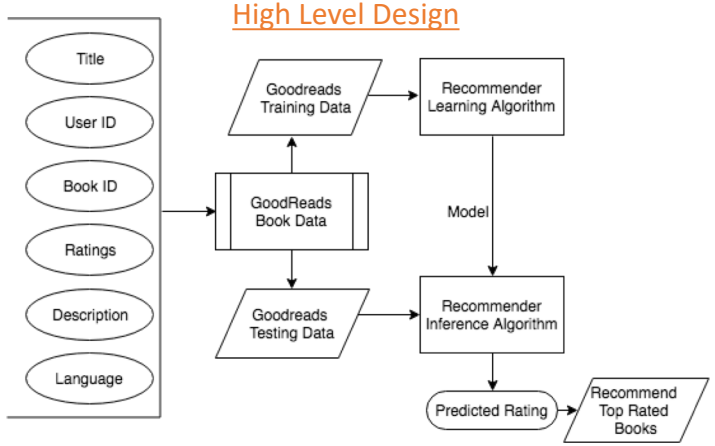
Goodreads Recommender System – What book you should read next?

Comparative Study of Recommender Systems, Understanding key desSign & functioning of different recommender models

Goal

To study recommender system algorithms to predict the rating of the books not read by the user and then recommend the top rated books which user should read next.

High Level Design



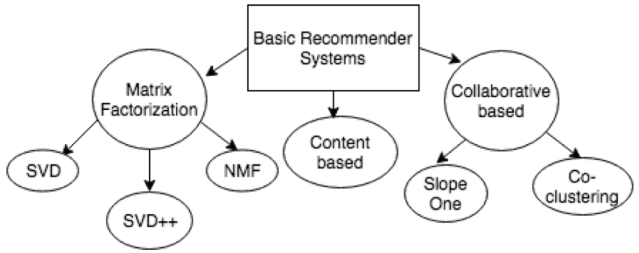
Data Description

Unnamed: 0	author	description0	Language	NumberofRatings	GoodReadAverageRating
0	J.K. Rowling	Harry Potter's life is miserable. His parents ...	English	4879247	4.44
1	Harper Lee	The unforgettable novel of a childhood in a ...	English	3368077	4.26

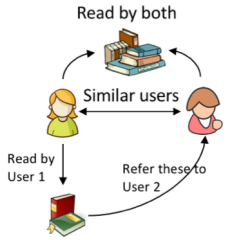
book_id_original	description1	title	book_id	goodreads_title
3	Harry Potter's life is miserable. His parents ...	Harry Potter and the Philosopher's Stone ...	2	Harry Potter and the Sorcerer's Stone (Harry ...
2657	Compassionate, dramatic, and deepl ...	To Kill a Mockingbird	4	To Kill a Mockingbird (To Kill a Mockingbird #1) ...

Books: 10000, # Users: 53424

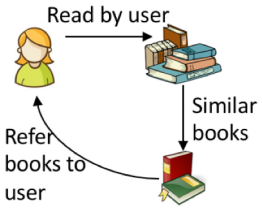
Common types of Recommenders



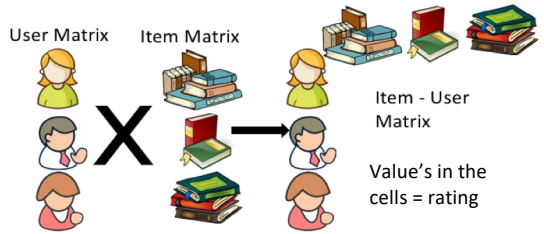
Collaborative based



Content based



Matrix Factorization based



Basic Recommender Algorithms

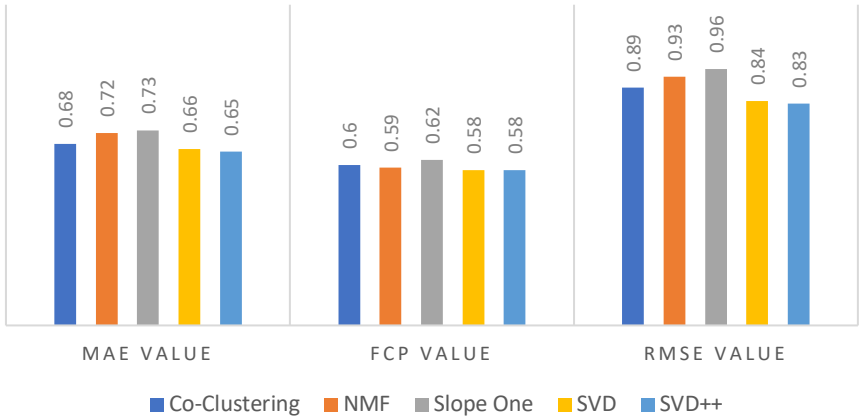
Models	RMSE
Matrix Factorization	1.064
Item Similarity Based	0.947
Content Based	3.967
SVD	0.8594
SVD++	0.8355
NMF	0.9327
Slope One	0.9633
Co-clustering	0.8959

$$MAE = \frac{1}{|\hat{R}|} \sum_{\hat{r}_{ui} \in \hat{R}} |r_{ui} - \hat{r}_{ui}|$$

$$RMSE = \sqrt{\frac{1}{|\hat{R}|} \sum_{\hat{r}_{ui} \in \hat{R}} (r_{ui} - \hat{r}_{ui})^2}.$$

FCP (Fraction of Concordant Pairs): A pair is concordant if the subject ranked higher on X also ranks higher on Y.

Evaluation Graph



Lower the RMSE value, better the model, SVD++ takes care of implicit ratings hence lower RMSE and better results.

Why content based did not work good?

- Not enough content to discriminate items precisely
- Over-specialization
- Not enough information to build solid profile for new user

Other models & Future Implementations

- Using Neural Networks along with collaborative filtering.
- Hybrid recommendations, using collaborative and content based together using a combiner.
- Using clustering techniques, recommended products go well together.
- Using Feature Weighted Linear Stacking for producing ensemble from a collection of heterogeneous models.

Conclusion

On the three evaluation criteria SVD++ gave the best results, these results can be improved using above mentioned methods. Thus Matrix factorization and collaborative system seems viable solution for GoodReads Recommender system. Content Based could be improved if taken more data for both users and Books.

Contact: Kanika Narang: knnarang@, Tehreem Ansari: tehreea@; **WIKI:** https://w.amazon.com/bin/view/GoodReadsRecommenderSystem/