Recommendation System for Video Games

By Devesh Gokalgandhi

November 18, 2019

Contents

5.2 Content-Based Filtering

5.3 Collaborative-Based Filtering

6. Further Research and Recommendations

6.1 Future Work

6.2 Recommendation to Clients

7. Conclusion

7.1 Data Exploration Conclusion

7.2 Modeling Conclusion

1. Executive Summary

- The objective of this project is to develop a recommendation system for gamers, who struggle to find games that are similar or just as enjoyable as their previously played games. Luckily, game content and other users/game critics, who have a rich history in reviewing many games, across many different titles, provide information that can help recommend different titles for a user who will have a high chance of enjoying those games.
- Using game review data, as input for the recommendation system, using content-based filtering and collaborative based filtering methods. 24023 game reviews and ratings were extracted from the video game database GiantBomb, along with the reviewed game's features, such as genre, themes, concepts, description, etc. To evaluate the content-based filtering method, an additional feature, or target variable, was extracted, similar games, which displays a list of games that the GiantBomb staff deemed similar to its respective game. There were a total of 6561 users reviewing 4223 games, spanning from July 2008 to September 2019.
- A relatively high percentage of games, in the similar games list, were of the same genre, theme, concept, platform, whereas an extremely low percentage of games were of the same character, franchise, publisher, and developer, when looking at the most popular values for these features.
- For content-based filtering, the best combination of features, when evaluated by the similar games list, was determined to be genres, themes, concepts, platforms, and short game summary, yielding a precision of 0.111 (average of 1 game in similar games list, out of 10 recommendations), a recall of 0.211 (average of 20% of the games in similar games list are recommended), and the harmonic mean of precision and recall, the F1 score, of 0.125.
- For collaborative-based filtering, the best algorithm that fit and predicted for the overall dataset was determined to be the matrix factorization algorithms, SVDpp and SVD. Both algorithms fitted the training data, and evaluated using the test data, for which the errors were: SVDpp RMSE and MAE of 0.8682 and 0.6662 respectively, SVD RMSE and MAE 0.8708 and 0.6679. However, the system overpredicts ratings, as a result of the uneven distribution of the ratings (heavily skewed to positive ratings). In this case, content-based filtering is the preferable method to recommending the games.

2. Introduction

Often times, it can be difficult for avid gamers to choose what titles they would want to play and enjoy next, as more and more games are released. Building a recommendation system can help to alleviate this problem and provide a number of titles that best suit a gamer, based on their gaming history, or the preferences of like-minded gamers. Luckily, descriptive game content and user critiques provide a rich assessment in reviewing many games, across many genres, themes, developers, etc., and provide information that can help recommend different titles for a user that they will have a high chance of enjoying. This information, along with user submitted ratings, can be utilized in both collaborative filtering (based on other users) and content based filtering (based on game features).

For content-based filtering, natural language processing (NLP) techniques, such as TF-IDF, are implemented to determine similarities between game features and descriptions, and assign recommendations that best suit the user's taste for the type of game they tend to enjoy. For collaborative filtering, model-based approaches, such as Matrix Factorization and Singular Value Decomposition will use machine learning to find user ratings to unrated games, based on the taste of similar users. An example of user-based collaborative filtering: If I like Halo, and you like Halo and Metal Gear Solid, I will be recommended Metal Gear Solid because we both enjoyed Halo. Content-based filtering will make use of the game's features, such as genre, platform, game description, publishers, and developers. An example of this: Because I like Halo, or games made by the developer company Bungie, I will be recommended Destiny and Destiny 2 because they are games developed by Bungie. These techniques will provide an efficient and accurate tool for gamers to determine which collection of games they should play to scratch their gaming itch!

3. Data Acquisition

3.1 Dataset Description

User-submitted reviews for games was collected using GiantBomb.com API queries, the documentation for which can be found here: https://www.giantbomb.com/api/documentation/. The review dataset contains a unique id corresponding to the game the user is reviewing, which was used to query metadata for the game. In total, there were 24023 user reviews, 6561 users reviewing 4223 games, spanning from July 2008 to September 2019. User features included in the dataset are: date of review, short review summary, full length review, reviewer, and score. The score will be crucial for the collaborative filtering model-based method, in predicting user ratings for unrated games. Game features include: short game description, full length game description, characters, concepts, developers, franchises, genres, platforms, publishers, themes, and similar games. The target variable ‘similar games’ feature will be crucial, as it is used to evaluate the content based filtering method by matching games recommended by the system.

3.2 Data Wrangling

All of the game feature names were nested within lists, which included API links and unique feature ids. Only the names were extracted from these lists, as there can be more than one feature name for each game, as these will be involved in the TF-IDF implementation in finding similarities between game content. Names with spaces between their words, such as the genre “First Person Shooter”, were connected with a dash to prevent splitting of descriptive words: “First-Person-Shooter”. HTML tags and stopwords (commonly used words) were removed from the user review features and game descriptions to keep the unique and informative words that describe each game. All descriptive feature values were converted to lower-case, to prevent duplicates of different case lettering. Any ‘s’ or ‘ies’ were removed from word endings, in order to convert plural cases to singular, and lessen the duplicates of plural and singular case letterings There were 1417 reviews missing a value for the ‘similar games’ feature, approximately 5% of the data, which were omitted.

While all game reviews were given a score from 0 to 5 (whole number), two reviews were given a score of 0.5, for which many of the content and user features were given the value ‘test’, and as a result, removed from the dataset. All other missing values were replaced with an empty space and ready to be merged to create a ‘bag of words’ instance, which will be used in the TF-IDF implementation. To ensure that the evaluator feature, ‘similar games’, list only included games from the reviews, games that were identified to not be part of the list of total unique games (games which cannot be recommended) were removed from each “similar games” list. This process was repeated until each review record’s ‘similar games’ list had at least one game that can be found in the list of total unique games in the dataset. After the removal process, approximately 1500 review records had no games within their ‘similar games’ feature, and were omitted from the final dataset. After the wrangling process, the final dataset contains 22866 reviews from 3592 users on 4220 games.

4. Data Exploration

4.1 Ratings and Reviews

Looking at the distribution of game ratings:

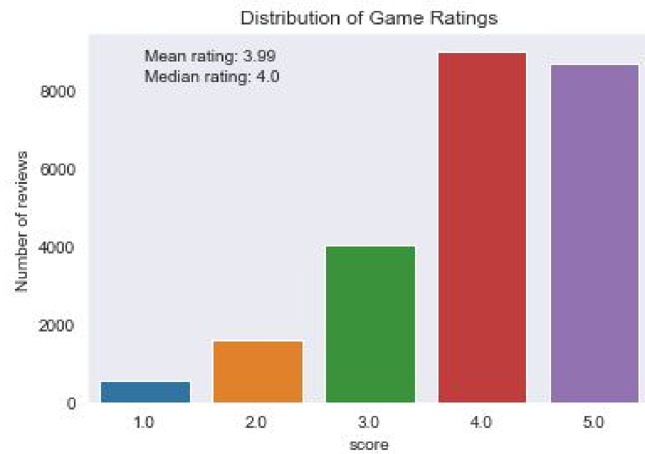


Figure 1. Distribution of ratings.

