



ОТКРЫТАЯ НАУКА
издательство

Международный журнал информационных технологий и
энергоэффективности

Сайт журнала:

<http://www.openaccessscience.ru/index.php/ijcse/>



УДК 004.83

СРАВНИТЕЛЬНЫЙ АНАЛИЗ РЕКОМЕНДАТЕЛЬНЫХ СИСТЕМ И МЕТОДОВ ОЦЕНКИ ИХ КАЧЕСТВА

¹Андреева Я. А., ²Василевский К. А.

Московский технический университет связи и информатики, Москва, Россия (123423, г. Москва, ул. Народного Ополчения, 32), e-mail: ¹andreevaya.00@mail.ru, ²alaxtver@yandex.ru

В последние годы рекомендательные системы становятся все популярней и применяются в различных сферах: от новостной ленты до интернет-магазинов. Такое разнообразие рекомендательных систем порождает вопрос оценки их качества. В данной статье подробно описаны различные типы рекомендательных систем, их плюсы и минусы, а также принцип их работы. Особое внимание уделено математическим методам оценки качества рекомендательных систем.

Ключевые слова: рекомендательные системы, коллаборативная фильтрация, метрики, оценка качества, ранжирование.

COMPARATIVE ANALYSIS OF RECOMMENDATION SYSTEMS AND METHODS FOR EVALUATING THEIR QUALITY

¹Andreeva Ya. A., ² Vasilevskii K. A.

Moscow Technical University of Communications and Informatics, Moscow, Russia (123423, Moscow, Narodnogo Opolcheniya str., 32), e-mail: ¹andreevaya.00@mail.ru, ²alaxtver@yandex.ru

In recent years, recommendation systems have become increasingly popular and are used in various fields: from news feeds to online stores. Such a variety of recommendation systems raises the question of assessing their quality. This article describes in detail the various types of recommendation systems, their pros and cons, as well as the principle of their operation. Special attention is paid to mathematical methods of evaluating the quality of recommendation systems.

Keywords: recommendation systems, collaborative filtering, metrics, quality evaluation, ranking.

Введение

Стремительный рост и разнообразие информации, доступной в Интернете, а также быстрое внедрение новых услуг электронного бизнеса (покупка товаров, сравнение товаров, аукцион и т.д.) часто ошеломляли пользователей, заставляя их принимать неправильные решения. Доступность выбора, вместо того чтобы приносить пользу, начала снижать благосостояние пользователей.

В связи с этим последние годы широкую популярность приобрело повсеместное использование интеллектуальных рекомендательных систем: в интернет-магазинах, в новостной ленте СМИ, социальных сетях, поисковых системах, стриминговых музыкальных и видео сервисах и т.д.

Рекомендательной системой или системой рекомендаций называется комплекс программ, который на основе различных данных о пользователе определяет его интересы и, в соответствии с ними, формирует различные предложения контента. Основной целью применения рекомендательных систем является персонализация контента, а также его автоматическая подстройка под текущие нужды конкретного пользователя.

Рекомендательные системы в первую очередь ориентированы на людей, которым не хватает достаточного личного опыта или компетенции, чтобы оценить огромное количество схожих элементов, которые, например, может предложить какая-либо платформа.

Релевантные рекомендации значительно сокращают время, необходимое пользователю для поиска товаров, услуг или контента, а также увеличивают возможность того, что пользователю попадутся другие объекты, которые смогут привлечь его интерес. Тем самым повышается лояльность пользователей и их удовлетворенность работой веб-сервисов. Как правило, в сервисах, где применяются рекомендательные системы, пользователи взаимодействуют с большим количеством товаров, что приводит к увеличению потребления и как следствие, росту прибыли. Благодаря тому, что область применения рекомендательных систем достаточно широкая и постоянно растет, их изучение не теряет актуальности. С каждым годом увеличивается количество рекомендательных систем и их качество.

Классификация рекомендательных систем

За последние годы технической революции основные задачи рекомендательных систем в целом не изменились. Удержание внимания пользователей и побуждение их к целевому действию все также актуальны. Для этого применяются различные прогностические методы, в основе которых лежит машинное обучение.

