

А. С. Хомяков, А. В. Паньков
(ГрГУ им. Я. Купалы, Гродно)

ИСПОЛЬЗОВАНИЕ МАШИННОГО ОБУЧЕНИЯ НА ПРИМЕРЕ АЛГОРИТМА КОЛЛАБОРАТИВНОЙ ФИЛЬТРАЦИИ ДЛЯ ПОСТРОЕНИЯ РЕКОМЕНДАТЕЛЬНЫХ СЕРВИСОВ

Рассматривается задача построения рекомендаций для пользователей с помощью алгоритма коллаборативной фильтрации. Пусть в системе есть пользователи A и объекты I . Пусть некоторые пользователи оценили некоторые объекты, оценка – это натуральное число от 1 до 5. Зафиксируем некоторого случайного пользователя a . Задача заключается в предсказании оценки, которую пользователь a поставил бы объекту i . Будем рассматривать только пользователя a и тех пользователей, которые оценили объект i . Алгоритм включает в себя 3 шага. **Первый шаг.** Для каждого пользователя u вычислим, насколько его интересы совпадают с интересами пользователя a . Каждому пользователю в матрице $R(A \times I)$ соответствует одна строка. Поэтому будем вычислять близость векторов строк пользователей. Используем косинусную меру схожести:

$$sim(u, a) = \frac{\sum_{z=1}^m r_{a_i} r_{u_i}}{\sqrt{\sum_{z=1}^m r_{a_i}^2} \sqrt{\sum_{z=1}^m r_{u_i}^2}},$$

где $sim(u, a)$ – мера близости (похожести) пользователей a и u , $r_{u,i}$ – элемент матрицы R , u – строка, i – столбец. **Второй шаг.** Выбор подмножества K множества пользователей A , наиболее похожих на пользователя a . Выбираем первые k элементов множества элементов A , упорядоченного по убыванию меры близости, где k – некоторая целая константа. **Третий шаг.** Имея множество K близких пользователей, нужно вычислить оценку, которую поставил бы пользователь a объекту i . Предсказываемая оценка пользователя a объекта i находится как ближайшее целое от $p_{a,i}$,

$$p_{a,i} = \frac{\sum_{u \in K} X sim(u, a)}{\sum_{u \in K} sim(u, a)}$$

Суть оценки – средняя схожесть по всем пользователям из множества K . Для того, чтобы сделать рекомендации для пользователя, достаточно предсказать оценки для всех не оцененных объектов и выбрать объекты с наибольшей предсказанной оценкой.

Информация, полученная в ходе построения рекомендаций, позволяет прогнозировать значения оценок и осуществлять автоматический выбор объектов, которые могут заинтересовать пользователя.

РЕПОЗИТОРИЙ ГГУ ИМЕНИ Ф. СКОРИНЫ