The majority of reviews within the dataset skew heavily to the positive side, with ratings of 4's and 5's. This could be attributed to a variety of reasons, such as:

1. Bias towards popular games
2. Bias towards publishers, developers, or franchises
3. Unwillingness to rate harshly

This positively skewed distribution will affect how the matrix factorization, and prediction of unrated games, will occur, as we expect more positively predicted ratings, based on other user's reviews, as part of collaborative filtering.

Looking at the distribution for the number of reviews as user has, as well as, the distribution for the number of reviews a game has:

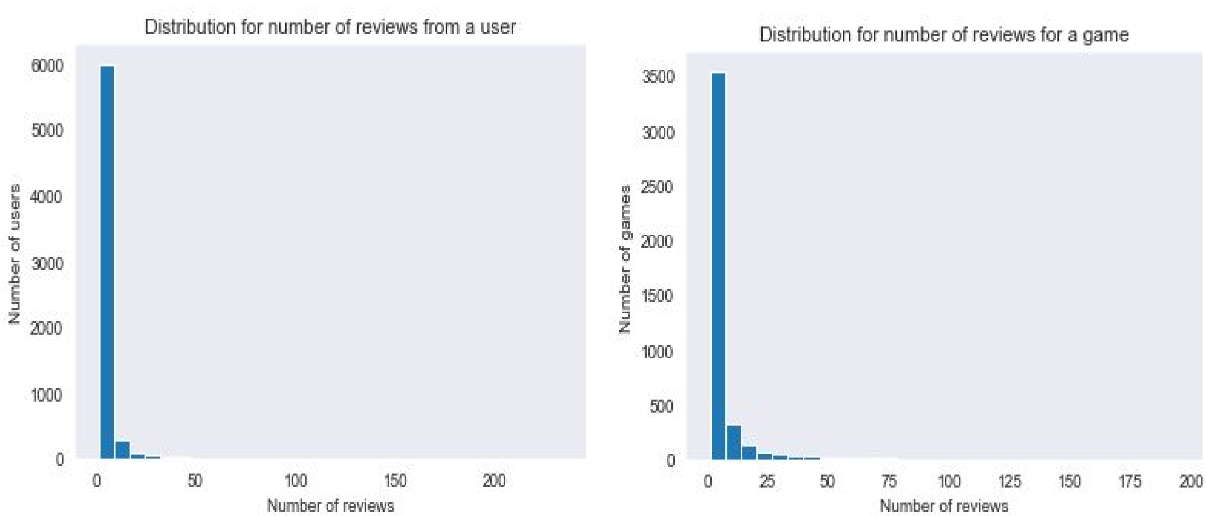


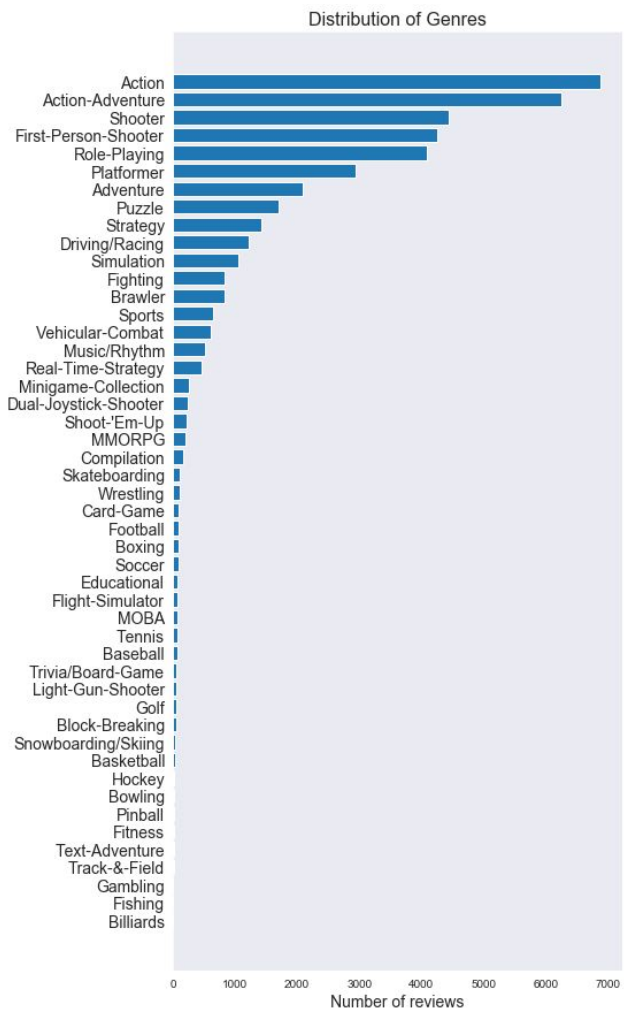
Figure 2. Distribution of number of reviews from a user and for a game.

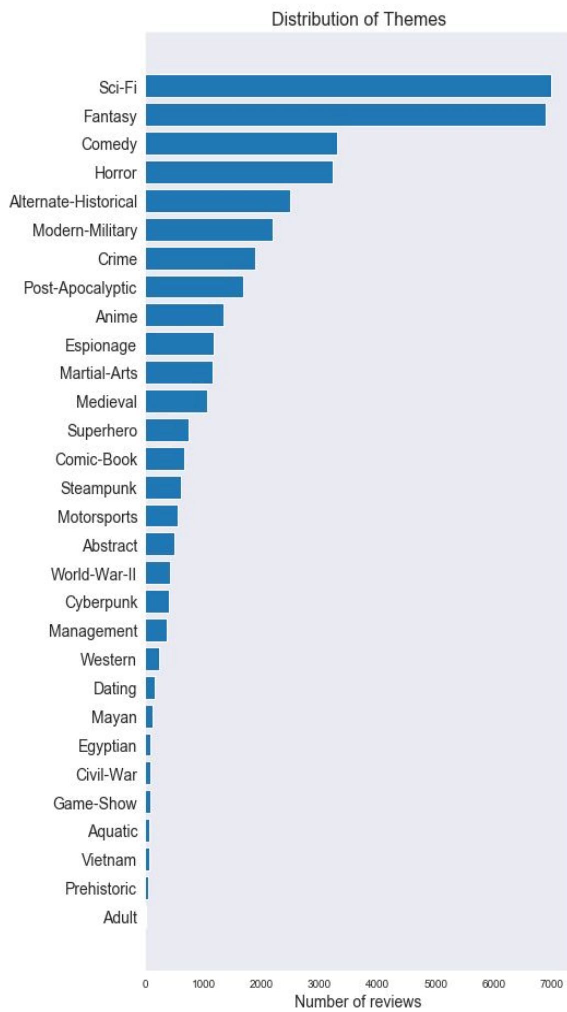
In both distributions, we can observe the presence of a long tail, indicating that many users have only reviewed one game, and many games only have only been reviewed once. This indicates that the data is highly sparse, as not every reviewer has played and rated every game, and we expect many cells in the user-item rating matrix to be empty. Knowing this:

1. We can expect, with the recommendations, that more popular games will tend to be recommended more, whereas games from the ‘long tail’ section might get ignored.
2. When querying a new video game to the system, the ‘cold start’ problem may occur, where unless a user has rated these new games, they won’t get recommended. The sparsity of the data adds on to that issue.

