравним близость покупательских свойств товаров В и С товару А. Пусть X и Y – покупательские свойства (характеристики) товара. В примере для простоты ограничимся всего двумя характеристиками, но в реальности это число обычно больше (для электроники среднее значение равно 10). На рис. 1 приведено положение товара А в пространстве характеристик. На рис. 2 показано, как товары В и С расположены относительно товара А, показатель близости к которому необходимо найти.

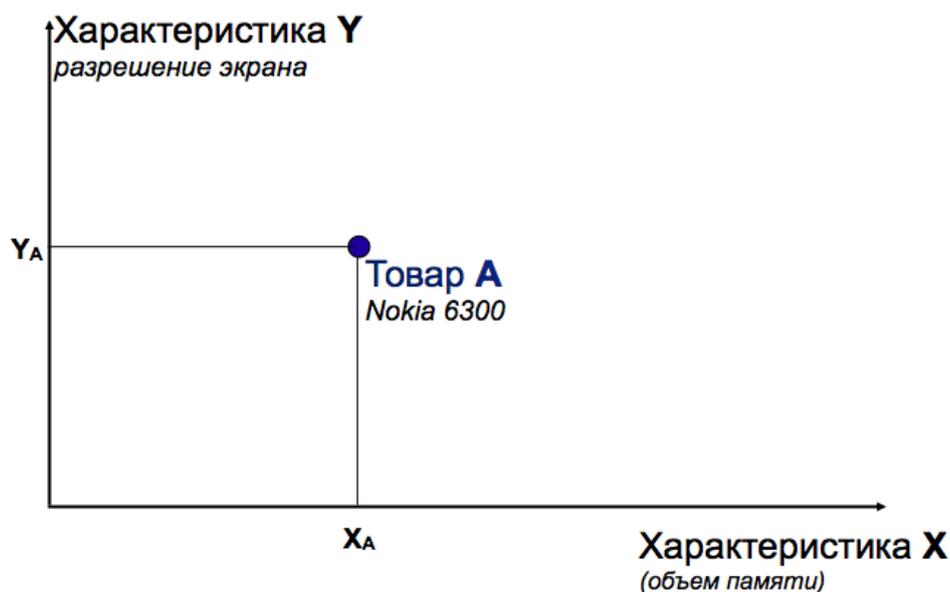


Рис. 1 – Характеристики товара А

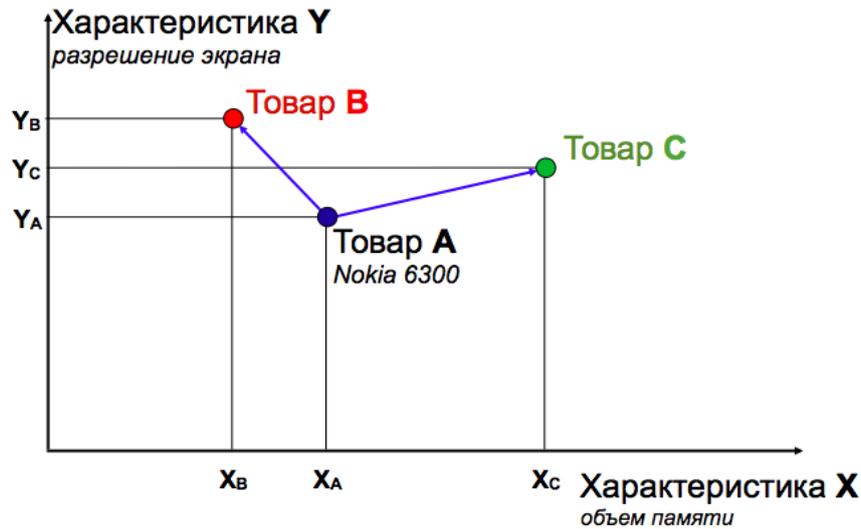


Рис. 2 – Взаимное расположение товаров А, В и С в пространстве признаков

В общем виде расчет показателя близости можно определить следующим образом:

$$D_{pr} = \sqrt{\sum_{i=1}^n k_i (c_i^p - c_i^r)^2} \quad (1)$$

D_{pr} – близость характеристик товара p к товару r ,

p – товар, к которому подбирается замена,

r – товар, сходство с которым вычисляется,

c – значение характеристики,

k – вес характеристики.

