

*Ерёмин Олег Юрьевич, кандидат технических наук, доцент кафедры
«Компьютерные системы и сети», Московский государственный технический
университет им Н.Э. Баумана, г. Москва*

*Моркулев Дмитрий Викторович, магистрант, Московский государственный
технический университет им Н.Э. Баумана, г. Москва*

E-mail morkulev.dim@yandex.ru

МЕТОДЫ РЕАЛИЗАЦИИ ГИБРИДНЫХ РЕКОМЕНДАТЕЛЬНЫХ СИСТЕМ

Аннотация: В статье рассматриваются современные методы построения гибридных рекомендательных систем, позволяющие решать такие актуальные проблемы, как проблема холодного старта и агрегирования информации из разных источников, преимущества гибридов перед отдельными рекомендательными моделями, а именно моделями совместной фильтрации и рекомендации на основе контента. Представлена обобщенная классификация гибридных рекомендательных систем, для каждого из типов приведены примеры реализации и описана сфера применимости. По результатам анализа и сравнения представляется возможным выбрать наиболее подходящий метод реализации гибридной рекомендательной системы при отличающихся условиях и требованиях.

Ключевые слова: рекомендательная система, совместная фильтрация, сходство, рекомендация на основе контента, гибридные рекомендаторы.

Annotation: The article discusses modern methods for implementing hybrid recommender systems that allow solving such urgent problems as the problem of cold start and aggregation of information from different sources, the advantages of hybrids over individual recommender models, namely collaborative filtering models and

content-based recommendations. A generalized classification of hybrid recommender systems is presented, implementation examples are given for each type, and the scope of applicability is described. Based on the results of the analysis and comparison, it seems possible to choose the most appropriate method for implementing a hybrid recommender system under different conditions and requirements.

Keywords: recommender system, collaborative filtering, similarity, content-based recommendation, hybrid recommenders.

Введение

Взрывной рост объема доступной цифровой информации и числа посетителей Интернета создал потенциальную проблему информационной перегрузки, которая препятствует своевременному доступу к интересующим элементам в Интернете. Информационно-поисковые системы, такие как Google, DevilFinder и Altavista, частично решили эту проблему, но приоритизация и персонализация (когда система сопоставляет доступный контент с интересами и предпочтениями пользователя) информации отсутствовали. Это увеличило спрос на рекомендательные системы больше, чем когда-либо ранее. Рекомендательные системы — это системы фильтрации информации, которые решают проблему информационной перегрузки [1] путем фильтрации фрагментов информации из большого объема динамически генерируемой информации в соответствии с предпочтениями, интересами или наблюдаемым поведением пользователя в отношении объекта [2]. Рекомендательная система имеет возможность предсказать, предпочтет ли конкретный пользователь какой-либо товар или нет, на основе профиля пользователя. Гибридные рекомендательные системы сочетают в себе различные методы рекомендаций, чтобы добиться лучшей оптимизации системы, позволяя избежать некоторых ограничений и проблем отдельных рекомендательных моделей [3]. Идея гибридных методов заключается в том, что комбинация алгоритмов обеспечивает более точные и эффективные рекомендации, чем один алгоритм, поскольку недостатки одного алгоритма могут быть преодолены другим

алгоритмом [4].

Рекомендательные системы выгодны как поставщикам услуг, так и пользователям [5]. Они снижают транзакционные издержки на поиск и выбор товаров в среде онлайн-покупок [6]. Системы рекомендаций также доказали, что они улучшают процесс и качество принятия решений [7]. В условиях электронной коммерции рекомендательные системы увеличивают доходы, поскольку они являются эффективным средством продажи большего количества продуктов [5]. В научных библиотеках рекомендательные системы поддерживают пользователей, позволяя им выйти за рамки поиска по каталогу. Таким образом, невозможно переоценить необходимость использования эффективных и точных методов рекомендаций в рамках системы, которая будет предоставлять релевантные и надежные рекомендации для пользователей.

