Implementing A Recommender System for BoardGameGeek

Introduction

BoardGameGeek is an online community of board game players. Classic games, such as "Monopoly", are listed; however, a myriad of lesser known games such as "Council of 4" are also available. Information is provided for each game including the game's ratings, number of players, description, type, category, etc. Some games have links to video tutorials on how to play. Within this community, you can buy and trade games from other users or find links to websites where you can purchase them.

Membership to this community is free, however, donations from users are encouraged. There are also ads on the margins of the webpages. One way to increase both these forms of revenue is to increase traffic to the website. A recommender system can help do this. It can also drive up sales of board games for which the website can charge a fee.

A recommender system predicts what rating a user would give an item or product. This information would then be used to suggest the higher rated items to the user for sales or other purposes. Recommender systems are used by many companies such as Spotify, Pandora, Netflix, and YouTube to create music or video playlists. Facebook, Instagram, and Twitter use recommender systems to suggest content and ads to their users. Amazon and eBay use it to suggest items that the user may want to purchase. In this capacity is how a recommender system can help BoardGameGeek increase traffic and sales. A user may not be aware of a game that they would be interested in. The recommender system would predict that the user may like the game and bring it to the user's attention. This increases the likelihood that the game will be purchased by the user. It also increases the likelihood that the user would return to the website for future board game purchases.

The DataSet

In order to build an accurate recommender system, data is required on all the board games listed on the website. Information about the users, the games, the games' ratings, etc. is provided by BoardGameGeek's API. Using Python code, the data will be downloaded from the API and wrangled into a dataset that can be used for analysis. This will be accomplished using the "xml.etree.ElementTree" Python tool. Later, sklearn tools will be used to build the recommender system from the dataset. The code and markup used to create the recommender system will be available on GitHub.

BoardGameGeek's API was used to scrape the board game data from the website. Each board game has a page with its characteristics, descriptions, users that rated the game, and the ratings given. Each page holds 100 ratings. Any additional ratings are recorded in subsequent pages. Some pages contain information of items that are not board games, such as expansion packs. These pages were ignored. Data for the first 50,000 pages was scraped for all the relevant board game data.

The API uses XML, therefore, xml.etree was used to parse the data. The following information was obtained for each board game: ID number, name, year published, minimum players, maximum players, playing time, minimum playing time, maximum playing time, board game category, board game mechanic, and board game designer. Once this information was obtained, the code assesses if there are ratings available for the board game. If there are, the code then assesses how many pages worth of ratings there are. Next, the code iterates through all those pages and retrieves all the ratings and users for that board game. This process is repeated until the code reads through the first 50,000 ID numbers, where each ID number corresponds to a unique item. Only those that for board games are stored in individual JSON files on the hard drive. This is wrangled data is now ready to be put in a data frame for analysis.

Exploratory Data Analysis

Exploratory data analysis, EDA, is one of the initial analyses that are performed on a dataset when searching for main trends, summarizing the data's characteristics, and using visual methods to get a better understanding of the data. John Tukey defined EDA as "Procedures for analyzing data, techniques for interpreting the results of such procedures, ways of planning the gathering of data to make its analysis easier, more precise or more accurate, and all the machinery and results of (mathematical) statistics which apply to analyzing data."

Since the goal of this recommender system is to provide a user with suggestions on similar games that are also favored by other users, the average rating for each game was the logical place to start when comparing board games. I checked to see if there were any games that had no ratings at all. There turned out to be 7,206 games that had no ratings. This came out to be almost 20% of the dataset. These board games were ignored when obtaining the mean and median average rating. The data then showed that even though 80% of the board games had ratings, a vast majority only had a handful or less ratings per game. Figure 1 below illustrates this finding.

```
Board games with 0 ratings: 7206 , 19.10 percent of the data
Board games with less than 2 ratings: 11792 , 31.25 percent of the data
Board games with less than 3 ratings: 15097 , 40.01 percent of the data
Board games with less than 4 ratings: 17427 , 46.19 percent of the data
Board games with less than 5 ratings: 19130 , 50.70 percent of the data
Board games with less than 10 ratings: 23731 , 62.90 percent of the data
Board games with less than 20 ratings: 27220 , 72.14 percent of the data
Board games with less than 30 ratings: 28890 , 76.57 percent of the data
Board games with less than 50 ratings: 30800 , 81.63 percent of the data
Board games with less than 100 ratings: 32978 , 87.41 percent of the data
```

