# СПОСОБЫ ПОСТРОЕНИЯ ГИБРИДНОЙ РЕКОМЕНДАТЕЛЬНОЙ СИСТЕМЫ НА ОСНОВЕ ДАННЫХ О ЗАКАЗАХ БИБЛИОТЕКИ

*Князева А.А., Колобов О.С., Турчановский И.Ю.*

Институт вычислительных технологий, Томский филиал, г. Томск

*DOI: 10.25743/ICT.2019.73.66.014*

В статье рассмотрены гибридные рекомендательные системы с точки зрения их применимости для использования в библиотеке университета. Приведены предложения по решению проблемы «холодного старта» при использовании методов коллаборативной фильтрации. В работе были использованы данные о выполненных заказах литературы в 2014-2015 гг. в Научно-технической библиотеке ТПУ.

*Ключевые слова: гибридные рекомендательные системы, коллаборативная фильтрация, контентные рекомендации, данные о заказах.*

# WAYS TO BUILD A HYBRID RECOMMENDATION SYSTEM BASED ON LIBRARY LOAN DATA

*Knyazeva A.A., Kolobov O.S., Turchanovsky I.Y.*

The methods of hybrid recommenders in terms of their applicability for using in an university library are considered in the paper. Suggestions are given for solving the problem of cold start when using collaborative filtering. The book loan data during 2014-2016 in Scientific and technical library of TPU were used in work.

*Ключевые слова на английском языке: hybrid recommender systems, collaborative filtering, content recommendations, loan data.*

**Введение.** В данной работе рассматривается задача построения рекомендаций для пользователей библиотеки при университете. Для такой библиотеки характерно относительно небольшое количество пользователей. В то же время, фонд библиотеки может быть достаточно обширным, особенно если учитывать возможности привлечения сторонних ресурсов и предоставление цифровых копий документов. В качестве примера в рамках данной работы была рассмотрена Научно-техническая библиотека ТПУ (НТБ ТПУ). Как указано на сайте библиотеки, её фонд составляет 2,4 млн. изданий технического, естественнонаучного, гуманитарного, социально-экономического профиля. Фонд включает научную и учебную литературу, в том числе периодические издания, информационные и реферативные журналы на бумажных и электронных носителях, диссертации, авторефераты диссертаций, нормативно-техническую и патентную документацию, редкие книги и рукописи [1].

Самым логичным способом на первый взгляд кажется построение рекомендаций по содержимому (метод content-based). Анализируем список документов, заказанных пользователем, и делаем для него подборку книг с похожим содержанием. Вариантов много: книги тех же авторов, книги с похожими заголовками, с теми же ключевыми словами в описаниях и т.п. Библиографические описания, хранящиеся в библиотечной информационной системе, содержат достаточно детальную информацию для большинства доступных документов. В то же время рекомендации по содержимому могут оказаться слишком однообразными: в биб-

лиотеке зачастую хранится несколько изданий одного и того же учебника, но это не значит, что мы должны рекомендовать их пользователю. Сама по себе схожесть описаний двух документов ещё не гарантирует интереса пользователя. Например, студент ВУЗа по специальности «Машиностроение» заказывает учебник «Высшая математика». Предлагать после этого для него в рекомендациях учебник с тем же названием для гуманитариев не обязательно будет хорошей идеей.

Коллаборативная фильтрация позволяет учесть схожесть в предпочтениях пользователей. Принцип подбора рекомендаций можно описать фразой: «Пользователи, заказавшие эту книгу, также заказывали…» Или «пользователи с похожими предпочтениями также просматривали…». По сравнению с рекомендациями по содержимому такой подход позволяет найти нетривиальные связи между книгами. Например, для студента-первокурсника, заказавшего в библиотеке книгу по основам физики, мы сможем порекомендовать учебник по линейной алгебре или основы теории вероятностей. И такая рекомендация будет более релевантной, чем, например, узкоспециализированная монография по физике. В то же время у коллаборативной фильтрации есть свои слабые стороны. Одна из наиболее частых проблем носит название «холодного старта». Мы не можем построить рекомендации для нового пользователя, о котором у системы ещё нет информации. Также, мы не можем подключить к рекомендациям новую книгу, которую ещё пока никто не заказывал. Ещё одной проблемой является «разреженность матрицы рейтингов». Как правило, пользователь заказывает ограниченное количество документов и не всегда возможно найти других пользователей, которые заказывали те же документы. Несмотря на перечисленные недостатки коллаборативная фильтрация остаётся одним из основных подходов формирования рекомендаций благодаря своей гибкости и возможности учёта предпочтений пользователей.

