Санкт-Петербургский Государственный Университет
Математико-механический факультет
Кафедра системного программирования

Рекомендации треков в социальных сетях

Александр Александрович Дзюба, 661 группа

Магистерская диссертация

Научный руководитель: к.ф.-м.н., Д.Ю. Бугайченко
Рецензент: к.ф.-м.н., доцент И.П. Соловьев

Рекомендательные системы

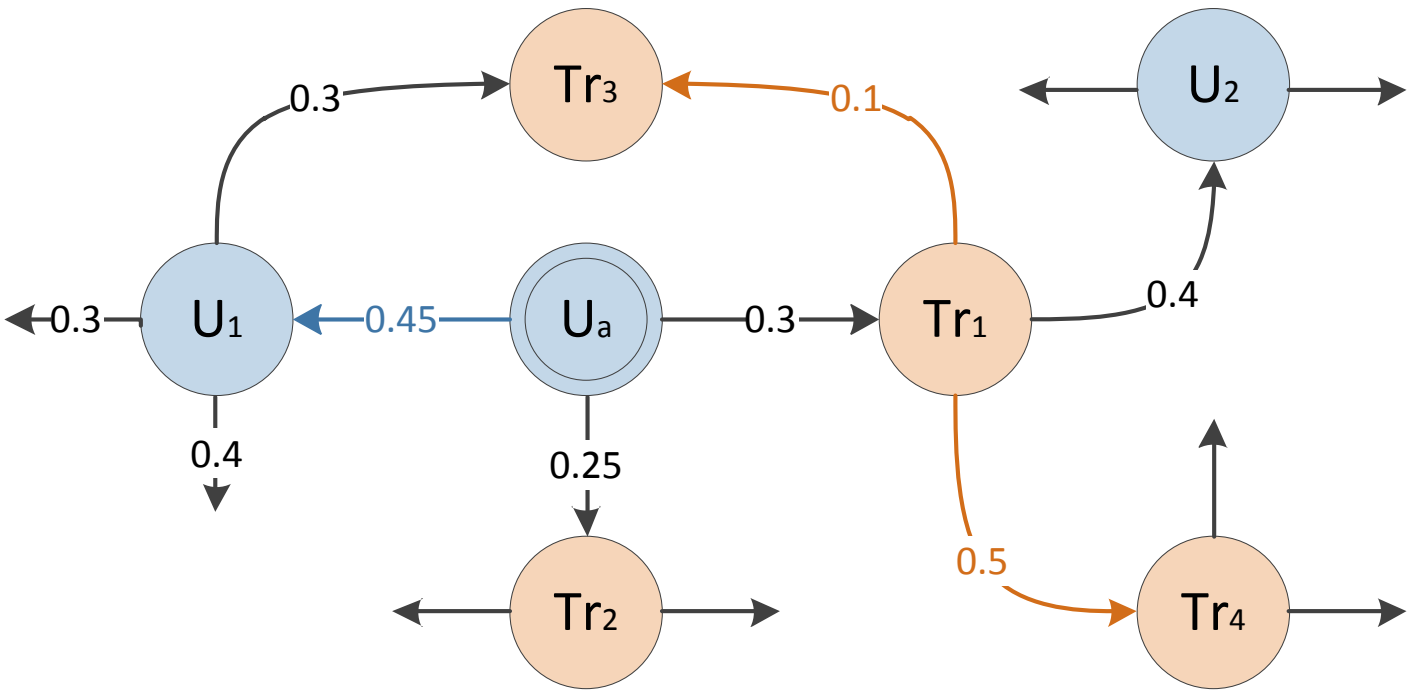
- Предсказывают, какой контент понравится пользователю
- Прогноз рассчитывается на основе сетевого профиля
  - С оценками предметов
  - Историей взаимодействия с предметами (количество прослушиваний, просмотров, фактов покупки)
  - С указанными предпочтениями (жанры, исполнители)

Классические методы

- Коллаборативная фильтрация работает с матрицей оценок предметов, выставленных пользователями
  - *User-based* – поиск похожих пользователей и использование их оценок для предсказания
  - *Item-based* – поиск преметов, которые нравятся пользователям вместе
- Рекомендация по содержимому сопоставляет известные интересы пользователя и описание контента
  - Применимость сильно ограничена

Алгоритм Random Walk with Restarts

- По качеству превосходит коллаборативную фильтрацию
- Граф социальной сети



- Вычисление степени «связанности» вершин с помощью случайного обхода

Алгоритм Random Walk with Restarts

Расчет рекомендаций:

$$\mathbf{p}^{(t+1)} = (1 - a)\mathbf{S}\mathbf{p}^{(t)} + a\mathbf{q}$$

$\mathbf{p}^{(t)}$ – вектор предпочтений на шаге t
$\mathbf{S}$ – матрица смежности социального графа
a – вероятность возврата в начало
$\mathbf{q}$ – вектор возврата

