Board Game Recommender System
DSC 478 Final Project
Steven Jordan
March 18th, 2019

Overview

I have built a recommender system that uses a content-based filtering approach to provide users board game recommendations. The recommender system creates and updates a user profile with their "favorited" board games, analyzes their favorited board games' attributes, and then identifies and prints the board games most similar to the user's favorites using the cosine similarity metric. Users have the ability to choose the number of recommendations they wish to see, and can choose to filter the recommendations to come from the top 500 (approx. 0.5%) board games of the 105,000+ board games found on www.boardgamegeek.com (BGG). I am an avid fan and player of board games, and have incorporated some of my own domain knowledge in the development process.

This recommender system is unique, as other recommender systems I have found online appear to be take solely a collaborative filtering method. The recommender system and all associated files can be found here: https://drive.google.com/open?id=1DDP2fslRaGztw4w_ca4L4NmAmgBqeHXp

Data Description & Pre-Processing

The data was downloaded from https://www.kaggle.com/mrpantherson/board-game-data, which was in turn sourced from the public API of www.boardgamegeek.com. From Kaggle, the dataset *bgg_db_1806.csv* was selected (renamed to *bgg.csv* in the project folder) and it contains the 4,999 highest ranked board games on the site as of June 08, 2018. This is just a small subset of the 105,736 board games listed on BGG as of March 18th, 2019. Board games that were expansions of other core games, or with missing or unknown values for any of the final attributes were removed. This reduced the number of board games by 424, to 4,575 board games.

Below is a list of the attributes contained in bgg.csv that were used in the recommender system, with their descriptions, and notes on why/how they were used.

Attributes used:	Description:	Notes:
bgg_url	The BGG URL of each board game in the dataset	This was kept as the main "label" for each board game since it is a unique identifier that is helpful to the user. This was partitioned from the main dataset, along with *names* and *geek_rating.* A list of the URLs used in the recommender system can be found in the attached *Acceptable_urls.txt*
names	The name of each board game in the dataset	Partitioned with *bgg_url* and *geek_rating,* and is used when the recommended board games are displayed. There are duplicate names, but those are either different versions of the same game (technically a

		separate game), or unrelated games with the same title.
geek_rating	BGG's rating of each game, that is used to determine their ranking on the BGG website. It is the Bayesian average of BGG's user ratings.	This was partitioned with *bgg_url* and *names*, and is used if the user decides to filter their recommendations to the top 500 board games in the final dataset. This was not used as an attribute to identify "similar" board games.
min_players	The minimum number of players for the board game	Used in the final model. Outliers with a *min_players* value greater than 4, had that value replaced with 4.
max_players	The maximum number of players for the board game	Used in the final model. Outliers with a *max_players* value greater than 9, had that value replaced with 9.
avg_time	The average playing time (in minutes), as provided by the board game publisher.	Used in the final model. Outliers with an *avg_time* value greater than 205, had that value replaced with 205.
age	The recommended minimum age, as provided by the board game publisher	Used in the final model. Outlier with an *age* value of 42, was determined to be an error, and was removed along with the missing values.
weight	The 'complexity' rating on a continuous scale between 1 and 5, as determined by polling BGG users. More complex games tend to have more rules, pieces, etc.	Used in the final model. Outlier with a *weight* value greater than 4.5945, had that value replaced with 4.5945
mechanic	The types or elements of game play (e.g. card drafting, dice rolling, tile placement)	There are 51 mechanics that were separated into unique dummy variables. Games have between 1 and 18 mechanics. A full list of the mechanics can be listed in Appendix 1.
category	The overarching theme (wargames, fantasy) or type (abstract strategy) of the board game.	There are 83 categories that were separated into unique dummy variables. One category, *Expansion for Base-game,* was removed along with any games with this attribute since they are not stand-alone board games. Games have between 1 and 12 categories. A full list of the mechanics can be listed in Appendix 1.

Below is a list of the attributes contained in bgg.csv that were **NOT** used in the recommender system, with their descriptions, and notes on why they were removed.

