Homework 4

Information, Impacts and Insights

E-FRI

Asha Mairh

Jeet Patel

Kavish Hukmani

Keshore Suryanarayanan

Neon Zhang

11-22-2021

Contents

1 Executive Summary

This report details a recommender system created as a proof of concept with a view of improving the firm's customer engagement and thereby retention. The system is developed using a survey dataset of ratings provided by 280 users for 50 movies. 3 collaborative filtering techniques are used: user-based filtering, item-based filtering and matrix factorization. Since matrix factorization accounts for both user bias and item bias, it is the collaborative filtering technique of choice for the rest of the report.

To determine the efficacy of the system, it is used to make predictions of ratings for existing movies in the dataset, by both current users and new users. The cold start predictions for new users are then used as a benchmark to compare the rating predictions made for the same movies, after including a few data points for the "new" users.

The recommender system is limited by response bias to the survey, sampling bias in the movies (all are Oscar winners), and the sparsity in responses due to a lot of movies being unwatched by many users. Absence of any user and movie metadata prevents inclusion of content-based filtering which would have further enhanced the recommender system.

It is finally recommended that the firm does incorporate a recommender system on their Over-the-top (OTT) platform, with a combination of collaborative filtering and content-based filtering to have multiple other facets in addition to ratings for clustering similar users and movies.

2 Introduction

We live in a world where we want to be told what we would like to watch or purchase. From searching on Google to watching on Netflix and purchasing on Amazon, we rely on recommendations. Therefore, for any firm that runs an OTT channel, recommender systems are instrumental in improving the user experience, engagement and thereby retention. Right now, the firm runs an OTT channel without a recommender system.

This report pitches a proof of concept of a recommender system developed by the analytics team. Different methods of collaborative filtering are explored and compared to arrive at the best performing model, which is then used to predict ratings of both existing and new users for certain movies. Based on the results, there are recommendations made on why recommender systems, in particular this one, are essential for the company.

3 Problem Formulation

Broadly, we do the analysis based on the following framework: build 3 recommender system models leveraging the survey data, choose the best model by accounting for biases, make cold start predictions of new users for specific movies, and re-create the predictions for the same users after inputting a few data-points for them. The idea is to iteratively develop the model and test it for existing and new users.

Using the available ratings data, we develop 3 different models to predict the ratings given by 5 users for 6 movies - Inception, Mad Max: Fury Road, The Social Network, Winter's Bone, A Serious Man and Son of Saul. The three models developed are: 1. User-based Collaborative Filtering 2. Item-based Collaborative Filtering [Sarwar et al., 2001] 3. Matrix Factorization. User-based filtering predicts ratings of users based on the ratings given for that movie by users with similar taste. Item-based filtering predicts the rating for that movie based on ratings received by "similar" movies, the similarity being identified by comparing the ratings received by that movie from other users, and those received by other movies in the dataset. Finally, Matrix Factorization takes the entire user-item matrix and decomposes it into 2 matrices, the product of which provides the predicted ratings for the user-movie combinations.

We arrive at the best model based on how the user and item biases are accounted for. Subsequently, we explore different approaches, both within and outside the scope of the dataset, to predict ratings of new users for a specific set of movies (called as a "cold start") [Lika et al., 2014].

Finally, we again predict ratings for these users and movies after inputting additional information of ratings given by these 3 users for the movies Precious, 12 Years a Slave, Mad Max: Fury Road, Black Swan and Toy Story 3.

4 Data Characteristics

The movie rating data consists of ratings from 280 users for 50 movies. The users are from 2 sources - 96 UC Davis MSBA students and 184 unnamed users. The rating score stretches from 1 to 5 indicating the least complimentary to the most, and includes NULL representing never watched by the user. In addition to the existing users, 3 users (Amy, Shachi and Camille) were included to make cold start rating predictions.

5 Model Development, Estimation and Results

Among the 3 collaborative filtering techniques experimented with, the user-based one has the lowest error, which is attributed to selection bias in the test dataset amplified by its small size. Since, intuitively, the Matrix Factorization model allows us to explain and make predictions without any user or item bias, we choose to use it for our recommendation system.

Table 1 contains the predicted movie ratings for our group members, that turned out to be very close to the actual ratings. Jeet is the kindest and Kavish is the harshest in rating.

