

МЕТОДЫ ОЦЕНКИ КАЧЕСТВА РЕКОМЕНДАТЕЛЬНЫХ СИСТЕМ

П.С. Павлов

Институт космических и информационных технологий
(Россия, г. Красноярск)

Аннотация. В данной статье описываются некоторые типы рекомендательных систем и методы их оценки. Рекомендательные системы являются неотъемлемой частью сервисов, которые предоставляют пользователю какой-либо выбор среди множества альтернатив. Пользуясь каким-либо сервисом, большинство людей даже и не задумываются о том, что же им предоставляет всё новые и новые рекомендации. Данные системы существенно облегчают использование таких сервисов, так как они предоставляют рекомендации на основе информации о пользователе. Особое внимание в статье уделено методам оценки качества рекомендательных систем. Оценка рекомендательных систем необходимо для того, чтобы понять, насколько эффективна та или иная система. Эффективность системы очень важна, так как, чем эффективнее она, тем лучше и качественней рекомендации она предоставляет. Статья разделена на две смысловые части: в первой говорится о том, что такое рекомендательная система, какие виды рекомендательных систем существуют, также описаны преимущества и недостатки каждого из видов, во второй описываются методы оценки рекомендательных систем. Во второй смысловой части говорится о двух методах оценки рекомендательных систем: метод оценки качества рекомендаций и метод оценки качества работы алгоритмов.

Ключевые слова: рекомендательная система, методы оценки, фильтрация, контентная, коллаборативная, Netflix.

Что такое рекомендательные системы? Каждая рекомендательная система направлена на взаимодействие с пользователем, которому предоставляется огромное количество объектов, среди которых он должен сделать свой выбор. Пользователю может не хватать знаний или опыта, или того и другого, чтобы выбрать то, что соответствует его потребностям. Пользователь явно или косвенно предоставляет системе информацию о своих предпочтениях. Таким образом, рекомендательная система представляется в виде системы (программы), использующей определенный алгоритм фильтрации и имеющуюся информацию о потребностях пользователя, чтобы рекомендовать ему набор объектов, которые он считает наиболее полезными для себя [1]. Рекомендательные системы используются в среде электронной коммерции [2], поисковых системах, системах электронного образования [3].

Рекомендательные системы классифицируются по способу отбора необходимого пользователю материала [4]. В

основном применяется два базовых подхода: коллаборативная фильтрация и контентная фильтрация. Также существует гибридная фильтрация, которая совмещает в себе как коллаборативную, так и контентную фильтрацию.

В рекомендательных системах, использующих контентную фильтрацию (фильтрация по содержанию), пользователи не зависят от других пользователей системы [5]. Для формирования рекомендаций системе необходим профиль пользователя с информацией о его интересах. В профиле в определенной форме хранится информация об интересующих пользователя объектах. Система также содержит информацию обо всех предметах, которые она может рекомендовать. Такая система использует описание объектов в профиле пользователя, находит похожие объекты в своей базе данных, а затем рекомендует их ему. Применение фильтрации такого рода очень уместно, когда пользователь имеет четко определенные, конкретные интересы и ищет схожие рекомендации. Преимущество контентной фильтрации

заключается в том, что для начала предоставления рекомендаций не требуется большое количество зарегистрированных пользователей, то есть рекомендации не зависят от других пользователей системы. Основным ограничением данного метода является невозможность системы с таким видом фильтрации рекомендовать новые объекты, не соответствующие интересам пользователя.

Для того, чтобы система имела высококачественные рекомендации, необходимо постоянно обновлять информацию об интересах пользователя. Но существует и другая проблема - пользователь всегда хочет минимизировать время, затрачиваемое на взаимодействие с системой, и неохотно делится информацией для обратной связи с системой. Это типичная проблема для всех рекомендательных систем, и поскольку рекомендации основаны исключительно на информации профиля пользователя, то есть чем меньше информации пользователь предоставляет, тем менее соответствующий набор рекомендаций пользователь получит. В таком случае для решения этой задачи используются автоматические системы сбора информации для поддержки профиля, но это затрудняет реализацию системы. Этот метод фильтрации используется в электронной коммерции. В сочетании с коллаборативной фильтрацией, используется в таких областях, как электронное образование, программирование и т. д.

