

СОДЕХАНИЕ

ВВЕДЕНИЕ.....	4
1 Рекомендательные системы recommender system(RS).....	5
1.1Определение:	5
1.2 Области применения	5
2 Критерии сравнения RS.....	6
2.1 Методы сравнения	6
2.1.1 Сравнение оффлайн.....	6
2.1.2 Сравнение онлайн	6
2.1.3Субъективное сравнение.....	7
2.2Индикаторы сравнения.....	7
2.2.1 Прогнозируемый показатель точности оценки	7
2.2.2 Точность классификации	8
2.2.3 Своевременность Real-time.....	8
2.2.4 Масштабируемость Scalability	9
2.2.5 Новизна Novelty	9
2.2.6 Сюрприз Serendipity	9
3 Классификация алгоритмов рекомендаций	10
3.1Рекомендуемый алгоритм Content-based Recommendations (CB)	10
3.2.Алгоритм синергического фильтрации collaboratIve filtering (CF)	12
3.3.Гибридные рекомендательные алгоритмыHybrid Recommendation.....	15
3.4 Сравнение преимуществ и недостатков трех алгоритмов	17
Заключение	19
СПИСОК ЛИТЕРАТУРЫ	20

ВВЕДЕНИЕ

В последние годы, в связи с бурным развитием Интернета и взрывным ростом объема информации, рекомендательные системы играют важнейшую роль в электронной коммерции, социальных сетях, музыке и видео[2]. В данной работе анализируются распространенные в повседневной жизни алгоритмы рекомендаций с помощью классификации рекомендательных систем, чтобы понять преимущества и недостатки различных алгоритмов рекомендательных систем. Затем анализируются распространенные методы оценки и показатели рекомендательных систем в современной индустрии, чтобы обеспечить основу для проектирования и оптимизации рекомендательных систем в практических приложениях.

Для достижения поставленной цели мне необходимо решить следующие задачи:

- Провести обзор рекомендательных систем для принятия решений.
- Сформировать сравнения критерии сравнения рекомендательных систем.
- Провести классификацию рекомендательных систем для принятия решений.
- Провести сравнительный анализ рекомендательных систем.

1 Рекомендательные системы recommender system(RS)

1.1 Определение:

Рекомендательная система определяется как стратегия принятия решений для пользователей в условиях сложной информационной среды [4]. Также рекомендательная система была определена с точки зрения электронной коммерции как инструмент, помогающий пользователям осуществлять поиск по записям знаний, которые связаны с интересами и предпочтениями пользователей [5]. Рекомендательная система была определена как средство помощи и дополнения социального процесса использования рекомендаций других людей для осуществления выбора при отсутствии достаточных личных знаний или опыта в отношении альтернатив [6]

1.2 Области применения

Применение рекомендательных систем предназначено именно для решения проблемы информационной перегрузки и улучшения пользовательского опыта.[7]

Например, в сфере электронной торговли, таких как ozon, рекомендуемая система может рекомендовать пользователю товары или рекламные кампании, основанные на истории покупок пользователем, поведении браузера. Такие персонизированные рекомендации могут не только увеличить покупательскую ставку и сумму, но и увеличить продажи и способствовать развитию электронной торговли.

2 Критерии сравнения RS

2.1 Методы сравнения

Что касается метода оценки RS, мы можем разделить и определить его в соответствии с его бизнес-процессом. Бизнес-процесс рекомендательной системы показан на рисунке 3:



Рисунок3-Бизнес-процесс системы рекомендаций

2.1.1 Сравнение оффлайн

Оффлайн-оценка в основном предназначена для оценки качества обученной модели рекомендаций (то есть того, хорошо ли и точно модель прогнозирует). Как правило, выборочные данные делятся на обучающий набор и тестовый набор. Обучающий набор используется для обучения модель, а тестовый набор используется для оценки ошибки прогнозирования модели (обычно также будет набор проверки, используемый для настройки гиперпараметров модели)

2.1.2 Сравнение онлайн

Онлайн-оценка - это процесс разработки нескольких сценариев, в которых могут участвовать онлайн-пользователи, и оценки сильных и слабых сторон с помощью отзывов пользователей. АВ (All-Between)

тестирование, как инструмент проверки, является распространенным методом онлайн-оценки.

