

Индивидуальное задание по проекту

[Code ▾](#)

Группа N23, afrakhmanova

Задание

Индивидуальное задание звучит следующим образом:

```
## [1] "Сделать рекомендацию для пользователя, который есть в системе, но хочет что-то новенькое, не похожее на его вкусы (даже если ему может не понравиться)"
```

Предыдущий код

Для выполнения данного задания я бы хотела рассмотреть два варианта возможных рекомендаций. Один из них основан на рекомендациях, использованных в групповом отчете, а второй – на логических решениях.

Тогда из предыдущего кода мне понадобится только часть с датасетами.

Решение

Вариант 1

Рекомендация строится аналогично привычному методу коллаборативной фильтрации. Разница и главная идея заключается в том, что оценки пользователя заменяются на противоположные по 5-балльной шкале методом вычитания по модулю.

В данных имеются нулевые оценки, которые могут олицетворять как хороший отзыв, так и плохой. А так как используемый метод основан только на оценках пользователя, стоит от таких избавиться.

Чтобы рекомендацию можно было сравнить со стандартной моделью, все изменения было решено оставить, как в групповом отчете: количество оценок, поставленных одним пользователем, больше 5, а количество оценок одного комикса при этом от 8 изначально.

Так как метод коллаборативной фильтрации делится на два метода, необходимо выбрать наилучший для получения адекватной рекомендации.

Оценки методов IBCF и UBCF выглядят так:

##	RMSE	MSE	MAE
##	1.2707599	1.6148307	0.9561138

##	RMSE	MSE	MAE
##	1.232180	1.518268	0.956435

IBCF рекомендует лучше, так как значение MAE ниже, хоть и не намного. Поэтому в дальнейшем будем использовать этот метод.

В итоге, сама рекомендация будет основана на наименьших оценках пользователя. То есть ему будут предложены комиксы, которые он еще не читал, но похожие на оцененные им на 1 или 2.

Вероятно, он может получить что-то сильно похожее на неинтересный для него комикс, что будет скорее плохой рекомендацией. А может, наоборот, получит что-то из той же “серии”, но гораздо интереснее, что не стал бы читать сам. Тогда это рекомендация будет хорошей. (Однако, зависит это все, конечно, от самого пользователя)

Если пользователь оценил менее 5 комиксов, то буквально любая рекомендация будет для него актуально новой. Поэтому, аналогично групповому проекту, он получит рекомендацию из 10 самых популярных комиксов по оценке и количеству отзывов (> 0.25 квантиль).

Вариант 2

Метод основан больше на логике, нежели на работе рекомендательных систем.

Суть заключается в том, что мы берем все оценки пользователя и удаляем из общих данных оцененные им комиксы. Кроме того, избавляемся от издателей этих же комиксов.

Из оставшихся комиксов мы выбираем те, которые имеют достаточно большое количество оценок (> 0.25 квантиль) и располагаем в порядке убывания средних оценок. Так как эта система не основна на привычных методах составления рекомендаций, а также дополнена смысловой нагрузкой, считаю возможным увеличить количество рекомендованных комиксов до 10 штук, полученных после фильтрации.

В данном случае кажется, что нет смысла фильтровать самих пользователей по количеству оцененных комиксов. Даже имея ровно одну оценку он вдруг может попросить “что-то новенькое”, тогда по логике нужно делать все те же действия, чтобы избавиться от прямого сходства с уже проставленной оценкой. Однако рекомендация может быть на что-то похожее, например, от другого издателя, что может иметь как плюсы, так и минусы, описанные в первом варианте.

Примеры

Для обоих вариантов рассмотрим два примера: пользователя со значительно большим количеством оцененных комиксов и пользователя с относительно маленьким количеством.

Вариант первый:

Пользователь “1d945500234cbc7a6138a4d017dbfe4b” оценил следующие комиксы:

title
<chr>
Invincible: Ultimate Collection, Vol. 1
Daredevil by Brian Michael Bendis & Alex Maleev Ultimate Collection, Book 2
Outcast, Vol. 1: A Darkness Surrounds Him
The Vision, Volume 1: Little Worse Than A Man
Alias, Vol. 4: The Secret Origins of Jessica Jones
Secret Six, Vol. 1: Unhinged
Irredeemable, Vol. 2
Northlanders, Vol. 4: The Plague Widow
Sweet Tooth, Volume 2: In Captivity
Batman Incorporated, Volume 1: Demon Star
1-10 of 45 rows
Previous 1 2 3 4 5 Next

Вот что выдает рекомендация:

title
<chr>
1 March: Book Three (March, #3)
2 The Absolute Sandman, Volume Two
3 Saga: Book Two
4 Skip Beat!, Vol. 26
5 Fullmetal Alchemist, Vol. 14 (Fullmetal Alchemist, #14)
6 The Sandman, Vol. 9: The Kindly Ones (The Sandman, #9)
7 The Complete Peanuts, Vol. 6: 1961-1962
8 Skip Beat!, Vol. 17
9 Fullmetal Alchemist, Vol. 7 (Fullmetal Alchemist, #7)
10 Daredevil by Brian Michael Bendis & Alex Maleev Ultimate Collection, Book 2
1-10 of 10 rows 1-3 of 4 columns

Тут мы видим новых издателей, не оцененных ранее, а также новые категории. Однако, есть и некоторые сходства. С одной стороны, это хорошо – рекомендованные комиксы могут понравиться пользователю из-за этих сходств. С другой стороны, наоборот, они могут быть слишком похожи, что для пользователя не окажется чем-то новым.