Необходимо также отметить ещё один распространённый подход к формированию рекомендаций: рекомендации по популярности. Этот подход является вычислительной дешёвым, поскольку он не предполагает персонализации рекомендаций: всем пользователям предлагается один и тот же набор из наиболее часто заказываемых документов.

В работе не рассмотрены методы построения рекомендаций на основе экспертных знаний или демографической информации, поскольку в данный момент такая информация нам недоступна.

**Данные.** В рамках работы были проанализированы данные о заказах документов за 2014-2016 гг. Под «документом» в данной работе подразумевается книга, журнал и любая другая публикация на бумажном носителе. Электронные документы, предоставленные пользователям, не были учтены в доступных нам данных.

Количество уникальных пользователей и документов за рассматриваемый период приведено в таблице 1.

Таблица 1. Количественное описание данных.

| | **2014** | **2015** | **2016** | **2014-2016** |
|---|---|---|---|---|
| **Уникальных пользователей** | 10 686 | 9 619 | 5 324 | 15 588 |
| **Уникальных документов** | 36 618 | 37 717 | 19 936 | 73 796 |

Информация о пользователях в располагаемом массиве данных о заказах закодирована, так что нет возможности получить какие-либо сведения о его специализации, поле, возрасте

и т.п. Заказанные документы представлены своими идентификаторами, что позволяет получить библиографическое описание с подробной информацией. Данные не содержат временных отметок за исключением года, что не позволяет учесть изменения предпочтений пользователей со временем.

**Гибридные рекомендательные системы.** Для того чтобы использовать сильные стороны различных алгоритмов и минимизировать ущерб от их недостатков, были придуманы гибридные рекомендательные системы.

Существует множество способов гибридизации [2]. Рассмотрим некоторые из них:

- Взвешенная комбинация (Weighted)
- Переключение (Switching)
- Смесь рекомендаций (Mixed)
- Конвейер (Cascade)
- Комбинирование признаков (Feature combination)
- Усиление признаков (Feature augmentation)

**Взвешенная комбинация (Weighted).** Интуитивно понятный способ объединения рекомендаций двух или более алгоритмов – присвоение каждому из них весового коэффициента и вычисление линейной комбинации рейтингов для каждого из объектов. Если используемые алгоритмы не вычисляют оценку рейтинга, а лишь представляют набор потенциально интересных пользователю объектов, то можно воспользоваться механизмом голосования.

Допустим, мы хотим объединить результаты коллаборативной фильтрации и метода на основе содержимого. В этом случае в начало списка рекомендаций попадут те объекты, которые были относительно высоко оценены и обоими методами.

**Переключение (Switching).** Под переключением подразумевается выбор одного из доступных методов построения рекомендаций в зависимости от обстоятельств [3]. Этот подход позволяет, например, решить проблему «холодного старта», которая является одним из слабых мест методов коллаборативной фильтрации. Для «холодного пользователя», предпочтения которого нам неизвестны, мы можем использовать рекомендовать наиболее популярные документы или подготовить «универсальный» набор.

**Смесь рекомендаций (Mixed).** Предположим, у нас есть несколько источников рекомендаций, которые мы хотели бы объединить. В этом случае можно задать правила смешивания, а затем показывать пользователю все рекомендации, чередуя их. Такой способ подходит в случае, когда источники рекомендаций для нас выступают в роли черных ящиков. Он не требует вычисления рейтинга объектов.

**Конвейер (Cascade).** Такой подход позволяет комбинировать алгоритмы последовательно. Результаты работы первого алгоритма в виде списка потенциальных рекомендаций поступает на вход второму алгоритму [4]. Например, с помощью анализа содержимого мы отсеиваем документы, которые с высокой долей вероятности не интересны для пользователя. А затем из оставшихся документов выбираем рекомендации с помощью коллаборативной фильтрации.