4.2 Genres, Themes, and Concepts

Most popular features in the reviews:





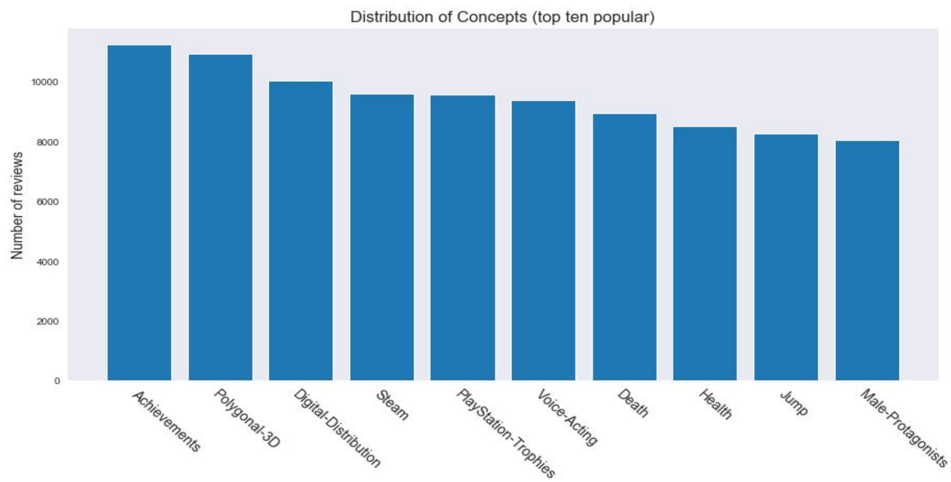


Figure 3. Distribution of genres, themes, concepts.

The majority of games reviewed are in the Action or Action-Adventure genres, and of the Sci-Fi or Fantasy theme. The significant amount of games that contain concepts, as evidenced by the concept frequency plot, are intuitive for an avid gamer, as many games provide in-game incentives such as Achievements (Xbox) and PlaysStation-Trophies (Playstation), be distributed digitally through Steam (popular digital distribution service platform), contain voice-acted non-playable characters, involve a health bar, jumping mechanics, and death sequence with respawn. For content based filtering, we expect that these popular genres, themes, and concepts to have less weight towards determining similarities among games, as most games contain these feature types. It is through associated features, whether it be other genres, themes, concepts, lesser-known features, etc. that similarities can become more filtered to fewer games, and better recommendations can be made.

Looking at the genre associations, or which pairs of genre labels are most likely to co-occur among the most popular genres:

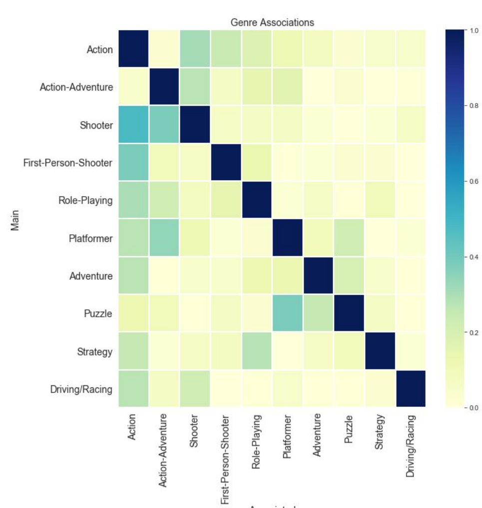


Figure 4. Genre Associations for most popular genres

Some examples of genres, with particularly strong associations, included in the plot above and for the less popular genres:

1. Shooter, Flight Simulator, and Vehicular Combat with Action
2. Baseball, Basketball, Fitness, Football, Golf, Hockey, Skateboarding/Skiing, Soccer, Tennis, Track and Field with other sports
3. Gambling with Card Games
4. Real Time Strategy with Strategy

Genres related to sports have moderate to strong associations to each other, which could be attributed to multi-sport games or to a subset of users who play primarily sports games. We expect the recommendation system to recommend titles related to the specific sport first, then widening the list to include other sports. We also expect the recommender system to recommend action games for games related to vehicular combat, shooter, and flight simulator, and card games for games related to gambling.

We look at the same associations in the popular themes:

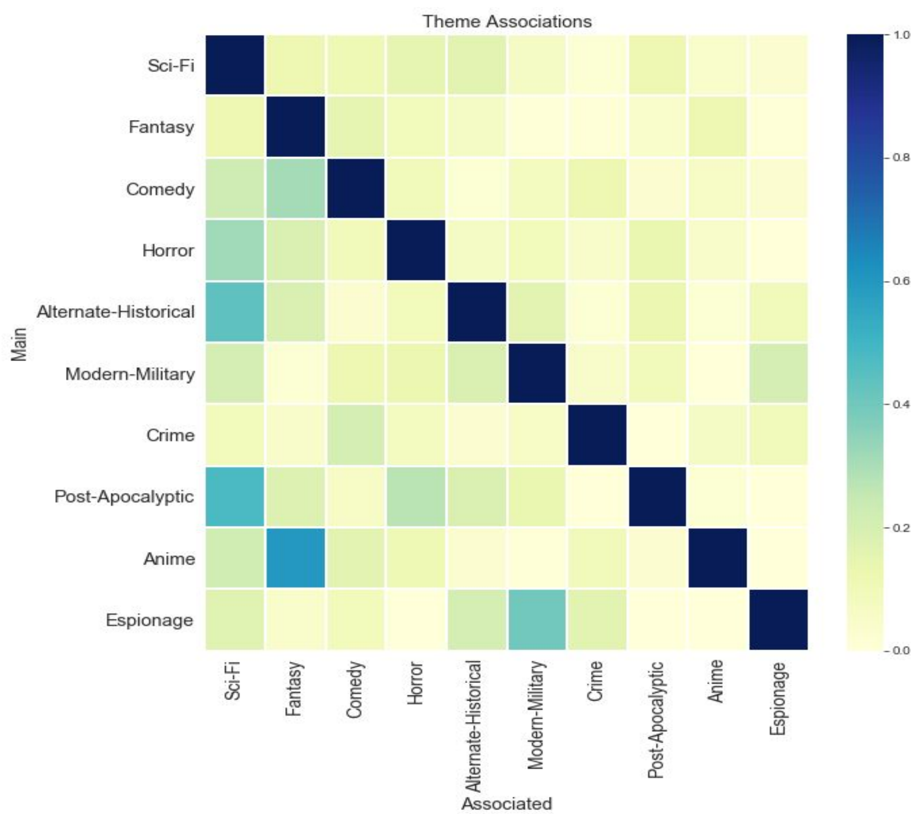


Figure 5. Themes Associations for most popular themes

Some examples of themes, with particularly strong associations, included in the plot above and for the less popular themes:

1. Aquatic, Dating, Medieval, and Steampunk with Fantasy
2. Fantasy with Anime
3. Crime with Superhero

Again, using the same logic for the genre associations, we expect the system to recommend games with the same associations as the respective game, then widening the list to the associated theme.

To evaluate the recommendation system, for content based filtering methods, the feature 'similar games' provides a list of at least one game that is highly similar to the respective reviewed game. It should be noted however, the recommendation system can still recommend similar games to the respective game, even though the recommender system's suggestions are not specifically listed in the 'similar games' field. The similar games field was created by avid gamers who use GiantBomb and found similarities between games. Reviews may only contain one similar game for the target game, when in reality a game can have various degrees of similarity with multiple other games not listed.

