Movie Rating Prediction: Implementation of Collaborative Filtering

Report for BAX-401 Business Analytics

Group 3

(Alphabetical)

Nasra Abdi

Xianglu He

Jiangfan Jing

Nora Luo

Xingxuan Zhang

Table of Content

Executive Summary

Collaborative filtering is a recommender system that produces personalized recommendations and has the effect of leading users towards their preferences. Its main goal is to recognize commonalities between users and create recommendations based on user or item comparison. The present report sought to predict movie ratings of 98 UC Davis students and faculties in the MSBA program by using collaborative filtering system to recommend movies. Moreover, it portrayed the limitations and the business values of using this recommender system.

The goal was to predict group three's rating for any three movies of choice, predict rating for three unrated movies for the group above, predict ratings for new customers, predict new customer's rating using added data about movies. The overall results showed that Group three gives high ratings. Cold start problems showed up as a limitation for the group three's unrated movies, and the new customers because there was no historical data. Without prior information, collaborative filtering does not provide predictions. In the cases above, average ratings were used to make predictions. Scenarios change when the new customer's preferences for five new movies is provided. The new results showed accuracy by altering average ratings. It is important to generate accurate predictions because it helps companies provide quality service to consumers.

The report recommends the following alternative to minimize cold start problems: use movie genre to find similar movies, identify possible relationship between rating and favorite actors, and gather initial information at the registration stage or other websites or platforms to help understand the consumer (demographic, pop-up surveys & movie preferences).

Introduction

Most internet based companies, such as Netflix and Amazon, utilize the recommendation systems to predict customers' preferences and provide personalized recommendations to promote consumption. Collaborative filtering (CF) is a technique used in the recommendation system based on an assumption that similar users would have similar preferences (Goldberg et al, 1992). In this report, we collected movie rating data from 98 students at UC Davis and performed a recommendation system to predict movie ratings for certain students and faculties in MSBA program. The report will also talk about the limitation and business value of the recommendation system.

Problem Formulation

There are four questions we plan to solve: (1) predict personal movie ratings (A Prophet, Amour, and Call Me By Your Name) for the students in group 3 of section 1; (2) predict the ratings of 3 unrated movies for the selected students, Winter's Bone, A Serious Man and Son of Saul specifically; (3) predict three new faculties' movie ratings for Avatar, The Wolf of Wall Street, and Inception; and (4) predict the three new users' movie ratings of Avatar, The Wolf of Wall Street and Inception given their ratings toward 5 movies.

Data Description

The analysis is based on the movie ratings generated by 98 MSBA students at UC Davis. 50 movies were rated according to a Likert scale from 1 (Bad) to 5 (Excellent). As we can see from Appendix A, majority of the ratings were in the range between 4 to 5. However, most of the 50 movies were unrated.

Model Development

Before predicting the ratings, we tested 18 collaborative filtering models (Appendix B). One certain model stands out as illustrated in Appendix C. Basically speaking, it is a user-based collaborative filtering model with specific parameters. It follows a core assumption that users would have similar preferences with other similar users and this is assumed to be constantly true. So, we first measure the similarities among users and then predict specific user's rating based on the existing ratings of similar users.

Results

By using the collaborative filtering model, the predicted ratings of the selected movies in question 1 were illustrated in Appendix D. Based on the results, we can see Jiangfan is the kindest person in rating while Nora is the harshest. However, the difference is small and all individuals in this group tend to give high ratings around 4 out of 5.

The cold start problem came up in question 2 (Appendix E) and 3 (Appendix F). This occurs when new users or new items come in, and there is no previous information. Finding similar users or movies to predict the ratings cannot work in this case. Thus, we use users' average ratings and movies' average ratings as our simple rules-of-thumb for prediction.

Additional information of these 3 new users were provided in question 4. Therefore, by incorporating their preferences in the 5 movies, more accurate predictions could be made. As shown in Appendix G, Avatar's average rating is 3.95, but Camile is a harsh person who often gives 2 or 3 points to movies. Without Camile's historical data, we thought Camile would rate 3.95 for Avatar. However, after knowing Camile's habit, the model predicted that Camile would rate Avatar as 2.92, significantly lower than the average of 3.95. This also demonstrated the negative effect of the cold start problem, producing not accurate predictions.