**Комбинирование признаков (Feature combination)** позволяет добавить в один алгоритм признаки, полученные с помощью другого алгоритма. Например, мы строим рекомендации на основе содержимого для конкретного пользователя и используем такие признаки документов, как название, автор, предметная рубрика и т.п. В то же время мы включаем в

рассмотрение рейтинги, выставленные данному документу пользователями, которые имеют схожие интересы. Таким образом, мы расширили перечень признаков для работы за счет привлечения коллаборативной фильтрации.

**Усиление признаков (Feature augmentation).** Этот подход похож на предыдущий. В нём можно выделить основной и дополнительный алгоритмы. Но, в отличие от комбинирования признаков, мы не добавляем новые признаки, а дополняем значения существующих [5]. В качестве примера рассмотрим коллаборативную фильтрацию в качестве основного алгоритма, а рекомендации по содержимому будем использовать как дополнительный алгоритм. Как правило, матрица рейтингов коллаборативной фильтрации является разреженной, то есть каждый пользователь выставляет небольшое число рейтингов. Мы можем взять результаты рекомендаций по содержимому и внести информацию о предполагаемых предпочтениях пользователя в матрицу рейтингов, сделав её более насыщенной. Далее полученная матрица с реальными и предполагаемыми рейтингами поступает на вход алгоритма коллаборативной фильтрации и на её основе формируются рекомендации.

Усиление признаков может использоваться также для гибкого подключения сторонних рекомендательных систем. Это может быть полезно в том случае, когда мы можем получить рекомендации для пользователя или список похожих документов от стороннего ресурса и учесть эту информацию в своей работе.

**Структура гибридной рекомендательной системы для НТБ ТПУ.** Использование гибридных подходов не ограничивается применением единственного способа комбинации алгоритмов. При необходимости можно использовать несколько гибридов в одной системе.

Предлагаемая архитектура гибридной рекомендательной системы для НТБ ТПУ представлена на рисунке 1. Первым задействованным элементом гибридизации является переключение. Для нового пользователя, о предпочтениях которого система не знает, она переключается на рекомендации по популярности. В случае если пользователь уже заказывал документы в библиотеке, система формирует для него рекомендации с помощью усиления признаков. В качестве основной рекомендательной системы при этом выступает коллаборативная фильтрация, а дополнительная система основана на оценке содержимого документов.

Таким образом, проблемы «холодного старта» для нового пользователя мы решаем с помощью переключения. В качестве алгоритма для нового пользователя неслучайно были выбраны рекомендации по популярности. Во-первых, для их вычисления не требуется дополнительной информации: достаточно оценить данные о заказах, которые уже есть в наличии. Во-вторых, рекомендации по популярности, частично представленные в таблице 2, по нашему мнению, отражают предпочтения студентов первого курса университета. Именно на эту группу пользователей библиотеки приходится наибольшее число новых пользователей каталога.