Looking at the percentage of similar games of the same genre (for top then most popular genres) as their respective game:

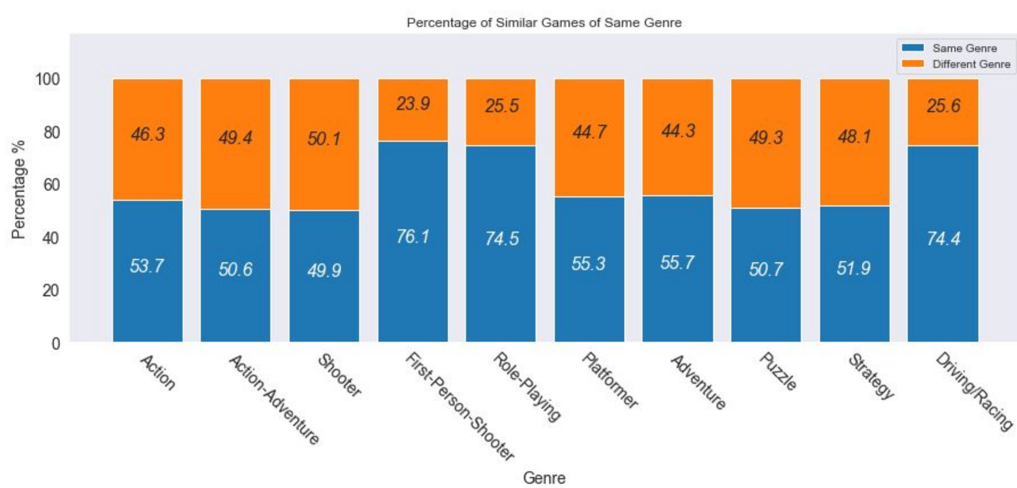


Figure 6. Percentage of similar games of same genre

Clearly at least 50%, or more, of the 'similar games' are of the same genre as their respective game, especially with First Person Shooter, Role Playing, and Driving/Racing games. This makes sense, as these genres are extremely specific with the type of game, or mechanics of the game, that are characteristic of the genre. Genres such as Action, Action-Adventure,

Adventure, etc. are more vague with the type of game mechanics involved. Even still, genre seems to be an important game feature to distinguish games, and will be utilized heavily when assigned importance by the TF-IDF algorithm.

We look at the same thing for themes:

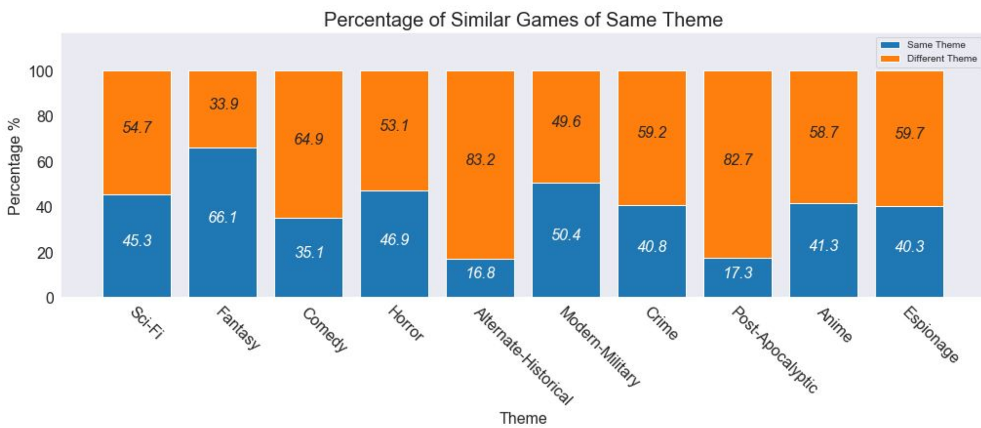


Figure 7. Percentage of similar games of the same theme.

With the exception of the Alternate-Historical and Post-Apocalyptic theme, at least 35%, or higher, similar games contains the same theme as their respective game, especially for Fantasy themed games. It seems that for Alternate-Historical and Post-Apocalyptic themed games, other features will play an important role in determining similarities between games. We can, however, state that genres seems to be a more important factor in determining the games, in the similar games list, over themes.

When analyzing the percentage of similar games that have the same concept as their respective game:

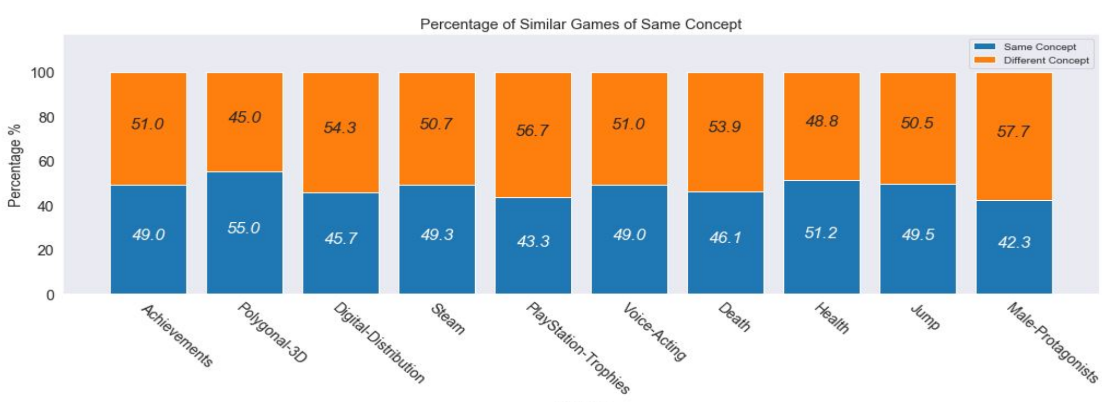


Figure 8. Percentage of similar games of same concept

The results are fairly consistent with the top most popular games, approximately 50% of the similar games have the same concept as the target game, which is significant. It should be noted however, based on the distribution of concepts in the games, the amount of games that these popular concepts lie in are extremely large. That leaves a wide array of games that could be recommended based solely on these popular concepts, so additional games features, coupled with these concepts will help narrow the games to be recommended more accurately. However, it can be concluded that the concepts feature contains more weight than the themes feature, when determining similar games for the target game.

4.3 Characters and Franchises

For the ten most popular characters and franchises reviewed, the percentage of similar games that are of the same character and the same franchise are shown below:

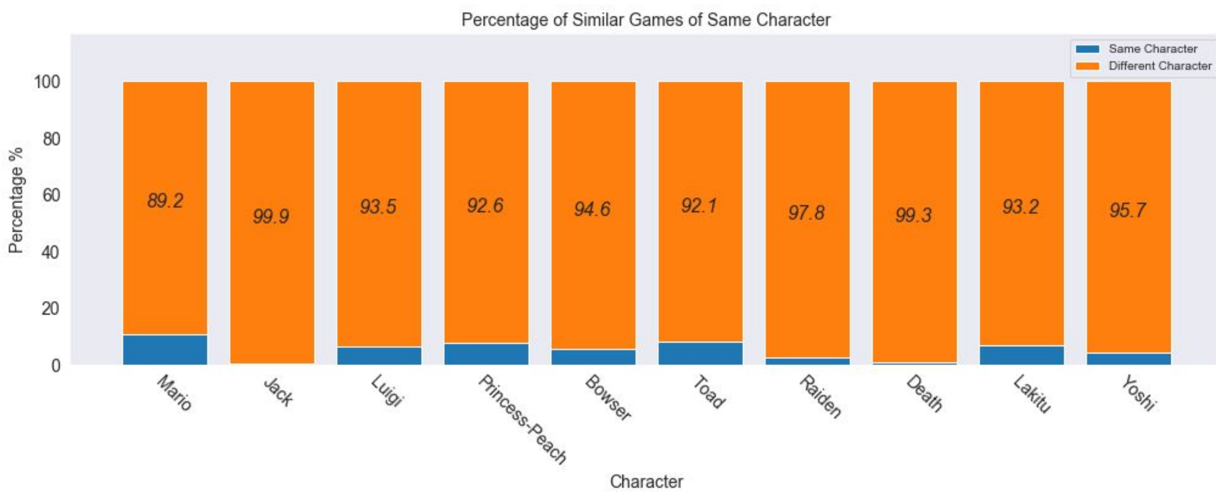


Figure 9. Percentage of similar games of the same character.

