



Кафедра системного программирования СПбГУ

Квалификационная работа
бакалавра информационных технологий

Предсказание предпочтений пользователя с использованием нейросетевого подхода

Автор:

Солодкая А. С.

Научный руководитель:

к. ф.-м.н. | Бугайченко Д. Ю.

Рецензент:

д. ф.-м.н. | Соловьев И.П.

8 июня 2012 г.

Системы рекомендаций

- Рекомендации неизвестных ранее пользователю объектов
- Применение
 - Интернет-магазины (ozon.ru, amazon.com, netflix.com)
 - Социальные сети (last.fm, vk.com, movielens.com)

Нейронные сети

- Одна из гипотетических моделей работы головного мозга
- Применение
 - Распознавание образов
 - Прогнозирование
 - Аппроксимация
 - Кластеризация
 - Сжатие данных

Постановка задачи

- Реализация нейронной сети для разделения пользователей по предпочтениям
- Реализация механизма предсказания рейтингов треков у каждого конкретного пользователя на основе группировки пользователей по предпочтениям
- Получение и подготовка данных для эксперимента
- Реализация подсчета метрик для анализа точности и эффективности работы системы предсказаний рейтингов

Эксперимент

- Данные о пользователях и артистах из базы last.fm
- 10 000 пользователей, 50 000 артистов
- 500 000 рейтингов

Инструменты

- Apache Mahout – библиотека для масштабируемого машинного обучения
- Last.fm-bindings – библиотека для работы с last.fm API

Подготовка данных

- Пересчет количества прослушиваний в рейтинги
 - Уменьшение влияния популярных объектов (Inverse user frequency)
 - Рейтинг по умолчанию (Default voting)
 - Пересчет количества прослушиваний в рейтинги
 - Стандартное значение (z-score)
 - Усреднение

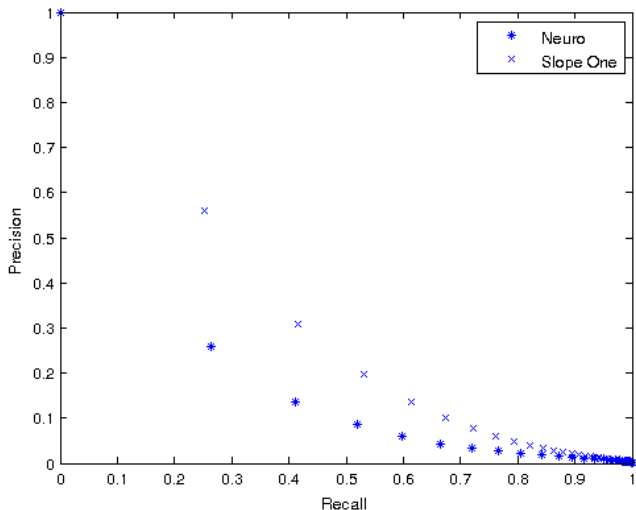
Группировка пользователей

- Уменьшение размерности на основе Locality Sensitive Hashing с k -квантователями
- Самоорганизующаяся карта Кохонена
 - Количество входов = количество артистов
 - Количество выходов = количество групп пользователей
 - Входные данные - вектор рейтингов соответствующих артистов у пользователя
 - Выходные данные - номер группы, к которой принадлежит пользователь

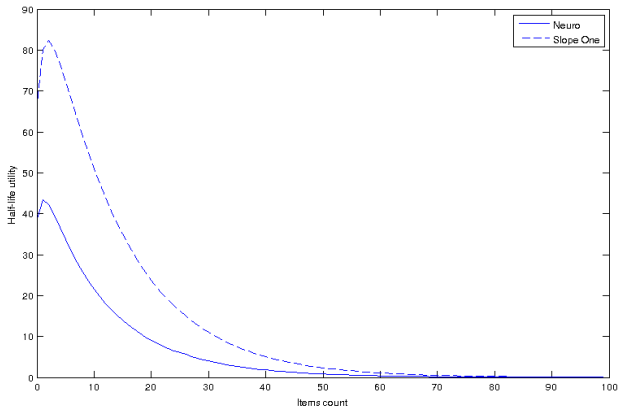
Рекомендации и анализ результатов

- Для каждого пользователя - агрегация рейтингов пользователей из его группы
- Анализ качества рекомендаций
 - Точность и охват (precision и recall)
 - Полезность полупериода (half-life utility)

Точность и охват



Полезность полупериода



Заключение

- Реализована нейронная сеть для разделения пользователей по предпочтениям
- Реализован механизм предсказания рейтингов треков у каждого конкретного пользователя на основе группировки пользователей по предпочтениям
- Реализован механизм получения и подготовки данных для эксперимента
- Реализован подсчет метрик для анализа точности и эффективности работы системы предсказаний рейтингов