

ФОРМИРОВАНИЕ РЕКОМЕНДАЦИЙ КЛИЕНТАМ ИНТЕРНЕТ-МАГАЗИНОВ

А. В. Шах

*Учреждение образования «Гомельский государственный технический
университет имени П. О. Сухого», Республика Беларусь*

Научный руководитель О. В. Лапицкая

С развитием электронной коммерции пользователи начали сталкиваться со сложностью выбора наиболее подходящего варианта из огромного разнообразия товаров и услуг, которые предоставляют вебсайты. Взрывной рост и разнообразие информации, доступной в интернете, и быстрое внедрение новых электронных сервисов часто перегружают пользователей информацией, что приводит к принятию ими неоптимальных решений.

В настоящее время рекомендательные сервисы являются ценным средством борьбы с проблемой информационной перегрузки и представляют собой программные инструменты и методы, которые обеспечивают предложения по предметам, которые, скорее всего, представляют интерес для конкретного пользователя.

Предложения относятся к различным процессам принятия решений, таким как, какие предметы купить, какую музыку послушать или какие новости читать.

Поскольку рекомендации, как правило, персонализированные, различные пользователи или группы пользователей могут извлечь выгоду из разнообразных рекомендаций. Кроме того, существуют также неперсонализированные рекомендации. Типичные примеры таких рекомендаций: самые продаваемые книги, компакт-диски и т. д.

В своей простейшей форме персональные рекомендации предлагаются в качестве списка предметов и их рейтингов. При вычислении этого рейтинга, сервис пытается определить наиболее подходящие продукты или услуги, в зависимости от предпочтений и ограничений пользователя. Для выполнения такой вычислительной задачи, необходимо собирать информацию о предпочтениях пользователей, которые либо выражены явно, например, в качестве оценки для продуктов или выводятся путем интерпретации действий пользователя. Например, можно рассматривать факт просмотра конкретной страницы веб-сайта как неявный признак предпочтения элементов, показанных на этой странице.

Актуальность данной темы заключается в том, что рекомендательные системы изменили способы взаимодействия веб-сайтов со своими пользователями. Вместо предоставления статической информации, когда пользователи ищут и, возможно, покупают продукты, рекомендательные системы увеличивают степень интерактивности для расширения предоставляемых пользователю возможностей, формируют предложения независимо для каждого конкретного пользователя на основе его прошлых покупок и поисков, а также на основе поведения других пользователей [1].

Выделяют следующие формы рейтингов:

- численные оценки, такие как оценка от 1 до 5;
- порядковые рейтинги, такие как «полностью согласен, согласен, нейтральный, не согласен, сильно не согласен», где пользователю предлагается выбрать термин, который лучше всего указывает на его мнение по поводу предмета;
- бинарные рейтинги, которые моделируют пользовательские решения (факт покупки или просмотра товара).

Выделяют следующие методы для формирования рекомендаций.

1. *Основанные на контенте.* Система учится рекомендовать элементы, которые похожи на те, что пользователю понравились в прошлом. Сходство элементов рассчитывается исходя из характеристик сравниваемых элементов. Например, если пользователь оценил фильм, который принадлежит к жанру комедии, то система может научиться рекомендовать другие фильмы из этого жанра. Классические методы рекомендаций на основе контента направлены на соответствие атрибутов профиля пользователя с атрибутами элементов. В большинстве случаев атрибутами являются просто ключевые слова, которые извлекаются из описания контента.

2. *Коллаборативная фильтрация.* Оригинальная и наиболее простая реализация этого подход дает рекомендации пользователю на основе предметов, которые понравились в прошлом пользователям с похожими вкусами. Сходство во вкусе двух пользователей рассчитывается на основе сходства рейтингов пользователей. Коллаборативная фильтрация считается самым популярным и широко применяемым методом формирования рекомендаций.

Коллаборативная фильтрация, в свою очередь, также разделяется на 2 основных подхода: основанный на соседстве и основанный на модели.

Подход, основанный на соседстве, является исторически первым в коллаборативной фильтрации и используется во многих рекомендательных системах. В данном подходе для активного пользователя подбирается подгруппа пользователей, схожих с ним. Комбинация весов и оценок подгруппы используется для прогноза оценок активного пользователя.