1 Классификация рекомендательных систем

Чтобы реализовать свою основную функцию, идентифицируя полезные для пользователя элементы, рекомендательная система (РС) должна предсказать, стоит ли рекомендовать элемент. Для достижения этой цели система должна иметь возможность прогнозировать полезность некоторых элементов или, по крайней мере, сравнивать полезность некоторых элементов, а затем решать, какие элементы рекомендовать на основе этого сравнения. Существует несколько различных типов рекомендательных систем, и таксономия, представленная в [8], которая стала классическим способом различения рекомендательных систем, позволяет выделить 3 основных класса рекомендательных систем.

1.1 Совместная фильтрация

Совместная фильтрация позволяет давать рекомендации активному пользователю на основе элементов, которые нравились другим пользователям со схожими вкусами ранее. Сходство во вкусах двух пользователей рассчитывается на основе сходства истории оценок пользователей. Совместная фильтрация считается наиболее популярной и широко применяемой техникой в РС.

1.2 Рекомендательные системы на основе контента

Рекомендательные системы на основе контента направлены на то, чтобы рекомендовать элементы, похожие на те, что нравились пользователю в прошлом. Сходство элементов вычисляется на основе признаков, связанных со сравниваемыми элементами. Например, если пользователь положительно оценил фильм, относящийся к жанру комедии, то система может научиться рекомендовать другие фильмы этого жанра [9].

Классические методы рекомендации на основе контента направлены на сопоставление атрибутов профиля пользователя с атрибутами элемента. В большинстве случаев атрибуты элементов представляют собой просто ключевые слова, извлеченные из описаний элементов.

1.3 Гибридные рекомендательные системы

Гибридные РС основаны на сочетании вышеупомянутых методов. Гибридная система, объединяющая методы А и В, пытается использовать преимущества А, чтобы исправить недостатки В [9]. Например, методы совместной фильтрации страдают от проблем холодного старта, то есть они не могут рекомендовать элементы, которые не имеют оценки, или новым пользователям. Это не ограничивает подходы на основе контента, поскольку предсказание новых элементов основано на их описании (функциях), которые обычно легко доступны. При наличии двух (или более) основных методов РС было предложено несколько способов их объединения для создания новой гибридной системы.

2 Классификация гибридных РС

Термин «гибридная рекомендательная система» используется для описания любой рекомендательной системы, которая объединяет несколько методов рекомендаций для получения результата [8]. Им следует уделять особое внимание, поскольку они наиболее часто реализуются и наиболее перспективны для решения проблемы холодного старта, позволяют агрегировать информацию из разных источников. Гибридными рекомендателями обобщенно называются типы рекомендаторов, показанные на рисунке 1.

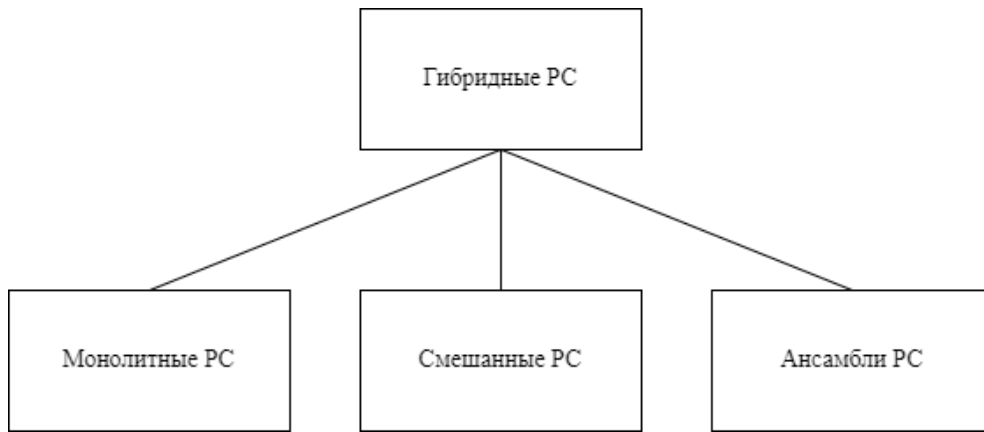


Рисунок 1 – классификация гибридных РС

Монолитные рекомендаторы берут компоненты различных рекомендаторов и по-новому склеивают их вместе. Ансамбль – это несколько работающих рекомендаторов, результаты работы которых комбинируются в одну рекомендацию. Смешанный рекомендатор возвращает результат работы сразу нескольких рекомендаторов [10].