AB (All-Between) тестирование, как инструмент проверки, является распространенным методом онлайн-оценки.

2.1.3 Субъективное сравнение

Пользователи имеют наибольшее влияние на опыт и ощущения от рекомендательной системы в процессе ее использования. В этом случае мы можем получить реальную оценку рекомендательной системы с помощью субъективной оценки.[17] Конкретными методами оценки могут быть анкетирование пользователей, телефонные интервью, непосредственное личное общение с пользователями и т. д., чтобы объективно понять отзывы и мнения пользователей о рекомендуемых продуктах. Субъективная оценка - очень важный способ оценки рекомендательной системы, который может быть использован для оптимизации всех аспектов рекомендательной системы, включая взаимодействие, опыт, эффект и так далее.

2.2 Индикаторы сравнения

2.2.1 Прогнозируемый показатель точности оценки

Прогнозируемый показатель точности оценки состоит из трех основных показателей: Средняя абсолютная погрешность (Mean Absolut Error, MAE), средняя погрешность (Mean Square Error, MSE) или равная погрешность по квадратному корню (Root Mean Square Error, RMSE), определенная формула:

$$MAE = \frac{\sum_{t=1}^n |\hat{y}_t - y_t|}{n} \quad (1)$$

$$MSE = \frac{\sum_{t=1}^n (\hat{y}_t - y_t)^2}{n} \quad (2)$$

$$RMSE = \sqrt{\frac{\sum_{t=1}^n (\hat{y}_t - y_t)^2}{n}} \quad (3)$$

2.2.2 Точность классификации

Точность классификации обычно измеряется с помощью точности и уровня отзыва, показателей F, AUC и других соответствующих показателей. Для произвольного пользователя u , в списке рекомендуемых длины N , точность u и коэффициент отзыва указаны следующим образом:

$$Precision = \frac{N_{rs}}{N_s} \quad (4)$$

$$Recall = \frac{N_{rs}}{N_r} \quad (5)$$

Где N_{rs} — число предметов, которые нравятся пользователям в рекомендованных предметах, N_s — рекомендуемое число предметов, и n представляет собой предыдущий n -рекомендационный компонент. N_r — это число предметов, которые нравятся пользователя.

2.2.3 Своевременность Real-time

Своевременность RS отражает текущие интересы и потребности пользователей. Она состоит из двух аспектов. Первый аспект - рекомендовать новые товары, добавленные в систему, пользователю в режиме реального времени; второй аспект - рекомендовать новые товары пользователю в режиме реального времени на основе поведения пользователя (покупка или клик) [31].

2.2.4 Масштабируемость Scalability

Когда рекомендательная система только начинает работать, объем ее данных (данные об объекте и данные о пользователе) обычно невелик, а некоторые из них могут быть аналоговыми, эту проблему также называют проблемой "холодного старта" рекомендательной системы

2.2.5 Новизна Novelty

Новизна рекомендуемой системы означает рекомендуемые предметы для пользователей, которых они не понимают.[18]В настоящее время статистическая новизна рекомендуемых систем проводится главным образом через опросы пользователей. Рекомендуемая система коммерциализации требует сбалансированного прогнозирования точности и новых показателей, таких как меткость, разнообразие и т.д.

2.2.6 Сюрприз Serendipity

В рекомендованной системе измерение неожиданности означает степень [10], в которой рекомендуемая система удивляет пользователей. Случайная рекомендация может удивить пользователей, но пользователь может снизить доверие к рекомендуемой системе, если рекомендует слишком много незначительных предметов. Таким образом, вопрос о том, как сбалансировать точность и неожиданность в последние годы, все больше привлекает внимание исследователей.

Таб.1 обобщение RS критериев оценки

Название критериев оценки	Содержание оценки	методы оценки
MAE	Прогнозируемая точность рейтинга	Офлайн оценка
MSE		
RMSE		
<i>Presicion</i>		

Recall	Точность рекомендуемых предметов	Онлайн оценка
Real-time	RS в реальном времени	
Scalability	Масштабируемость RS	
Novelty	Рекомендуемая новинка	Субъективная оценка
Serendipity	Рекомендуемый сюрприз	

3 Классификация алгоритмов рекомендаций

В зависимости от рекомендованных объектов, алгоритм делится на три алгоритма:

3.1 Рекомендуемый алгоритм Content-based Recommendations (CB)

Алгоритм рекомендаций на основе элементов - это распространенный подход к рекомендательным системам, который дает рекомендации, анализируя сходство между элементами. Этот алгоритм основывается на прошлом поведении или предпочтениях пользователя, определяет другие товары, похожие на те, которым отдает предпочтение пользователь, и рекомендует эти товары пользователю.