Современные интеллектуальные рекомендательные системы должны обладать рядом свойств:

- системе необходимо приспосабливаться под конкретного пользователя, потому что предпочтения разных людей могут кардинально отличаться;
- системе необходимо учитывать текущие предпочтения конкретного пользователя, адаптируясь под него с течением времени;
- система должна все время искать новые области информации и предоставлять их пользователю.

Всего выделяют 4 типа рекомендательных систем по методу поиска необходимого пользователю материала (рисунок 1).

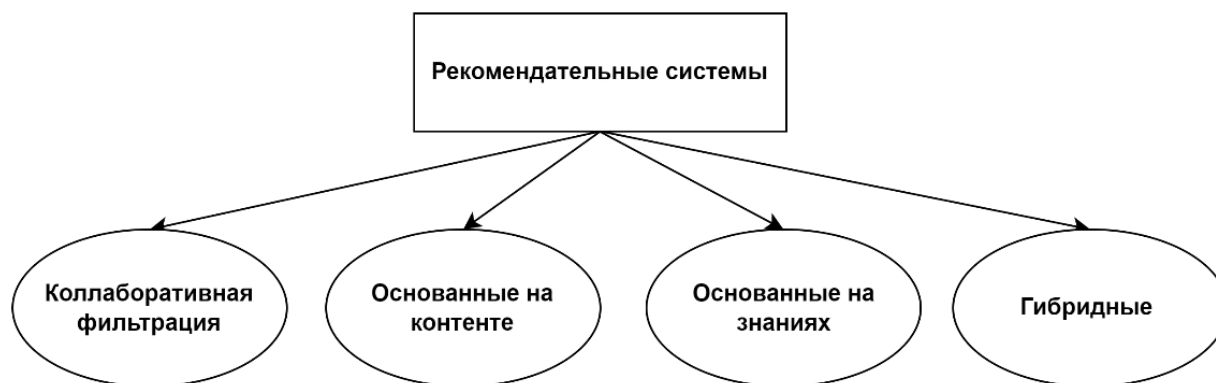


Рисунок 1 – Типы рекомендательных систем.

1. Метод коллаборативной фильтрации

Метод коллаборативной фильтрации (англ. collaborative filtering) построен на истории оценок как самого пользователя, так и других пользователей [1, 2].

В коллаборативной фильтрации используется два типа входных данных: множество объектов интереса и множество пользователей. Отношения между ними чаще всего выражаются при помощи оценок, выставленных пользователями.

К преимуществам систем, использующих коллаборативную фильтрацию относятся:

- простота алгоритмов;
- простота объяснения;
- высокая точность;
- стабильность.

Существенным минусом такого типа рекомендательных систем является проблема «холодного начала». Рекомендательная система, построенная на методе коллаборативной фильтрации не будет предоставлять пользователю подходящие рекомендации, если необходимое количество пользователей системы не сообщит о своих интересах.

2. Фильтрация, основанная на контенте

Рекомендательные системы, основанные на контенте (англ. contentbased), строятся на данных о каждом конкретном объекте [3]. Пользователю предлагаются объекты, схожие с теми, которыми он интересовался до этого. Схожесть этих объектов оценивается по их содержанию.

Среди преимуществ таких рекомендательных систем можно выделить тот факт, что для начала генерации системе не требуется большое число зарегистрированных пользователей, ведь рекомендации не зависят от других пользователей этой системы [4].

Основными проблемами данного метода являются:

- отсутствие возможности рекомендовать объекты, которые не соответствуют интересам конкретного пользователя (зависимость от выбранной области);
- снижение точности;
- ограниченность полезности таких рекомендаций.

3. Фильтрация, основанная на знаниях

Рекомендательные системы, в основе которых лежат знания о предметной области (англ. knowledge-based), а не об отдельных объектах имеют достаточно высокую точность, предоставляя пользователю именно то, что ему нужно. Помимо этого, такие системы изучают и производят анализ взаимосвязей между различными объектами, а также учитывают дополнительную информацию, относящуюся к индивидуальным данным конкретного пользователя [5].

Рекомендации основываются на множестве объектов и множестве правил, которые в зависимости от информации, заданной пользователем, описывают, какие объекты ему необходимо рекомендовать.

