

H&M Personalized Fashion Recommendations

Zharova Maria

Daniil Merkulov

zharova.ma@phystech.edu daniil.merkulov@skoltech.ru

Project Proposal

В проекте рассматривается создание рекомендательной системы на основе алгоритма матричной факторизации, конкретно - метода ASVD (Asymmetric singular value decomposition). Это один из наиболее известных эффективных методов, получивший особую популярность после соревнования Netflix Prize. В данной работе предлагается реализовать приведенный алгоритм и в качестве бонуса сравнить его с результатом работы классического градиентного спуска.

1 Идея

Помимо явных признаков на "рекомендации" могут влиять также другие скрытые факторы, которые тоже необходимо учитывать. Для решения этой проблемы применяются SVD-подобные разложения более низкого ранга матрицы рейтингов, в частности - модель ASVD, которая является улучшением обычного SVD, учитывающим неявную информацию. Перечислим ещё ряд дополнительных преимуществ по сравнению с SVD [1]:

- уменьшение сложности за счёт использования меньшего числа параметров (происходит замена характеристик пользователей признаками объектов);
- возможность обрабатывать новых пользователей без повторного обучения модели и оценки новых параметров вследствие отсутствия параметризации;
- прогнозы хорошо ообъяснимы, т.к. являются прямой функцией ответов прошлых пользователей - такая структура позволяет определить, какие из прошлых действий пользователя больше всего влияют на вычисляемый прогноз;

Таким образом, используя ASVD, можно предоставить пользователям более быстро и точно качественные рекомендации.

1.1 Problem

Рассмотрим модель, которую даёт нам обычное сингулярное разложение: каждый пользователь u связан с вектором факторов $p_u \in R_f$, а каждый элемент i — с вектором факторов элементов $q_i \in R_f$. Тогда предсказание делается по правилу:

$$r_{ui} = b_{iu} + p_u^T q_i.$$

Расширим модель, учитывая также неявную информацию, для этого используем следующее правило прогнозирования:

$$r_{ui} = b_{iu} + p_u^T (|R(u)|^{\frac{1}{2}} \sum_{j \in R(u)} (r_{uj} - b_{uj}) x_j + |N(u)|^{\frac{1}{2}} \sum_{j \in R(u)} y_j)$$

- здесь каждому элементу i ставится в соответствие три фактор-вектора q_i , x_i , $y_i \in R_f$, т.е. вместо явной параметризации пользователей, мы представляем их через элементы, которые они могли бы предпочесть [1].

2 Outcomes

Результатом проекта будет реализация рекомендательной системы товаров для пользователей интернет-магазина H&M на языке Python, которая будет основана на работе алгоритме ASVD; а также оценка её качества по описанным далее метрикам.

3 Литературный обзор

Краткая теория и необходимые формулы для написания кода описываются в статьях "[Factorization meets the neighborhood: a multifaceted collaborative filtering model](#)" и "[Collaborative Filtering Ensemble](#)".

Более подробно механизм устройства рекомендательных систем можно посмотреть в книге "[RecommenderSystem](#)" и лекциях курса "Introduction to recommender systems" от Сколтеха.

Для примера можно посмотреть дипломную работу выпускника ВШЭ "[Factorization meets the neighborhood: a multifaceted collaborative filtering model](#)".

В качестве дополнительного материала для лучшего понимания проблемы и альтернативных способов её решения можно изучить краткий курс на Coursera "[Advanced Recommender Systems](#)" и статью "[Understanding and improving relational matrix factorization in recommender systems](#)".

Также сохраним себе небольшую [шпаргалку по факторизации](#).

4 Метрики качества

Оценивать модель будем по метрике, указанной в условиях соревнования на Kaggle [2] (откуда возьмём данные для построения рекомендательной системы) - Mean Average Precision @ 12 (MAP@12):

$$MAP@12 = \frac{1}{U} \sum_{u=1}^U \frac{1}{\min(m, 12)} \sum_{k=1}^{\min(n, 12)} P(k) \times rel(k),$$

где U - количество клиентов, $P(k)$ - точность при отсечке k , n - количество прогнозов на одного клиента, m - количество основных значений истинности на клиента, $rel(k)$ - индикаторная функция: равна 1, если элемент ранга является релевантной меткой, в противном случае — 0.

5 Примерный план

- Исследование и предобработка датасета (выделение нужных признаков, очистка данных) - 23 апреля
- Построение алгоритма матричной факторизации ASVD - 27 апреля
- Применение алгоритма к поставленной задаче и его оценка - 2 мая
- Если успею, проведу сравнение построенной системы на основе ASVD с алгоритмом, построенным на градиентном спуске - 5 мая
- Доработка проекта и финальная отсылка результатов на Kaggle - 9 мая

6 References

- 1 Yehuda Koren, Factorization Meets the Neighborhood: a Multifaceted Collaborative Filtering Model, 2008.
- 2 Kaggle competition "HM Personalized Fashion Recommendations", <https://www.kaggle.com/competitions/h-and-m-personalized-fashion-recommendations/overview/evaluation>