Основная идея алгоритма рекомендации товаров заключается в том, что если пользователь отдает предпочтение какому-либо товару, то он, скорее всего, будет заинтересован и в других товарах, похожих на этот товар. Поэтому сначала мы оцениваем сходство между элементами. Существует несколько распространенных методов расчета сходства, таких как косинусоидальное сходство, коэффициент корреляции Пирсона и коэффициент сходства Жаккара.

На следующем этапе, в сеансе рекомендаций, мы отфильтруем объекты, которые нравятся пользователям на основе их прошлого поведения или предпочтений, и на основе сходства между этими объектами выберем для

рекомендации объекты, которые больше похожи на пользователя. Можно выбрать Top-N элементы для рекомендации на основе рейтинга сходства или выбрать элементы с более высокими оценками для рекомендации путем взвешивания оценок на основе прошлого поведения и предпочтений пользователя.

Исходя из принципов и концепций алгоритма, алгоритм должен понимать только поведение или предпочтения рекомендующего пользователя и не имеет прямой связи с поведением или предпочтениями других пользователей, что делает результаты более персонализированными и независимыми. Аналогичным образом, связь между поведением пользователя и историческими данными, элементами и рекомендациями, а также использование контентных признаков для рекомендаций делает данный рекомендательный алгоритм обычно обладающим высокой интерпретируемостью результатов, однако существуют и некоторые ограничения. Так, например, рекомендации даются только в пределах диапазона похожих товаров, игнорируя тот факт, что пользователи могут быть заинтересованы в других типах товаров. Что касается проблемы "холодного старта", то данный рекомендательный алгоритм может решить проблему "холодного старта" товара, но страдает от проблемы "холодного старта" нового пользователя. Достоинства и недостатки контентных методов рекомендаций подробно рассмотрены в табл. 1.

табл.2 Достоинства и недостатки контентных методов рекомендаций подробно рассмотрены

Рекомендуемый алгоритм Content-based Recommendations (CB)	
Достоинства	<ol style="list-style-type: none"> 1. высокая независимость пользователя 2. высокая интерпретируемость проблем 3. может решить проблему "холодного старта" нового проекта
Недостатки	<ol style="list-style-type: none"> 1. существует проблема холодного старта для новых пользователей

2. ограниченный объем рекомендаций, не позволяющий изучить потенциальный интерес пользователей
3. отсутствие своевременности

3.2.Алгоритм синергического фильтрации collaborative filtering (CF)

Она сопоставляет пользователей с родственными интересами и предпочтениями, вычисляя сходство между их профилями для выработки рекомендаций [8].

Методы коллаборативной фильтрации можно разделить на две категории: классификация на основе памяти (CF) и классификация на основе модели [8].

Основные этапы совместной фильтрации, основанной на использовании пользователей, следующие:

1. Построение матрицы оценок "пользователь - элемент": построение данных оценок "пользователь - элемент" в виде двумерной матрицы, строки которой соответствуют пользователям, столбцы - элементам, а элементы матрицы представляют собой оценки элементов пользователями.
2. Вычислить сходство между пользователями: использовать соответствующие меры сходства.
(например, косинусоидальное сходство или коэффициент корреляции Пирсона) для расчета сходства между пользователями.
3. найти K наиболее похожих пользователей на целевого пользователя: на основе рассчитанных значений сходства найти K наиболее похожих пользователей на целевого пользователя.
4. составление рекомендаций на основе данных о рейтингах K пользователей: составление рекомендаций целевому пользователю с использованием

элементов, которые понравились К похожим пользователям и еще не были оценены целевым пользователем.

Эти шаги являются основной идеей алгоритма коллаборативной фильтрации на основе пользовательских данных. Анализируя исторические данные о поведении пользователя, алгоритм способен найти других пользователей, имеющих схожие с целевым пользователем интересы, и выработать рекомендации на основе предпочтений этих пользователей. Конкретная реализация может быть соответствующим образом адаптирована и усовершенствована в соответствии с потребностями набора данных и рекомендательной системы.