2.1 Монолитные РС

Монолитный рекомендатор состоит из частей различных типов рекомендательных алгоритмов. Обычно и сам рекомендатор состоит из множества различных компонентов, например, алгоритма расчета сходства, отбора кандидатов и так далее. Монолитный рекомендатор берет компоненты из разных рекомендаторов, вероятно добавляет что-то новое для повышения общей производительности. На рисунке 2 показан монолитный рекомендатор, в котором взят алгоритм сходства элементов из рекомендатора 1 и отбор кандидатов с прогнозированием из рекомендатора 2.

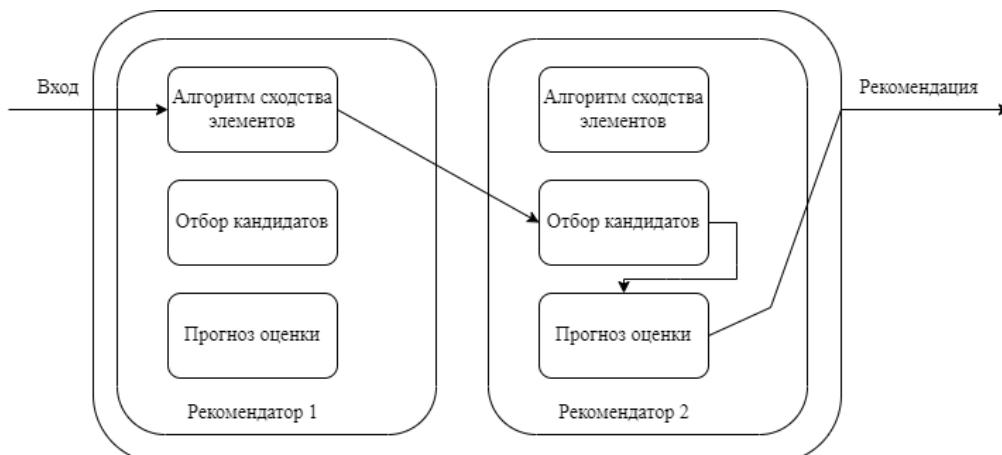


Рисунок 2 – монолитный гибридный рекомендатор, состоящий из частей разных РС

Так, в монолитном рекомендаторе может использоваться подход на основе контента, в котором находятся все подобные элементы, комбинированный с подходом совместной фильтрации для расчета прогноза оценок.

Примером монолитного гибридной РС может быть рекомендатор, основанный на совместной фильтрации, с одной дополнительной стадией предварительной обработки, где в матрицу оценок будут добавлены оценки для дополнительных связей [10]. Это особенно применимо в случаях, где невозможно найти сходство между элементами, что не позволяет генерировать рекомендацию.

Однако, реализация монолита требует немалой работы, и для превращения обычного рекомендатора в монолитный в целом потребуется много усилий. Если уже есть готовые рекомендаторы, рациональнее попробовать смешанные гибриды или ансамбли.

2.2 Смешанные гибридные РС

Смешанный рекомендатор возвращает комбинацию результатов всех входящих в него рекомендаторов. Если рассматривать рекомендаторы с точки зрения персонализации, то можно сделать первый рекомендатор максимально персонализированным, а затем продолжать в сторону неперсонализированных рекомендации, пока не будет получен рекомендатор по популярности элементов. Часто наиболее персонализированный рекомендатор дает только одну или две рекомендации, следующий рекомендатор дает несколько больше и таким образом система всегда будет давать хорошее количество рекомендаций, при этом не в ущерб качеству. На рисунке 3 показан смешанный гибридный рекомендатор.

