

# Обзор подходов, использующихся при построении рекомендательных систем

*Кутайсов Дмитрий Вадимович*

КРЫМСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ ИМЕНИ В.И. ВЕРНАДСКОГО

ТАВРИЧЕСКАЯ АКАДЕМИЯ

ФАКУЛЬТЕТ МАТЕМАТИКИ И ИНФОРМАТИКИ

КАФЕДРА ИНФОРМАТИКИ (ГРУППА 601-И)

e-mail: [dmitry.light@gmail.com](mailto:dmitry.light@gmail.com)

*В работе рассматриваются рекомендательные системы и использующиеся при их построении подходы.*

Рекомендательные системы - это программы и технологии, которые формируют рекомендации независимо для каждого конкретного пользователя на основе его прошлых покупок и поисков, а также на основе поведения других пользователей.

Рекомендательные системы заработали популярность в Интернете, помогая пользователям справляться с огромным количеством информации, и

стали одним из самых распространенных инструментов в электронной коммерции. Сегодня большинство крупных коммерческих веб-сайтов рекомендует своим пользователям различные предложения, например, товары или людей, которые могут представлять интерес. Рекомендательные механизмы сортируют огромные объемы данных для выявления потенциальных предпочтений пользователей [1].

Известные компании, веб-сайты и сервисы, использующие рекомендательные системы [2, 3]:

- *Amazon, eBay, Ozon* — крупные коммерческие интернет-сервисы. Они рекомендуют книги, фильмы, товары, основываясь на том, что покупал, просматривал, оценивал и какие оставлял отзывы пользователь;
- *Last.fm, Pandora, Spotify, Яндекс.Музыка* — рекомендации музыкальных композиций. *Last.fm* использует данные о музыке — автор, стиль, дата, тэги и т. п. *Pandora* основывается на «содержании» музыкальной композиции, используя проект «Music Genome Project», в котором профессиональные музыканты анализируют композицию по нескольким сотням атрибутов;
- *Google, Yahoo!, Яндекс* — рекомендуют пользователям веб-сайты. Поисквики пытаются предсказать, насколько документ отвечает данному запросу;
- *SurfingBird* - российская рекомендательная поисковая система;
- Социальные сети (*Facebook, ВКонтакте, Одноклассники*) — рекомендации друзей.

Также простейшими примерами рекомендательных «систем», используемых повсеместно, являются статьи экспертов, рейтинги популярных продуктов, хит-парады и т. д. Но в дальнейшем будем рассматривать только подходы, вырабатывающие индивидуальные рекомендации для пользователя.

В большинстве рекомендательных систем применяется один из двух базовых подходов: коллаборативная фильтрация (collaborative filtering) и контентная фильтрация (content-based filtering). Гибридные подходы совмещают обе эти концепции [3].

Контентная фильтрация составляет рекомендации на основе поведения пользователя в прошлом. Например, этот подход может использовать информацию о просмотрах (какие блоги читает пользователь и характеристики этих блогов). Если пользователь обычно читает статьи на определенную тему или оставляет комментарии и отзывы, то контентная фильтрация может использовать эту ретроспективную информацию для выявления подобного контента

и предлагать похожий контент в качестве рекомендованного для этого пользователя. Этот контент может быть определен в ручном режиме или извлечен автоматически.

**Коллаборативная фильтрация** считается самой популярной и широко используемой концепцией, применяемой в рекомендательных системах. Коллаборативная фильтрация — это процесс отфильтровывания информации, в котором принимают участие несколько агентов, пользователей, участников. В узком смысле в рекомендательных системах так называют подход, при котором система предлагает пользователю продукты, которые понравились другим пользователям с похожими вкусами. Сходство вкусов двух пользователей вычисляется из сходства поставленных ими рейтингов или оценок в прошлом.

Пусть задано множество пользователей и множество продуктов, и некоторые пользователи как-то оценили некоторые продукты. Данные состоят из троек вида  $(u, i, r_{u,i})$ , где  $i$  обозначает продукт,  $u$  — пользователя, а  $r_{u,i}$  — рейтинг, который пользователь  $u$  поставил продукту  $i$ .

Представим эти данные как матрицу, каждый столбец которой соответствует пользователю, а строка — продукту. Необходимо предсказать неизвестные элементы матрицы, а именно — какие из неизвестных элементов будут максимальными в своём столбце, то есть какие продукты больше всего понравятся тому или иному пользователю [2].

фильм/пользователь	Анна	Сергей	Елена	Николай
Любовь до конца	5	5	0	0
Романтическое путешествие	5	?	?	0
Милые щенки	?	5	0	?
Автомобильные погони	0	0	3	5
Боевые искусства	0	0	3	?

Пример таблицы рейтингов фильмов

#### Формальная постановка задачи коллаборативной фильтрации.

Будем использовать следующие обозначения [4]:

- $U$  — множество пользователей (субъектов);
- $I$  — множество продуктов (объектов);
- $Y$  — пространство описаний транзакций;
- $D = (u_t, i_t, y_t)_{t=1}^m \in U \times I \times Y$  — транзакционные данные;
- $\|r_{ui}\|$  — матрица кросс-табуляции размера  $|U| \times |I|$ , где  $r_{ui} = \text{aggr}\{(u_t, i_t, y_t) \in D \mid u_t = u, i_t = i\}$ ;

Пусть задано множество пользователей  $U$ , множество продуктов  $I$ . Задачи: спрогнозировать значения незаполненных ячеек  $r_{ui}$ ; оценить сходство  $p(u, u')$ ,  $p(i, i')$ ,  $p(u, i)$ ; построить список рекомендаций для  $u$  и для  $i$ .

Среди проблем коллаборативной фильтрации можно выделить разреженность данных и проблему холодного старта. Реальные коммерческие рекомендательные системы должны работать с огромным массивом данных, а пользователи могут поставить оценки лишь небольшому количеству продуктов. Тогда матрица пользователей-продуктов будет сильно разрежена, а качество рекомендаций снизится. В связи с этим возникает также проблема холодного старта: неясно, что рекомендовать новому пользователю в системе и кому рекомендовать новый товар. Для решения этих проблем были предложены разные подходы, в частности, снижение размерности матрицы субъектов-объектов (SVD, PCA), гибридные алгоритмы (content-boosted CF) и другие [3].

Таким образом, рекомендательные системы используются в большинстве коммерческих веб-сайтов, приложений и сервисов. Рассмотренные подходы являются самыми распространенными и успешными, они достаточно просты для понимания. На основе коллаборативной фильтрации и алгоритма факторизации низкоранговой матрицы планируется создать рекомендательную систему, прогнозирующую оценки фильмов и формирующую список предложений для конкретного пользователя.

#### СПИСОК ЛИТЕРАТУРЫ

- [1] Ricci F., Rokach L., Shapira B., Kantor P.B. *Recommender Systems Handbook..* – Springer, – 2011.
- [2] Николенко С. *Рекомендательные системы: постановка задачи.* – 2012 – <http://habrahabr.ru/company/surfingbird/blog/139022/>
- [3] Su X., Khoshgoftaar T.M. *A Survey of Collaborative Filtering Techniques.* – Advances in Artificial Intelligence – 2009.
- [4] Воронцов К.В. *Курс лекций по машинному обучению. Коллаборативная фильтрация.* – 2011 – <http://www.machinelearning.ru/wiki/images/9/95/Voron-ML-CF.pdf>