

К. Е. Кудак — студент кафедры вычислительной техники и программирования

М. Д. Поляк — научный руководитель

МЕТОДЫ ГИБРИДИЗАЦИИ РЕКОМЕНДАТЕЛЬНЫХ АЛГОРИТМОВ

Современный мир перенасыщен информацией. Для того, чтобы найти действительно что-то полезное для себя в море источников информации, человек зачастую вынужден проделывать огромную работу по поиску, анализу и выборке самых разнообразных данных и сведений. Рекомендательные системы были созданы для решения этих задач. Анализируя данные о предпочтениях пользователей, алгоритмы рекомендательных систем находят наиболее подходящие конкретному пользователю объекты (книги, музыка, фильмы и т. д.). Основой любой рекомендательной системы является, в первую очередь, алгоритм фильтрации (предсказания, рекомендации). Именно он определяет эффективность работы рекомендательной системы. В рекомендательных системах, как в подклассе систем фильтрации информации, под процессом фильтрации понимается удаление избыточной или нежелательной информации из информационного потока, используя (полу)автоматизированные и компьютеризированные методы, до представления её человеку.

Гибридный алгоритм фильтрации в рекомендательной системе комбинирует в себе две или более рекомендательные техники для получения более высокого результата, чем даёт каждая в отдельности. Чаще всего коллективная фильтрация комбинируется с какой-либо другой техникой для решения проблем нового пользователя или объекта, имеющих пока что малое количество оценок. В табл. 1 приведены методы комбинирования рекомендательных техник [1].

Таблица 1

Метод гибридизации	Краткое описание
Взвешенный	Оценки (голоса) нескольких рекомендательных техник комбинируются вместе для получения единой рекомендации.
Переключение	Система переключается между рекомендательными техниками в зависимости от текущей ситуации.
Смешанный	Рекомендации нескольких техник выводятся одновременно.
Комбинация признаков	Признаки из различных рекомендательных источников данных объединяются вместе и подаются на вход единого рекомендательного алгоритма.
Каскадный	Одна рекомендательная техника уточняет результат работы другой.
Расширение множества признаков	Результат одной рекомендательной техники используется как входные данные для другой.
Мета-уровень	Модель, обученная одной рекомендательной техникой, используется как входные данные для другой.

Рассмотрим методы гибридизации более подробно.

1. Взвешенный метод

Взвешенный метод гибридизации предполагает, что оценка рекомендуемого объекта вычисляется как результат всех рекомендательных техник, используемых в системе. Например, простейший вариант взвешенной гибридизации представляет собой линейную комбинацию предсказанных оценок [2].

Преимуществами такого метода можно назвать использование всех возможностей рекомендательных техник в полной мере, простоту реализации, легкость дальнейшего анализа работы алгоритма и его корректирования. Метод предполагает, что относительные результаты рекомендательных техник

более или менее универсальны на всём пространстве возможных объектов. Однако, известно, что в большинстве случаев это не так. Например, коллективная фильтрация плохо работает с объектами, имеющими малое количество оценок.

2. Метод переключения

Метод переключения предполагает внесение некоторой чувствительности гибридного алгоритма к разного рода объектам, на основе которой будет производиться выбор между использованием той или иной рекомендательной техники. Например, если фильтрация по содержимому не может предсказывать оценку объекта с определённой точностью, то используется коллективная фильтрация [3].

Серьёзным недостатком метода переключения является дополнительная сложность, возникающая от необходимости определить критерий переключения между различными рекомендательными системами, что требует ещё одного уровня параметризации. Если же такая задача будет решена, то гибридный алгоритм будет успешно использовать сильные и слабые стороны различных рекомендательных техник.

3. Смешанный метод

В случаях, когда производится большое количество рекомендаций одновременно, имеет смысл использовать смешанный гибридный алгоритм [4]. В результате конечному пользователю предлагаются рекомендации независимых рекомендательных техник. Смешанный метод не позволяет компенсировать недостатки одной из рекомендательных техник достоинствами другой, но чрезвычайно прост в реализации.

4. Комбинация признаков

Одним из способов добиться слияния фильтрации по содержимому и коллективной фильтрации является рассмотрение коллективной информации, как простого дополнительного признака рекомендуемого объекта, и дальнейшее применение фильтрации по содержимому к этому объекту [5].

Комбинация признаков позволяет гибриднему алгоритму использовать коллективную информацию и в тоже время набор признаков объектов, что существенно расширяет возможности фильтрации по содержимому.

5. Каскадный метод.

В отличие от описанных выше методов гибридизации каскадный метод предполагает поэтапный процесс рекомендации. На первом этапе одна из рекомендательных техник применяется для более грубого определения рекомендаций, затем на остальных этапах каждая следующая техника используется для уточнения и выборки результатов предыдущей [1].

Каскадный метод позволяет системе избегать применения второй низко-приоритетной техники для уже хорошо классифицированных объектов или же плохо оценённых объектов, которые никогда не будут предложены в качестве рекомендаций. Поскольку каждый следующий этап каскадного гибридного алгоритма нацелен только на такие объекты, для которых необходимо дополнительное различение, этот алгоритм более эффективен, чем, например, взвешенный гибрид, применяющий все свои техники ко всему множеству объектов. К тому же каскадный метод мало чувствителен к «шуму», производимому низко-приоритетными техниками, так как последние могут только корректировать результаты высоко-приоритетных техник, но не могут менять их коренным образом.