Recommendations and Implications

Recommendation systems are significantly valuable for companies and users. Research shows that these recommendations would effectively improve user conversion rate, customer loyalty and cross-selling (Schafer et al, 2001). By solving problems similar to 1 and 4, we help companies understand their consumer's preference and consequently deliver higher quality of service or product by personalizing recommendations. Companies use this predictive model to help consumers navigate the abundant options by presenting what they are likely to enjoy. It minimizes the complexity of the process, and it facilitates successful transactions. There is a great added value for businesses because a satisfied consumer is more likely to become loyal.

Beyond movie ratings, other web-based businesses such as ecommerce and mobile commerce in general also use CF to understand buyer behavior, and to segment them into groups that share similar purchase habits. For instance, they match a purchase habit of one customer to other similar customers, or based on purchase history from their web browser. They know what consumer's interests are and make recommendations that entices them to buy. Retailers go as far as advertising outside their website to attract these buyers. Moreover, CF helps minimize balance assortments issues for retailers due to the personalized recommendations which can allow retailers to improve their assortment planning by showing relevant item to each consumer, meeting consumer demand and ultimately maximize profit.

Cold-start problem is a well-known issue that happens when there are no prior data, such as ratings, are known for certain users or items (Gantner 2010). When solving question 2 and 3, we were facing the cold start problem for new items and new users. To better predict ratings for new movies and new users, more data need to be collected.

For the new movies, we can use the movie genre to find similar movies, and by doing so, we can predict the rating more accurately. We can also study the relationship between ratings and actors since people may rate movies higher when they see their favorite actors. We can also do a survey for the new movies by randomly inviting people to watch the movies and asking for ratings. For the new users, we can collect demographic information from the new users, such as their age and gender, assuming that the same demographic group may have similar preferences. We can ask for new users' preferences by doing a pop-up survey when they are registering. Or, if possible, we can try to obtain users' behavior data from other websites or platforms, such as social media. More solutions for cold start problem are shown in Appendix H.

The model we used in this report contains some limitations, such as the item bias (Appendix C). In recent research, there are some methods can improve the collaborative filtering model we used (Appendix I).

Conclusion

We built up a simple collaborative filtering recommendation system and experienced the cold start problem. The cold start problem can be solved with more information and data collected from the new users and new items. The recommendation system can not only work in movies but also perform well in other industries, such as e-commerce. Amazon launched its item-based collaborative filtering recommendation system in 1998 (Smith 2017) and used this algorithm across the websites, such as its homepage recommendations. According to McKinsey (2013), 35% of Amazon's revenue is generated by its recommendation system. Thus, it is worthy to build an effective recommendation system for the company.

Reference

Goldberg, D., Nichols, D., Oki, B.M., Terry, D. (1992) "*Using Collaborative Filtering to Weave an Information Tapestry*", Comm. ACM, Vol.35 (1992), pp. 61 - 70.

Schafer JB, Konstan JA, Riedl J (2001). "*E-Commerce Recommendation Applications*." Data Mining and Knowledge Discovery, Vo.5(1/2), pp.115–153.

MacKenzie, I., Meyer, C., and Noble, S. (2013). "*How retailers can keep up with consumers*". Available at: https://www.mckinsey.com/industries/retail/our-insights/how-retailers-can-keep-up-with-consumers. [Retrieved December 1, 2019]

Smith, B., and Linden, G. (2017). "*Two Decades of Recommender Systems at Amazon.com*". IEEE Internet Computing, Vol.21(3), pp. 12-18.