К плюсам такого подхода можно отнести отсутствие «холодного старта», а к минусам – более высокую сложность разработки системы и необходимость дополнительных данных.

4. Гибридные рекомендательные системы

Помимо базовых подходов, описанных выше, существуют гибридные рекомендательные системы (hybrid), которые объединяют в себе возможности этих методов [6]. Это позволяет нейтрализовать или минимизировать недостатки, свойственные предыдущим типам рекомендательных систем.

Методы оценки качества рекомендательных систем.

В последние годы с бурным развитием рекомендательных систем важным вопросом становится оценка их качества. Это необходимо для того, чтобы понять, насколько эффективна та или иная рекомендательная система и насколько качественные и точные рекомендации она может предложить. На сегодняшний день существует множество метрик, позволяющих оценить качество рекомендательных систем.

Метрики качества принято делить на три группы (рисунок 2).

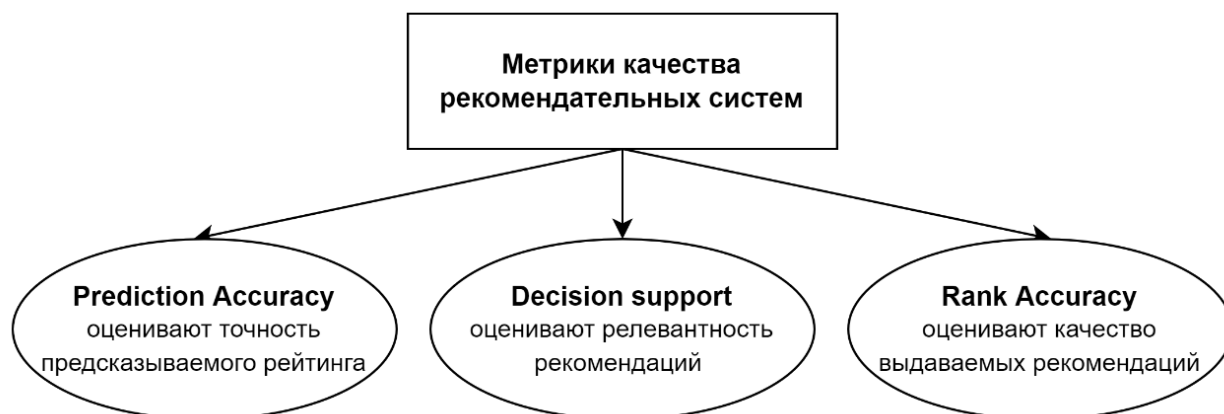


Рисунок 2 – Метрики оценки качества рекомендательных систем.

Рассмотрим подробнее некоторые из них.

Prediction Accuracy

Когда рейтинги оцениваются по непрерывной шкале (0-10), как правило, достаточно метрик класса Prediction Accuracy. Данные метрики дают возможность оценить разницу между реальной оценкой пользователя и оценкой, предсказанной системой [7].

Наиболее популярной среди метрик данного типа стала MAE (Mean Absolute Error, пер. средняя абсолютная ошибка). Помимо нее используются другие схожие метрики, например, MSE (Mean Squared Error, пер. средняя квадратичная ошибка), RMSE (Root Mean Squared Error, пер. средняя квадратичная ошибка) и другие. Метрика RMSE стала особенно популярной в последнее время после применения в конкурсе Netflix Prize.

- MAE (Mean Absolute Error)

Средняя абсолютная ошибка оценивается как абсолютная разность между предсказанием алгоритма и реальной оценкой по модулю:

$$MAE = \frac{\sum_{i \in n} |P_i - R_i|}{n} \quad (1)$$

Статистические свойства и простота вычислений сделали данную метрику наиболее популярной в оценке рекомендательных систем, хотя и существуют некоторые ограничения при оценке систем, ориентированных на рекомендации определенного количества объектов.

- MSE (Mean Squared Error)

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

MSE почти никогда не равна нулю, а происходит это из-за присутствия элемента случайности в данных или из-за невозможности учитывания всех факторов, позволяющих увеличить предсказательную способность.

