

РАЗРАБОТКА НОВОСТНОГО ПОРТАЛА С ИСПОЛЬЗОВАНИЕМ ФРЕЙМВОРКА SPRING

М. Р. Чечёткин

Белорусский государственный университет, г. Минск;

frm.chechetk@bsu.by;

науч. рук. – Е. Д. Рафеенко, доц.

Увеличивающийся объём, интерес к информации создают спрос на сервисы, главная цель которых обеспечить простой и быстрый доступ к данным. В последние годы просматривается тренд – персонализация маркетинга. Цель новостного портала – создание информационного пространства, позволяющего людям делиться публикациями любой тематики. Важной частью работы является система рекомендаций, которая призвана улучшить опыт взаимодействия с приложением. Было проведено изучение существующих алгоритмов построения персонализированных и неперсонализированных рекомендаций, а также произведена их адаптация.

В результате проведённой работы был разработан новостной портал, обладающий широким списком возможностей как для создания, так и чтения публикаций.

Ключевые слова: новостной портал; рекомендации; Spring.

ПЕРСОНАЛИЗИРОВАННЫЕ РЕКОМЕНДАЦИИ

Неотъемлемой частью любой персонализированной рекомендательной системы является матрица предпочтений – матрица, на одной из осей которой находятся объекты, а на другой – пользователи [1]. Некоторые её элементы, являющиеся парой (пользователь, объект) заполнены оценками – показателем заинтересованности, выраженном по заданной шкале, например, от 1 до 10.

Таблица 1

Матрица предпочтений

| | Объект 1 | Объект 2 | Объект 3 | Объект 4 | Объект 5 |
|----------------|----------|----------|----------|----------|----------|
| Пользователь 1 | | 5 | | 8 | |
| Пользователь 2 | 1 | | 7 | | |
| Пользователь 3 | | | | 1 | 7 |
| Пользователь 4 | | 10 | | 4 | 5 |
| Пользователь 5 | | 4 | 6 | 9 | |
| Пользователь 6 | 5 | | 3 | | 7 |

Обычно пользователи оценивают лишь малую часть объектов, поэтому матрица может содержать пустые ячейки, свидетельствующие об отсутствии информации о заинтересованности пользователя объектом. Задача

рекомендательной системы – заполнить все пустые ячейки на основе имеющейся информации.

Для построения персонализированных рекомендаций порталом собирается информация о просмотрах и оценках конкретных публикаций.

Отметим, что сбор метрик, необходимых для построения рекомендаций, незначительно сказывается на скорости работы системы. Это достигается при помощи определённых оптимизаций.

Теперь же опишем адаптированный алгоритм использования полученной информации о пользователе для построения рекомендаций.

Прежде всего находится набор посетителей со схожими интересами. Отметим, что пользователь a считается схожим с пользователем b , если первый оценил или же просмотрел определённое число раз некую новость n , которая также была оценена пользователем b . Заметим, что в определении схожести двух посетителей факт просмотра некоторой публикации используется лишь при достижении определённого значения. Это сделано для того, чтобы минимизировать влияние на конечную подборку сценариев, которые не дают никакой информации об интересности публикации для пользователя. Для каждого найденного схожего посетителя высчитывается вес, значение которого прямо пропорционально зависит от оценки.

Матрица предпочтений строится следующим образом: для каждой новости при использовании оценённых схожими пользователями публикаций формируется рейтинг, который является суммой произведений веса посетителя и оценки новости. Если число слагаемых рассмотренной суммы больше минимально допустимого значения, то полученный рейтинг делится на их число. В противном случае рейтинг обнуляется, поскольку для оценки рассматриваемой новости недостаточно информации.

В исходную подборку попадают лишь публикации с ненулевым рейтингом и располагаются в порядке его уменьшения.

НЕПЕРСОНАЛИЗИРОВАННЫЕ РЕКОМЕНДАЦИИ

Базовый алгоритм можно сформулировать следующим образом [2]:

- для заданного объекта n найти строку R в матрице схожести S , соответствующую ей,
- из строки R выбрать множество NR наиболее похожих на n объектов,

Точность рекомендаций зависит лишь от метода определения схожести двух разных объектов. В разработанной системе используется метод ассоциаций.