Также можно проверить пользователя с малым количеством оцененных комиксов – “6baf45d03466a5858403d892286ff222”:

title	rating	publisher
<chr>	<dbl>	<chr>
V for Vendetta	4	Vertigo
1 row		

И вот его рекомендация:

title
<chr>
1 March: Book Three (March, #3)
2 The Absolute Sandman, Volume Two
3 Saga: Book Two
4 Skip Beat!, Vol. 26
5 Fullmetal Alchemist, Vol. 14 (Fullmetal Alchemist, #14)
6 The Sandman, Vol. 9: The Kindly Ones (The Sandman, #9)
7 The Complete Peanuts, Vol. 6: 1961-1962
8 Skip Beat!, Vol. 17

title
<chr>
9 Fullmetal Alchemist, Vol. 7 (Fullmetal Alchemist, #7)
10 Daredevil by Brian Michael Bendis & Alex Maleev Ultimate Collection, Book 2
1-10 of 10 rows 1-3 of 4 columns

Ожидаемо, это самые популярные комиксы. Ведь он поставил слишком мало оценок.

Вариант второй:

Для начала изучим рекомендацию для первого пользователя – “1d945500234cbc7a6138a4d017dbfe4b” – с большим количеством оценок:

title	average_rating
<chr>	<dbl>
1 March: Book Three (March, #3)	4.71
2 Skip Beat!, Vol. 26	4.61
3 Fullmetal Alchemist, Vol. 14 (Fullmetal Alchemist, #14)	4.60
4 The Complete Peanuts, Vol. 6: 1961-1962	4.59
5 Skip Beat!, Vol. 17	4.57
6 Fullmetal Alchemist, Vol. 7 (Fullmetal Alchemist, #7)	4.57
7 Yotsuba&!, Vol. 04 (Yotsuba&! #4)	4.55
8 黒執事 XX [Kuroshitsuji XX]	4.55
9 東京喰種トーキョーグール 14 [Tokyo Guru 14] (Tokyo Ghoul, #14)	4.55
10 Turma da Mônica: Lições	4.55
1-10 of 10 rows 1-3 of 4 columns	

Очевидно, что тут не будет тех же издателей. Однако могут встретиться случаи, когда издатели отличаются одним словом, но по факту имеют один смысл, поэтому что-то похожее может встретиться, хоть и довольно редко. А вот категории могут пересекаться, но, скорее всего, с меньшей вероятностью, чем в первом варианте.

Для другого пользователя – “6baf45d03466a5858403d892286ff222” – с меньшим количеством оценок система работает аналогично:

title
<chr>
1 March: Book Three (March, #3)
2 The Absolute Sandman, Volume Two
3 Saga: Book Two
4 Skip Beat!, Vol. 26
5 Fullmetal Alchemist, Vol. 14 (Fullmetal Alchemist, #14)
6 The Complete Peanuts, Vol. 6: 1961-1962

title
<chr>
7 Skip Beat!, Vol. 17
8 Fullmetal Alchemist, Vol. 7 (Fullmetal Alchemist, #7)
9 Daredevil by Brian Michael Bendis & Alex Maleev Ultimate Collection, Book 2
10 Yotsuba&!, Vol. 04 (Yotsuba&! #4)
1-10 of 10 rows 1-3 of 4 columns

Рекомендация из комиксов различных издателей и категорий, отличных от прочитанных – то что нужно. Сравнивать более конкретно особо не с чем, так как пользователь оценил только один комикс, что слишком мало для анализа его вкусовых предпочтений.

Выводы

Я рассмотрела два варианта решения поставленной задачи. К сожалению, в силу человеческих издержек, ни одна из них не может быть названа идеальной. Однако можно сказать, что обе системы работают довольно адекватно – выдают что-то действительно новое и не сильно похожее, как в стандартных случаях с рекомендациями.

Обобщим информацию:

> первая модель работает аналогично привычной нам системе коллаборативной фильтрации, однако основываясь при этом на низких оценках пользователя

> вторая модель основана на логике: чтобы выдать что-то новое, надо убрать все старое/схожее

Таким образом, обе модели могут дать как хорошую, так и плохую рекомендацию, что может быть оценено скорее только самим пользователем или нами, но довольно субъективно. Связано это с тем, что в том и другом случае могут встретиться сходства с уже оцененными комиксами. Тогда эти сходства могут понравиться пользователю, как отголоски уже прочитанных, интересных ему комиксов. Но могут и не понравиться в силу того, что тогда рекомендация не такая уж и новая для него.

Можно сказать, что обе идеи работают не совсем корректно. А точнее, сильно зависят от воли случая и реальных предпочтений пользователя.