User	Movie	Original	Predicted		
			User-based	**Item-based**	**MF**
Kavish	The Social Network	No Rating	3.9831	3.9384	3.7108
	Mad Max: Fury Road	No Rating	3.3223	3.6461	3.4481
	Inception	4	4.6628	3.6461	4.2112
Jeet	The Social Network	5	4.8654	4.1116	4.3565
	Mad Max: Fury Road	4	4.0907	3.9807	4.0825
	Inception	5	4.7749	4.7007	4.8499
Neon	The Social Network	No Rating	4.0989	4.1291	3.9500
	Mad Max: Fury Road	No Rating	3.8807	3.7849	3.6970
	Inception	5	4.609	4.5863	4.4570
Keshore	The Social Network	4	4.2933	4.1168	3.9759
	Mad Max: Fury Road	3	3.4304	3.9005	3.7004
	Inception	5	4.8222	4.3886	4.4816
Asha	The Social Network	4	4.3813	4.1168	4.230101
	Mad Max: Fury Road	5	4.5297	3.9916	3.973779
	Inception	5	4.8478	4.6898	4.743422

Table 1: Original vs Predicted Ratings

For Matrix Factorization, we use the Singular Value Decomposition (SVD) algorithm [Salakhutdinov and Mnih, 2008] [Hug, 2020], which allows us to decompose our user-item rating matrix into two smaller matrices, such that their product gives us the original matrix. These two latent matrices are used to infer the user and item biases. To avoid overfitting our SVD model, we use k-fold cross-validation to find the optimal number of epochs for training and size of the latent vectors.

After choosing this model, as a proof of concept, we predict ratings for the following three movies, not included in the initial survey but included in the DBMI data. These ratings can be seen in Table 2.

While this model works well, upon the joining of new users, we do not have any data to use. To tackle this cold start problem, we remove the respective user biases in the available ratings for the selected movies and then calculate the average rating, which provides the true

User	A Serious Man	Son of Saul	Winter's Bone
Asha	4.0552	4.1399	4.3041
Jeet	4.1740	4.2482	4.4206
Kavish	3.5088	3.6469	3.7771
Keshore	3.7246	3.8713	4.0427
Neon	3.7064	3.8628	4.0193

Table 2: New Movie Predicted Ratings

rating of the movies without any user biases. Ideally, the average rating should be added with the respective user biases of Amy, Shachi and Camille, but since we do not have any data on them, we assume their user biases to be 0. The ratings present in Table 3 compare the before and after of the inclusion of their data-points, giving a benchmark to compare the cold start predictions with.

User	Avatar		The Wolf of Wall Street		Inception	
	Before	**After**	**Before**	**After**	**Before**	**After**
Amy	3.9989	3.996	4.0324	4.4148	4.4181	4.034
Camille	3.9989	3.998	4.0324	4.4205	4.4181	4.0328
Shachi	3.9989	3.998	4.0324	4.4205	4.4181	4.0328

Table 3: New User Predicted Ratings

After adding the new data points, the ratings for Avatar and Inception have decreased by 0.04% and 8.71% respectively, whereas those of The Wolf of Wall Street have increased by 9.58%. The changes are because of incorporation of user bias and alterations in item bias due to the new data points.

6 Recommendations and Managerial Implications

Based on the results, we see that collaborative filtering effectively predicts ratings for a given user-item pair, provided the users have rated at least one movie. In case of new users, existing movie rating data could be used to recommend content for them by discounting the individual user bias from the movie ratings. Recommending a combination of highly rated movies and the most popular movies could be a good starting point. Age, gender, geographical demography, etc. could be factored in to refine the above recommendations. Alternatively, while onboarding new users, data on their preferred genres could be collected and used in conjunction with the suggested approach to provide more relevant recommendations. Over time, search history and users' interaction with the content could be used as supplementary information to refine the recommendations provided.

Providing good recommendations to users could lead to an increase in customer engagement, customer retention and thereby in customer lifetime value[Jannach and Jugovac, 2019]. In music recommendations, [Domingues et al., 2013] have established that using both content and usage based recommendations has resulted in 50% higher activity levels among the users, than using the individual strategies alone. Despite the empirical evidence of the value added by the recommendation systems, estimating the precise monetary benefit turns out to be a non-trivial task because of the longitudinal effects associated with the process that are difficult to assess. To summarise, no single approach is better than the other, and recommendations using a combination of both collaborative and content based filtering have provided promising results [Tewari et al., 2014].

7 Conclusion

We devise a recommender system using the survey data of movie ratings, adopting matrix factorization as the final technique since it accounts for both user bias and item bias. We use the model to predict ratings for existing users, make cold start predictions for new users, and compare those predictions with those made after obtaining a few moving ratings from the same users. The efficacy of the system in making these recommendations advocate for the incorporation of this system onto the OTT platform.

The analysis is limited by the sparsity in the dataset. Also, cold start predictions with collaborative filtering are limited in their accuracy and content-based filtering (genre, director for the movie and demographics for the user) could make these predictions more accurate. Furthermore, all the movies in the dataset are Oscar winners and in general highly rated, causing sampling bias in the training data.

These limitations could be addressed by incorporating metadata on users and movies since these pieces of factual information provide more facets to identify "similar" movies and users. Lastly, by including in the dataset movies that are not critically acclaimed, the recommender system could better predict ratings for any new movies of similar nature in the dataset.

References

[Domingues et al., 2013] Domingues, M. A., Gouyon, F., Jorge, A. M., Leal, J. P., Vinagre, J., Lemos, L., and Sordo, M. (2013). Combining usage and content in an online recommendation system for music in the long tail. *International Journal of Multimedia Information Retrieval*, 2(1):3–13.