Fig 1. Percentage of data per number of ratings

Since so much of the data had such few ratings. I compared the mean and median average rating and the mean and median number of ratings for each grouping. The data was grouped into games that had 1 or more ratings, less than five ratings, less than ten ratings, and less than 30 ratings. Figure 2 below shows the summary results.

```
Board games with 1 or more ratings:   Board games with less than 10 ratings:
Mean average-ratings: 5.455           Mean average-ratings: 5.873
Median average-ratings: 5.61          Median average-ratings: 5.93
Mean number of ratings: 223.182       Mean number of ratings: 482.67
Median number of ratings: 8.0         Median number of ratings: 48.0


Board games with less than 5 ratings: Board games with less than 30 ratings:
Mean average-ratings: 5.742           Mean average-ratings: 6.068
Median average-ratings: 5.82          Median average-ratings: 6.11
Mean number of ratings: 364.914       Mean number of ratings: 754.42
Median number of ratings: 26.0        Median number of ratings: 113.0
```

Fig 2. Summary results per group

The mean and median of the average ratings are close for all four groups. Since this is the case, it can be said that the dataset of board games with at least 30 ratings per game would reasonably reflect the games with fewer ratings.

Data visualization tools were used to illustrate characteristics of the data. The average rating box plot below, Figure 3, shows that most of the average ratings users gave the games are between 3.5 and 8, with a mean average rating of about 6. Outside these bounds there are many outliers, however, most of the games are rated within this range. The histogram in Figure 3 shows that the data is normally distributed. The QQ plot, in Figure 4 below, shows this as well. However, it can be seen in the histogram that the data is slightly left skewed. This left skewness is more clearly seen in the QQ plot.