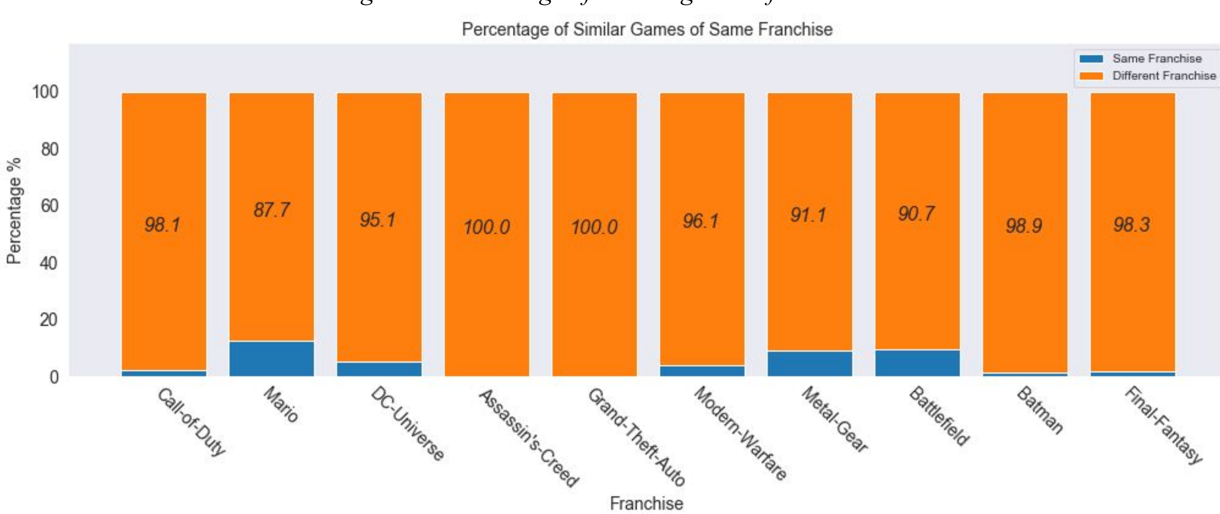


Figure 10. Percentage of similar games of the same franchise.

As shown, a significant amount of similar games do not feature the same character as the target game, and is not of the same franchise as the target game. This suggests that the characters and franchises features, for the purposes of the TF-IDF method in content-based filtering, should not be utilized, as these feature values will tend to recommend games of the same franchise and with the same characters, due to the similarities. Therefore, multiple combinations of the aggregation of content based features will be tested for TF-IDF, and we expect the most accurate combination, evaluated by the similar games list feature, will not include the franchise and characters features. Once again, this does not mean games recommended that are of the same franchise and characters are wrong to be recommended, but for the sake of the evaluation by the similar games feature, they seem to be not important.

4.4 Platforms, Publishers, and Developers

Looking at the association of the top ten common platforms for each game:

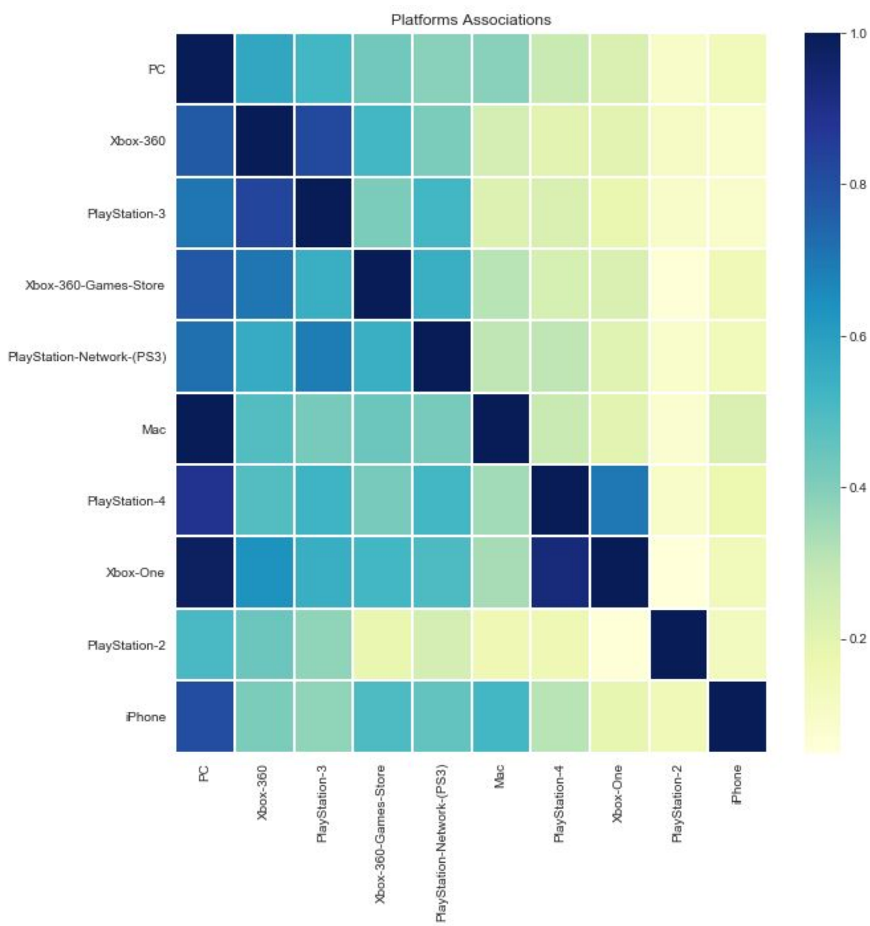


Figure 11. Platform Associations for most popular platforms.

This heat-map indicates that many games are cross-platform, especially for games available on PC. This makes sense, as digital distributors are expanding, providing access to games for gamers who don't own, or need to purchase a console. Many games are shared between Xbox (360 and One) and Playstation (3 and 4), with a few console-exclusive exceptions, which also makes sense, as publishers would want the highest possible reach, or audience, to play their games. We expect the recommender system to utilize the platforms feature by recommending console exclusive games for gamers who have enjoyed the same console's exclusive games.