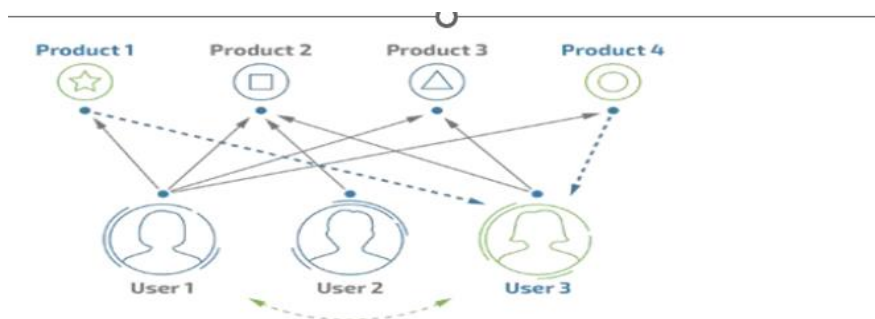


Рисунок 1 –Совместная фильтрация на основе пользователей

Разница между синтаксическим алгоритмом фильтрации, основанным на предметах и пользователем, заключается в том, что объекты отличаются друг от друга в расчетах. Аналогичные объекты, основанные на пользователях, являются людьми, а затем объекты, основанные на схожести, являются предметами. Рассчитайте степень сходства между различными предметами и после этого рекомендуйте Один и тот же предмет в соответствии с таблицей пользовательских товаров. Рисунок 2-3, показанный на рисунке, основан на симпозитивном алгоритме фильтрации, основанном на предметах:

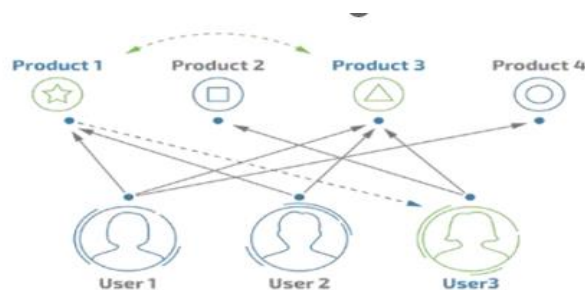


Рисунок 2 - Совместная фильтрация на основе элементов

Алгоритм коллективной фильтрации на основе элементов состоит из следующих 4 основных шагов:

1. Построение матрицы оценок "элемент-пользователь": данные о пользовательских оценках элементов организуются в двумерную матрицу, где строки представляют элементы, а столбцы - пользователей, и каждый элемент матрицы представляет собой пользовательскую оценку элемента.

2. Вычисление сходства между элементами: вычисление сходства между элементами с помощью некоторой меры сходства (например, косинусоидального сходства, коэффициента корреляции Пирсона и т.д.).

3. Найти объекты, оцененные целевыми пользователями: найти объекты, оцененные целевыми пользователями, из матрицы оценок "объект-пользователь".

4. Формирование списка рекомендаций: на основе сходства между элементами и элементами, оцененными целевыми пользователями, найти элементы с высокой степенью сходства, не оцененные целевыми пользователями, и сформировать список рекомендаций. Список рекомендаций может быть отсортирован в соответствии с определенными

правилами (например, сортировка по сходству, по высокому или низкому рейтингу и т.д.) для получения окончательного результата рекомендаций.

В результате выполнения четырех описанных выше этапов алгоритм совместной фильтрации на основе элементов может сформировать индивидуальный список рекомендаций для пользователя на основе сходства между элементами.

Для более точного сравнения двух различных типов алгоритмов коллаборативной фильтрации в табл. 2 построена таблица :

табл. 2 Сравнение методов совместной фильтрации на основе пользователей и элементов

особенности	Пользователь(user)	Элемент(item)
Требования к данным	Матрица рейтингов пользователей	Матрица рейтингов пользователей
Вычисление подобия	Основываясь на сходстве между пользователями	Основываясь на сходстве между вещами
Рекомендованный процесс	Найти других пользователей с подобными интересами для целевого пользователя и рекомендовать предметы, которые ему нравятся	Найти другие предметы, похожие на те, которые нравятся целевым пользователям
Рекомендовать объяснение	Низкая интерпретация результатов рекомендаций, основанная на предпочтениях других пользователей	Сильная интерпретация результатов рекомендаций, основанных на аналогичных предметах
Проблема холодного старта	Трудности с новыми пользователями, поскольку недостаточно данных о поведении пользователей.	Трудности с новыми предметами, поскольку недостаточно данных о сходстве предметов.

