Book Recommendation System

—

Worthless Without Coffee

Decision Parameters

1. Genre: 'Fantasy', 'Fiction', 'Self-Help', 'Drama', 'Romance', '', 'Thriller', 'Biography', 'Erotic', 'Kids', 'Poetry', 'Horror', 'History', 'Academic', 'Comedy', 'Classic', ' Thriller', 'Domestic Fiction', 'SciFi', 'Crime Fiction', 'Psychological Fiction', 'Young-Adult Fiction'
2. Author: Given twice the weight
3. $\alpha = 0.4$

Other possible parameters:

- Publisher for academic books,
- Short/Long description
- Language.

Model: Hybrid

- Combination of Popularity Model, Collaborative Filtering and Content based recommendation.
- Increases efficiency of suggestions.

Popularity Model

- Sort according to average of all given user ratings for a particular book.
- Recommends top 25 books of different genres to a new user.
- Solves cold start problem.
- Recommendation is user independent.
- Gives the user the choice of wide variety of genres in the beginning.

Collaborative Filtering

- Recommendation is user dependent.
- Equation to be optimized:

$$\hat{r}_{ui} = p_u \cdot q_i.$$

Content based Recommendation

- Solves the problem of first-rater.
- “Soup” made for each book - author + genres
- Countvectorizer applied on matrix
- Cosine similarity

$$\text{similarity} = \cos(\theta) = \frac{A \cdot B}{\|A\|\|B\|} = \frac{\sum_{i=1}^{n} A_i \times B_i}{\sqrt{\sum_{i=1}^{n} (A_i)^2} \times \sqrt{\sum_{i=1}^{n} (B_i)^2}}$$

Content based Recommendation

- User profile - vector of ratings given by user to books
- Prediction -

$$P_{u,i} = \frac{\sum_{\text{all similar items, N}} (s_{i,N} * R_{u,N})}{\sum_{\text{all similar items, N}} (|s_{i,N}|)}$$

Hybrid Model

- It gives the predicted rating as weighted combination of the above described methods
- $\alpha = 0.4$

$$R_{hybrid} = (1-2\alpha)*R_{popularity} + \alpha*R_{collaborative} + \alpha*R_{content}$$

Training - Testing

- The data was split into 20%-80% ratio, 20%used for testing purpose and 80% for training purpose.
- Ratings ~ 31,00,000; Users ~ 53,000; Books ~ 1000.
- Training algorithms :
  - For Collaborative, Stochastic Gradient Descent
  - For Content based, countvectorizer's output used for cosine similarity
  - For Hybrid, SGD can be used to minimize RMSE and find optimal value of a.
- Accuracy metric : RMSE
- RMSE for hybrid : 0.696

Console

```
------------------------------Welcome to the Book Recommendation Engine--------------------------

1. Book Recommendation for New User.
2. Book Recommendation for Existing User.
2

Please Enter User Id: 3

----------------Welcome User3-------------------
Evaluating RMSE, MAE of algorithm SVD.

------------
Fold 1
RMSE: 0.8654
MAE:  0.6716
------------
Fold 2
RMSE: 0.8662
MAE:  0.6730
```

Recommended Books

	Book ID	Title	Genres	Collaborative Rating
29	260	How to Win Friends and Influence People	[Self-Help]	5.0
32	301	Heart of Darkness	[History]	5.0
16	80	The Little Prince	Classic, Young-Age, Psychological]	5.0
18	109	Les Misérables	[Drama, History, Classic]	5.0
26	193	Outliers: The Story of Success	[Self-Help]	5.0
22	127	The Tipping Point: How Little Things Can Make ...	[Self-Help]	5.0
27	211	Blink: The Power of Thinking Without Thinking	[Self-Help]	5.0
28	236	Into Thin Air: A Personal Account of the Mount...	[Biography, History]	5.0
23	128	Steve Jobs	[Biography]	5.0
4	15	The Diary of a Young Girl	[Biography]	5.0
37	630	Lean In: Women, Work, and the Will to Lead	[Biography]	5.0
25	160	Great Expectations	[Classic, Fiction]	5.0
17	82	Into the Wild	[Biography]	5.0
3	10	Pride and Prejudice	[Drama, Romance]	5.0
5	18	Harry Potter and the Prisoner of Azkaban (Harr...	[Fantasy, Fiction, Young-Age]	5.0
14	54	The Hitchhiker's Guide to the Galaxy (Hitchhik...	[Fantasy, Fiction]	5.0
19	114	Tuesdays with Morrie	[Biography, Fiction]	5.0

Thank-you!