

**Трофимова Е.В., Туральчук К.А.** ©  
Магистрант, СПбГПУ, ИИТУ, СПб

## **РАЗРАБОТКА РЕКОМЕНДАТЕЛЬНОЙ СИСТЕМЫ НА ОСНОВЕ АНАЛИЗА ТОНАЛЬНОСТИ ТЕКСТА**

### *Аннотация*

*Всемирная сеть содержит огромное количество информации, и пользователи часто сталкиваются с проблемой поиска информации. В этой статье, мы описываем рекомендательную систему фильмов, которая использует гибридный способ фильтрации и строит систему рейтингов, основываясь на тональности сообщений пользователей социальных сетей.*

**Ключевые слова:** web-mining, рекомендательные системы, анализ тональности текста, холодный старт, матрица рейтингов, коллаборативная фильтрация, контентные методы.

**Keywords:** web-mining, recommender system, sentiment mining, coldstart, user-ratings matrix, collaborative filtering, content-based methods.

Интернет в настоящее время содержит огромное количество знаний и информации. Пользователи могут просматривать всевозможные документы, аудиофайлы и видеофайлы. Однако это многообразие информации влечет за собой проблемы, которые могут возникнуть как при анализе, так и при поиске необходимой информации.

Проблема поиска нужной информации связана с тем, что пользователь не всегда сразу может найти необходимые ему электронные ресурсы. Проблема изучения потребителей связана с предоставлением пользователю информации, которая оказалась бы ему интересна. Это особенно актуально для электронных торговых порталов, которые могли бы предлагать пользователю продукты, которые могли бы его заинтересовать.

Рекомендательные системы — программы, которые пытаются предсказать, какие объекты (фильмы, музыка, книги, новости, веб-сайты) будут интересны пользователю, имея определенную информацию о его профиле. Рекомендательные системы анализируют интересы пользователей и пытаются предсказать, что именно будет наиболее интересно для конкретного пользователя в данный момент времени.[1]

Большинство рекомендательных систем использует методы коллаборативной фильтрации, методы, основанные на контексте или гибридные методы.

Пусть есть множество пользователей и множество продуктов, причём некоторые пользователи как-то оценили некоторые продукты. Можно представить себе эти данные как матрицу, каждая строка которой соответствует пользователю, а столбец – продукту. Задача рекомендательной системы предсказать неизвестные элементы матрицы, и какие из неизвестных элементов будут максимальными в своей строке, то есть, какие продукты больше всего понравятся тому или иному пользователю.[2, 187]

Контентная фильтрация формирует рекомендацию на основе поведения пользователя. Например, этот подход может использовать ретроспективную информацию о просмотрах (какие блоги читает пользователь и характеристики этих блогов). [3]

Методы, основанные на контенте могут однозначно характеризовать пользователя, но коллаборативная фильтрация имеет некоторые преимущества (коллаборативная фильтрация может работать в областях, где контент сложен для компьютерного анализа – идеи, мнения и т.д.).

У систем коллаборативной фильтрации существует две проблемы:

1. Матрица рейтингов, как правило, очень разреженная (sparse). Очень часто и пользователей, и продуктов много, а рейтингов гораздо меньше, чем их произведение, ведь средний пользователь оценивает совсем немного продуктов

2. Проблема холодного старта (coldstart). Когда в системе появляется пользователь, который оценил очень мало продуктов, то матрица рейтингов для него очень разреженная. Проблема холодного старта для продуктов заключается в том, что нужно определить сколько нужно рейтингов для нового продукта, прежде чем его можно будет уверенно рекомендовать. [2,188]

Гибридные методы объединяют подходы, используемые в коллаборативной и контентной фильтрации. Они объединяют в себе преимущества этих двух методов.

Разрабатываемая рекомендательная система должна будет предлагать пользователям фильмы, удовлетворяющие их запросам, с наилучшим рейтингом. Некоторые социальные сети, такие как Facebook, Twitter и Google Plus, имеют открытый API. В них с помощью мета-тэгов можно достать нужную информацию со страницы социальной сети. В данном случае, это комментарии пользователей. Система будет анализировать тональность этих комментариев, чтобы оценить отношение пользователя к данному фильму, с помощью методов Sentiment Mining (анализа тональности). Приложение интегрировано с инструментом NClassifier, который представляет собой Байесовский классификатор для .NET. Байесовский классификатор NClassifier проанализирует веса слов в комментариях. Слово с негативной окраской имеет отрицательный вес, а с позитивной – положительный вес. Таким образом будет подсчитываться рейтинг каждого фильма. Затем с помощью гибридного метода будут получены фильмы, имеющие хорошие рейтинги и удовлетворяющие вкусам пользователя.

### **Литература**

1. Рекомендательная система – Наука [Электронный ресурс]. – Режим доступа: [http://ru.science.wikia.com/wiki/Рекомендательная\\_система](http://ru.science.wikia.com/wiki/Рекомендательная_система)
2. P.Melville, R.J. Mooney, R.. Nagarajan Content-Boosted Collaborative Filtering for Improved Recommendations // University of Texas, USA, : Материалы конф. / AAAI-02, Austin, TX, USA – 2002, – С. 187-192.
3. Рекомендательные системы: Часть 1. Введение в подходы и алгоритмы [Электронный ресурс]. – Режим доступа: <http://www.ibm.com/developerworks/ru/library/os-recommender1/>