Я считаю, что проблема малой заинтересованности в книгах людей набирает обороты. Возникает она потому, что люди не знают, какое произведение почитать, не могут найти подходящие для них книги. Поэтому в Сириусе я хочу реализовать программу, рекомендующую с помощью описанного мною ниже алгоритма произведения, которые пользователю будет интересно прочитать.

Примечание:Человеком, который хочет получить рекомендации, буду я.

Алгоритм сначала выбирает пользователей моей возрастной категории, после оценивает мою схожесть с ними на основе оценок книг, которые мы(я и эти пользователи) прочитали. Используя полученные данные, он выполняет некоторые математические операции, после дает рекомендации для прочтения, а также прогнозирует оценку, которую я, скорее всего, поставил бы произведению.

Более подробное описание алгоритма:

1.Поиск пользователей моей возрастной группы.(Опционально)

Алгоритм должен отсеять людей отличной от моей возрастной группы. Данный пункт опциональен из-за проблем с возрастной периодизацией.(Когда, например, ребенок 12 лет от роду мыслит как 16-летний).

Возрастные категории(стратификация Эрика Эриксона) :

(Первая возрастная категория тоже хочет получить рекомендации!)

1)Ранее детство(1-3 лет);

2)Дошкольных возраст (4-7 лет);

3)Школьный возраст (7-12 лет);

4)Юность (13-19 лет);

5)Молодость (19-35 лет);

6)Взрослость (35-60 лет);

7)Старость (от 60 лет);

2.Определение коэффициента корреляции с другими пользователями.

Сначала алгоритм должен выбрать людей, которые оценили хотя бы 5 книг, которые критиковал я. После определить их коэффициент схожести со мной.

Для этого я буду использовать среднее арифметическое расстояние от точек до прямой y=x.

Для наглядности используется плоскость в Декартовой системе координат.На ней помечаются точки P(X;Y), где Р - первая буква названия книги, но так как называть точки принято латинскими буквами,то русские придется переводить в латинские.(например, "Горе от ума"А.С. Грибоедова- G(X;Y)), X - моя оценка, Y - оценка другого пользователя, с которым меня сравнивает алгоритм.Например, "13 Апостол" Дмитрия Быкова был оценен мною на "2", а другим пользователем на "5", тогда точка имеет название А, Х равен 2, а У равен 5,то есть А(2;5). Я "отметил" еще три точки:B(5;1), М(4,3), Е(1;1).

Я выбрал формулу прямой у=х, потому что, если я и другой пользователь оценили книгу одинаково, например, на "1", то точка будет принадлежать прямой y=x, а значит расстояние между ними будет равняться нулю. Чтобы найти расстояние от точки P(X;Y) до прямой y=x, нужно воспользоваться формулой |aX+bX+c|/✓(a^2+b^2). Так как y=x не вида ax+bx+c=0, то приведем её к этому виду - 1y-1x+0=0. Теперь подставим значения a и b в формулу - |1x-1y+0|/✓(1^2+1^2) или |x-y|/✓2. Глупо делить на ✓2, т.к. ✓2 - иррациональное число, равное 1.414...Просто округлим его до 2. Итак, формула для нахождения расстояния - d=|x-y|/2.

Чтобы найти корреляцию меня и другого пользователя, нужно найти среднее арифметическое всех расстояний от точки до прямой. Пусть N - количество расстояний, тогда формула для нахождения коэффициента схожести - K = (D1+D2+D3...)/N. Значение будет тем меньше, чем мой вкус схоже со вкусом сравниваемого человека, но мне нужно, чтобы значение было тем больше, чем вкусы схоже.Чтобы этого добиться, мы возьмем обратное значение коэффициента корреляци и добавим единицу, чтобы никогда не делить на 0, то есть 1/(К+1).

На рисунке 3 изображены примеры вычислений.

3. "Некоторые математические операции"(ранжирование книг)

Теперь, когда мы нашли коэффициенты корреляции у всех пользователей, можно было бы посмотреть, какие книги понравились людям со схожими с моим вкусом и отобрать из них те, которые я еще не читал, но тогда можно было бы наткнуться на критиков, не оценивших книги, которые могли бы мне понравиться. Можно также отобрать схожего со мною пользователя, который почему-то поставил высокую оценку книге, получившей низкие оценки от всех остальных людей. Для избежания этих проблем, нужно вычислить взвешенную сумму оценок критиков. Нужно, чтобы оценка схожих со мною критиков, вносила больший вклад в общую оценку. Чтобы добиться этого, нужно перемножить К критика на оценку книги, которую я еще не читал. (см. табл)

|  |  |  |  |
| --- | --- | --- | --- |
| Пользователь | Коэффициент корреляции | "Вино из одуванчиков" Рэя Боэдбери | Результат умножения оценок на коэф. |
| Пользователь 13 | 0,7 | 4,5 | 3,15 |
| Ненавистник Брэдбери | 0,25 | 2 | 0,5 |
| глаза\_царицы | 0,4 | 3,5 | 1,4 |
| Сумма | 1,35 |  | 5.05 |
| Частное суммы оценок и суммы коэффициентов.(прогноз оценки, которую я поставил бы) | 3,7 |  |  |

(Табл.1)

Можно было бы использовать саму эту сумму для выдачи ранжирования, однако в таком случае книга, оцененная большее количество раз, имела бы приемущество.Чтобы исправить это, нужно полученную величину-сумму оценок разделить на сумму коэффициентов подобия.

Итак, я полностью описал алгоритм, который, как я считаю, поможет повысить интерес людей к книгам, так как они смогут, наконец, легко и просто находить те произведения, которые будут им по вкусу.