- RMSE (Root Mean Squared Error)

$$RMSE = \sqrt{\frac{\sum_{i \in n} (P_i - R_i)^2}{n}} \quad (3)$$

Ошибки, обнаруженные с помощью метрики RMSE, могут оказать большое влияние на решение пользователя. Именно поэтому данная метрика довольно часто применяется в оценке рекомендательных систем.

Однако, у данной метрики есть и несколько минусов:

- У ошибки в предсказании высокой оценки вес такой же, что и у ошибки в предсказании низкой оценки;
- Кроме предсказания рейтинга также важно преподносить пользователям объекты в необходимом порядке, то есть требуется учитывать ранжирование, а этого данная метрика не умеет.

Decision Support

Метрики класса Decision Support позволяют узнать, насколько качественно система может отличать плохие объекты от хороших. Наибольшей популярностью среди метрик данного типа пользуются такие метрики как точность, полнота и ROC-показатели.

Данные показатели можно применять для задачи поиска наиболее релевантных объектов, особенно в том случае, если предпочтения пользователей определяются бинарными оценками (0 и 1, да и нет).

В ситуации, когда рейтинги изначально откладываются на непрерывной шкале, их можно перевести в бинарный формат, применив четкое правило. К примеру, если оценка меньше 3.5, то она считается «плохой», а если больше, то «хорошей» [8]. И наоборот, если оценки пользователей находятся в широком числовом диапазоне, тогда с помощью этих показателей нельзя оценивать правильный порядок объектов в списке рекомендаций.

Данные метрики показывают хороши или нет рекомендуемые объекты и не учитывают, какой объект лучше. В этих случаях данная метрика является не лучшим вариантом решения.

- Accuracy (пер. точность)

В простейшем случае метрикой может быть доля объектов, по которым система приняла правильное решение.

$$Accuracy = \frac{P}{N} \quad (4)$$

Где P – количество объектов, по которым система приняла правильное решение,
 N – размер обучающей выборки.

Главная особенность данной метрики заключается в том, что при ее использовании всем объектам присваивается одинаковый вес, что может быть не корректно в том случае, когда распределение объектов в обучающей выборке довольно сильно смещено в сторону какого-либо одного или нескольких классов.

- Precision (пер. точность) и recall (пер. полнота)

Точность и полнота являются метриками, которые применяются при оценке большей части алгоритмов извлечения информации.

Точность рекомендательной системы в пределах определенного класса – это доля объектов, действительно принадлежащих этому классу относительно всех объектов, которые система причислила к данному классу.

Полнота рекомендательной системы – это доля найденных системой объектов, принадлежащих классу относительно всех объектов этого класса в тестовой выборке.

Для расчета вышеупомянутых метрик используются бинарные матрицы ошибок.

Бинарная матрица ошибок состоит из следующих значений:

- истинно положительные (TP): Фактическое значение положительное, модель предсказывает положительное.
- ложноотрицательные (FN): Фактическое значение положительное, модель предсказывает отрицательное.
- ложноположительные (FP): Фактическое значение отрицательное, модель предсказывает положительное.
- истинно отрицательные (TN): Фактическое значение отрицательное, модель предсказывает отрицательное.

Тогда метрики точность и полнота определяются следующим образом:

$$Precision = \frac{TP}{TP + FP} \quad (5)$$

Данная метрика показывает долю рекомендаций, понравившихся пользователю.

$$Recall = \frac{TP}{TP + FN} \quad (6)$$

Метрика recall определяет долю интересных пользователю товаров, которая была показана.

- F-мера

F-мера является гармоническим средним между полнотой и точностью. Она также стремится к нулю, когда полнота или точность стремятся к нулю.

$$F1 = \frac{2PR}{P + R} \quad (7)$$

В этой формуле точности и полноте придается одинаковый вес, в связи с чем F-мера будет одинаково падать при уменьшении и точности и полноты.

- ROC-кривая

Данная метрика чаще всего используется как альтернатива точности/полноте, и обычно применяется в теории сигналов. ROC-кривая наглядно показывает поведение системы при классификации релевантных и нерелевантных объектов.

Кривая ROC является графическим представлением верной и ошибочной классификации объектов.

Rank Accuracy

Обычно рекомендации представляют собой список из определенного количества позиций. Например, список рекомендованных товаров в интернет-магазине или список фильмов, подобранных для пользователя на стриминговой площадке.