When we observe the percentage of similar games that are of the same platform as the target game:

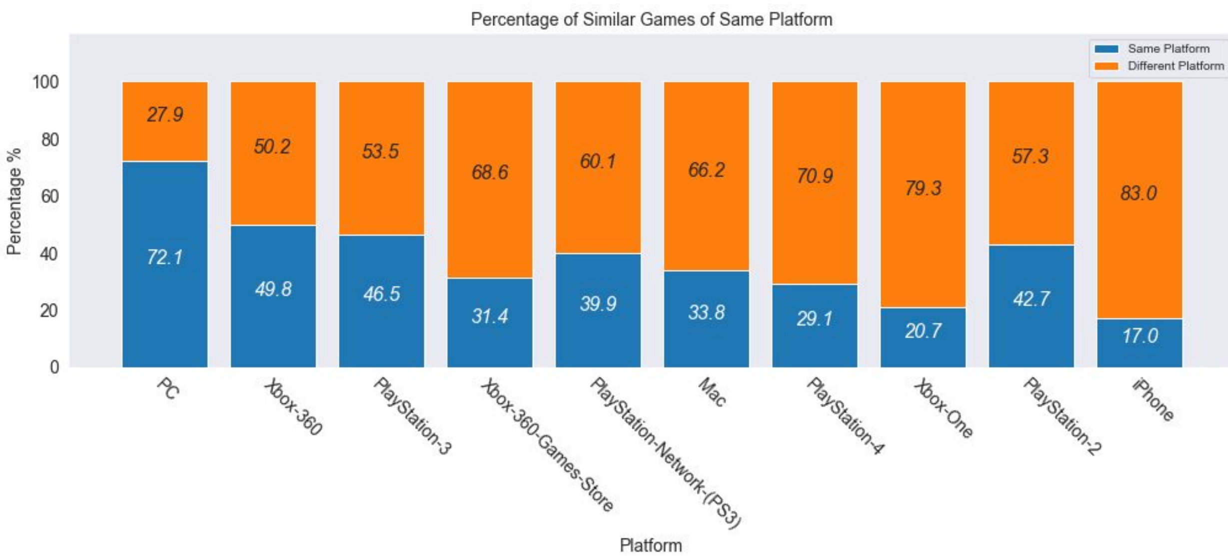


Figure 12. Percentage of similar games of same platform

The similar games tend to be available on the same platform as their target game, especially for PC games, which makes sense, as the majority of games are available on PC. We expect the system to recommend PC games more often for target PC games. For the rest of the platforms, we expect that additional content-based information will help strengthen the accuracy of the system to recommend games that are within the similar-games list, providing an array of games that are both available and not available on the same platform. Therefore, the platform of the games will play a role in determining similarities for the TF-IDF method.

We look at the percentage of similar games that are of the same publisher and developer, for the top ten most popular publishers and developers:

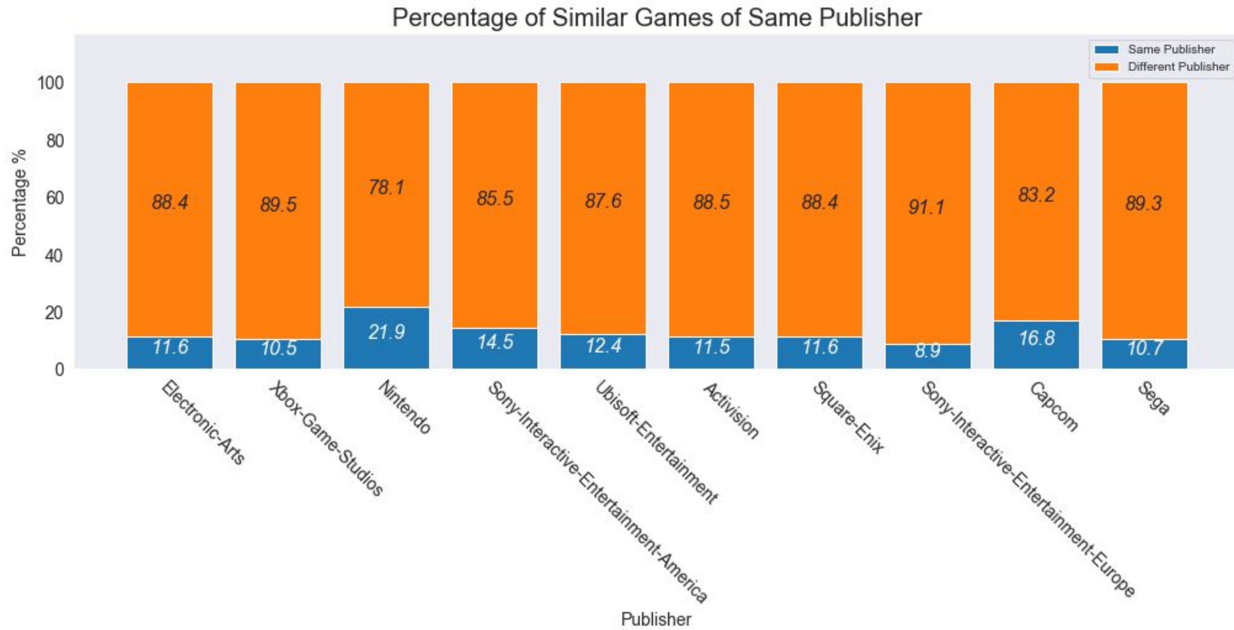


Figure 13. Percentage of similar games of same publisher.

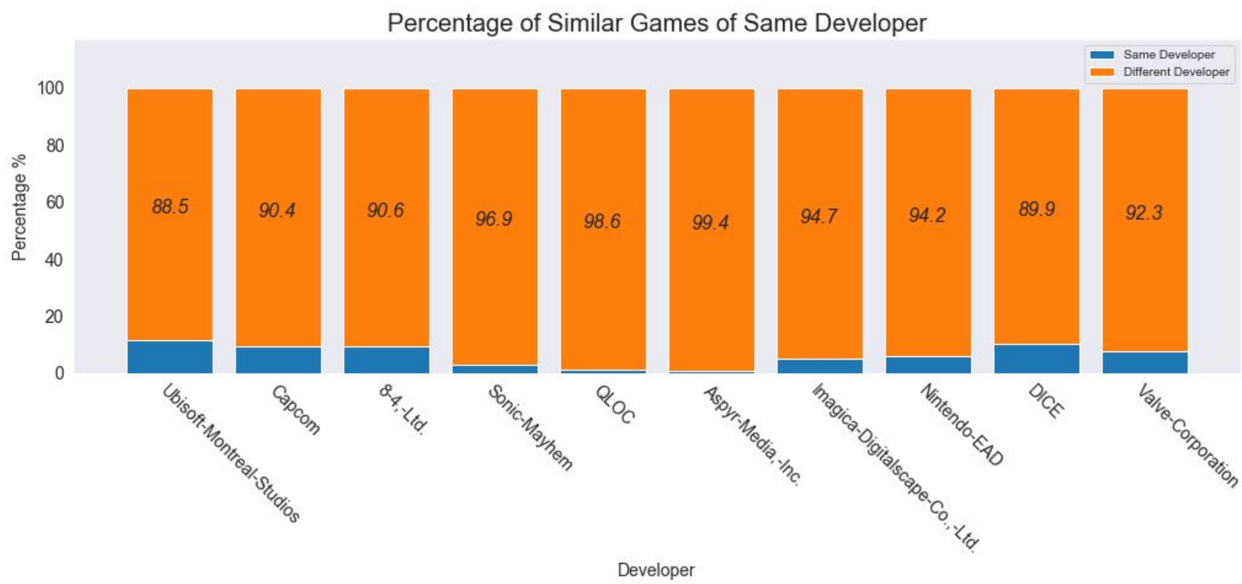
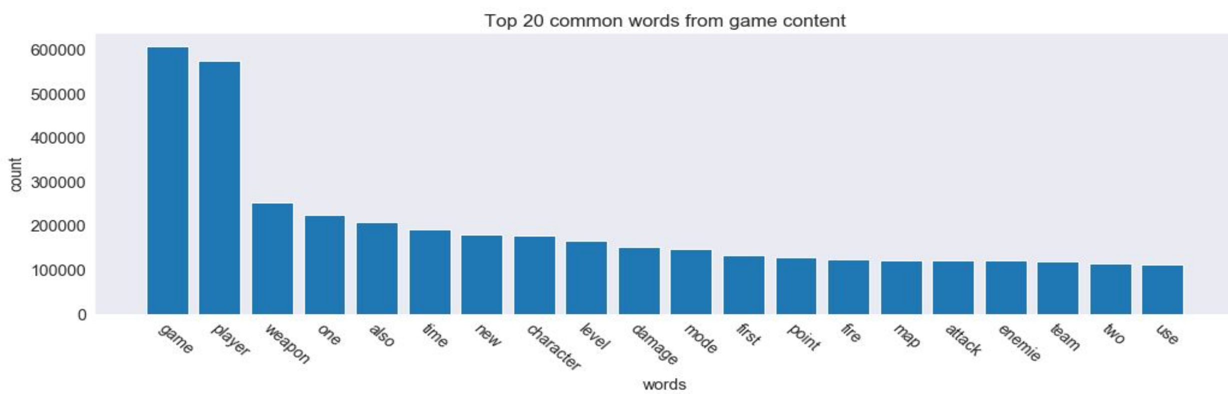


