

## МЕТОДЫ КОЛЛАБОРАТИВНОЙ ФИЛЬТРАЦИИ РЕКОМЕНДАТЕЛЬНЫХ СИСТЕМ

Белорусский государственный университет информатики и радиоэлектроники  
г. Минск, Республика Беларусь

Жевлакова С.И.

Волосевич А. А. – кандидат физ.-мат. наук, доцент

Методы коллаборативной фильтрации являются наиболее популярными при построении прогнозов в рекомендательных системах. На вход подаются данные об оценках пользователями продуктов в прошлом. Основная идея заключается в сравнении интересов пользователей или объектов на основе этих оценок.

Входные данные представлены сильно разреженной матрицей, состоящей из рейтингов, которые пользователи (строки матрицы) присвоили продуктам (столбцы матрицы):

$$R = (r_{i,a})_{i=1,a=1}^{N,M}$$

Необходимо предсказывать оценки  $r_{i,a}$ , зная некоторые уже расставленные в матрице оценки  $r_{i',a'}$ . Получив наилучшие предсказания  $\hat{r}_{i,a}$ , необходимо выбрать один или несколько продуктов, для которых  $\hat{r}_{i,a}$  максимальны.

Общий алгоритм выглядит следующим образом:

1. Найти, насколько другие пользователи (продукты) похожи на данного пользователя (продукт).
2. По оценкам других пользователей (продуктов) предсказать, какую оценку даст данный пользователь продукту, учитывая с большим весом тех пользователей (продукты), которые больше похожи на данный.

Термин «похожий» в данном случае означает коэффициент корреляции двух векторов вещественных чисел на основе элементов, которые присутствуют в обоих векторах.

Коэффициент корреляции Пирсона равен

$$w_{ij} = \frac{\sum_a (r_{i,a} - \bar{r}_i)(r_{j,a} - \bar{r}_j)}{\sqrt{\sum_a (r_{i,a} - \bar{r}_i)^2} \sqrt{\sum_a (r_{j,a} - \bar{r}_j)^2}}$$

$\bar{r}_i$  — средний рейтинг объекта пользователя  $i$ .

Далее вычисляется предположительная оценка пользователя конкретному предмету как сумма средней оценки данного пользователя и среднего отклонения других пользователей от их средней оценки.

$$\hat{r}_{i,a} = \bar{r}_i + \frac{\sum_j (r_{j,a} - \bar{r}_j) w_{ij}}{\sum_j |w_{ij}|}$$

При вычислении оценки используют данные пользователей только множества  $K$  наиболее похожих на данного пользователя, т.к. большое количество пользователей с непохожими или несильно пересекающимися интересами отрицательно влияют на точность предсказания оценки.

Существуют следующие подходы:

- «рекомендации, основанные на пользователях», при котором происходит поиск похожих пользователей;
- «рекомендации, основанные на продуктах», при котором производится поиск похожих объектов.

Алгоритм, основанный на сравнении объектов, имеет несколько преимуществ. Т.к. интересы пользователей часто меняются, а средняя оценка объектов остается постоянной, то при сравнении объектов достаточно посчитать меру близости каждого с каждым и сохранить в памяти. В частом обновлении этих данных нет необходимости.

Используя меру близости и алгоритмы кластеризации пользователей или объектов можно разделить на группы, которые будут храниться рядом. В этом случае на первом шаге пользователя или объект можно сравнивать с пользователем или объектом соответственно из определенного кластера.

Матрица рейтингов, как правило, очень разреженная, т.к. средний пользователь оценивает немного. Для объектов с малым количеством рейтингов можно использовать методы, основанные на знаниях.

Присутствует проблема холодного старта для пользователей и объектов, когда недостаточно данных для рекомендации. Проблема холодного старта для пользователей может быть решена с использованием демографических данных, которые сами пользователи указывают при регистрации. Проблема холодного старта для объектов может быть решена с использованием алгоритмов анализа содержимого.

При достаточном количестве оценок объектов пользователями методы коллаборативной фильтрации эффективны, а при объединении различных методов фильтрации эффективность рекомендательных систем повышается.

Список использованных источников:

1. А.Г.Гомзин, А.В.Коршунов, Системы рекомендаций: обзор современных подходов, Труды Института системного программирования РАН, 2012
2. F. Ricci, L. Rokach, B. Shapira, P. Kantor, Recommender Systems Handbook, 2010
3. P. Melville, V. Sindhvani; Recommender systems, Encyclopedia of Machine Learning, 2010