Nguyen, H.T, Mary, J., and Preux, P. (2014) "*Cold-start Problems in Recommendation Systems via Contextual-bandit Algorithms*", Available at: https://arxiv.org/pdf/1405.7544.pdf

Isinkaye, F., Folajimi, Y., and Ojokoh, B. (2015). "*Recommendation systems: Principles, methods and evaluation*". Egyptian Informatics Journal, 16(3), 261–273. doi: 10.1016/j.eij.2015.06.005

Appendix

Appendix A: Data Exploration

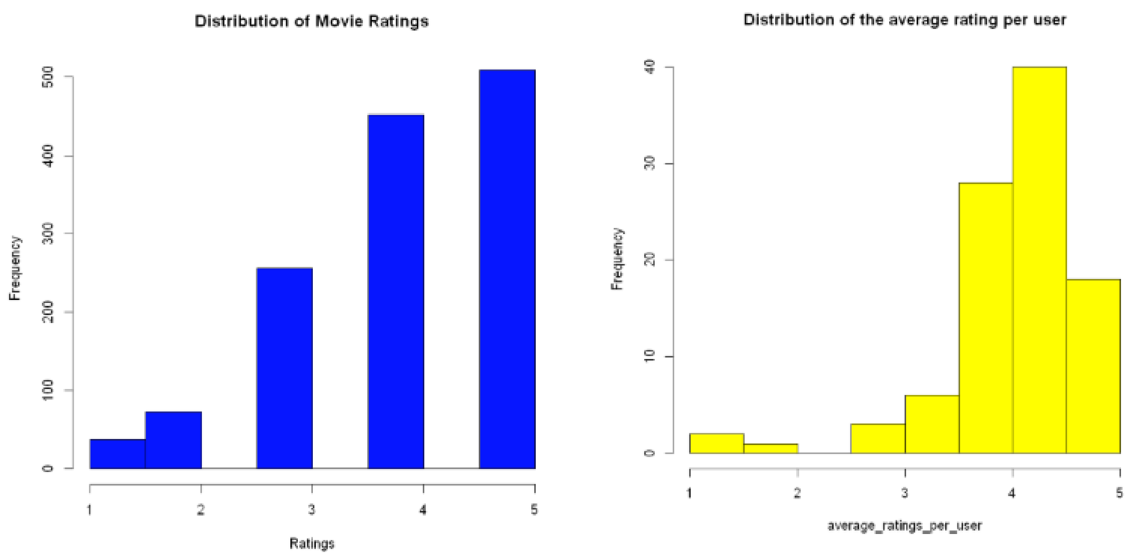


Figure 1. Distributions of Movie Ratings

The data were collected from 98 MSBA students at UCD. The distribution of movie ratings shows that most ratings are 5 or 4. The distribution of the average rating per user also shows that most students' average ratings are concentrated at 4 or 5 points.

Appendix B: Test of 18 Collaborative Filtering Models

	RMSE	MSE	MAE
UBCF_C_P	0.7984235	0.6374801	0.6882476
UBCF_Z_P	0.7984322	0.6374941	0.6898722
UBCF_Z_C	1.0316313	1.0642632	0.8391653
UBCF_C_E	1.0316735	1.0643502	0.8472380
UBCF_C_C	1.0324599	1.0659735	0.8453428
UBCF_Z_E	1.0340969	1.0693564	0.8473650
IBCF_N_E	1.0773371	1.1606553	0.8213670
IBCF_N_C	1.0855458	1.1784096	0.8346155
IBCF_N_P	1.0919247	1.1922996	0.8502663
IBCF_Z_E	1.1135451	1.2399827	0.8663676
IBCF_C_E	1.1198831	1.2541381	0.8491322
IBCF_C_P	1.1657573	1.3589902	0.9246453
IBCF_Z_P	1.2059010	1.4541971	0.9777855
IBCF_C_C	1.2498311	1.5620779	0.9629664
IBCF_Z_C	1.2508593	1.5646490	0.9609594
UBCF_N_P	2.4904894	6.2025372	2.1835239
UBCF_N_E	2.5622408	6.5650781	2.1757835
UBCF_N_C	2.6099730	6.8119588	2.2548503

Figure 2. Results of the Model Test

Where

- IBCF stands for Item-Based Collaborative Filtering, which recommends an item because you like other similar items.
- UBCF stands for User-Based Collaborative Filtering, which recommends an item because people like you purchase this item.
- The second letter:
    - _N_ stands for no adjustment
    - _C_ stands for mean-centered adjustment
    - _Z_ stands for Z-score adjustment
- The third letter:

- E stands for Euclidean distance
- C stands for Cosine Similarity
- P stands for Pearson Correlation

