

# РАЗРАБОТКА СИСТЕМЫ РЕКОМЕНДАЦИЙ

**В. О. Рунец**

*Белорусский государственный университет, г. Минск;*

*runets99@gmail.com;*

*науч. рук. – В. Ю. Сакович, ст. преп.*

В современном быстро развивающемся мире, где существует большое разнообразие различных видов продуктов, товаров, услуг, ценность рекомендательных систем растет. Цель рекомендательных систем – проинформировать пользователя об объекте, который может его заинтересовать. Разработка рекомендательных сервисов не ограничивается использованием небольшого количества алгоритмов. Сейчас для разработки зачастую используются нейросети, а алгоритмы совершенствуются очень быстро. Однако до сих пор не существует универсального метода для предсказания симпатий человека.

В результате проведенной работы построен алгоритм, который работает с данными и дает удовлетворительный результат. Точность работы построенного алгоритма была проанализирована с помощью оценивающих метрик.

**Ключевые слова:** рекомендательные системы; поиск; объект; метрики; факторизация матриц.

## ПОДХОДЫ В РАЗРАБОТКЕ РЕКОМЕНДАТЕЛЬНЫХ СИСТЕМ

Рекомендательные системы – системы, которые подсказывают пользователю, какие сущности (фильмы, музыка, книги, новости, веб-сайты, товары) будут ему интересны, и насколько они ему понравятся. Работа рекомендательной системы заключается в выборе объектов на основе его персональных предпочтений и вкусов. Каждый пользователь получает свою персональную рекомендацию, основываясь на различных видах информации о пользователе.

Две основные стратегии создания рекомендательных систем – контентные методы и методы коллаборативной фильтрации. Часто используется объединение этих двух стратегий.

В контентно-ориентированных методах для каждого пользователя и объекта необходим профиль с информацией о них. Профиль пользователя должен содержать характеристики товаров, которые ему нравятся. Профиль объектов заполняется при добавлении объекта в базу. Характеристики могут быть различными, в зависимости от категории объекта. Контентно-ориентированные методы предполагают поиск сходства между товарами по объективным характеристикам [2].

Коллаборативная фильтрация использует информацию об активности пользователей в сети:

- посещение определенных веб-сайтов;
- отзывы об объектах;
- оценка объектов.

В этом случае учитываются субъективные свойства объектов, которые сложно получить, не имея статистики определенной активности от пользователя. Плюс данной фильтрации в том, что не имеет значения, с какими типами объектов ведётся работа, но при этом могут учитываться неявные характеристики, которые сложно было бы учесть при использовании контентно-ориентированных методов [3].

## ОЦЕНИВАЮЩИЕ МЕТРИКИ

Рассмотрим метрики, оценивающие точность предсказываемого рейтинга. В качестве статистических метрик обычно используются средняя абсолютная ошибка (MAE), среднеквадратическая ошибка (RMSE) и корреляция. Самой популярной является метрика MAE. MAE рассчитывается по формуле (1), а RMSE рассчитываются по формуле (2).

$$MAE = \frac{1}{N} \sum |predicted - actual| \quad (1)$$

$$RMSE = \sqrt{\frac{1}{N} \sum (predicted - actual)^2} \quad (2)$$

Чем меньше значения метрик, тем точнее метод, используемый для прогноза рекомендаций [1].

## ИНТЕРФЕЙС РАЗРАБОТАННОГО ПРИЛОЖЕНИЯ

Пользовательская часть разработана на языке Java с использованием библиотеки Swing. Интерфейс представлен оконным приложением, которое наглядно показывает отличие реальной оценки пользователя от предсказываемой алгоритмом.

Рассмотрим пошагово работу приложения:

Шаг 1 (представлен на рисунке 1). Необходимо ввести id пользователя, для которого нужно подобрать список рекомендаций.

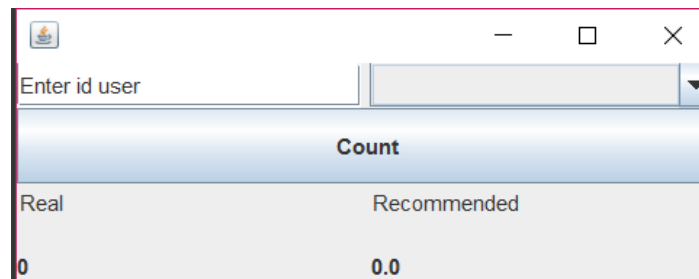


Рис. 1. Первый шаг работы приложения

Шаг 2 (представлен на рисунке 2). После ввода id в выпадающем списке отображаются фильмы, которые данный пользователь уже посмотрел и оценил. Необходимо выбрать фильм, для которого будет высчитываться предсказываемая оценка.