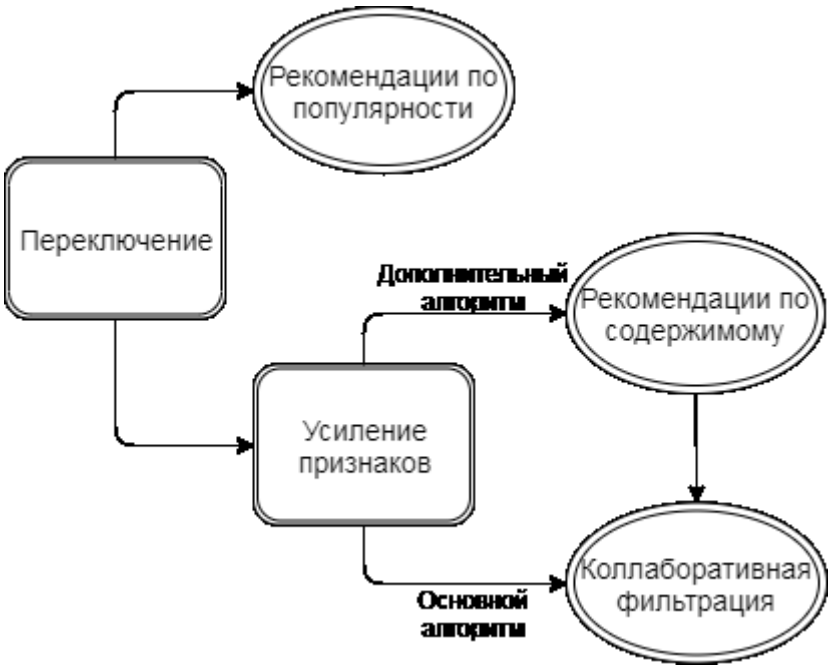


Рис. 1. Архитектура рекомендательной системы.

Таблица 2. Топ-5 рекомендаций по популярности.

| № | Название | Автор |
|---|---|---|
| **1** | Курсовое проектирование деталей машин | С. А. Чернавский |
| **2** | Лабораторный практикум по общей и неорганической химии | Н. Ф. Стась и др. |
| **3** | Справочник по общей и неорганической химии | Н. Ф. Стась |
| **4** | Сборник задач по физике. Механика. Молекулярная физика. Термодинамика | И. П. Чернов и др. |
| **5** | Отечественная история | Н. В. Трубникова |

Рекомендации по популярности не являются единственным возможным решением проблемы новых пользователей. Альтернативным подходом может быть создание специалистами специальных наборов рекомендаций, например, для каждой специализации. С одной стороны, это логичный механизм в традициях предложения списков литературы для тех же первокурсников. С другой стороны, в этом случае рекомендательная система должна определить номер группы или специальность, по которой обучается пользователь, что входит в противоречие с требованием защиты персональных данных.

Работать с новым документом нам позволяет механизм усиления признаков: благодаря использованию рекомендаций по содержимому новый документ попадает в качестве предполагаемого рейтинга в таблицу рейтингов, которая поступает на вход алгоритму коллаборативной фильтрации. Метод усиления признаков был выбран как наиболее перспективный для решения поставленной задачи, поскольку он хорошо зарекомендовал себя в похожих условиях [2]. Для того, чтобы убедиться в его эффективности для рекомендаций документов на основе данных о заказах планируется провести ряд экспериментов.

**Заключение.** В работе были рассмотрены различные способы объединения рекомендательных алгоритмов в гибридную рекомендательную систему с точки зрения их применимости для данных о заказах. В качестве исходных данных использовались данные о заказах до-

кументов Научно-технической библиотеки ТПУ за 2014-2016 гг. В результате работы был представлен проект архитектуры гибридной рекомендательной системы, удовлетворяющей доступным в настоящее время данным о предпочтениях пользователей. Данная система позволит решить такие проблемы рекомендательных алгоритмов как «холодный старт», разреженная матрица рейтингов, недостаточное разнообразие рекомендаций. В рамках дальнейшей работы планируется реализация гибридной рекомендательной системы и проведение экспериментов с целью выбора оптимального механизма гибридизации методов коллаборативной фильтрации и рекомендаций на основе содержимого.

## ЛИТЕРАТУРА


[1] О библиотеке / *НТБ ТПУ*. https://www.lib.tpu.ru/today/about.html (дата обращения 09.09.2019).

[2] *Burke R.* Hybrid web recommender systems // The adaptive web / Lecture Notes In Computer Science. 2007. V. 4321. P. 377-408.

[3] *Tejeda-Lorente Á., Porcel C., Peis E., Sanz R., Herrera-Viedma E.* A quality based recommender system to disseminate information in a university digital library // Inf. Sci. 261, 2014. P. 52-69.

[4] *Covington P., Adams J., Sargin E.* Deep Neural Networks for YouTube Recommendations. // Proc. of the 10th ACM Conference on Recommender Systems (RecSys '16). ACM, New York, NY, USA, 2016. P. 191-198.

[5] *Melville P., Mooney R.J., Nagarajan R.* Content-boosted collaborative filtering for improved recommendations // 8th national conf. on Artificial intelligence. Menlo Park, CA, USA, 2002. P. 187-192.