UU	UTr	UTg
UTr	0	TgTr
UTg	TgTr	0

Постановка задачи

- Реализовать рекомендательную систему на основе алгоритма Random Walk with Restarts
- Использовать традиционные методы расчёта рекомендации для составления графа социальной сети
- Оценить качество рекомендаций для разных способов построения графа, используя данные реальных социальных сетей
- Предложить метод ускорения расчёта рекомендаций для применения в крупных социальных сетях
- Оценить время и качество рекомендаций при использовании предложенного метода

Тестовые данные

Last.fm

- 3 тыс. пользователей
- 30 тыс. треков
- 800 тыс. связей пользователь-трек
- 27 млн. прослушиваний

Улучшение качества

- Замена социальных связей близостью вкусов
  - Расчет схожести векторов-рейтингов
  - Косинус угла между векторами, коэффициент Пирсона и другие
  - Реализованы в Apache Mahout
- Добавление дуг между схожими треками
- Раздельное нормирование столбцов S
- Оценки качества
  - Recall-Precision кривые
  - Half-life utility

Оценка качества, Half-life utility

- Разделение истории прослушиваний на тренировочное и проверочное множество
- Элементы проверочного множества (Q) заранее неизвестны и используются для определения релевантности результата

$$HL_{a,c} = \frac{\sum_j r_{a,j}}{2^{(j-1)(c-1)}}, \text{ где } r_{a,j} = \begin{cases} 0, & \text{если } v_j \notin Q_a; \\ 1, & \text{если } v_j \in Q_a. \end{cases}$$

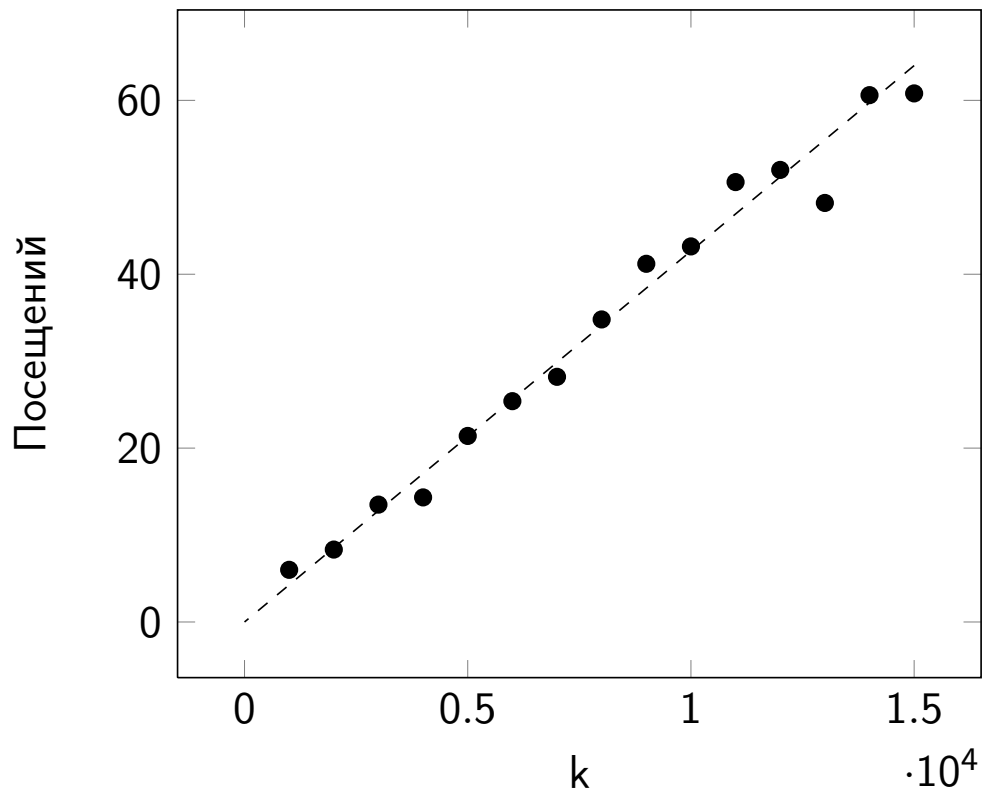
c – «период полураспада»

v_j – j-й рекомендованный трек

- Удалось существенно улучшить качество рекомендаций согласно данной метрике
  - для $c = 10$: с 2.675 до 4.15