Attributes removed:	**Description:**	**Notes:**
rank	The ranking of each game on BGG	Directly correlated with *geek_rating*
game_id	The BGG id number of each game	Not useful or descriptive

min_time	The minimum time (in minutes) a game should take to play, as provided by the publisher.	Along with *max_time*, this value was often not provided, and its values were identical to *avg_time*
max_time	The maximum time (in minutes) a game should take to play, as provided by the publisher.	Along with *min_time*, this value was often not provided, and its values were identical to avg_time
max_players	The maximum number of players for the board game	Used in the final model. Outliers with a *max_players* value greater than 9, had that value replaced with 9
year	The year the game was published	Not a useful characteristic, at least in my system
num_votes	Number of users that rated this game	Not a useful characteristic
img_url	The URL of the board game's image.	Not a descriptive characteristic, and unhelpful in the label (for this version of the recommender system)
owned	The number of users that own this game	Not a useful characteristic
designer	The names of the designer(s) of the game	This could be useful in a future version of the recommender system but was removed due to the fact this would create 1000+ dummy variables since games may have a high number of designers.

Algorithms and Techniques

The recommendation engine can be broken down into four components: the function that loads the data *'load_data()'*, the function manages the user's profile and favorites '*userfile()'*, the function that determines and prints the boardgame recommendations '*get_recs()'*, and the user interface with input commands.

load_data()

This function does not take in an input, but when is called (by *get_recs())*, it searches the current directory for a file called *bgg.csv*. This function was created in this way, so an administrator can update the csv file with new games, new data, and assuming the table has the same structure, the recommender system should still function. Using *pandas* and *NumPy*, it removes the unnecessary columns listed earlier, as well as the games with missing data. It creates the dummy variables for the *mechanic* and *category* attributes, and using *sklearn*, it scales each of the variables. It returns a data frame of the labels, and an array of the scaled attribute data.

userfile()

This function takes in the URL of the new favorited game by the user in string format, as well as the user's username in string format. It is called by *get_recs().* First, it tries to open and read a .txt file titled by the username. If it doesn't exist, it creates one, and returns the new URL in a list as the sole favorited game. If the username does exist, it searches the contents of the txt file for the new URL to prevent "double favoriting" which would impact the results of the system. If that URL already exists as a favorite,

it returns the existing favorites. If the URL doesn't already exist as a favorite, it appends it, and returns the full list.

get_recs()

This function is called by the user interface and has four input parameters: the URL of the favorited game (in string format), the user's username (in string format), whether the user wants to filter their recommendations to the top 500 board games (1 if True, 0 if False), and the number of recommendations to display.

After calling *load_data()*, and obtaining the labels and scaled data, it checks the user inputted URL to make sure it is a valid input. If the URL is for a game that was not in the initial 4,999 games, for a game that was removed for having missing values, or simply isn't a valid BGG URL, it will let the user know and restart the process.

It then calls *userfile()*, which returns the list of the existing favorited games for the user, and their index numbers from the labels data frame. Using those indices, it creates an array of the scaled attribute data for the favorited games, and using the *KMeans* clustering function from *sklearn* (but just a single cluster), it finds the centroid values of the favorited games.

Then, again using *sklearn* Cosine Similarity function, we obtain similarity of the centroid to every game in the database. It zips the similarity scores to the games' URLs, names, and geek rating values, sorting it by similarity.

The function then prints out the user-defined number of URLs, names, and similarity scores of the most similar board games – filtering out games that were already favorited, and games not in the top 500 (if the user selected that option).

User Interface

The user interface asks the user for the username, favorited URL, whether to filter the results, and how many recommendations to provide. After the recommendations are provided, the user is able to provide another recommendation.

Evaluation

By the nature of the application, and the lack of user data to test the dataset, it is challenging to evaluate. However, I will evaluate it by A) manually examining the characteristics of recommended board games and B) build out a list of favorites for myself and see if it starts identifying other games I love before I put them in myself.

Manual Examination of Characteristics

For this first test (full results in Appendix 2), I input the URL game *Pandemic Legacy: Season 1* without filtering, and the first 5 recommendations are other games in the Pandemic series:

Pandemic
Pandemic: Iberia
Pandemic Legacy: Season 2
Pandemic: Reign of Cthulhu

Pandemic: The Cure

This is an indicator that the recommender is functioning since they're within the same family (even though their designers and names were not used as part of the algorithm. Like *Pandemic Legacy: Season 1*, they have similar categories (*Environmental, Medical),* they have the *Co-operative Play* mechanic, and have the same minimum number of players (2), and similar maximum number of players (4 or 5).