- Measurements for error:

Mean squared error	$\mathrm{MSE} = \frac{1}{n}\sum_{t=1}^{n} e_t^2$
Root mean squared error	$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n} e_t^2}$
Mean absolute error	$\mathrm{MAE} = \frac{1}{n}\sum_{t=1}^{n} \lvert e_t \rvert$

Excluding group 3 in section 1 and three new faculties, we trained and tested 18 models that using different calculations for similarity on the rest of the dataset. We used three measurements for model error (MSE, RMSE, and MAE) to test the quality of the models. The test result shows that the user-based collaborative filtering with a mean-centered Pearson correlation adjustment model performs the best job with minimum errors. More details about the test can be found in https://cran.r-project.org/web/packages/recommenderlab/vignettes/recommenderlab.pdf

The models we tested is limited within memory-based collaborative filtering. However, there are more techniques can be used in recommendation systems. For example, the model-based collaborative filtering employs some machine learning techniques, such as Support Vector Machine (SVM) and Neural Networks (Isinkaye, 2015).

Appendix C: User-Based Collaborative Filtering - Mean-Centered Pearson Correlation

The user-based collaborative filtering based on an assumption that users have similar preferences with other similar users. The Pearson correlation is a method we used to measure how similar the two users are. The formula of the Pearson correlation is shown below.

$$s(a,u) = \frac{\sum_{i=1}^{n}(r_{a,i}-\overline{r_a})(r_{u,i}-\overline{r_u})}{\sqrt{\sum_{i=1}^{n}(r_{a,i}-\overline{r_a})^2}\sqrt{\sum_{i=1}^{n}(r_{u,i}-\overline{r_u})^2}}$$

where s(a,u) is the similarity between two users a and u, $r_{a,i}$ is the rating given to item i by user a, $\overline{r_a}$ is the mean rating given by user a while n is the total number of items in the user-item space.

Figure 3. Pearson Correlation

Prediction is a weighted average of the movie rating on the similar users (Isinkaye, 2015), and the formula is shown in Figure 4.

$$p(a,i) = \overline{r_a} + \frac{\sum_{i=1}^{n}(r_{u,i}-\overline{r_u}) \times s(a,u)}{\sum_{i=1}^{n} s(a,u)}$$

Figure 4. Prediction Formula

However, there are 2 certain biases in recommendation systems. One is user bias which means certain users are inclined to give high ratings than others. The mean-centered adjustment considered users' bias that some people like to provide high scores while some people like to give low scores.

The other one is item bias which indicates the difference in average ratings among movies. For example, 3 stars for a successful movie is quite different from the 3 stars for a bad movie. The selected model does not address the item bias.

Appendix D: Results - Question 1

Question 1: Predict Group 3 in Section 1's Ratings for Three Movies

Serial Number	**First Name**	**Last Name**	**Call me by your name**	**A Prophet**	**Amour**
6	Xingxuan	Zhang	4.138959	3.989832	4.011330
26	Nan	Luo	4.055472	3.961750	3.977047
50	Xianglu	He	4.195401	4.107223	4.138832
55	Jiangfan	Jing	4.328772	4.217288	4.247349
85	Ashna	Sandhu	4.138393	4.018787	4.130601

The three movies we predicted are Call Me by Your Name, A Prophet, and Amour. The result shows that Jiangfan is the kindest person in group 3, but all people's ratings are close to 4 out of 5.

Appendix E: Results - Question 2

Question 2: Predict Group 3 in Section 1's Ratings for Three Movies That Do Not Have Historical Rating Yet

Serial Number	First Name	Last Name	Winter's Bone	A Serious Man	Son of Saul
6	Xingxuan	Zhang	4	4	4
26	Nan	Luo	4	4	4
50	Xianglu	He	4.090909091	4.09090909	4.09090909
55	Jiangfan	Jing	4.230769231	4.230769231	4.230769231
85	Ashna	Sandhu	4.137931034	4.137931034	4.137931034

Since we do not have historical rating data for Winter's Bone, A Serious Man, and Son of Saul, the collaborative filtering models cannot work well (most of them returns 0s). The model we used in question 1 would return the average ratings of each person for these three movies. Before we obtain more information about the movies, it is reasonable to predict that people would rate these movies at their average ratings.