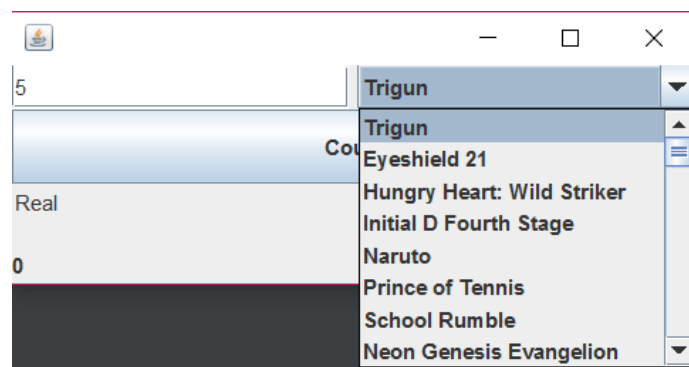


Рис. 2. Второй шаг работы приложения

Шаг 3 (представлен на рисунке 2.3). При нажатии на кнопку Count алгоритму подается на вход выбранный пользователь. В полученном списке рекомендаций находится выбранный фильм и его предсказываемая оценка. И как результат работы приложения, на экране отображается реальная оценка пользователя и предсказываемая.



Рис. 3. Третий шаг работы приложения

## РАБОТА РАЗРАБОТАННОГО АЛГОРИТМА

В данном приложении оценка выставляется пользователем от 1 до 10. Для нахождения каких-либо рекомендаций используется коллаборативная фильтрация.

В разработанном приложении данные о пользователях, фильмах и просмотрах хранятся в базе данных MySQL, данные извлекаются с помощью стандарта JDBC.

Алгоритм предоставления расчета рекомендуемых оценок разработан на языке Java. На вход ему подается пользователь с информацией о его просмотрах, на выходе мы получаем список фильмов с предсказываемым рейтингом.

Имея информацию о просмотренных пользователями фильмах, составляется матрица, представленная в таблице 1.

Таблица 1

Матрица оценок

|        | Item 1 | Item 2 | Item 3 | Item 4 | Item 5 | Item 6 |
|--------|--------|--------|--------|--------|--------|--------|
| User 1 | 5      | 4      | 5      |        |        |        |
| User 2 | 4      |        | 5      |        |        |        |
| User 3 |        | 3      | 5      |        | 4      |        |
| User 4 |        |        |        | 3      | 4      |        |
| User 5 |        |        | 4      | 2      | 4      |        |
| User 6 | 3      |        |        |        |        | 5      |

В данной матрице в столбце указаны пользователи, а в строке – фильмы. На пересечении  $i$ -ого пользователя с  $j$ -ым фильмом указана соответствующая оценка. Далее формируются группы, называемые кластерами, где и объединяются похожие пользователи. Такая группировка производится путем вычисления средней разности между оценками пользователей. У кого такая разность входит в диапазон от 0 до 2,5, считаются похожими пользователями.

Оценку пользователя объекту будем предсказывать как среднюю оценку кластера этому объекту по формуле (3), и пользователю рекомендуются фильмы с самыми высокими предположительными оценками.

$$\hat{r}_{ui} = \frac{1}{|F(u)|} \sum_{v \in F(u)} r_{vi} \quad (3)$$

где  $r_{ui}$  – оценка пользователя  $u$  фильму  $i$ , а  $F(u)$  – функция распределения по кластерам.

Для выборки с количеством пользователей равным 1000 и количеством фильмов равным 11876 был получен RMSE 1.22. Для сравнения: собственный алгоритм Netflix в 2006 году предсказывал оценки пользователей с качеством 0.9514 по метрике RMSE. Сейчас эта рекомендательная система является самой технологичной в мире и является комбинацией из 27 рекомендательных алгоритмов.

#### **Библиографические ссылки**

1. Карауш, А. С. Рекомендательные системы в публичных библиотеках / А. С. Карауш // Библиосфера. – 2009. – № 1. – 41 с.
2. Квятковская, И. Ю. Информационно-рекомендательная система, адаптированная к облачному рекламному сервису / И. Ю. Квятковская, П. Н. Савченко // Вестн. Астрах. гос. техн. ун-та. Сер.: Управление, вычисл. техника и информатика. – 2015. – № 2. – 88 с.
3. Клеменков, П. А. Построение новостного рекомендательного сервиса реального времени с использованием NoSQL СУБД / П. А. Клеменков // Информатика и ее применения. – 2013. – Т. 7, № 3. – 14 с.