For the same username, I then input the game Twilight Struggle, a 2-player game about the Cold War, in which the two players (US and USSR) attempt to win by influencing other countries on the game map. My top five recommendations shifted to:

Labyrinth: The War on Terror, 2000 - ?
Pandemic: Iberia
Pandemic
1989: Dawn of Freedom
1960: The Making of the President

This is also an indicator of the success of my recommender. Both *Pandemic* and *Pandemic: Iberia* remained in the top 5, though their placement switched, indicating that there are features in *Pandemic: Iberia* that are closer than *Pandemic* to *Twilight Struggle*. *Labyrinth: The War on Terror, 2000 - ?* is a 1-2 player game about the War on Terror with similar mechanics to *Twilight Struggle*. *1989: Dawn of Freedom* is about the end of the Cold War, as the two sides try to influence the nations around the iron curtain. *1960: The Making of the President* is a 2-player game about the election between Nixon and Kennedy, during the time of the Cold War, in which both players are attempting to influence the different states of the US. The themes, mechanics, and player counts are very similar to *Twilight Struggle*, indicating the recommender system is functioning correctly.

Evaluation of Personal Recommendations

For this evaluation (full results in Appendix 3), I put in my personal favorite three board games - *Pandemic Legacy: Season 1*, *Chess*, *and Santorini*. Immediately in the results, I saw the second and sixth recommendations were respectively the original *Pandemic* and *Onitama*, both games that I own and love. After adding *Onitama*, the next list contained *Hive*, another personal favorite of mine that I own. Because my recommender system is suggesting games I already know and love (but have not yet favorited), it indicates that the recommender system is functioning as expected.

Next Steps

There are many, many improvements I know I can make to this recommender system, most of which I would like to implement due to my love of the hobby. These are some of the next steps I have planned:

- Obtain user feedback, as well as user's "dislikes" to incorporate into algorithm
- See recommendations without entering a URL
- Connect to Board Game Geek API for up-to-date data and the ability to import favorites in bulk
- Graphical user interface with board game images
- Search/add board games by name instead of URL (with automatic correction)
- Hybridize the algorithm with other users' likes
- Use feedback to weight different characteristics

Appendix 1 – List of mechanic and category dummy variables

Mechanics		
Acting	Grid Movement	Route/Network Building
Action / Movement Programming	Hand Management	Secret Unit Deployment
Action Point Allowance System	Hex-and-Counter	Set Collection
Area Control / Area Influence	Line Drawing	Simulation
Area Enclosure	Memory	Simultaneous Action Selection
Area Movement	Modular Board	Singing
Area-Impulse	Paper-and-Pencil	Stock Holding
Auction/Bidding	Partnerships	Storytelling
Betting/Wagering	Pattern Building	Take That
Campaign / Battle Card Driven	Pattern Recognition	Tile Placement
Card Drafting	Pick-up and Deliver	Time Track
Chit-Pull System	Player Elimination	Trading
Co-operative Play	Point to Point Movement	Trick-taking
Commodity Speculation	Press Your Luck	Variable Phase Order
Crayon Rail System	Rock-Paper-Scissors	Variable Player Powers
Deck / Pool Building	Role Playing	Voting
Dice Rolling	Roll / Spin and Move	Worker Placement

Categories		
Abstract Strategy	Fantasy	Pike and Shot
Action / Dexterity	Farming	Pirates
Adventure	Fighting	Political
Age of Reason	Game System	Post-Napoleonic
American Civil War	Horror	Prehistoric
American Indian Wars	Humor	Print & Play
American Revolutionary War	Industry / Manufacturing	Puzzle
American West	Korean War	Racing
Ancient	Mafia	Real-time
Animals	Math	Religious
Arabian	Mature / Adult	Renaissance
Aviation / Flight	Maze	Science Fiction
Bluffing	Medical	Space Exploration
Book	Medieval	Spies/Secret Agents
Card Game	Memory	Sports
Children's Game	Miniatures	Territory Building
City Building	Modern Warfare	Trains
Civil War	Movies / TV / Radio theme	Transportation
Civilization	Murder/Mystery	Travel
Collectible Components	Music	Trivia
Comic Book / Strip	Mythology	Video Game Theme
Deduction	Napoleonic	Vietnam War
Dice	Nautical	Wargame
Economic	Negotiation	Word Game
Educational	Novel-based	World War I
Electronic	Number	World War II
Environmental	Party Game	Zombies
Exploration		

