

Секция «Вычислительная математика и кибернетика»

Настройка линейной модели рекомендательной системы

Кириллов Александр Николаевич

Студент

*Московский государственный университет имени М.В. Ломоносова, Факультет
вычислительной математики и кибернетики, Москва, Россия*

E-mail: arhipisk@gmail.com

Рекомендательные системы – область, активно развивающаяся в последнее десятилетие. Большим толчком в развитии послужил конкурс Netflix prize [5], в котором требовалось построить рекомендации для пользователей, которые оценивают фильмы из огромной базы. В качестве данных в этом конкурсе были предоставлены только оценки пользователей. Для подобной задачи было разработано большое количество методов коллаборативной фильтрации, например, метод slope one [2], однако при построении реальных систем рекомендаций, как правило, используется большое число различных источников данных.

В данной работе рассматривается задача построения рекомендательной системы для интернет-магазина, которая должна рекомендовать к конкретному товару список подходящих. В качестве источников данных были доступны следующие: история покупок пользователей, история просмотров страниц, информация о товарах (название, описание, рейтинг и т. д.). Используя эти данные, были получены различные признаки, основанные на статистике покупок и просмотров (slope one, SVD и т. д.), и признаки, основанные на информации о товарах, такие как TF-IDF.

Сформулируем задачу формально. Определим множество товаров $X = \{x_1, \dots, x_L\}$. Решается задача показа рекомендаций для товара x_i . Признаковое пространство состоит из признаков вида $f_k(x_i, x_j)$, $k = 1, \dots, K$. Необходимо для каждого товара x_i определить список товаров-рекомендаций. Будем сопоставлять каждому товару некоторый вес

$$W_i(j) = w_1 \cdot f_1(x_i, x_j) + \dots + w_K \cdot f_K(x_i, x_j).$$

После получения всех весов $W_i(j)$, $j = \{1, \dots, n\} \setminus \{i\}$ товаров x_j , искомый список товаров-рекомендаций определим, как 10 товаров с максимальным весом.

Для настройки весов w_1, \dots, w_K использовался алгоритм «LENKOR» [1]. Настройка производилась на функционал качества nDCG [4]. Полученная модель оценивалась при помощи A/B - тестирования [3] на реальных пользователях. По сравнению с системой рекомендаций, основанной на одном источнике данных, которая раньше использовалась интернет-магазином, прирост конверсии визитов пользователей в заказы составил 46%.

Литература

1. D'yakonov A. Two Recommendation Algorithms Based on Deformed Linear Combinations // ECML-PKDD 2011 Discovery Challenge Workshop. 2011. Pp. 21-27.
2. Lemire D., Maclachlan A. Slope One Predictors for Online Rating-Based Collaborative Filtering // In SIAM Data Mining (SDM'05), Newport Beach, California, April 21–23, 2005.

3. A/B-testing: http://en.wikipedia.org/wiki/A/B_testing
4. NDCG: http://en.wikipedia.org/wiki/Discounted_cumulative_gain
5. Netflix: <http://www.netflixprize.com/>

Слова благодарности

Работа выполнена при финансовой поддержке РФФИ, проект 12-07-00187-а.