

Санкт-Петербургский государственный университет

Математическое обеспечение и администрирование информационных  
систем

Кафедра информационно-аналитических систем

Смирнов Александр Львович

# Построение гибридной рекомендательной системы новостей с применением методов оптимизации

Дипломная работа

Научный руководитель:  
к.ф.-м.н., доц. Михайлова Елена Георгиевна

Санкт-Петербург  
2020

# Оглавление

Введение	3
1. Постановка задачи	4
2. Обзор	5
2.1. Методы . . . . .	5
2.2. Подходы . . . . .	5
3. Решение	6
4. Дальнейшие планы	7
Список литературы	8

# Введение

Невозможно представить современное приложение, оперирующее большими объёмами информации, без рекомендательной системы. Информации становится настолько много, что пользователи не имеют возможности найти интересный себе контент. Для решения данной проблемы применяются рекомендательные системы.

Рекомендательная система – совокупность алгоритмов, которые принимают на вход данные об активности пользователей и информацию о рекомендуемых предметах (дата создания, источник и т.д.) и на выходе выдают список предметов, отранжированный по мере того, насколько он релевантен конкретному пользователю.

Компании IT Service требуется рекомендательная система новостей, для того, чтобы увеличить вовлечённость пользователей путём рекомендации релевантного контента. Ключевым фактором является то, что отсутствует информация о действиях пользователей, вследствие чего не представляется возможным оценить работу алгоритмов на реальных данных.

# 1. Постановка задачи

Целью данной работы является построение и оптимизация гибридной рекомендательной системы новостей. Для её достижения необходимо решить следующие задачи:

- провести обзор существующих решений
- реализовать каждое уникальное решение и оценить его
- соединить различные решения в единую систему
- найти оптимальный вклад каждого отдельного решения в результирующий список рекомендаций

## 2. Обзор

### 2.1. Методы

В области рекомендательных систем различают два основных подхода:

- collaborative filtering
  - рекомендации на основе действий других пользователей
- content-based filtering
  - рекомендации на основе содержимого

Оба подхода обладают своими плюсами и минусами. К плюсам коллаборативной фильтрации можно отнести отсутствие необходимости знать что-либо о рекомендуемом предмете, так как рекомендации строятся на основе действий других пользователей. К минусам относится необходимость иметь данные об активности других пользователей, что называется "холодный старт".

В свою очередь рекомендации на основе контента не требуют данных о действиях других пользователей, но необходимы признаки, которыми обладает рекомендуемый предмет, для построения вектора предпочтений пользователя.

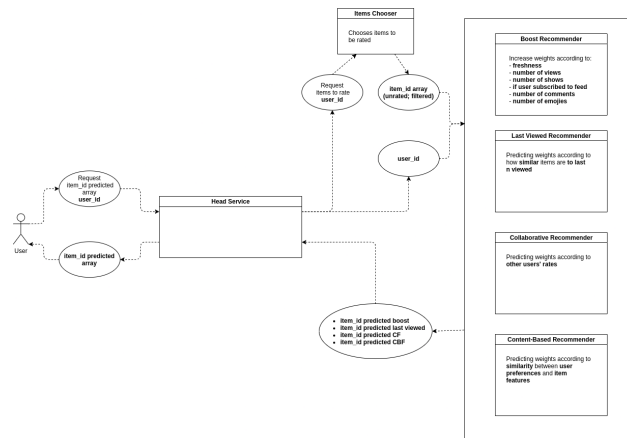
### 2.2. Подходы

Для коллаборативной фильтрации используют метод kNN [1] и корреляцию Пирсона для нахождения похожих векторов.

Для рекомендации на основе содержимого используются байесовский классификатор, решающие деревья, кластерный анализ.

### 3. Решение

В качестве решения предлагается комбинация подходов:



Подходы:

- Content-based recommender
- Collaborative recommender
- Session recommender
  - рекомендации на основе последних просмотренных элементов
- Boost recommender
  - рекомендации на основе информации об элементе

Каждый из сервисов уже был реализован, остаётся решить задачу оптимизации, то есть найти подходящие веса для каждого из подходов.

## 4. Дальнейшие планы

В начале января планируется релиз приложение и масштабная пиар-компания, вследствие чего будет доступно очень много логов пользовательской активности.

Собирая и используя эту информацию мы сможем оценить качество построенных моделей.

Также планируется проводить А / В тестирование.

## Список литературы

- [1] Wikipedia. k-nearest neighbors algorithm // Википедия, свободная энциклопедия. — 1992. — URL: [https://en.wikipedia.org/wiki/K-nearest\\_neighbors\\_algorithm](https://en.wikipedia.org/wiki/K-nearest_neighbors_algorithm) (online; accessed: 21.12.2020).