Для определения схожести двух публикаций используется модифицированный метод вычисления коэффициента Жаккара:

$$J(A, B) = \frac{\frac{|A \cap B|}{|A|}}{\frac{|\bar{A} \cap B|}{|A|}}, \quad (1)$$

где A – множество пользователей, которые просматривали объект a ,
 B – множество пользователей, которые просматривали объект b ,
 \bar{A} – множество пользователей, которые не просматривали объект a .

Отметим, что если объект b очень популярен, то знаменатель в формуле будет достаточно большим, а, значит, значение схожести будет меньше. В данном случае рекомендации будут релевантными.

При построении неперсонализированных рекомендаций используются такие же, как и для персонализированных, метрики: рейтинг и просмотры. Однако, в этом случае алгоритм формирования упрощается. Для каждой новости высчитывается средний рейтинг и учитывается общее число просмотров, если оно превышает минимальное количество. Схожесть публикаций на основе просмотра пользователями определяется по формуле (1). Опять же, фильтрация числа просмотров необходима для включения в подборку только популярных новостей. Причём эта метрика имеет меньший, чем рейтинг, вес и её вклад ограничен. То есть невозможен сценарий, при котором рейтинг публикации стремится к бесконечности при сверхбольшом числе просмотров.

РАЗРАБОТАННОЕ ПРИЛОЖЕНИЕ

В ходе разработки серверной части приложения был использован фреймворк Spring, включая такие структурные элементы как Core, Data, Web MVC, AspectJ, Security [3]. Вся информация, включая используемые при построении рекомендаций метрики, хранится в базе данных PostgreSQL.

Клиентская часть реализована в виде отдельного приложения при помощи библиотеки React.

Особое внимание было уделено защите персональных данных пользователей. Каждый ресурс на сервере защищён и может быть получен лишь при наличии у роли пользователя нужной привилегии. Помимо этого, пароли передаются и хранятся в зашифрованном виде. Соответствующая защита ресурсов реализована и на стороне клиента, что улучшает опыт использования для обычных пользователей.

Новостной портал полностью адаптирован под три языка: русский, английский, французский. Смена происходит путём нажатия соответствующей кнопки в шапке страницы без необходимости перезагрузки.

В разработанном приложении есть два различных типа рекомендаций, которые применяются в зависимости от сценария использования. В случае, когда пользователь авторизован, система будет формировать для него персонализированную подборку на основе его оценок и просмотров, в противном случае – неперсонализированную.

Помимо рекомендательной системы разработанное приложение предоставляет целый ряд дополнительных возможностей как для читателей, так и авторов.

Читатели являются основными потребителями любого новостного портала, поэтому при проектировании приложения особый упор был сделан именно на эту группу.

Общий список новостей представлен в пагинируемом формате, который может быть сконфигурирован конкретным пользователем: возможно изменение числа новостей, отображаемых на одной странице. При этом каждая публикация отображается особым способом, в зависимости от её прочитанности конкретным посетителем. Портал предоставляет возможности поиска новости как по тегам, так и по автору. Помимо этого, на отдельной странице доступны история просмотра новостей, подборка рекомендаций. Каждый пользователь имеет профиль, где может быть получена статистика просмотров новостей и изменена личная информация.

Основным отличием возможностей автора от обычного читателя является допустимость первого создавать и редактировать новости. При необходимости автор может изменить созданную ранее публикацию, причём это действие будет отображено в системе изменением даты последнего редактирования.

Администратор обладает самым широким набором возможностей на портале. Основной его функцией является модерирование всей системы.

Интерфейс пользователя зависит непосредственно от его роли. Например, у администратора появляется панель модерирования, которая отсутствует у других посетителей.

Библиографические ссылки

1. Фальк, К. Рекомендательные системы на практике / К.Фальк; пер. с англ. Д.М. Павлова. – Москва: ДМКПресс, 2020. – 448 с.
2. Recommender Systems Handbook / Ricci [и др.]. – New York: Springer Science + Business Media, 2017. – 842 с.
3. Spring 5 для профессионалов / И. Козмина [и др.]; пер. с англ. И.В. Берштейна. – Санкт-Петербург: ООО “Диалектика”, 2019. – 1120 с.