Building a Recommendation System

Using Collaborative Filtering

BAX-401-002 Group D

Ashi Malik

Jason Stergiou

Shushmita Mirazkar Subash

Thea Castañeda

Vinit Hasija

Table of Contents

Executive Summary

We were tasked with building recommendation systems for a movie streaming service company, and provided with data to use in various problem scenarios to present a business case for the use of recommendation systems. We established a recommendation system to predict ratings for existing (our group) and new users (Camille, Shachi, and Amy) using collaborative filtering and user-based and item-based approaches. We note that these are limited and other approaches can be taken by the company for future recommendation systems, though we find that recommendation systems would be good in lowering costs as well as improving customer experience by providing insight into potential customer response to movies within the streaming service's catalogue.

Introduction

In the use cases, we establish a recommendation system for a streaming service company to enhance user experience in the film arena by using a predictive model which recommends movies based on user ratings to improve the overall effectiveness and tailor response to each individual experience. Current data is available to create such predictions through the use of customer moving ratings. For the purpose of this report we used a Collaborative Recommender system, which is the method used to determine the best prediction for current customers by predicting which films they are likely to watch, in addition to new films and potential customers. Creating this predictive model to recommend new films to current and prospective users will enhance the customer experience and as a result, improve business value for the streaming company. Our objective is to justify the use of a Collaborative recommendation system and we assume that the streaming company currently recommends based on the average user ratings, which can create several bias and might not be as effective.

Data Description

As can be seen from the excel that there are a total of 51 movies with the respective movie ratings for 95 users. The user bias and item bias is removed using the mean centered and

z-score approach. Our aim is to check when the user wants to check the ratings based on the ratings of the other users and how would the user rate a particular movie based on the rating of the other movies. The absolute cosine similarity is calculated using the concept of weighted average of how similar the ratings of the user that is to be predicted is with the other users and how those users have rated the movie that we want to predict.

Problem Formulation

Firstly, our group aims to build a movie recommender system for a streaming platform and to justify the recommendation system by this use case. To begin with, based on the training dataset given to us we need to predict our group members ratings for three movies that none of the group members have seen, which are: *Zero Dark Thirty, Spotlight, and Up In The Air.*

We will use Cosine Similarity to determine similarities between the rating pattern of users, which we will ultimately use to predict the ratings for the above mentioned movies . We would use both user-based and item based techniques. As the user-based approach has issues with user bias since every user rates differently, we account for the biases using Mean Centered and Z-score approach, by calculating the difference between predicted ratings and the real ratings, and we chose the one with the least difference as our predicted result.

Secondly, based on the training dataset we will predict the ratings of our group members, for the following three movies that no user has seen: *Winter's Bone, A Serious Man, and Son of Saul.* This problem is similar to the first one, though instead of choosing the three movies for which we predict the ratings for they are pre-selected, and hence we will follow the same approach as we did in the first problem.

For the third problem, we will predict the rating for new customers, Camille, Shachi, and Amy, who joined the streaming service, though these are customers we do not have any past

information for. For the fourth problem, we obtained some ratings for the new users, Camille, Shachi, and Amy, we add these ratings to the train dataset we used before in problem 1 and 2 and then we will use the user ratings from the updated data set, to compare the ratings we predicted in problem 3 and 4 .

Prediction Results

Objective I. Existing Customer: Member Ratings for Zero Dark Thirty, Spotlight & Up in the Air

Table 1. Item-Item Z-Scored Prediction Rating

****Excel in Appendix***

Film	Thea	Shushmita	Jason	Vinit	Ashi
Spotlight	4.2265	4.408315258	4.24646384	3.925257	4.7166066
Up In The Air	3.785	4.032884431	3.70155137	2.597937	4.524466
Zero Dark Thirty	3.68255012	3.867123956	3.75893601	3.624945	4.0915328

Table 2. Item-Item Mean-Centered Prediction Rating

Film	Thea	Shushmita	Jason	Vinit	Ashi
Spotlight	3.693741244	3.8645822	3.7762	3.649567232	4.06959135
Up In The Air	3.8714705	3.82011144	3.621359844	3.621359844	3.97056994

Zero Dark Thirty	*4.186210926*	*4.3511029*	*4.288*	*4.276546729*	*4.49445651*

Table 3. Average Movie Ratings

DELETE?

