

# Выпускная квалификационная работа

## Построение ансамбля алгоритмов рекомендаций

**Выполнил:**  
студент 417 группы  
Кудрявцев Георгий Алексеевич

**Научный руководитель:**  
д.ф-м.н., профессор  
Дьяконов Александр Геннадьевич

3 мая 2016 г.

# Неформальная постановка задачи

Требуется улучшить качество работы алгоритмов ранжирования при помощи ансамблирования уже существующих методов.

Рассматривается задачи ранжирования по данным с двоичной релевантностью.

На практике данная задача решается при помощи алгоритмов машинного обучения.

В данной работе рассматриваются факторизационные методы и их линейные ансамбли.

Построение рекомендаций для:

- социальных сетей
- сайтов знакомств
- интернет магазинов

Цель: разработать метод ансамблирования, который стабильно улучшает качество ранжирования.

Задачи:

- Составить обзор современных факторизационных методов ранжирования.
- Предложить эффективный метод ансамблирования.
- Реализовать методы и провести их сравнительный анализ.

# Формальная постановка задачи

## Входные данные:

матрица  $R$  размера  $M \times N$ , где  $M$  - количество пользователей,  $N$  - количество предметов.  $R_{ui} = 1$ , если пользователи  $u$  взаимодействовал с предметов  $i$ . В противном случае  $R_{ui} = 0$ .

## Выходные данные:

Для каждого пользователя  $u$  ранжированный список предметов, которые не лежат в тренировочной выборке.

# Обзор существующих методов

- **CLiMF**<sup>1</sup> – Факторизационный метод, который оптимизирует сглаженную версию метрики MRR.
- **MPR\_MF**<sup>2</sup> – Факторизационный метод, который оптимизирует AUC.
- **TFMAP**<sup>3</sup> – Факторизационный метод, который оптимизирует сглаженную метрику MAP.
- **iMF**<sup>4</sup> – Факторизационный метод, который оптимизирует взвешенную квадратичную ошибку.

---

<sup>1</sup>Yue Shi, Alexandros Karatzoglou, Linas Baltrunas. CLiMF: learning to maximize reciprocal rank with collaborative less-is-more filtering. 2012

<sup>2</sup>Steffen Rendle, Christoph Freudenthaler, Zeno Gantner. BPR: Bayesian Personalized Ranking from Implicit Feedback. 2009

<sup>3</sup>Yue Shiao, Alexandros Karatzoglou, Linas Baltrunas. TFMAP: Optimizing MAP for Top-N Context-aware Recommendation. 2012

<sup>4</sup>Yifan Hu, Yehuda Koren, Chris Volinsky. Collaborative Filtering for Implicit Feedback Datasets. 2008

Пусть имеется множество базовых алгоритмов  $b_i(x)$ .

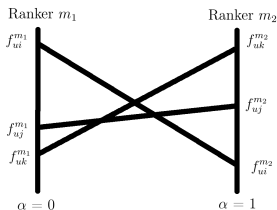
Необходимо подобрать такие веса  $\alpha_i$ , чтобы линейная комбинация алгоритмов  $\hat{b}(x) = \sum_i \alpha_i b_i(x)$  показывала лучший результат по какому-нибудь заданному функционалу.



# Оптимальное взвешенное голосование

Рассмотрим линейную комбинацию двух алгоритмов ранжирования в следующем виде:

$$\hat{f}_{ui} = \alpha f_{ui}^{m_1} + (1 - \alpha) f_{ui}^{m_2}, \text{ где } 0 \leq \alpha \leq 1$$



**Рис. :** Две вертикальные линии обозначают размер величины  $f$ . Чем выше точка, тем больше величина. Невертикальные линии обозначают линейные комбинации величин  $f$ . Если  $\alpha = 0$ , то линейная комбинация учитывает только Ranker  $m_1$ , если  $\alpha = 1$ , то линейная комбинация учитывает только Ranker  $m_2$