Appendix 2 – Evaluation – Manual Examination of Characteristics Results

```
Hello, what is your username? If new, please create one now. sajordan

Please paste here the URL of your favorited game from BoardGameGeek.com: h
ttps://boardgamegeek.com/boardgame/161936/pandemic-legacy-season-1

How many recommendations would you like to see? 10

Would you like to limit your recommendations from the highest 500 rated bo
ard games? (Y/N) n

New username created.

Hello sajordan, your top 10 board game recommendations are:

Game Title:  Pandemic
URL:  https://boardgamegeek.com/boardgame/30549/pandemic
Similarity Score 0.9438

Game Title:  Pandemic: Iberia
URL:  https://boardgamegeek.com/boardgame/198928/pandemic-iberia
Similarity Score 0.9434

Game Title:  Pandemic Legacy: Season 2
URL:  https://boardgamegeek.com/boardgame/221107/pandemic-legacy-season-2
Similarity Score 0.7997

Game Title:  Pandemic: Reign of Cthulhu
URL:  https://boardgamegeek.com/boardgame/192153/pandemic-reign-cthulhu
Similarity Score 0.7567

Game Title:  Pandemic: The Cure
URL:  https://boardgamegeek.com/boardgame/150658/pandemic-cure
Similarity Score 0.749

Game Title:  Kings of Israel
URL:  https://boardgamegeek.com/boardgame/142084/kings-israel
Similarity Score 0.7341

Game Title:  Black Orchestra
URL:  https://boardgamegeek.com/boardgame/156858/black-orchestra
Similarity Score 0.6299

Game Title:  Marvel Heroes
URL:  https://boardgamegeek.com/boardgame/14808/marvel-heroes
Similarity Score 0.5979

Game Title:  Dead Panic
URL:  https://boardgamegeek.com/boardgame/133437/dead-panic
Similarity Score 0.5917

Game Title:  Train Heist
URL:  https://boardgamegeek.com/boardgame/146880/train-heist
```

Similarity Score 0.5911

Would you like to add in another favorite? (Y/N) Y
Hello, what is your username? If new, please create one now. sajordan

Please paste here the URL of your favorited game from BoardGameGeek.com: https://boardgamegeek.com/boardgame/12333/twilight-struggle

How many recommendations would you like to see? 10

Would you like to limit your recommendations from the highest 500 rated board games? (Y/N) n
Hello sajordan, your top 10 board game recommendations are:

Game Title: Labyrinth: The War on Terror, 2001 – ?
URL: https://boardgamegeek.com/boardgame/62227/labyrinth-war-terror-2001
Similarity Score 0.7907

Game Title: Pandemic: Iberia
URL: https://boardgamegeek.com/boardgame/198928/pandemic-iberia
Similarity Score 0.7382

Game Title: Pandemic
URL: https://boardgamegeek.com/boardgame/30549/pandemic
Similarity Score 0.7353

Game Title: 1989: Dawn of Freedom
URL: https://boardgamegeek.com/boardgame/26997/1989-dawn-freedom
Similarity Score 0.6923

Game Title: 1960: The Making of the President
URL: https://boardgamegeek.com/boardgame/27708/1960-making-president
Similarity Score 0.6621

Game Title: We the People
URL: https://boardgamegeek.com/boardgame/620/we-people
Similarity Score 0.6582

Game Title: Pandemic Legacy: Season 2
URL: https://boardgamegeek.com/boardgame/221107/pandemic-legacy-season-2
Similarity Score 0.658

Game Title: Here I Stand
URL: https://boardgamegeek.com/boardgame/17392/here-i-stand
Similarity Score 0.6573

Game Title: Pandemic: The Cure
URL: https://boardgamegeek.com/boardgame/150658/pandemic-cure
Similarity Score 0.6523

Game Title: 13 Days: The Cuban Missile Crisis
URL: https://boardgamegeek.com/boardgame/177590/13-days-cuban-missile-crisis

