Evaluating the Accuracy and Coverage Performance of Collaborative Filtering, Content-based, and Hybrid Recommender Systems

Anh Hoang, Malavika Kalani, Mike Remezo

INTRODUCTION

- Recommender systems are a personalized information filtering technology employed to make predictions on potential likeness of users towards different products and provide them with recommendations for a set of items that are of interest to them.
- On the user's side, given the wide availability of products online, recommender systems solve the problem of how to filter, prioritize, and efficiently deliver suggestions on items that best fit consumers' needs, reducing the risks of information overload and increasing users' satisfaction. On the service provider's side, effective recommender systems can increase revenue and build consumers' loyalty to the business.
- User-based and Item-based Collaborative Filtering are algorithms that generates recommendations for a user through a systematic analysis of data collected from other users sharing similar tastes or information needs
- Content-based Filtering is a recommender algorithm that looks for items that are similar to the items that users have liked in the past through analyzing items' descriptions and features.
- Hybrid Method is a recommender algorithm that combines both content-based filtering and collaborative filtering methods to maximize their strengths and minimize their weaknesses.

THESIS

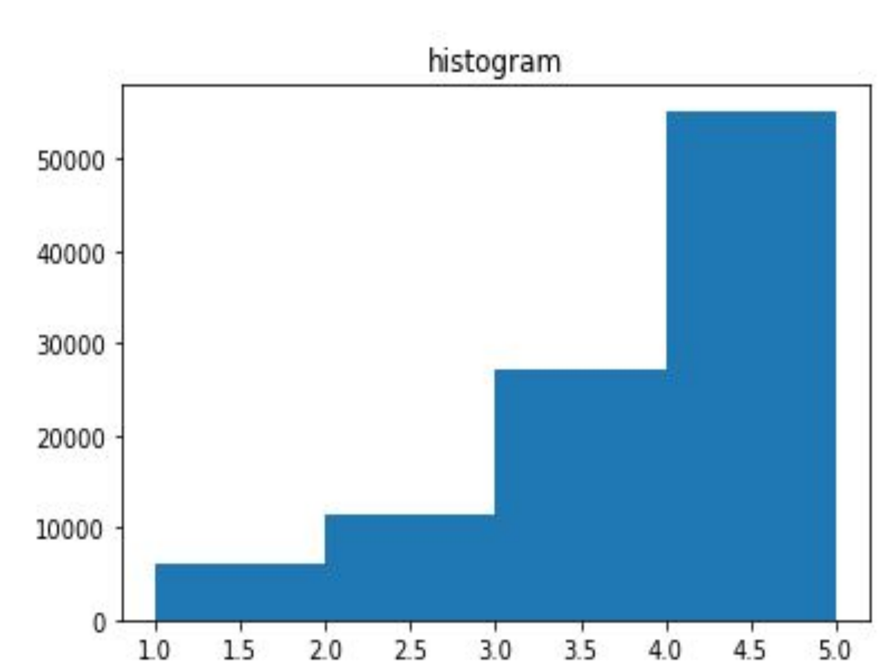
- Our research will analyse the performance of the three different recommender systems approaches using the different algorithms below:
  - User-based Collaborative filtering (using Distance and Pearson similarity measures)
  - Item-based Collaborative filtering (using Distance and Pearson similarity measures)
  - Matrix Factorisation using Stochastic Gradient Descent (MF-SGD)
  - Matrix Factorisation using Alternating Least Squares (MF-ALS)
  - Content-based method using Feature Encoding
  - Content based method using TFIDF with cosine similarity measure
  - Hybrid approach combining item-based collaborative filtering (using Distance and Pearson similarity measure) and content based method using TFIDF.
- For our collaborative filtering methods, we analyze the similarity threshold and significance weighting.
- For our MF-SGD algorithm,, we analyze the normalisation rate, learning rate and the number of factors. For our MF-ALS algorithm, we analyze the normalisation rate and the number of factors.
- For our content-based algorithms, we analyze the similarity threshold for our cosine similarities.
- With the best set of parameters that we empirically tested for each algorithm, our goal is to find the best recommendation algorithm based on the metrics: coverage and accuracy.

RELATED WORKS

- Herlocker et. al *1999* lays a strong foundation for our findings on user and item-based collaborative filtering algorithms.
- Koren et. al 2009 gave a clear description of matrix factorisation techniques which characterizes both items and users of factors inferred from item rating patterns. In order to minimise the regularized squared error, the paper detailed two algorithms which have been analysed in our research: stochastic gradient descent and alternating least squares.
- Zhou et. al 2008 observed that the alternating least squares with weighted regularization (ALS-WR) does not overfit the data if the number of iterations and item features are increased. We have analysed similar behaviour in our research to compare our findings.
- Adomavicius et.al 2005 highlighted an important limitation of content-based systems using feature encoding: the features of the recommended items must be easily parsed or assigned to items which can be a tedious process.
- Adomavicius et.al 2005 observed that hybrid approaches produce more accurate recommendations in comparison to pure approaches. We have analysed our results for two such hybrid systems using item-based collaborative filtering approaches with content-based TFIDF technique.
- Burke 2002 provides a comprehensive view of different hybrid models of recommender systems, gives in-depth examples of those hybrid models in practice, and discusses their benefits and drawbacks, which aids us greatly in our analysis of hybrid recommender models.

EXPERIMENTAL DESIGN

We used the ML-100k data which is organized in a nested dictionary, where each row is a user and each column is the movie that particular user has rated. We then tested the accuracy of each algorithm by using LOOCV for content based filtering and Holdout CV for SGD and ALS algorithms. Predictions were made based on the similarity matrix for collaborative filtering, content-based and hybrid approaches. We measured the accuracy using MSE and RMSE. A lower error metric would mean higher accuracy.



Ratings Distribution of the ML-100k Dataset

Number of users: 943
Number of items: 1664
Number of ratings: 99693
Overall average rating: 3.53 out of 5, and std dev of 1.13
Average item rating: 3.08 out of 5, and std dev of 0.78
Average user rating: 3.59 out of 5, and std dev of 0.44
User-item Matrix Sparsity: 93.65%
Average number of ratings per users: 105.718982, and std dev of 100.567291
Max number of ratings per users: 736
Min number of ratings per users: 19
Median number of ratings per users: 64.000000

Popular items – most rated		
Title	No. of Ratings	Average Rating
Star Wars (1977)	583	4.36
Contact (1997)	509	3.80
Fargo (1996)	508	4.16
Return of the Jedi (1983)	507	4.01
Liar Liar (1997)	485	3.16

Popular items – highest rated		
Title	Average Rating	No. of Ratings
Entertaining Angels: The Dorthy Day Story(1996)	5.00	1
Someone Else's America (1995)	5.00	1
Aiqing wansui (1994)	5.00	1
Star Kid (1997)	5.00	3
Marlene Dietrich: Shadow and Light (1996)	5.00	1

Summary of Statistics For ML-100k Dataset

DAVIDSON