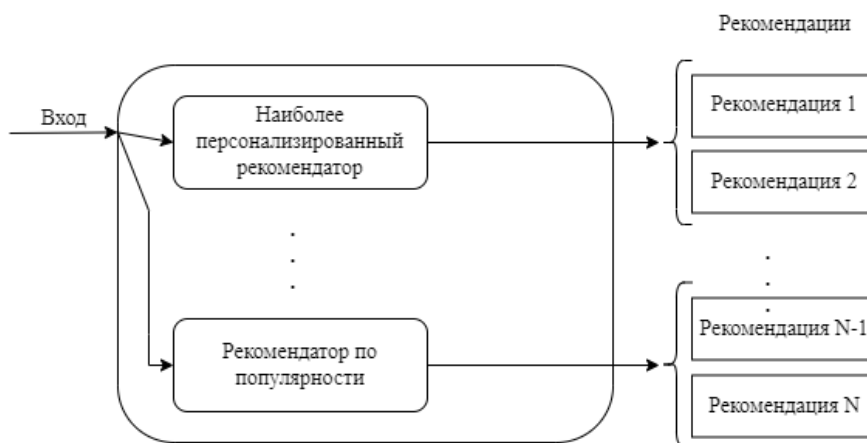


Рисунок 3 – смешанный гибридный рекомендатор, складывающий выходы нескольких рекомендаторов, начиная с самого персонализированного, заканчивая менее персонализированным

Если имеется несколько рекомендаторов, каждый из которых возвращает рейтинг рекомендации в результате, то можно вернуть соответствующим образом упорядоченный список. Однако, важно помнить, что рейтинги необходимо нормировать, чтобы все результаты имели один масштаб.

2.3 Ансамбли РС

Ансамбль – это группа вещей, рассматриваемых как единое целое, а не по отдельности, что аналогично и для ансамбля рекомендаторов: берутся результаты работы нескольких рекомендаторов и объединяются. Разница между ансамблем и смешанным рекомендатором заключается в том, что ансамбль может и не показать результат одного из рекомендаторов, в то время как смешанный гибрид всегда показывает все результаты. Наиболее популярным подходом в ансамбле является использование рекомендаторов на основе контента и с совместной фильтрацией, запущенных одновременно, которые путем сочетания дают возможность получить лучший результат, что показано на рисунке 4.

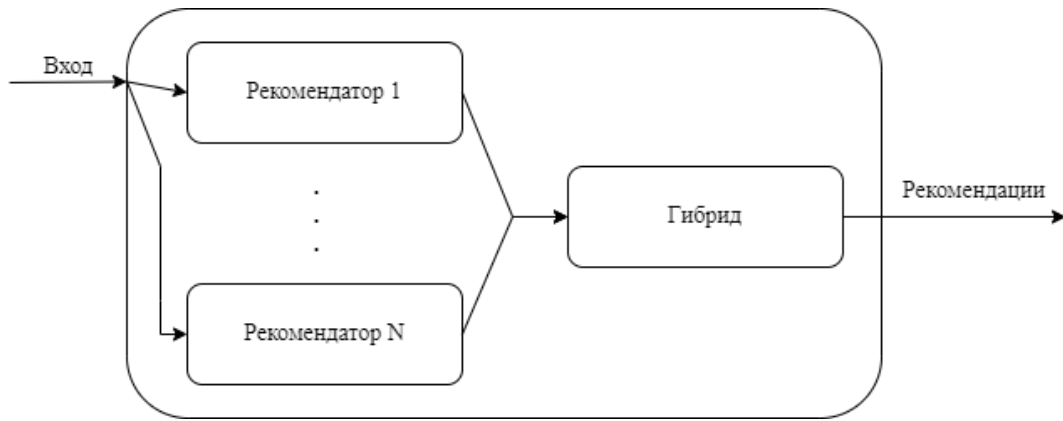


Рисунок 4 – ансамбль – это ряд рекомендаторов, чьи результаты объединяются в одну рекомендацию

Идея ансамблей заключается в использовании нескольких полноценных рекомендаторов, чьи результаты каким-то образом объединяются. Можно взять результат нескольких рекомендаторов и превратить их в один многими способами. Например, прибегнуть к голосованию и отсортировать элементы по его результатам [10], что показано на рисунке 5.