[Hug, 2020] Hug, N. (2020). Surprise: A python library for recommender systems. *Journal of Open Source Software*, 5(52):2174.

[Jannach and Jugovac, 2019] Jannach, D. and Jugovac, M. (2019). Measuring the business value of recommender systems. *ACM Transactions on Management Information Systems (TMIS)*, 10(4):1–23.

[Lika et al., 2014] Lika, B., Kolomvatsos, K., and Hadjiefthymiades, S. (2014). Facing the cold start problem in recommender systems. *Expert Systems with Applications*, 41(4, Part 2):2065–2073.

[Salakhutdinov and Mnih, 2008] Salakhutdinov, R. and Mnih, A. (2008). Probabilistic matrix factorization. In *Advances in Neural Information Processing Systems*, volume 20.

[Sarwar et al., 2001] Sarwar, B., Karypis, G., Konstan, J., and Riedl, J. (2001). Item-based collaborative filtering recommendation algorithms. In *Proceedings of the 10th international conference on World Wide Web*, pages 285–295.

[Takács et al., 2008] Takács, G., Pilászy, I., Németh, B., and Tikk, D. (2008). Matrix factorization and neighbor based algorithms for the netflix prize problem. New York, NY, USA. Association for Computing Machinery.

[Tewari et al., 2014] Tewari, A. S., Kumar, A., and Barman, A. G. (2014). Book recommendation system based on combine features of content based filtering, collaborative filtering and association rule mining. In *2014 IEEE International Advance Computing Conference (IACC)*, pages 500–503.

A Appendix I - Models

In User-Based Collaborative Filtering recommendation system, the average of the similar users' ratings is used as a predictor indicating a certain customer's rating. This similarity is represented by the cosine similarity represented in equation 1, to portray the resemblance of each pair of users. Considering the different impact that variate degree of similarity has on the average, the approach where the rating of the most similar user matters more than the second most similar user is adopted by choosing the top 10 most similar users and using following function to add weights to the ratings so that the heavier the weight, the more the rating would matter.

$$similarity(x, y) = \frac{\sum_{i=1}^{n} x_i y_i}{\sqrt{\sum_{i=1}^{n} x_i^2 \times \sqrt{\sum_{i=1}^{n} y_i^2}}} \tag{1}$$

Similarly, in the Item-Based Collaborative filtering recommendation system, the average of the similar items' ratings is used as a predictor indicating a certain item's rating.

In the Matrix Factorization based approach, we used Single Value Decomposition (SVD), an algorithm popularized by its use in the Netflix Prize competition [Takács et al., 2008].

The prediction $\hat{r}_{ui}$ is a function of the average rating μ, user bias b_u, item bias b_i and the product of the latent product and item matrices p_u and q_i, as seen in equation 2

$$\hat{r}_{ui} = \mu + b_u + b_i + q_i^T p_u \tag{2}$$

To estimate the unknown, we minimize the error equation 3 using Stochastic Gradient Descent (SGD)

$$\sum_{r_{ui} \epsilon R_{train}} (r_{ui} - \hat{r}_{ui})^2 + \lambda(b_i^2 + b_u^2 + ||q_i||^2 + ||p_u||^2) \tag{3}$$

We can see the Root Mean Square Error(RMSE) of the test set in table 4

Model Type	**RMSE**
User-based Rating	0.3519
Item-based Rating	0.5698
MF Rating	0.4968

Table 4: Root Mean Square Errors for the various models

B Appendix II - Paths for Cold Start Recommendations

To tackle the cold start problem, we develop three methods to provide recommendations (and not predict ratings) to new users based solely on the data available to us -

1. Recommend highly rated items - We identify the movies that are highly rated, i.e. movies with the highest values of item bias and recommend them to new users.

2. Recommend most watched items - We identify the movies that are the most watched based on the available dataset and recommend them to new users.

3. A combination of the two - We devise a metric that takes the product of A) item bias and B) min-max normalized version of number of users who watched a movie.

The results have been summarized in table 5.

Movie Name	**Popularity Score**	Item Bias	Combined Score
Inception	0.9136	0.4371	0.3993
The Imitation Game	0.6749	0.2014	0.1359
The Grand Budapest Hotel	0.5144	0.1613	0.0830
Inglorious Basterds	0.3992	0.2016	0.0805
The King's Speech	0.5514	0.1309	0.0722

Table 5: Recommendations for New Users

To tackle the cold start problem with new movies (recommendations and not ratings), an issue which is not addressed in the main report, we develop the following methods purely based on the data available to us:

1. Based on the available information, we recommend the new movies to users who have high user bias, i.e. they tend to rate movies highly in general.

2. We recommend the new movies to users who have watched a lot of movies. They are likelier to watch another movie if they tend to watch a lot of movies in general.

3. We devise a metric similar to the combined score in the cold start for new users piece, which accounts for user bias and number of movies watched by a user.

In addition to the available data, we could add metadata of users and movies such as demographic information, user preferences, movie genres, directors and actors, enabling us to do content-based filtering, which are particularly handy in handling cold starts.