Симуляция

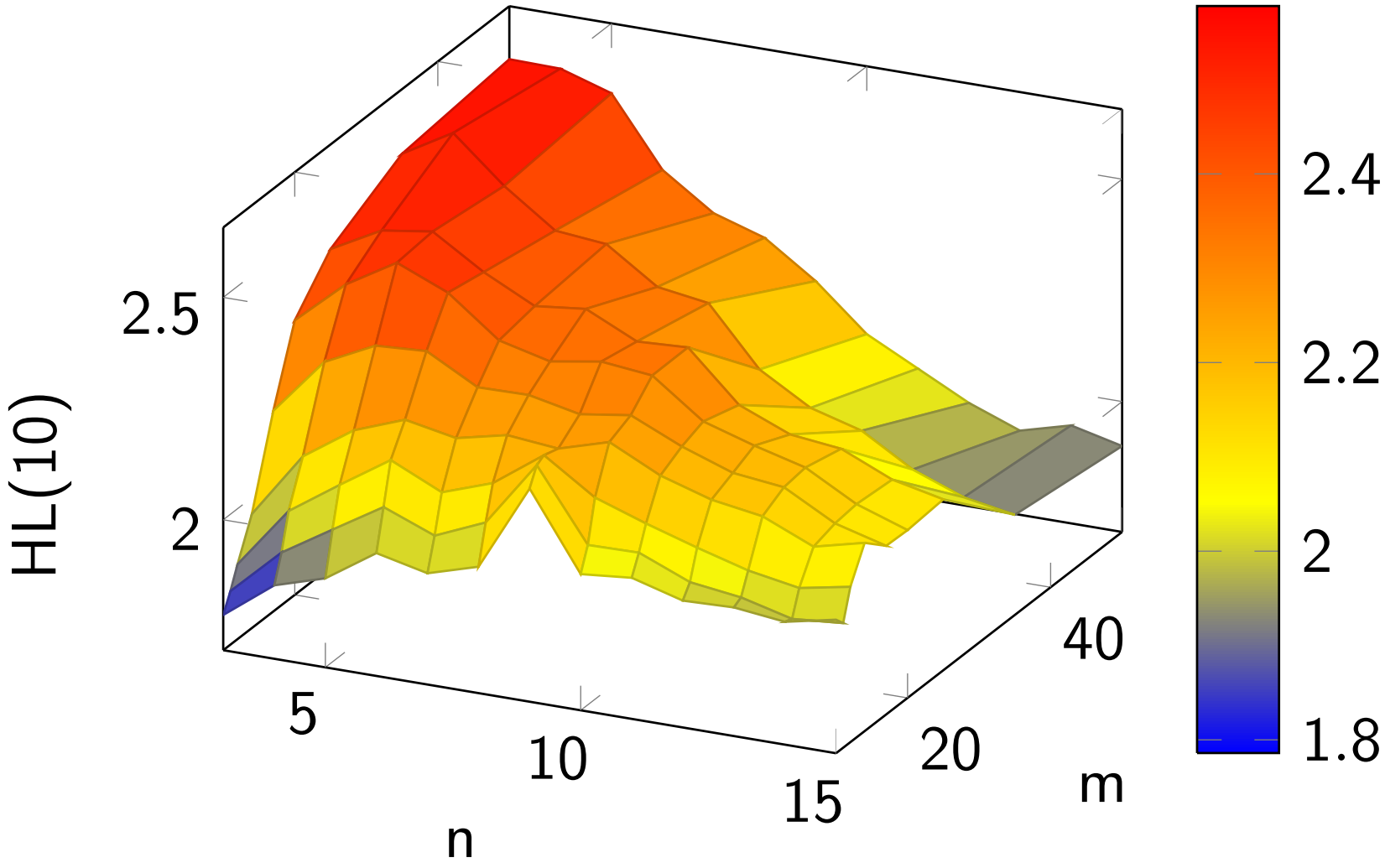
- Считается количество посещений вершин-треков при случайном обходе из k шагов
- При $k < 10^5$ необходимая точность не достигается



Персональные подграфы

- Подграф строится сразу после запроса рекомендаций:
  1. Включаются наиболее близкие пользователю объекты
  2. Аналогичные действия для объектов, добавленных на 1 шаге
- Время и качество расчёта рекомендаций зависит от размера включения вершин и дуг

Зависимость качества от размера включения вершин



n включаемых вершин на 1 шаге

m включаемых вершин на 2 шаге

Результаты

- Реализована Java-платформа для конфигурирования и тестирования рекомендательных систем
- Предложена модификация алгоритма Random Walk with Restarts, дающая более качественные рекомендации
- Предложен способ упрощения оригинального алгоритма, проведено исследование качества предложенного метода
- Приведены результаты симуляции алгоритма Random Walk with Restarts
- Описанные методы протестированы на наборе данных социальной сети Last.fm