Расчет близости производится для всех товаров из базы, при этом для товаров со статусом «архивный» похожие товары определяются, а сами они из расчета исключаются в целях экономии ресурсов.

Значения характеристик должны быть нормализованы, качественные значения необходимо предварительно перевести в числовой эквивалент.

Вес k позволяет учесть важность характеристики с точки зрения пользовательских предпочтений и должен определяться эмпирически для каждого набора покупательских свойств категории. Подход к расчету этих коэффициентов описан ниже.

Как упоминалось в предыдущем разделе, для расчета также используются данные об онлайн-активности пользователей. Предполагается, что на сайте размещен дополнительный код Google Analytics, который в разрезе товарных категорий собирает информацию о том, какими фильтрами в подборе по параметрам чаще всего пользуются посетители интернет-магазина. На каталоге товарной категории размещен блок с фильтрами по характеристикам, используя которые, покупатель может сразу уточнить поиск интересующего его товара (например, цена, производитель и другие характеристики, специфичные для конкретной

категории). На основании данных Google Analytics о популярности характеристик и формируются весовые коэффициенты k . Сам механизм сбора данных об онлайн-активности посетителей в рамках этой статьи рассматриваться не будет.

Расчет матрицы схожести для всех товаров – это основной, но не последний этап формирования списка похожих товаров. В момент, когда на сайт для конкретного товара необходимо вывести блок со списком похожих, предварительно нужно выполнить дополнительную фильтрацию: из списка убираются товары со статусами «не в наличии», «архивный», а также те товары, которые уже были куплены данным пользователем. Для фильтра по статусу наличия используются актуальные данные из базы, а для отсеечения уже купленных товаров – информация из Google Analytics.

На рис. 3 приведена общая схема работы автоматической системы подбора похожих товаров: вся необходимая для расчета информация экспортируется в облачное хранилище из общей базы товаров (характеристики товаров), базы отдела логистики (актуальные статусы товаров) и серверов Google Analytics (данные об онлайн-активности пользователей), после чего распределенно обрабатывается в облаке, и итоговые результаты загружаются на сайт магазина в виде рекомендаций похожих товаров.

Отдельное внимание необходимо уделить расписанию вызова задачи по пересчету матрицы схожести. Так как этот процесс достаточно ресурсоемкий, то обновление матрицы должно происходить лишь тогда, когда список товаров для данной категории существенно изменится; при добавлении в базу нового товара похожие товары нужно находить только для него; полагаем, что коэффициенты k не изменяются в течение достаточно длительного периода времени.