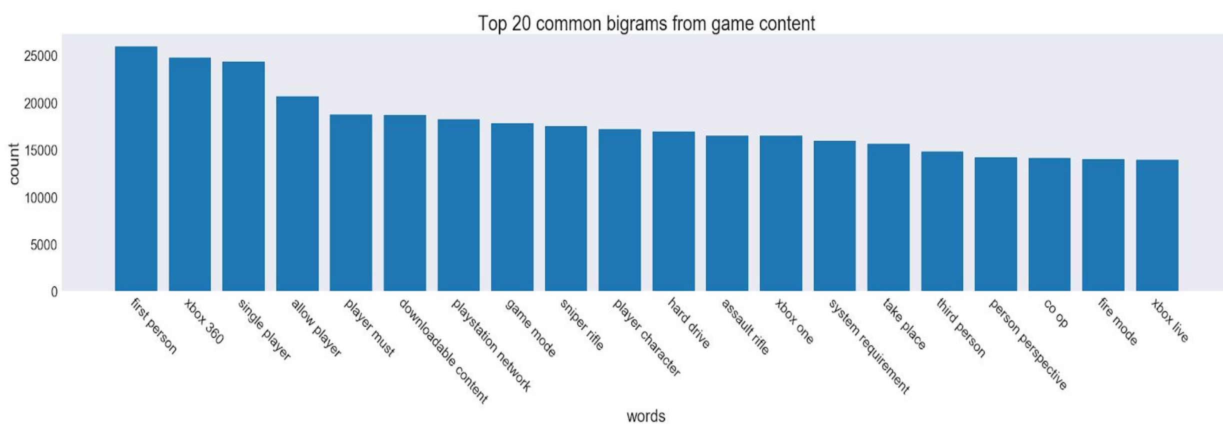
Figure 14. Percentage of similar games of same developer.

As shown, a significant amount of similar games are not under the same publisher and developer as the target game. This suggests that the publisher and developer features, for the purposes of the TF-IDF method in content-based filtering, should not be utilized, as these feature values will tend to recommend games of the same publisher and developer, due to the similarities. However, just as with the characters and franchises feature, this does not mean that the system recommending games of the same publisher and developer is wrong, but these games will most likely have no matches with the similar games list used to evaluate the system.

4.5 Term Frequency

First, we observe the top terms, bigrams (two words), and trigrams (three words) from the bag of words, which is composed of all of the content-based features:





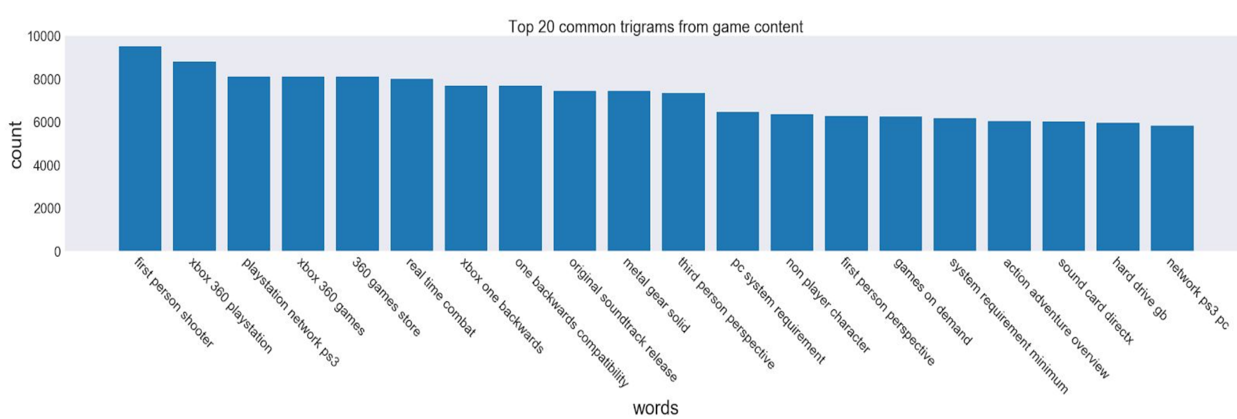


Figure 15. Most popular terms, bigrams, and trigrams.

Clearly, each of the popular terms, bigrams, and trigrams refer to common game words and concepts such as "game", 'player', 'first person', 'single player', "xbox 360 games", "games on demand", etc. We expect the system to not assign high weight to these terms and phrases, when determining what factors will most accurately recommend games that are high similar to the target game, as they are most likely associated with the majority of games in the dataset. Coupled with additional information, such as genres, themes, concepts, etc., will filter out less similar games and provide a more accurate list of similar games for the target game.

Next, because genre and themes seems to be the significant features that would steer the recommendation system towards recommending games that would match the games provided by the evaluating feature list, similar games, we look at how often certain terms are used for games of a certain genre. To observe some examples of how the system will use TF-IDF method to match games of the same genre/theme or to other games of a different genre, based on the terms involved in all of the content-based features, we look at a scattertext plot, which displays term frequency, with a baseline of 500 uses from the bag of words, from two different genres. Specifically, for genre pairs such as Action and Action-Adventure, as well as, Shooter and First-Person-Shooter, we would expect that both genres would relate to a high degree, judging from the name alone, but want to know what specific terms distinguish or relate for both genres.

