Урок 3. Коллаборативная фильтрация

1. В чем принципиальные отличия item-item рекомендаций от ALS?

Принципиальное отличие item-item рекомендаций от ALS (AlternatingLeastSquares) в том, что первый использует детерминированный алгоритм - Метод KNN - k -ближайших соседей, где предсказываются те числа, которые стоят в самой матрице user-item (покупки, количество покупок, рейтинг, взвешенные метрики и т.д.), а второй – использует принцип разбиения матрицы user-item (где воспринимает ее как матрицу из 0 и 1) на матрицу латентных факторов юзеров и матрицу латентных факторов товаров, значения в которых (или латентные факторы, или они же эмбеддинги, или они же - весовые коэффициенты) подбираются путем обучения методом градиентного спуска, где в процессе обучения сперва обновляются значения латентных факторов для матрицы латентных факторов товаров, а потом - значения латентных факторов для матрицы латентных факторов юзеров, и затем после обучения произведение полученных матриц латентных факторов юзеров и товаров и образуют результирующую матрицу предсказаний user-item, где предсказанные значения лишь только некоторые score, но не сами целевые значения и чем они больше, те больше вероятность правильного предсказания.

1. Приведите 3 примера весов user-item матрицы для задачи рекомендаций товаров

Существуют TF-IDF взвешивание, BM25 взвешивание, косинусное взвешивание, которые генерируют веса каждый по своим алгоритмам.

1. Какие ограничения есть у ALS? (Тип информации, линейность/нелинейность факторов и т д)

Результат, выдаваемой ALS – некоторые score,

а не целевые параметры (факт покупки, количество покупок и т.д.) или их вероятность, причем чем больше они, тем более релевантен товар, т.е. больше вероятность покупки.   
Т.к. алгоритм работает методом градиентного спуска для подбора факторов латентных товаров и юзеров, то факторы – это коэффициенты линейного уравнения, где в результате их перемножения и сложения(перемножения - матрицы латентных факторов юзеров на матрицу латентных факторов товаров) появляются результирующие значения в матрице user-item.

Также не учитывается сезонность.

Не учитываются характеристики самих пользователей (пол, возраст и т.д.) и товаров (категория, цена и т.д.)