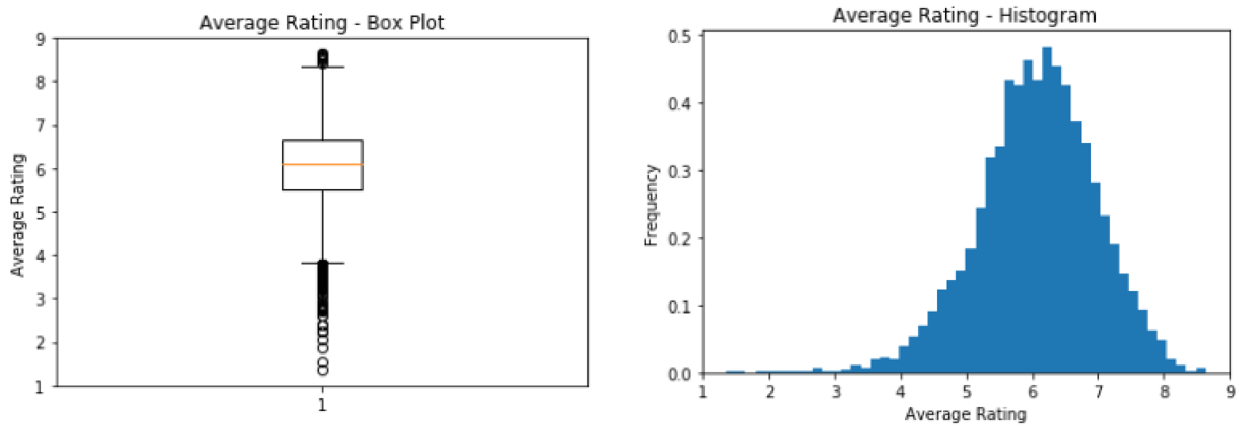

Fig 3. Box plot and Histogram

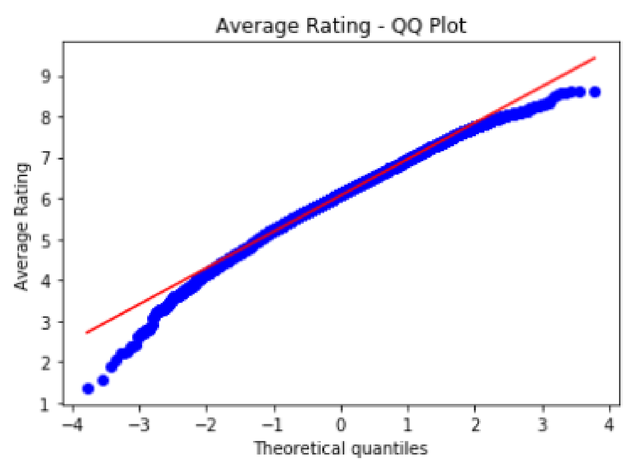

Fig 4. QQ plot

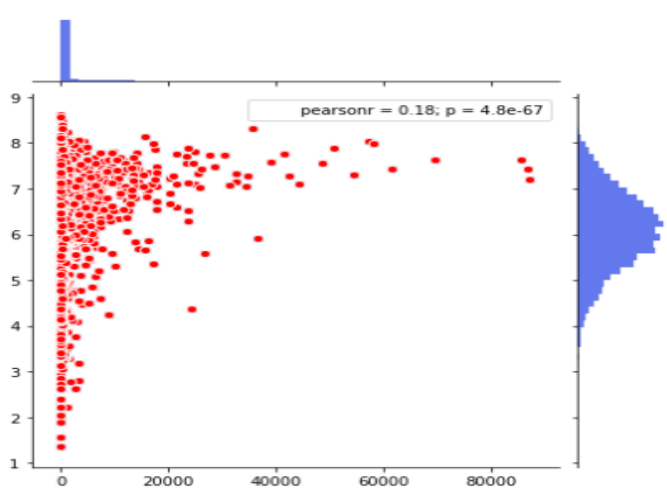

Fig 5. Joint plot

The joint plot in Figure 5 is a combination of the number of ratings vs. average ratings scatter plot, the histogram of the average ratings, and the bar graph of the number of ratings. The scatter plot shows the number of ratings vs the average rating of a board game. The plot shows that games with a higher number of ratings tend to receive a higher average rating. The histogram on the right demonstrates that the average, average-rating is around 6. The Bar graph on top illustrates that a vast majority of the board games have few ratings. The joint plot also contains a correlation analysis. The two variables are shown to have an r of .18 and a p-value of zero.

The correlation was then analyzed independent of the joint plot. First, the correlation between the average rating and the number of ratings was taken for games with at least 30 ratings. It was found to have a positive correlation of .182. The p-value for the correlation coefficient is zero, meaning that it is statistically significant. The correlation between the two variables was then taken without considering the number of ratings. Here, the correlation coefficient decreases slightly to .105.

The top twenty games in terms of average rating and number of ratings are shown in Figure 6 and 7 respectively. The top twenty includes only games that had at least thirty ratings.

```
                      RPGQuest: Greek Mythology
            The Battle of Fontenoy: 11 May, 1745
                                        RPGQuest
                                Connection Games
                      Prague: The Empty Triumph
             Sports Action Canadian Pro Football
Crusade and Revolution: The Spanish Civil War,...
                   Axis Empires: Totaler Krieg!
                              1844: Switzerland
                              Twilight Struggle
                 The Penguin Book of Card Games
                              Where Eagles Dare
                   RPGQuest: Oriental Adventures
                             Funkenschlag: EnBW
                            D-Day at Omaha Beach
                                            DAK2
                                       Case Blue
                                        Manassas
                          International Cricket
                                Face To The Mat
```

Fig 6. Top 20 Board games with the highest average rating

```
                 Catan
           Carcassonne
              Pandemic
              Dominion
        Ticket to Ride
              Agricola
           Puerto Rico
           Small World
            Power Grid
Ticket to Ride: Europe
              Citadels
                 Dixit
   Race for the Galaxy
             Stone Age
              Munchkin
     Twilight Struggle
          Arkham Horror
              Bohnanza
           Lost Cities
        The Resistance
```

Fig 7. Top 20 rated games

The list of games that received the greatest number of ratings has games that are widely popular and can be found at most toy and bookstores. It is interesting to note, however, that none of those games made it to the top twenty list of games by average rating. Evidently the number of ratings is not the only factor related to a game's average rating.