У данного подхода можно выделить следующие основные шаги:

- присвоить вес каждому пользователю с учетом схожести его оценок и активного пользователя;

– выбрать несколько пользователей, которые имеют максимальный вес, т. е. максимально похожи на активного пользователя. Данная группа пользователей и называется соседями;

– высчитать предсказание оценок активного пользователя для неоцененных им предметов с учетом весов и оценок соседей.

Подход, основанный на модели, предоставляет рекомендации, измеряя параметры статистических моделей для оценок пользователей, построенных с помощью таких методов как метод байесовских сетей, кластеризации, латентной семантической модели, такие как сингулярное разложение, вероятностный латентный семантический анализ и др. Модели разрабатываются с использованием интеллектуального анализа данных, алгоритмов машинного обучения, чтобы найти закономерности на основе обучающих данных. Число параметров в модели может быть уменьшено в зависимости от типа с помощью метода главных компонент.

Этот подход является более комплексным и дает более точные прогнозы, так как помогает раскрыть латентные факторы, объясняющие наблюдаемые оценки.

Данный подход имеет ряд преимуществ. Он обрабатывает разреженные матрицы лучше, чем подход, основанный на соседстве, что в свою очередь помогает с масштабируемостью больших наборов данных.

Недостатки этого подхода заключаются в «дорогом» создании модели, проблеме «холодного старта», возникающей по причине отсутствия информации о рейтингах для новых пользователей и товаров [2].

Разработанное приложение позволяет формировать следующие типы рекомендаций: персонализированные рекомендации товаров пользователям на основе их явного и неявного отклика, неперсонализированные рекомендации товаров, которые наиболее схожи с текущим товаром (продуктом), который пользователь, например, просматривает в данный момент; рекомендации товаров пользователю, полученные путем кластеризации пользователей на основе их характеристик и поведения.

Последние два типа рекомендаций применяются в случае так называемой проблемы «холодного старта»: ситуации, в которой система не может формировать рекомендации новым пользователям в связи с отсутствием или небольшим количеством явного и неявного отклика от этих пользователей – проблема нового пользователя. Приложение позволяет задавать значение порога по достижению которого пользователь перестает считаться новым.

Для формирования персонализированных рекомендаций пользователям на основе их явного и неявного отклика предлагается использовать коллаборативную фильтрацию с использованием латентной модели факторов на основе матричного разложения с учетом неявного отклика (SVD++).

Пусть $N_{(u)}^1$ – множество товаров, которые пользователь u просмотрел; $N_{(u)}^2$ – множество товаров, которые пользователь u купил; $N_{(u)}^3$ – множество товаров, которые пользователь u добавил в wish list.

Тогда значение рейтинга вычисляется по формуле

$$\overline{r_{u,i}} = m + b_i + b_u + q_i^T \left(p_u + |N_{(u)}^1|^{-0,5} \sum_{j \in N_{(u)}^1} y_j^{(1)} + |N_{(u)}^2|^{-0,5} \sum_{j \in N_{(u)}^2} y_j^{(2)} + |N_{(u)}^3|^{-0,5} \sum_{j \in N_{(u)}^3} y_j^{(3)} \right).$$

Минимизация осуществляется стохастическим градиентным спуском.

Рекомендательные механизмы сортируют огромные объемы данных для выявления потенциальных предпочтений пользователей.

Результатом данного проекта является приложение для формирования рекомендаций клиентам интернет-магазинов.

Возможности сбора данных, которые предоставляет Интернет, существенно упростили формирование рекомендаций с помощью коллаборативной фильтрации. С другой стороны, огромное количество доступных данных усложняет реализацию этой возможности. К примеру, поведение некоторых пользователей вполне поддается моделированию, однако другие пользователи не демонстрируют типичного поведения. Наличие таких пользователей может приводить к смещению результатов рекомендательной системы и к снижению ее эффективности. Кроме того, пользователи могут воздействовать рекомендательную систему для повышения предпочтительности одного продукта относительно другого продукта – например, посредством отправки позитивных отзывов об одном продукте и негативных отзывов о его конкурентах. Хорошая рекомендательная система обязана справляться с этими проблемами [1].

Л и т е р а т у р а

1. Рекомендательные системы. Часть 1. Введение в подходы и алгоритмы. – Режим доступа: <https://www.ibm.com/developerworks/ru/library/os-recommender1>. – Дата доступа: 05.04.2018.
2. Коллаборативная фильтрация. – Режим доступа: <https://intellect.ml/kollaborativnaya-filtratsiya-4778>. – Дата доступа: 05.04.2018.