

**Коллаборативная фильтрация методом сингулярных вейвлетов**

Романчак В.М., Лаппо П.М.

Белорусский национальный технический университет  
Белорусский государственный университет

С помощью коллаборативной фильтрации можно предсказать, какой продукт понравится пользователю, имея неполную информацию о его предпочтениях.

**Пример.** Допустим, у нас имеется матрица оценок, выставленных пользователями различным продуктам. Строки матрицы соответствуют пользователям, столбцы матрицы - продуктам. Не все элементы матрицы заполнены – имеются пропуски. Необходимо предсказать какие из продуктов будут востребованы пользователями.

Расстояние между объектами может быть просто евклидовым, если все данные измерены в абсолютной шкале или можно использовать коэффициент корреляции Пирсона для шкалы интервалов.

В основе алгоритма лежит предположение, что реальные таблицы содержат похожие между собой объекты (строки и столбцы). Если же избыточность отсутствует (как, например, в таблице случайных чисел), то предпочесть один прогноз другому невозможно. Наша задача аппроксимировать функцию  $U$  двух переменных  $X$  и  $Y$ , каждая из переменных которой является, в свою очередь, вектором:  $U=U(X_i, Y_j)$ , где  $X_i$  – строка с номером  $i$ ,  $Y_j$  – столбец с номером  $j$ .

Алгоритм тестировался на реальных данных. Таблица наблюдений была полностью заполнена ответами интервьюеров. Далее, часть данных считалась неизвестной и восстанавливалась с помощью метода сингулярных вейвлетов и по методу ближайших соседей. Результаты аппроксимации представлены в таблице 1.

Таблица 1

Фрагмент результатов аппроксимации

3	2	(2,9) 4 (2,3)	4
4	(2,5) 1 (2,6)	3	(4,2) 4 (4,3)
3	2	(2,5) 3 (2,1)	5
5	2	1	(4,4) 5 (4,7)

В таблице 1 приведены наблюдаемые ответы интервьюеров, в скобках - результаты аппроксимации по двум методам.