3.3. Гибридные рекомендательные алгоритмы Hybrid Recommendation

Из приведенного выше анализа следует, что алгоритмы CB и CF имеют свои преимущества и недостатки. Как же использовать их достоинства и избежать недостатков? В повседневной жизни мы часто используем гибридные алгоритмы рекомендательных систем. Под ним понимается комбинация нескольких рекомендательных алгоритмов в надежде получить лучшие результаты рекомендаций.

Конкурс Netflix Prize, [10]объявленный компанией Netflix в 2006 году, является одним из самых известных примеров в истории развития рекомендательных систем, улучшающих рекомендации с помощью гибридных рекомендательных алгоритмов. Победителем конкурса в 2009 году стала компания Bellkor's Pragmatic Chaos, объединившая алгоритмические силы трех команд и объединившая более 500 алгоритмических моделей с помощью модели GBDT. Этот метод интеграции представляет собой гибридный рекомендательный алгоритм.

К распространенным гибридным алгоритмам относятся следующие:

Взвешенный гибрид: результаты рекомендаций, генерируемые различными алгоритмами, линейно объединяются в соответствии с определенными весами для получения окончательных результатов рекомендаций. Этот метод прост, интуитивно понятен и легко реализуем, и часто используется для смешивания результатов коллаборативной фильтрации и рекомендаций на основе контента.

Каскадное смешивание: используя выход одного алгоритма в качестве входа другого алгоритма, несколько алгоритмов последовательно соединяются, образуя каскадное смешивание. Например, алгоритм рекомендаций на основе содержания используется для формирования списка кандидатов, а затем алгоритм совместной фильтрации используется для сортировки и фильтрации списка кандидатов.

Параллельное смешивание: использование нескольких алгоритмов для одновременной генерации результатов рекомендаций и их объединение или выбор одного из них в качестве окончательного результата рекомендации при определенных условиях. Этот метод позволяет в полной мере использовать преимущества каждого алгоритма и часто используется для

решения проблемы, связанной с тем, что одному алгоритму сложно охватить все ситуации.

Контекстное смешивание: динамический выбор подходящего алгоритма рекомендации на основе контекстной информации о пользователе (например, время, местоположение, устройство и т.д.). Например, в разные периоды времени выбираются разные алгоритмы, или соответствующие алгоритмы выбираются в зависимости от среды, в которой находится пользователь.

Гибридные алгоритмы могут компенсировать недостатки одного алгоритма и повысить охват и точность рекомендательной системы. Однако при разработке и реализации гибридных алгоритмов необходимо всесторонне учитывать методы слияния, компромиссные соотношения и требования к реальному времени между различными алгоритмами, а в практических приложениях требуются эффективные компромиссы и корректировки.

3.4 Сравнение преимуществ и недостатков трех алгоритмов

Независимо от комбинации, чтобы по-настоящему реализовать преимущества рекомендательных алгоритмов, необходимо следовать принципу использования сильных сторон и дополнения слабых [5]. Для того чтобы интуитивно сравнить преимущества и недостатки трех алгоритмов (алгоритм рекомендаций на основе контента, алгоритм рекомендаций на основе коллаборативной фильтрации и гибридный алгоритм рекомендаций). Для сравнения преимуществ и недостатков трех рекомендательных алгоритмов была создана следующая таблица 3.

таблица 3 Сравнительная таблица
преимуществ и недостатков трех алгоритмов

название метода	Название критерия сравнения				
	Accuracy	Coverage	Real-time	Cold-start Problem	User- Satisfaction
CB	высокая	низкая	хорошая	есть	высокая
CF (user+item)	Средняя	высокая	средняя	User есть, item нет	средняя
Hybird	высокая	высокая	хорошая	нет	высокая

Заключение

Проведя вышеизложенный анализ, мы можем сделать вывод:

1. Выбор алгоритма:

CB:RS имеет богатую информацию о содержании товара, но данные о поведении пользователя более скудны

CP: RS имеется много данных о поведении пользователя, но информация о содержании товара более скудная.