Метрики типа Rank Accuracy позволяют оценить качество ранжирования выдаваемых рекомендаций в таком списке. Рассмотрим подробнее некоторые из них.

- MRP (Mean Reciprocal Rank, пер. среднеобратный ранг)

$$MRP = \frac{1}{|Q|} \sum_{i=1}^{|Q|} \frac{1}{rank_i} \quad (8)$$

Где $rank_i$ обозначает положение первого релевантного ответа для некоторого запроса i .
Данная метрика позволяет выяснить, на какой позиции из всего списка рекомендаций пользователь находит первую полезную.

- MAP (Mean Average Precision, пер. средняя точность)

Данная метрика является средним значение показателей точности для каждого объекта.

$$MAP = \frac{\sum_{q=1}^Q AveP(q)}{Q} \quad (9)$$

- nDCG (Normalized Discounted Cumulative Gain, пер. нормированный дисконтированный совокупный доход)

$$nDCG = \sum \frac{R(i)}{\max(1, \log(i))} \quad (10)$$

Метрика nDCG показывает информативность выдачи с учетом ранжирования рекомендаций.

Выводы

При выборе метрики оценки качества рекомендательных систем нужно учитывать множество факторов. Например, тип решаемой задачи, тип используемой рекомендательной системы и т.д.

В статье рассмотрены различные типы рекомендательных систем и дан подробный анализ существующих методов оценки их качества: приведена их классификация и описание. Также в работе описаны основные плюсы и минусы использования той или иной метрики.

Результаты проведенного исследования могут быть полезны при разработке собственной рекомендательной системы и позволяют облегчить выбор метода оценки ее качества.

Список литературы

1. Ekstrand M. D., Riedl J. T., Konstan J. A. Collaborative Filtering Recommender Systems // Foundations and Trends® in Human–Computer Interaction, 2011. Vol. 4, No. 2. – P. 81-173.
2. Billsus D., Pazzani M.J. Learning Collaborative Information Filters // Proceeding 15th International Conference on Machine Learning, 1998. – P. 46-54.
3. Jannach D., Zanker M., Felfernig A., Friedrich G. Recommender Systems – An Introduction. Cambridge University Press, 2010. – 360 P.
4. Pazzani M., Billsus D. Learning and revising user profiles: The identification of interesting web sites // Machine Learning - Special issue on multistrategy learning, 1997. Vol. 27, Issue 3. – P. 313–331.
5. Berry M.W. Large scale singular value computations // International Journal of Supercomputer Applications, 1992. No. 6(1). – P. 13–49.
6. Николенко С.А. Рекомендательные системы. СПб: Изд-во Центр Речевых Технологий, 2012. – 53 с.
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. 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.

References

1. Ekstrand M. D., Riedl J. T., Konstan J. A. Collaborative Filtering Recommender Systems // Foundations and Trends® in Human–Computer Interaction, 2011. Vol. 4, no. 2. - P. 81-173.
 2. Billsus D., Pazzani M.J. Learning Collaborative Information Filters // Proceeding 15th International Conference on Machine Learning, 1998. - P. 46-54.
 3. Jannach D., Zanker M., Felfernig A., Friedrich G. Recommender Systems - An Introduction. Cambridge University Press, 2010. - 360 P.
 4. Pazzani M., Billsus D. Learning and revising user profiles: The identification of interesting web sites // Machine Learning - Special issue on multistrategy learning, 1997. Vol. 27, Issue 3. - P. 313-331.
 5. Berry M.W. Large scale singular value computations // International Journal of Supercomputer Applications, 1992. No. 6(1). – P. 13–49.
 6. Nikolenko S.A. recommender systems. St. Petersburg: Center for Speech Technologies, 2012. - 53 p.
 7. Vorontsov K.V. Combinatorial approach to assessing the quality of learning algorithms. Mathematical questions of cybernetics / Ed. ABOUT. Lupanova. - M.: Fizmatlit, 2004. - Т. 13. - P.5-36.
 8. Vakhrusheva M.Yu., Evdokimov I.V. Development of software for analytical information systems // Proceedings of the Bratsk State University. Series: Economics and Management. 2014. V. 1. No. 1. - S. 196-199.)
-