Appendix F: Results - Question 3

Question 3: Predict Three New Faculties' Ratings for Three Movies

Serial Number	**First Name**	**Last Name**	**Inception**	**Avatar**	**The Wolf of Wall Street**
99	Shachi	Govil	4.48	3.95	4.04
100	Amy	Russell	4.48	3.95	4.04
101	Camille	Mack	4.48	3.95	4.04

Similar to question 2, we predicted three new users' ratings for Inception, Avatar, and The Wolf of Wall Street. Before we obtain more information about the three faculties, it is reasonable to predict that people would rate these movies at the movies' average ratings.

Appendix G: Results - Question 4

Question 4: Predict Three Faculties' Rating for Three Movies After Receiving Their Ratings For Other Five Movies

Serial Number	**First Name**	**Last Name**	**Inception**	**Avatar**	**The Wolf of Wall Street**
99	Shachi	Govil	4.669794	4.195536801	4.3131363
100	Amy	Russell	3.442421	2.887575774	3.158912996
101	Camille	Mack	3.289055	2.917554876	2.882770184

After having three faculties' ratings for other five movies, we can use the collaborative filtering model to predict. The result is different from question 3's. According to those five movies' ratings, Camille is the harshest person who likes to rate movies with low scores. The result from the model adjusted her habit when predicting the ratings.

Appendix H: Cold Start Problem

Recommendations system might not be accurate for new customers without knowing their previous preference, like question three in this report. To solve this cold start problem, 4 approaches could be used.

(1) Representative based approach - Just like question four, we can ask the new users to first rate a small portion of items and then make predictions based on them. Or we can use the preference of a similar subgroup to infer their preferences, such as using the group of females who worked in higher education institutions to address question three.

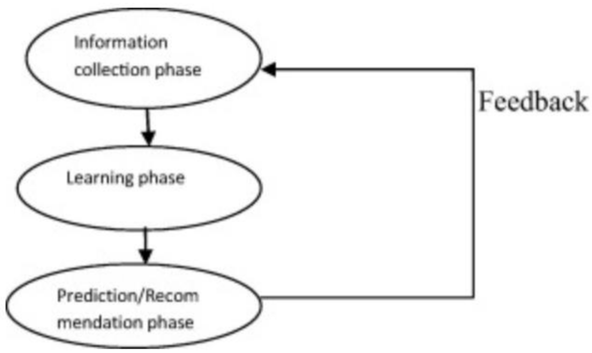
(2) Content based approach - we can use their social media to infer their preference for items through data mining their text content. Nowadays, users are encouraged to use their social media account to login because it allows for the collection of their metadata.

(3) Deep learning approach - build a deep network to incorporate every information as features to predict the preferences of new users.

(4) Moreover, if we do not have data for new movies, there is also an item cold start problem. This can be remedied by bandit approach - we can do experiments to a small portion of users and then iterate the process the find the true preference (Nguyen et al, 2014).

Appendix I: Methods to Improve the Model

According to Isinkaye et al (2015), the recommendation process has the following phases. (1) Information collection phase, including explicit feedback, implicit feedback and hybrid feedback, (2) learning phase, (3) prediction/recommendation phase. A graph is also illustrated below.



For the information collection phase, we can use how sparse the rating matrix is across all users and all items. The less sparse rating matrix is, the more information we can obtain, thus higher quality the data have. So, we can use the sparseness of the rating matrix as a value function in this phase.

For the learning phase, all the algorithms begin with finding the similarities between users. This would be much more complex with the increasing of the number of users, say 100 million. So, the computation efficiency is a value function in this case. By employing Alternating Least Squares (ALS), matrix factorization can be performed to reduce computation time.

For prediction/recommendation phase, the root mean squared error (RMSE) is used as a value function to measure how accurate the prediction is. The less RMSE, the better the model will be. By using these 3 value functions as our criteria, we can objectively judge and refine the recommendation process.

Appendix J: Recommendation System Code in R

The code would be attached as a pdf file.