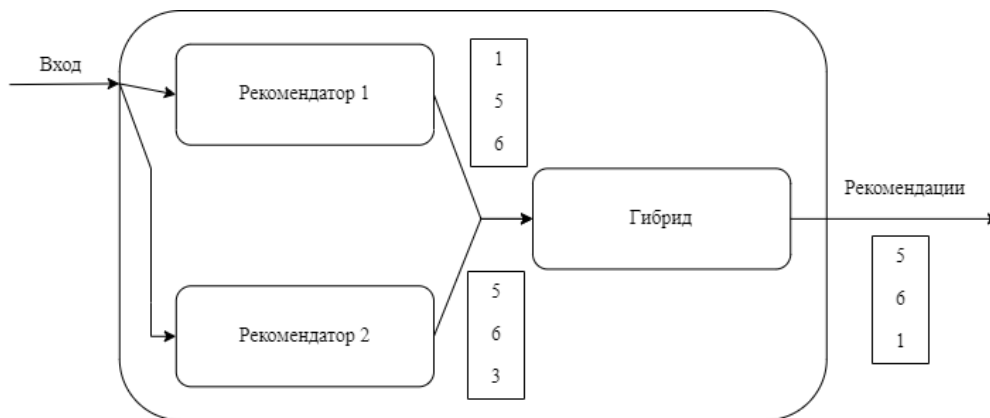


Рисунок 5 – пример работы ансамбля рекомендаторов с голосованием

Рекомендатор 1 возвращает рекомендацию для топ-3 элементов 1, 5 и 6. Рекомендатор 2 возвращает 5, 6 и 3. Затем гибрид возвращает 5, 6 и 1 в зависимости от того, как будет рассчитан приоритет. Элементы 5 и 6 точно попадут в выдачу, так как оба рекомендатора отметили их, но 6 будет чуть ниже. Затем будет рекомендован элемент 1, потому что у рекомендатора 1 он на первом месте.

Переключаемый ансамбль рекомендаторов подразумевает под собой выбор лучшего инструмента для работы. Имея два или более рекомендаторов, переключающийся ансамбль будет выбирать, какой из них использовать, с учетом контекста запроса.

Например, в системе может быть два разных рекомендатора для двух разных стран [10]. Когда заходит пользователь из одной страны, выводятся результаты одного рекомендатора, и, если заходит кто-то из другой страны, используется второй рекомендатор. Помимо этого, система может разделять их и по времени суток: один работает по утрам, а второй – вечером. Или, например, страница национальных новостей в газете заполняется последними новостями, а на культурную страницу выводятся контентные рекомендации для конкретных книг.

Система может переключаться между рекомендаторами в зависимости от раздела сайта, где находится пользователь. Или в самом простом виде в системе может быть два рекомендатора для пользователей, поставивших более 20 или менее 20 оценок, как показано на рисунке 6, где пользователи с более чем 20 оценками получают рекомендации от рекомендатора совместной фильтрации, а пользователи с меньшим количеством оценок получают рекомендации от рекомендатора на основе контента.

