



**Кафедра системного программирования СПбГУ**

Квалификационная работа  
бакалавра информационных технологий

## **Агрегирование из различных источников рекомендаций для аудиозаписей**

**Автор:**

Лебедев Д. Ю.

**Научный руководитель:**

к. ф.-м.н. | Бугайченко Д. Ю.

**Рецензент:**

д. ф.-м.н. | Соловьев И.П.

8 июня 2012 г.

# Системы рекомендаций

- Социальные сети
  - Системы рекомендаций (youtube.com, last.fm)
- Подходы к предоставлению рекомендаций
  - Content-based
  - Collaborative filtering

# Постановка задачи

- Реализовать алгоритм расчета схожести пользователей
- Реализовать механизм агрегации рейтингов
- Произвести оценку предоставленных рекомендаций

# Схожесть пользователей

- Схожесть пользователей  $u$  и  $v$ :

$$W_{u,v} = \alpha * \text{connected}(u, v) + \beta * \text{tracks}(u, v) + \gamma * \text{group}(u, v) + \delta * \text{friends}(u, v)$$

- $\text{connected}(u, v) \in \{0, 1\}$  - связь между пользователями  $u$  и  $v$
- $\text{group}(u, v) \in \{0, 1\}$  - принадлежность к одной группе (нейросетевая группировка)
- $\text{tracks}(u, v) \in [0, 1]$  - схожесть по количеству общих аудиозаписей
- $\text{friends}(u, v) \in [0, 1]$  - схожесть по количеству общих друзей

- $\cos(T, S) = \frac{T \cdot S}{\|T\| \|S\|}$

# Агрегация

- Расчет рейтинга производится по следующей формуле:

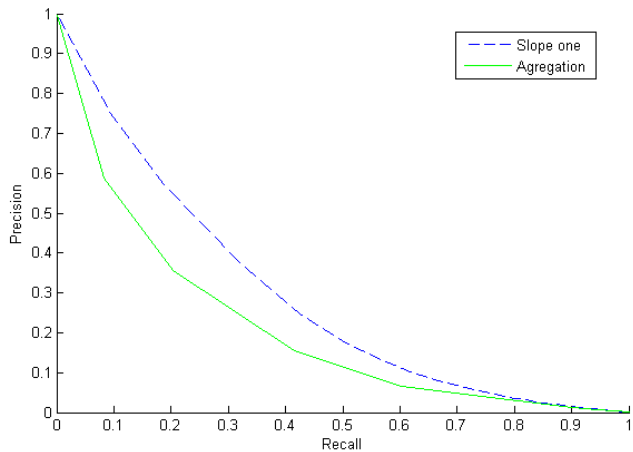
$$r_{u,i} = \bar{r}_u + \frac{\sum_{v \in N(u)} W_{u,v} * (r_{v,i} - \bar{r}_v)}{\sum_{v \in N(u)} W_{u,v}}$$

- $\bar{r}_u$  - средний рейтинг пользователя  $u$
- $N(u)$  - соседи пользователя  $u$  (нейросетевая группировка)
- $W_{u,v}$  - весовой коэффициент схожести пользователей  $u$  и  $v$

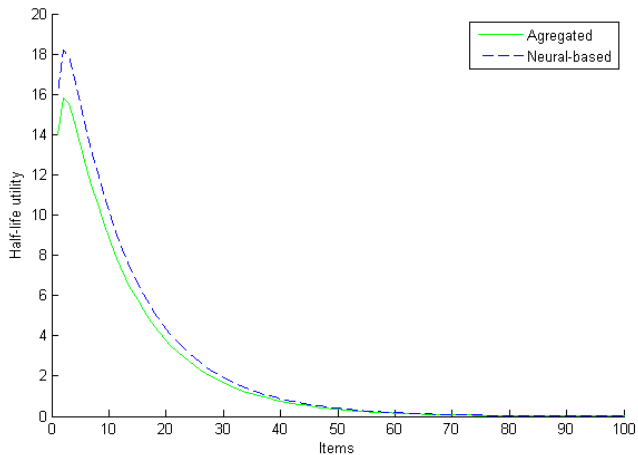
# Инструменты и данные для эксперимента

- Apache Mahout – библиотека машинного обучения
  - SparseMatrix - работа с разреженными данными
- Данные last.fm:
  - 10 000 пользователей
  - 50 000 артистов
  - 500 000 рейтингов

## Recall-Precision



# Half-life utility





# Заключение

- Реализован алгоритм расчета схожести пользователей на основе социальных связей и музыкальных предпочтений
- Реализовать механизм агрегации рейтингов на основе нейросетевой группировки пользователей
- Произведена оценка предоставленных рекомендаций