Thea	*Shushmita*	*Jason*	*Vinit*	*Ashi*

Objective II. Movie Ratings: Winter's Bone, A Serious Man & Son of Saul

Table 3 Item-Item Mean-Centered Rating Prediction

	Winter's Bone	*A Serious Man*	*Son of Saul*
Thea	*2.892*	*3.897*	*2.321*
Shushmita	*3.1000*	*2.8631*	*2.7826*
Jason	*3.3471*	*3.454*	*2.8661*
Vinit	*3.7064*	*3.2326*	*2.9878*
Ashi	*2.7621*	*3.8788*	*2.331*

Table 4 Item-Item Z-Scored Prediction Rating

	Winter's Bone	*A Serious Man*	*Son of Saul*
Thea	*2.899*	*3.788*	*2.3457*
Sushmita	*3.1342*	*2.7898*	*2.667*
Jason	*3.238*	*3.451*	*2.1234*
Vinit	*3.698*	*3.2377*	*3.229*
Ashi	*2.799*	*3.7621*	*2.285*

Objective III. New Customer: Camille, Shachi, & Amy

&

Objective IV. New Customers Other Ratings: New Movies for Camille, Shachi, & Amy

Table 5 Item-Item Mean-Centered Prediction Rating

	Avatar	***The Wolf of Wall Street***	***Inception***
Shachi	*4.12*	*4.19*	*4.67*
Amy	*3.99*	*4.01*	*4.47*
Camille	*3.96*	*4.06*	*4.48*

Table 6 User - User Mean-Centered Prediction Rating

	Avatar	***The Wolf of Wall Street***	***Inception***
Shachi	*4.123*	*4.13*	*4.20*
Amy	*3.2*	*2.90*	*2.98*
Camille	*2.67*	*2.70*	*2.63*

Table 7 Item- Item Z-Score Prediction Rating

	Avatar	***The Wolf of Wall Street***	***Inception***
Shachi	*4.09*	*4.15*	*3.9*
Amy	*4.15*	*4.19*	*3.94*

Camille	*4.01*	*4.1*	*3.8*

Table 8 User - User Z-Score Rating Prediction

	Avatar	**The Wolf of Wall Street**	**Inception**
Shachi	3.19	2.75	2.88
Amy	4.12	4.16	4.20
Camille	2.64	2.70	2.6

Recommendation and Managerial Implications

Problem 1 provides value to the firm in using recommendation systems to engage existing customers, predicting their ratings when other customers have ratings for the movies. Problem 2 provides the firm with the ability to predict for existing customers when no other customers have ratings for the movies by using item-based predictions.

Problem 3 shows that recommendation systems allow for firms to strategize for new customers when there's no prior information on their preferences or behavior to increase chances of them subscribing to the streaming service. Problem 4 provides value in situations where potential customers, for whom there is prior information, may use the streaming service and gives predictions on what they might prefer to improve upon situations such as those in Problem 3.

Recommendation systems increase business value by decreasing costs through better prediction models. A more refined prediction model with more complete data, as well as running tests of accuracy on the predictions would help improve the model over time.

In cold start situations such as those in problem 3, the firm could use other approaches such as establishing a Matching Matrix. Further, the use of other information beyond ratings within one firm, such as other rating systems data for potential new users or demographic data would allow for different strategies in matching users by similarity and predicting unknown ratings.

The most obvious goal of a recommendation system is to recommend relevant products to the user. The recommendation system used is based on collaborative filtering of population built systems where we use the explicit data to suggest popular movies based on the average 5 star rating or using implicit data based on the number of times the movie has been played. Thus using these systems is advantageous because the user could start resistant and can suggest products without any information on the user. Additionally it can also be used in environments with a lesser number of users. Methods like content based filtering can be used to avoid any cold starts issues on item based profiles. Hybrid systems can be used that combine both content based and collaborative filtering methods. This solution is more effective than any of the two methods separately. The hybrid system is a result of both comparing the similar habits of similar users (collaborative filtering) and finding movies that have similar characteristics like

movies that users liked in the past(content based).

The business value a recommendation system has will allow the end result of improving customer benefit. This creates an impact which is key through the use of data analytics tools like predictive models. The ability to garner such methods improve revenue growth through new customers. Expense reduction by eliminating precious resources that involve individuals doing this in a cumbersome manner. In addition to risk mitigation that involves any business such as human error. This impacts the business model by making into a customer centric approach to benefit the consumer and business alike.

Conclusion

We attempted to build recommendation systems for a movie streaming service company using data from our class in various problem scenarios. We established a recommendation system to predict ratings for customers, namely ourselves and Camille, Shachi, and Amy using collaborative filtering and user-based and item-based approaches, and presented a business case for the use of recommendation systems. We suggest other recommendation system approaches for the company in the future, with more information. We do suggest the use of recommendation systems as they could lower costs to improve customer experience, as well as improving the system of the streaming service, increase user interaction along with attracting and retaining customers. Recommendation systems are excellent machine learning solutions which help increase customer satisfaction leading to a significant increase of your business revenue.