Рекомендательные системы, использующие коллаборативную фильтрацию, ищут пользователей с общими интересами и рекомендуют объекты, которые были высоко оценены другими пользователями [5, 6]. Интересы пользователя сохраняются в виде графика предпочтений, который показывает, насколько пользователю нравится тот или иной предмет. Такой системы достаточно для оценки пользователей по выбранной тематике. Преимущество этой системы заключается в том, что система должна знать интересы многих пользователей. Эта система не может создавать рекомендации, основанные только на интересах одного человека.

Главным недостатком коллаборативной фильтрации является «холодное начало». Система не способна эффективно генерировать рекомендации, если большая часть пользователей не укажут свои интересы в профиле.

Гибридные системы обычно сочетают коллаборативную фильтрацию и контентную фильтрацию. Это позволяет решить ряд проблем, которые возникают при использовании этих методов по отдельности. В гибридной системе информация об интересах пользователей представлена в профиле в двух видах - как перечень свойств определенного объекта и как его оценка пользователем. Эта характеристика является как преимуществом, так и недостатком системы. Преимуществом является полнота информации, что позволяет использовать эффективные алгоритмы фильтрации и формировать необходимые рекомендации. Недостатком является то, что пользователь должен ввести больше информации в свой профиль, которая всегда очень неохотно, или не делается вообще.

Качество рекомендательных систем можно определить, оценив качество работы алгоритмов, качество рекомендаций.

Методика оценки качества рекомендаций рекомендательных систем.

Пусть исходные данные: список пользователей $U = (u_1, \dots, u_m)$ и объекты $I = (i_1, \dots, i_n)$ представлены объектно-признаковой таблицей. Каждому пользователю в таблице соответствует одна строка. В ходе взаимодействия пользователей с системой осуществляется заполнение (формирование) данной таблицы. Бинарное отношение $T \subseteq U \times I$ показывает, оценил ли пользователь $u \in U$ объект $i \in I$, которое можно рассматривать как рейтинг $r_{u,i}$ объекта i у пользователя u .

Для оценки качества рекомендаций с точки зрения точности и полноты существует несколько стандартных подходов [5]. Одним из таких подходов является метод скользящего контроля [7]. Использование данного метода заключается

ся в разбиении исходного набора множества U на два подмножества: тестовое U_{test} и обучающее $U_{training}$. Полнота и точность поиска оценивается на тестовой выборке. Каждый элемент u тестовой выборки U_{test} разбивается на две части: на оцененные признаки $I_{visible}$ и не оцененные I_{hidden} .

Алгоритм рекомендаций по оценкам пользователей использует сходство объектов из тестовой и обучающей выборки для формирования рекомендаций. Каждый пользователь из U_{test} получает рекомендации, как множество фиксированного размера $r_n(u) = \{i_1, i_2, \dots, i_n\}$. Далее точность и полнота рассчитываются следующим образом:

$$recall = \frac{|r_n(u) \cap u^I \cap I_{hidden}|}{|u^I \cap I_{hidden}|}, \quad (1)$$

$$precision = \frac{|r_n(u) \cap u^I \cap I_{hidden}|}{|r_n(u) \cap I_{hidden}|}, \quad (2)$$

где u^I - совокупность всех оцененных пользователем u объектов из I . Значения этих мер вычисляются для каждого пользователя и усредняются. Эксперимент проводится для разбиения исходного множества, из которого производится выбор, на тестовое и обучающее множество несколько раз. Значения мер также усредняются. В дополнение к стандартным способам вычисления точности и полноты в статье [8] предлагаются нетрадиционные варианты раскрытия неопределенностей, возникающих при их оценке: $|u^I \cap I_{hidden}| = 0$, то $recall = 1$. Если $|r_n(u) \cap I_{hidden}| = 0$ и $u^I = 0$, то $precision = 1$, иначе $precision = 0$.

Методика оценки качества работы алгоритмов.

Существует множество метрик для оценки качества алгоритмов. В основном это метрики для оценки точности ожидаемого и реального значения. Рассмотрим некоторые из них: MAE, RMSE.