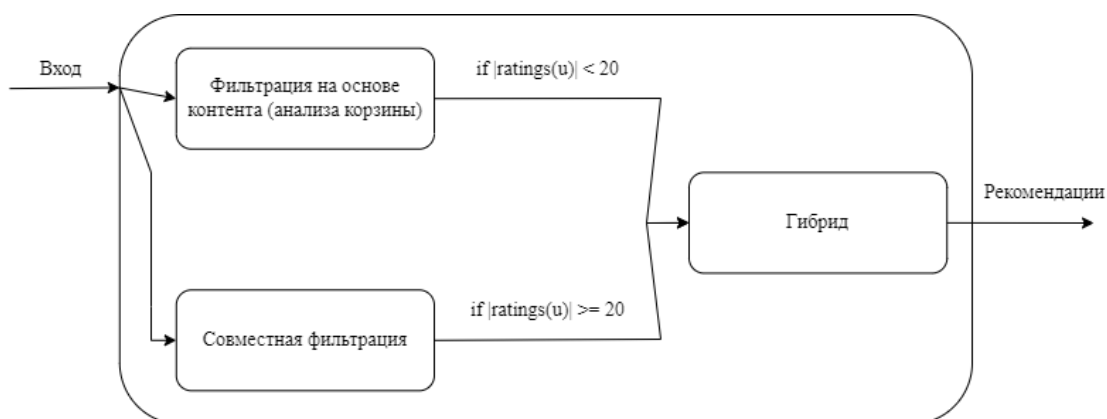


Рисунок 6 – переключаемый ансамбль, где пользователь с более чем 20 оценками получает результаты от одного рекомендатора, а с менее чем 20 – от другого

Заключение

Охвачено пространство возможных гибридных рекомендательных систем, доступных с двумя базовыми алгоритмами рекомендаций: совместная фильтрация и рекомендация, основанная на контенте. Было исследовано три типа гибридных рекомендательных систем: монолитная РС, смешанная РС, ансамбль РС. Рассмотренные типы гибридных рекомендательных систем позволяют объединить усилия различных рекомендаторов, чтобы получить лучшие результаты. Преимущества использования гибридных рекомендательных систем состоят в том, что можно решить такие проблемы, как проблема холодного старта и агрегации данных из разных источников с использованием моделей, подходящих для извлечения данных из определенного источника. Комбинация методов приводит к повышению точности предсказаний. Улучшение результата достигается за счет использования преимуществ одного из рекомендаторов для исправления недостатков другого.

Библиографический список:

1. Констан Дж.А., Ридл Дж. Рекомендательные системы: от алгоритмов к пользовательскому опыту. Пользовательская модель взаимодействия с адаптацией к пользователю // Springer. 2012. Том. 22, № 101. С. 23-27.
2. Пан С., Ли В. Рекомендация по исследовательской работе с тематическим анализом // Computer Design and Applications IEEE. 2010. Том. 4. С. 254-264.
3. Штерн Д.Х., Хербрих Р., Грпель Т. Крупномасштабные онлайн-байесовские рекомендации: материалы 18-й международной конференции по всемирной паутине. Нью-Йорк, США. 2009. С. 111–120.
4. Шафер Дж. Б., Франковски Д., Херлокер Дж. Рекомендательные системы совместной фильтрации. Берлин: Springer, 2007. 324 с.
5. Пу П., Чен Л., Ху Р. Ориентированная на пользователя структура оценки для рекомендательных систем: материалы пятой конференции ACM по рекомендательным системам (RecSys'11). Нью-Йорк, США. 2011. С. 57–164.
6. Ху Р., Пу П. Потенциальные проблемы принятия ASED-

персонализированных рекомендательных систем: материалы конференции АСМ по рекомендательным системам (RecSys'09). Нью-Йорк, США. 2009. С. 22–25.

7. Патхак Б., Гарфинкель Р., Гопал Р. Эмпирический анализ влияния рекомендательных систем на продажи // J Manage Inform Syst. 2010. Том. 27, № 2. С. 159–188.

8. Берк Р. Гибридные рекомендательные веб-системы. Берлин: Springer, 2007. 408 с.

9. Справочник по рекомендательным системам. Риччи Ф., Рокач Л., Шапира Б., Кантор П.Б. Нью-Йорк: Springer, 2011. 1053 с.

10. Фальк К. Рекомендательные системы на практике. Москва: ДМК-Пресс, 2020. 448 с.