Гибридная рекомендация: RS необходимо сбалансировать данные об интересах и поведении пользователей и информацию о содержании предмета.

2. проведение оценки:

Выбор соответствующих методов оценки и показателей основывается на Таблице 4 и требованиях к дизайну RS.

СПИСОК ЛИТЕРАТУРЫ

1. Кутянин А.Р. «Рекомендательные системы: обзор основных постановок и результатов».2018-2023
2. GUO Jinghan. Personalized Music Intelligent Recommendation System Oriented to User Preference [J]. Automation Technology and Application,2023
3. Guo yanhong. Рекомендуемые системы для совместных алгоритмов фильтрации и прикладных исследований (D)Политехнический университет, 2008.
4. Rashid AM, Albert I, Cosley D, Lam SK, McNee SM, Konstan JA et al. Getting to know you: learning new user preferences in recommender systems. In: Proceedings of the international conference on intelligent user interfaces; 2002. p. 127–34
5. Schafer JB, Konstan J, Riedl J. Recommender system in e-commerce. In: Proceedings of the 1st ACM conference on electronic commerce; 1999. p. 158–66.
6. P. Resnick, H.R. Varian Recommender system's CommunACM, 40 (3) (1997), pp. 56-58,
7. Хуан Сяоин, Ли Чэнлун. Персонализированная система рекомендаций для информации традиционной китайской медицины на основе анализа больших данных [J]. Технологии и приложения автоматизации, 2023

8. J.L. Herlocker, J.A. Konstan, L.G. Terveen, J.T. Riedl Evaluating collaborative filtering recommender systems ACM Trans Inform Syst, 22 (1) (2004), pp. 5-53
9. WU Jinkun. Research on User interest-based Recommendation Algorithm [D]. Nanjing University of Posts and Telecommunications,.2022.
10. Zeng Xiuqin, He Meng, Shen Mengli et al. Research on Subjective evaluation indicators of music recommendation System -- A case study of NetEase Cloud Music [J]. News and Communication Review,2019,
11. FENG Xing-Jie, Zhang He. Dialog recommendation method combining User's long and short Term Preferences [J]. Computer Engineering and Design,2023, 7024.2023.10.037
12. N. Friedman, D. Geiger, M. Goldszmidt Bayesian network classifiers
a) Mach Learn, 29 (2–3) (1997), pp. 131-163
13. Wan Xin, FENG Yun, Xiao Yan et al. Design of Personalized Learning resource recommendation system [J]. Fujian Computer,2
14. GAO Fei, Chen Deli, YAN Tao. Personalized evaluation based on content recommendation and collaborative filtering algorithm [J]. Journal of the University of Anhui (Natural Science Edition),2022, 1.
15. Han Di, Chen Yi-Jun, LIAO Kai et al. Effective coupling and application of accuracy, novelty and Diversity in recommendation systems [J]. Journal of

16. CHU Hong-Lin. Research and Analysis of Collaborative Filtering Recommendation Algorithm [J]. Fujian Computer,2021 -
17. Bi Jiehao. Recommendation Effect evaluation of Personalized recommendation system based on user perception [D]. China University of Petroleum (Beijing),2021.
18. ZENG Xiuqin, Jiang Chen, LI Canyu et al. Research on Evaluation index system of Personalized recommendation system of mobile Video client [J]. China Network Communication Research,2017,
19. Yan H. Research on novelty and diversity recommendation algorithms and design and implementation of evaluation system [D]. Beijing University of Posts and Telecommunications,2020.
20. Guo Jinghan. Personalized Music Intelligent Recommendation System Oriented to User Preference [J]. Automation Technology and Application,2023
21. Wang Miao, Li Dawei. Application of Hybrid recommendation Algorithm based on Content and Collaborative Filtering in Digital Science Museum [J]. Network Security Technology and Application,2023
22. Breese J, Heckerman D, Kadie C. Empirical analysis of predictive algorithms for collaborative filtering. In: Proceedings of the 14th conference on uncertainty in artificial intelligence (UAI-98); 1998. p. 43–52.