6. Расширение множества признаков

Расширение множества признаков подразумевает использование оценки или классификации объекта и информации, включённой в него, как входные данные для следующей рекомендательной техники. Например, существует рекомендательная система, использующая фильтрацию по содержимому для книг, размещённых в интернет-магазине Amazon.com [6]. В информации о книге также присутствуют поля «Связанные работы» и «Связанные авторы», которые были сгенерированы уже системой коллективной фильтрации интернет-магазина. Эти поля-признаки вносят существенный вклад в качество рекомендаций.

Расширение множества признаков очень привлекательно, потому что позволяет увеличить производительность рекомендательной системы без её фактической модификации. Дополнительная функциональность добавляется с помощью рекомендательных систем-посредников, которые могут использовать совершенно другие техники и подходы. Необходимо отметить, что такой метод гибридиза-

ции отличается от комбинирования признаков, где совмещаются входные, необработанные данные различных техник.

7. Мета-уровень

Ещё один способ комбинировать несколько рекомендательных техник – это использовать модель, созданную с помощью одной из техник, как входные данные для другой техники [7]. Этот метод отличается от расширения множества признаков тем, что в последнем используется обученная модель для генерации признаков на вход следующего алгоритма, а в гибриде мета-уровня на вход подаётся вся модель целиком. Обученная модель является сжатым представлением пользовательских интересов. Коллективной фильтрации, например, проще работать с таким информационно-плотным представлением, чем с сырыми данными об оценках пользователей.

Гибридизация позволяет решить некоторые проблемы, ассоциированные с той или иной техникой фильтрации. Коллективная фильтрация и фильтрация по содержанию, независимо от типа, всегда страдают от проблемы нового пользователя и нового объекта и всегда требуют набор оценок для своей работы. Тем не менее, гибридные алгоритмы, использующие эти рекомендательные техники, широко распространены, потому что оценки объектов или уже есть, или их можно получить из данных об объекте. Метод мета-уровня обходит проблему разреженности сжатием оценок множества объектов в единую модель. Алгоритмы фильтрации, основанные на знаниях, также являются хорошими кандидатами на гибридизацию, так как они не подвержены проблемам нового пользователя и нового объекта.

Основываясь на описании методов гибридизации, можно разделить их на две группы: чувствительные к порядку применения рекомендательных техник и не чувствительные (табл. 2). Рекомендации, предлагаемые пользователю каскадным гибридным алгоритмом фильтрации, будут существенно отличаться, если поменять порядок применения той или иной техники. В тоже время смешанный гибридный алгоритм будет формировать одни и те же рекомендации независимо от порядка применения рекомендательных техник.

Таблица 2

Чувствительные методы	Каскадный, расширение множества признаков, мета-уровень.
Нечувствительные методы	Взвешенный, смешанный, переключение, комбинация признаков.

Необходимо заметить, что не все методы гибридизации могут быть применимы к некоторым комбинациям рекомендательных техник и к некоторым техникам вообще. Так, например, фильтрация, основанная на знаниях, может использовать любую информацию об объекте, поэтому применение метода комбинации признаков для гибридизации с другой рекомендательной техникой не имеет никакого смысла.

Выбор того или иного метода гибридизации сильно зависит от условий использования рекомендательной системы. Необходимо правильно оценить поведение совмещаемых алгоритмов. Если один из алгоритмов обладает преимуществом по сравнению с другим, то следует обратить внимание на чувствительные к порядку методы гибридизации, в которых этот алгоритм займёт лидирующую роль, иначе – на нечувствительные, в которых рекомендация будет получена в результате компромисса.

Библиографический список

1. Burke R. (2002) Hybrid Recommender Systems: Survey and Experiments Journal User Modeling and User-Adapted Interaction, Volume 12 Issue 4, Ноябрь 2002. Pp. 331–370
2. Claypool, M., Gokhale, A., Miranda, T., Murnikov, P., Netes, D. and Sartin, M.: 1999, 'Combining Content-Based and Collaborative Filters in an Online Newspaper'. SIGIR '99 Workshop on Recommender Systems: Algorithms and Evaluation. Berkeley, CA.
3. Tran, T. and Cohen, R.: 2000, «Hybrid Recommender Systems for Electronic Commerce». In Knowledge-Based, Electronic Markets, Papers from the AAAI Workshop, AAAI Technical Report WS-00-04. страницы 78-83. Menlo Park, CA: AAAI Press.
4. Smyth, B. and Cotter, P.: 2000, «A Personalized TV Listings Service for the Digital TV Age». Knowledge-Based Systems 13: 53-59. California State University, Fullerton Department of Information Systems and Decision Sciences
5. Basu, C., Hirsh, H. and Cohen W.: 1998, «Recommendation as Classification: Using Social and Content-Based Information in Recommendation». Proceedings of the 15 National Conference on Artificial Intelligence, Madison, WI. Pp. 714–720.

6. Mooney, R. J. and Roy, L.: 1999, «Content-Based Book Recommending Using Learning for Text Categorization». SIGIR '99 Workshop on Recommender Systems: Algorithms and Evaluation. Berkeley, CA.
7. Rocchio, Jr., J.: 1971, «Relevance feedback in information retrieval». The SMART System – Experiments in Automatic Document Processing, New York: Prentice Hall. Pp. 313–323.