The correlation coefficient between average rating and number of ratings was .182. The correlation between the other attributes in comparison to average rating was calculated to search for any attributes that may be highly correlated. Figure 8 shows the attributes that had a positive or negative correlation greater than .1.

```
Correlation between a boardgame's average rating and bg mechanic Route/Network Building = 0.121
Correlation between a boardgame's average rating and bg mechanic Set Collection = -0.102
Correlation between a boardgame's average rating and bg mechanic Campaign / Battle Card Driven = 0.105
Correlation between a boardgame's average rating and bg mechanic Chit-Pull System = 0.100
Correlation between a boardgame's average rating and bg mechanic Hex-and-Counter = 0.278
Correlation between a boardgame's average rating and bg mechanic Area Control / Area Influence = 0.113
Correlation between a boardgame's average rating and bg mechanic Roll / Spin and Move = -0.279
Correlation between a boardgame's average rating and bg mechanic Dice Rolling = 0.159
Correlation between a boardgame's average rating and bg mechanic Simulation = 0.228
Correlation between a boardgame's average rating and bg category World War II = 0.215
Correlation between a boardgame's average rating and bg category Action / Dexterity = -0.104
Correlation between a boardgame's average rating and bg category Civil War = 0.108
Correlation between a boardgame's average rating and bg category Movies / TV / Radio theme = -0.181
Correlation between a boardgame's average rating and bg category Napoleonic = 0.133
Correlation between a boardgame's average rating and bg category Wargame = 0.367
Correlation between a boardgame's average rating and bg category Party Game = -0.123
Correlation between a boardgame's average rating and bg category Miniatures = 0.148
Correlation between a boardgame's average rating and bg category World War I = 0.239
Correlation between a boardgame's average rating and bg category Trivia = -0.160
Correlation between a boardgame's average rating and bg category Card Game = -0.115
```

Fig 8. Correlations greater than .1/ less than -.1

The list above shows that there is a high, positive correlation between wargames, World War I, and World War II games. There is a strong, negative correlation between a roll-spin-move games and that game's average rating. Hex-and-counter games, on the other hand, have a strong positive correlation. From the correlation results, it can be concluded that games that simulate war tend to receive higher ratings. In addition to this, games that do not rely on luck or "the roll of the die" tend to be more well-liked by users.

Hypothesis Testing

Board games that are categorized as "Wargame" have the highest correlation with the average rating. A two-sided t-test was conducted to see if the difference in the mean of the average rating between games categorized as "Wargame" and those that are not is statistically significant.

Category	Amount	Avg-Rating Mean	Avg-Rating Variance
Wargame	2,412	6.6013	5.8673
Not Wargame	6,428	0.5667	0.731

Table 1. Number of Games Per Category

A two-sided t-test was performed to see if the difference in means of the average rating of games categorized as "Wargame" versus those not in this category is statistically significant. The null and alternate hypothesis are as follows:

- Null-Hypothesis: The mean of the average ratings of board games in the wargame category is equal to that of the non-wargame category
- Alt-Hypothesis: The mean of the two categories are not equal
- Alpha: .05

t-statistic	39.3026
p-value	8.89E-294

Table 2. Results of Two-Sided T-Test

The p-value of the two-sample t-test was far below .05 in the Frequentist Approach. Hence, we reject the null hypothesis that the mean of the average ratings of board games in the "Wargame" category is equal to that of board games not in the "Wargame" category. There is evidence to suggest that board games in the "Wargame" category do indeed receive higher average ratings than board games that are not in the category.