MAE (Mean Absolute Error, пер. средняя абсолютная ошибка) – ошибка оценивается как разница между расчетной и фактической оценкой по модулю:

$$MAE = \frac{\sum_{i \in n} |P_i - R_i|}{n}. \quad (3)$$

RMSE (Root Mean Squared Error, пер. средняя квадратичная ошибка) – ошибка вычисляется как корень из суммы квадратов разниц между предполагаемым и действительным значением:

$$RMSE = \sqrt{\frac{\sum_{i \in n} (P_i - R_i)^2}{n}} \quad (4)$$

Другие характеристики рекомендательных систем также могут быть оценены, но на основе всего перечня рекомендаций. Для этого необходимо иметь информацию о том, какие элементы в списке рекомендаций соответствуют запросу, а какие нет. Исходя из этого, характеристику точности можно представить в виде формулы:

$$Accuracy = \frac{|T|}{L}, \quad (5)$$

где, L - длина всего списка рекомендаций, T - множество соответствующих рекомендаций, F - множество несоответствующих элементов в списке - F .

Характеристику полноты, в свою очередь, можно представить в виде следующей формулы:

$$Recall = \frac{|T|}{|T| + |G|}, \quad (6)$$

где, G – множество элементов, которые должны были быть порекомендованы, но не оказались в списке.

Вычислительный эксперимент

В октябре 2006 года Netflix, крупная фирма по аренде DVD, проводила открытое соревнование на лучший алгоритм предсказания оценки, которую зритель поставит фильму, на основе предыдущих оценок этого и других зрителей. Главный приз составлял \$1000000. В Netflix используется алгоритм, основанный на линейной регрессии, который называется Cinematch.

Тривиальный алгоритм, который ставит фильму среднюю оценку, даёт среднеквадратичное отклонение (СКО) 1,0540. Алгоритм Cinematch, который

использует данные, предоставленные участникам соревнования, даёт СКО 0,9525 на тестовом множестве. Это значит, что он на 10% лучше тривиального алгоритма.

Конечно, были споры по поводу того, действительно ли уменьшение ошибки на 10% поможет пользователям? По мнению одного из участников соревнования, улучшение ошибки даже на 0,01 существенно изменит список 10 рекомендованных фильмов.

Всем участникам были предоставлены 4 файла:

- Тренировочные данные, в котором были оценки каждого клиента, о каждом фильме;

- Квалификационные данные, которые содержали информацию о том какой пользователь, какой фильм, когда оценил;

- Названия и годы выхода фильмов;

- Фрагмент тренировочной базы данных, распределённый так же, как квалификационные данные.

Никакой информации о пользователях, в базах данных, не было предоставлено участникам в целях конфиденциальности пользователей. Некоторые квалификационные и тренировочные данные были намеренно изменены в тех же целях.

Каждый год лучшей команде выдавался приз в \$50000, если не было достигнуто главной цели.

Спустя почти три года (26.07.2009) соревнование было прекращено, победа и приз были присвоены команде «Bellkor's Pragmatic Chaos», которая улучшила алгоритм на 10,09% и уменьшила СКО до 0,8554. Церемония награждения и объявление результатов состоялись 21 сентября 2009.

Из всего этого следует, качество рекомендательной системы зависит от качества алгоритмов. Если улучшить качество алгоритмов, то улучшится качество рекомендаций, а это значит, что

рекомендательная система станет более эффективной. В рекомендательной системе фирмы Netflix, используется коллаборативная фильтрация. Несмотря на все недостатки такой фильтрации, если улучшить её алгоритм, можно получать достаточно точные рекомендации.

Проведём эксперимент, оценим качество рекомендаций рекомендательной системы NetFlix, для этого зарегистрируемся на сайте «КиноПоиск» [www.kinopoisk.ru]. Для оценки качества рекомендаций по полноте используем формулу:

$$recall = \frac{|r_n(u) \cap u^I \cap I_{hidden}|}{|u^I \cap I_{hidden}|}$$

Так как мы только зарегистрировались и ничего не заказывали и не искали, то $u^I = 0$. Из этого следует, что $|u^I \cap I_{hidden}| = 0$, а это значит, что $recall = 1$.