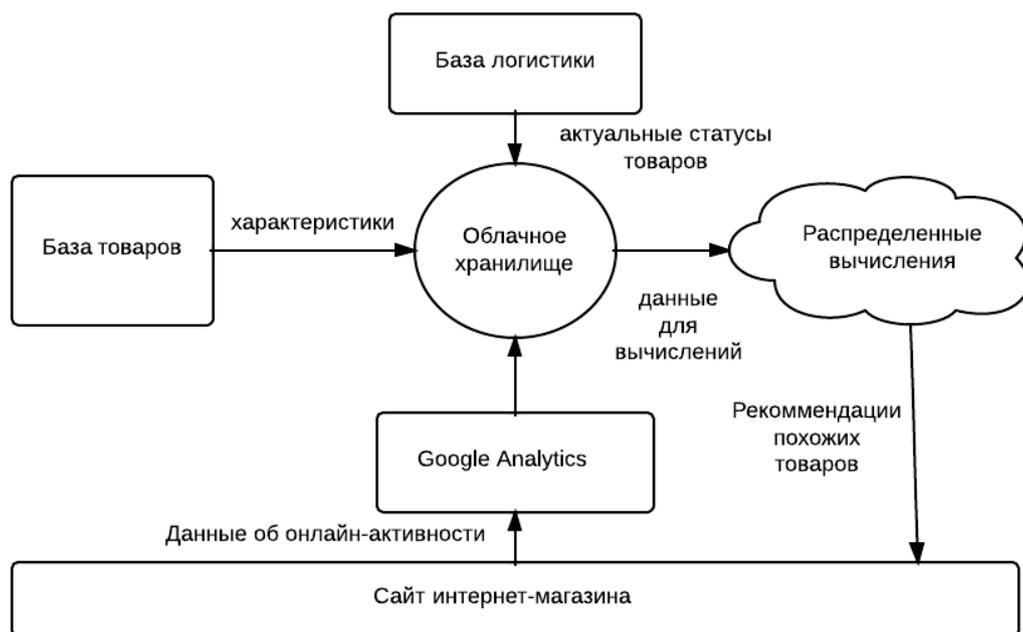


Рис. 3 – Общая схема работы системы

Выводы и дальнейшая работа

В данной статье была описана автоматическая система подбора похожих товаров на основании близости их покупательских свойств (характеристик). Для отбора товаров, релевантных конкретному покупателю, в Google Analytics дополнительно собираются данные об онлайн-активности пользователей, которые используются системой для построения фильтров.

Данный метод может быть применен в интернет-магазинах с:

1. Большим ассортиментом (более 1000 товарных позиций);
2. Большой долей товаров со схожими покупательскими свойствами (например, электроника);
3. Регулярным обновлением ассортимента.

В дальнейшем планируется внедрить разработанную систему в функционирующий интернет-магазин и оценить эффективность ее работы. Рекомендации предлагается отображать на карточке товара в виде блока «Похожие товары». Эффективность предполагается оценить с помощью A/B теста: одной части посетителей показывать старые карточки товара, а второй – обновленные, с блоком «Похожие товары». С помощью средств Google Analytics для вышеописанных сегментов пользователей сравнить ключевые показатели эффективности:

- показатель отказов – отношение количества посещений, в которых пользователь просмотрел только одну страницу сайта, к общему числу посещений;
- коэффициент конверсии – отношение количества посещений, в которых был оформлен заказ, к общему числу посещений.

Список использованной литературы

1. Owen S., Anil R., Dunning T., Friedmann E. Mahout in Action. Manning, 2012.
2. Sarwar B.M., Karypis G., Konstan J.A., Riedl J. Item-based collaborative filtering recommendation algorithms. [In Proceedings of the 10th International World Wide Web Conference (WWW10)]. Hong Kong, May 2001.
3. Schafer J.B., Konstan J.A., and Riedl J. E-Commerce Recommendation Applications. [Data Mining and Knowledge Discovery]. Kluwer Academic, 2001, issue 5.
4. Кошик А. Веб-аналитика 2.0 на практике / А. Кошик. – Диалектика, 2011.

Стаття надійшла до редакції 24.02.14 російською мовою

© Л.В. Флакс

МЕТОД ПІДБОРУ ТОВАРІВ-ЗАМІННИКІВ НА ПІДСТАВІ БЛИЗЬКОСТІ КУПІВЕЛЬНИХ ВЛАСТИВОСТЕЙ

У даній статті розглядалася задача підбору схожих товарів для великих інтернет-магазинів з широким та часто оновлюваним асортиментом. Метою була розробка автоматичної системи пошуку товарів-замінників на підставі близькості їх характеристик. Запропоноване рішення дозволяє оптимізувати існуючі механізми формування груп схожих товарів за рахунок заміни ручної праці автоматично створеними рекомендаціями.

© L.V. Flaks

CUSTOMER FEATURES PROXOMITY BASED SUBSTITUTE SELECTION METHOD

This paper addresses the problem of choosing similar goods for large online-stores with big and frequently updated stock. The goal was to develop an automated search system of substitutes based on proximity of their features. The proposed solution optimizes existing methods for creating groups of similar goods, by replacing manual labor with automaticly generated recommendations.