In-Depth Analysis

The technique used to build this recommender system is collaborative filtering. There are two approaches used with this technique, memory-based and model-based. In collaborative filtering, the machine learning algorithm predicts the preference of user A, for example, to an unrated item by comparing user A's historical preferences to those of many other users. The algorithm will then assume that user A's rating for the unrated item will be similar to the ratings given to that item by all the other users that have similar preferences to user A. For the analysis of BoardGameGeek's data, the dataset consists of every username, every board game's ID number, and the rating given to that game by that user. Table 3 below is five rows of the dataset.

username	game_ID	rating
MossGrande	49454	3.0
crostino	49454	3.0
wolfzell	49454	2.0
scatman	49454	2.0
Nicodemus42	49454	1.0

Table 3. Five rows of the dataset to be converted to matrix format.

The data is then transposed into a matrix format where each row is a user, each column is a board game, and the fields are filled with the user's rating for that board game. Table 4. below is an example of the format. Given that not every user rates every game, the data produces a sparse matrix.

	game 1	game 2	game3	game 4	game 5
user 1	8			4	
user 2		7			
user 3	6		5		
user 4				6	
user 5		2			3

Table 4. Example of the matrix format.

The dataset consisting of username, board game ID, and ratings had 6,669,077 rows. This large a dataset was too much for the hardware available. Therefore, a random sampling of 50,000 rows was taken from the full dataset then transposed to the matrix format mentioned above.

A benchmark analysis of the recommender system library "Surprise" was ran using RMSE as the test statistic. The lower the RMSE, the better the results are said to be. The results below are provided after a 3-fold cross validation.

Algorithm	test_rmse	fit_time	test_time
SVD	1.503394	2.814424	32.666214
BaselineOnly	1.507441	0.393118	0.155054
KNNBaseline	1.518484	44.611629	73.806944
SVDpp	1.527421	5.471261	0.221407
KNNBasic	1.665548	51.948159	33.955794
CoClustering	1.776849	5.688466	0.109350
KNNWithMeans	1.779789	61.708028	40.270545
KNNWithZScore	1.783237	46.179942	15.265567
SlopeOne	1.784331	0.729425	0.121677
NMF	2.033936	5.771079	0.205073
NormalPredictor	2.322187	0.070478	0.156901

Table 5. RMSE results of different algorithms with time shown in seconds.

The BaselineOnly algorithm gives a random rating to the empty fields based on the distribution of the dataset (assumed to be normal). The RMSE of this algorithm will be the standard to beat as it just predicts ratings based on normal distribution. The results from the benchmark analysis show that only the SVD algorithm provided the better, albeit marginally, RMSE.

Model-Based Approach

Since SVD did best, a train-test-split analysis was conducted after a grid search was performed to find the best parameters for SVD. The grid search provided an RMSE of 1.4958 with a 3-fold cross validation. The best parameters tested are in the figure below.

```
1.4958252328910397
{'n_epochs': 150, 'lr_all': 0.005, 'reg_all': 0.025}
```

Fig 9. SVD GridSearchCV result.

The parameters above were then used on a train-test-split analysis. This resulted in an RMSE of 1.4874 which beats the RMSE of the BaselineOnly algorithm RMSE of 1.507441.

```
RMSE: 1.4874

1.487407426480512
```

Fig 10. RMSE on the SVD test set.

Memory-Based Approach

A grid search was performed to find the best parameters of KNNWithMeans algorithm. However, the RMSE result was higher than both the SVD and BaselineOnly algorithms.

```
Best RMSE: 1.623663575518912
Best params: {'sim_options': {'name': 'msd', 'min_support': 4, 'user_based': False}}
```

Fig 11. KNNWithMeans GridSearchCV result.

The best parameters were then used to run an estimator that produces the estimated rating any chosen user gave to any chosen game.

Conclusion

The collaborative filtering method that provided the best results for the recommender system was the SVD algorithm. Its RMSE was better than that of the BaselineOnly algorithm which was used as the base model. KNNWithMeans did not provide a better result than either of the other two. The results of the algorithms, however, can be improved greatly by using more data, which is available. The full dataset was over 6.6 million rows of data for board games that had 30 ratings or more. This amount of data unfortunately is too large to be feasible for the hardware available. Only 50,000 rows were able to be used in the analysis.