Оценим качество рекомендательной системы по точности. При первом входе в систему количество рекомендаций и не оценённых объектов, то есть объекты, которые пользователь не просматривал, равно количеству объектов в этой системе, то есть $r_n(u) = 107$ и $I_{hidden} = 107$. Из этого следует, что $|r_n(u) \cap I_{hidden}| \neq 0$. Так как мы только зарегистрировались и ничего не заказывали и не искали, то $u^I = 0$. Значит, $precision = 0$.

Если посмотреть на результаты оценки качества рекомендательной системы, использующей коллаборативную фильтрацию, то можно понять, что в ней есть проблема "холодного начала". Система не способна эффективно генерировать рекомендации, если большая часть пользователей не укажут свои интересы в профиле. То есть, в данном случае пострадала точность, так как при создании аккаунта не были указаны интересы пользователя и схожих пользователей не обнаружилось.

Библиографический список

1. *Jannach D., Zanker M., Felfernig A. Friedrich G. Recommender Systems. An Introduction.* New York: Cambridge University Press 32 Avenue of the Americas, 2011. 352 P.
2. *Вахрушева М.Ю., Евдокимов И.В. Разработка программного обеспечения аналитических информационных систем // Труды Братского государственного университета. Серия: Экономика и управление. 2014. Т. 1. № 1. С. 196-199.*
3. *Кокташев В.В., Макеев В.В., Михалев А.С. Применение систем управления инцидентами для разработки программного обеспечения в образовании // Современные информационные технологии. 2017. №4 (30). С. 129-135.*
4. *Джонс Т.М. Рекомендательные системы. Часть 1. Введение в подходы и алгоритмы. [Электронный ресурс]: <https://www.ibm.com/developerworks/ru/library/os-recommender1/> (дата обращения: 11.05.2018).*
5. *Melville P., Sindhvani V. Recommender systems. Encyclopedia of Machine Learning. 2010.*
6. *Su X., Khoshgoftaar Survey of Collaborative Filtering Techniques. Advances in Artificial Intelligence. 2009.*
7. *Воронцов К.В. Комбинаторный подход к оценке качества обучаемых алгоритмов. Математические вопросы кибернетики / Под ред. О.Б. Лупанова. – М.: Физматлит, 2004. – Т. 13. – С.5-36.*
8. *Игнатов Д.И., Каминская С.Ю., Магизов Р.А. Метод скользящего контроля для оценки качества рекомендательных Интернет-сервисов // Труды 12-й национальной конференции по искусственному интеллекту, М., Физматлит, Т. 1., С.183-191, 2010.*
9. *Netflix. [Электронный ресурс]: <https://www.netflix.com/ru/> (дата обращения: 11.05.2018).*
10. *Евдокимов И.В. Процедура идентификации как этап создания систем управления и принятия решений // Проблемы социально-экономического развития Сибири. 2012. № 4. С. 14-18.*

METHODS OF ASSESSMENT OF THE QUALITY OF RECOMMENDATION SYSTEMS

P.S. Pavlov

**Institute of space and information technology
(Russia, Krasnoyarsk)**

***Abstract.** This article describes some types of recommendation systems and methods for their evaluation. Recommendation systems are an integral part of services that provide the user with a choice among a variety of alternatives. Using a service, most people do not even think about what gives them more and more new recommendations. These systems greatly facilitate the use of such services, since they provide recommendations based on user information. Particular attention is paid in the article to methods for assessing the quality of reference systems. Evaluation of advisory systems is necessary in order to understand how effective this or that system is. The effectiveness of the system is very important, since the more effective it is, the better and more qualitative recommendations it provides. The article is divided into two semantic parts: the first one deals with what is the recommendation system, which types of recommendation systems exist, the advantages and disadvantages of each species are also described, the second describes methods for evaluating the recommendation systems. In the second semantic part, two methods for evaluating recommendatory systems are described: the method for assessing the quality of recommendations and the method for evaluating the quality of work of algorithms.*

***Keywords:** advisory system, assessment methods, filtering, content, collaborative, Netflix.*