Similarity Score 0.6343

Appendix 3 – Evaluation - Personal Recommendations Results

Hello, what is your username? If new, please create one now. sajordan2

Please paste here the URL of your favorited game from BoardGameGeek.com: https://boardgamegeek.com/boardgame/194655/santorini

How many recommendations would you like to see? 10

Would you like to limit your recommendations from the highest 500 rated board games? (Y/N) y
Hello sajordan2, your top 10 board game recommendations are:

Game Title: Pandemic: Iberia
URL: https://boardgamegeek.com/boardgame/198928/pandemic-iberia
Similarity Score 0.6929

Game Title: Pandemic
URL: https://boardgamegeek.com/boardgame/30549/pandemic
Similarity Score 0.6864

Game Title: TZAAR
URL: https://boardgamegeek.com/boardgame/31999/tzaar
Similarity Score 0.669

Game Title: ZÈRTZ
URL: https://boardgamegeek.com/boardgame/528/zertz
Similarity Score 0.6674

Game Title: DVONN
URL: https://boardgamegeek.com/boardgame/2346/dvonn
Similarity Score 0.6662

Game Title: Onitama
URL: https://boardgamegeek.com/boardgame/160477/onitama
Similarity Score 0.6452

Game Title: Pandemic: Reign of Cthulhu
URL: https://boardgamegeek.com/boardgame/192153/pandemic-reign-cthulhu
Similarity Score 0.623

Game Title: Pandemic Legacy: Season 2
URL: https://boardgamegeek.com/boardgame/221107/pandemic-legacy-season-2
Similarity Score 0.6148

Game Title: The Duke
URL: https://boardgamegeek.com/boardgame/36235/duke
Similarity Score 0.5894

Game Title: The Downfall of Pompeii
URL: https://boardgamegeek.com/boardgame/13004/downfall-pompeii

Similarity Score 0.5813

Would you like to add in another favorite? (Y/N) y
Hello, what is your username? If new, please create one now. sajordan2

Please paste here the URL of your favorited game from BoardGameGeek.com: https://boardgamegeek.com/boardgame/160477/onitama

How many recommendations would you like to see? 10

Would you like to limit your recommendations from the highest 500 rated board games? (Y/N) y
Hello sajordan2, your top 10 board game recommendations are:

Game Title: DVONN
URL: https://boardgamegeek.com/boardgame/2346/dvonn
Similarity Score 0.7349

Game Title: TZAAR
URL: https://boardgamegeek.com/boardgame/31999/tzaar
Similarity Score 0.7342

Game Title: ZÈRTZ
URL: https://boardgamegeek.com/boardgame/528/zertz
Similarity Score 0.7017

Game Title: The Downfall of Pompeii
URL: https://boardgamegeek.com/boardgame/13004/downfall-pompeii
Similarity Score 0.6824

Game Title: YINSH
URL: https://boardgamegeek.com/boardgame/7854/yinsh
Similarity Score 0.6246

Game Title: The Duke
URL: https://boardgamegeek.com/boardgame/36235/duke
Similarity Score 0.6212

Game Title: Torres
URL: https://boardgamegeek.com/boardgame/88/torres
Similarity Score 0.6185

Game Title: Hive
URL: https://boardgamegeek.com/boardgame/2655/hive
Similarity Score 0.6091

Game Title: Pandemic: Iberia
URL: https://boardgamegeek.com/boardgame/198928/pandemic-iberia
Similarity Score 0.6079

Game Title: Pandemic
URL: https://boardgamegeek.com/boardgame/30549/pandemic
Similarity Score 0.6021

Appendix 4 – Project Attachments

1. *ReadMe.txt – Provides basic instructions and overview*
2. *acceptable_urls.txt – Provides a list of acceptable URLs in the final dataset*
3. *FinalProjectReport_DSC478.pdf - Final Project Report*
4. *BG_Recommender_Start.ipynb – Clone of BG_Recommender_Start.py in iPython format*
5. *Data_Exploration.py – Exploration of the data with visualizations*
6. *Data_Exploration.ipynb – Clone of Data_Exploration.py in iPython format*
7. *Video Demo - https://youtu.be/xHhrMdqcDoQ*

BG_Recommender_Start.py – File to initiate the recommender system
bgg.csv – The original dataset from www.boardgamegeek.com, but pulled from https://www.kaggle.com/mrpantherson/board-game-data