Here is the scattertext plot of the terms for Action and Action-Adventure games:

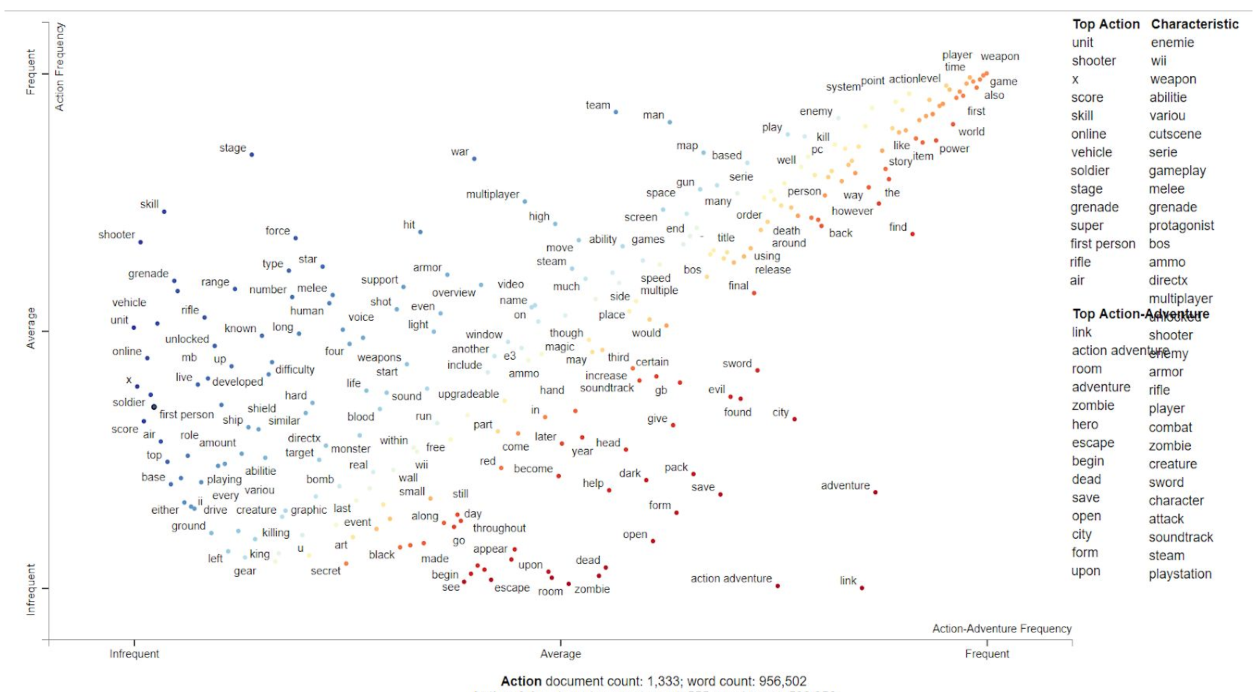


Figure 16. Scattertext for Action vs. Action-Adventure games.

The visual indicates:

- Terms used frequently among both genres include commonly used game words such as: games, player, system, time, multiplayer, etc. Reference to violence (combat, damage, kill, attack, etc.), weapons (grenade, ammo, shotgun, pistol, etc.), and player mechanics (jump, shoot, cover, etc.) seem to be dominant among both genres. Interestingly, the term ‘action’ is heavily used by both genres, whereas the terms ‘action-adventure’ and ‘adventure’ are exclusive to the Action-Adventure genre.We expect the system to use these terms to recommend both Action and Action-Adventure games.
- Terms more frequently used for Action games, but not Action-Adventure, include shooter and first person (Shooter genre most associated with Action genre), and various war, weapons and vehicular references (rifle, soldier, melee, vehicle, grenade, shield, etc.). We expect these terms to steer the system to recommend more Action games.
- Terms more frequently used for Action-Adventure, but not Action, include fantasy references (hero, zombie, etc.) and player objectives (escape, help, form, see, pack). We expect the system to use these terms to recommend more Action-Adventure games.

Comparing terms from First-Person-Shooter and Shooter games:

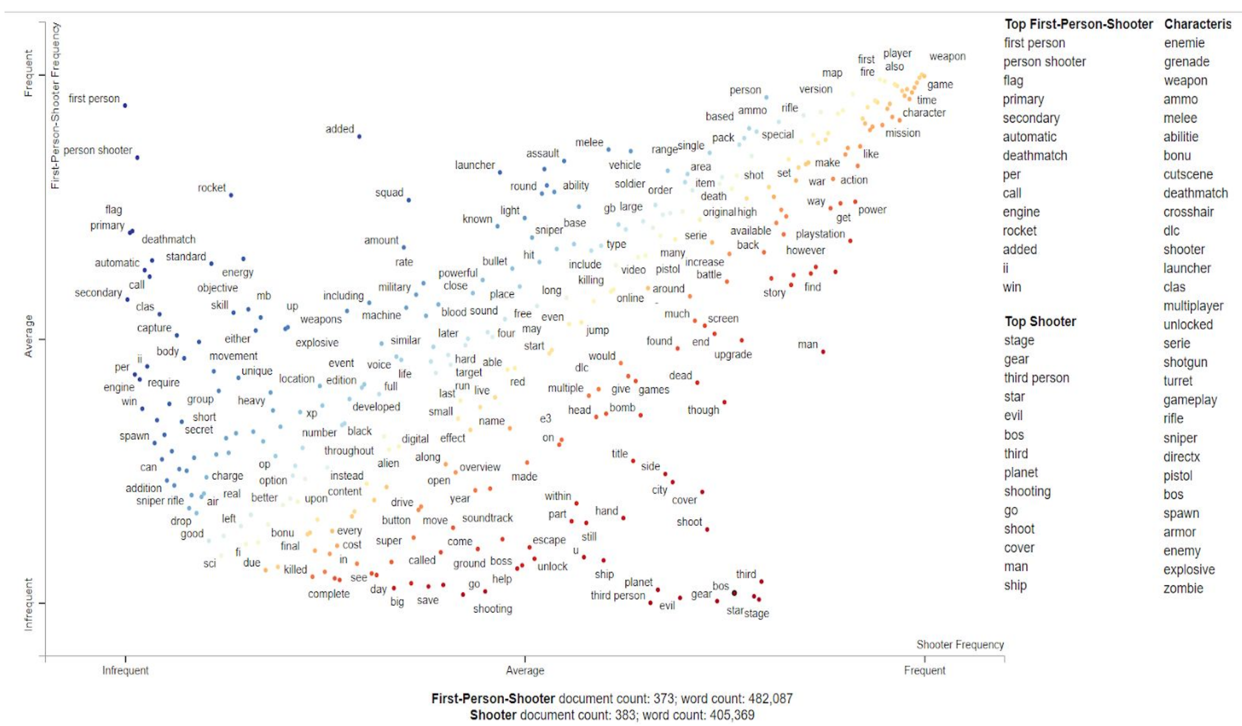


Figure 17. Scattertext for First-Person-Shooter vs. Shooter games.

The visual indicates:

- Terms used frequently among both genres include commonly used game words such as: games, player, enemy, gameplay, multiplayer, etc. Large amount of references to violence (combat, damage, kill, attack, etc.) and weapons (grenade, ammo, shotgun, pistol, etc.) seem to be dominant among both genres especially. We expect the system to use these terms to recommend both Shooter and FPS games.
- Terms more frequently used for FPS games, but not Shooter, include specific weapon and game mode terms (first-person, fire mode, fully automatic, deathmatch, etc.). References to popular FPS games such as Halo and Call of Duty are made. We expect these terms to steer the system to recommend more FPS games.
- Terms more frequently used for Shooter, but not FPS, include more references to popular Shooter games such as Uncharted, Metal Gear Solid, Resident Evil. Interestingly, terms such as ‘shoot’, ‘shooting’, etc. are more tied to the Shooter genre rather than the FPS genre. The term “third person” is also widely used, as expected, in this genre, as it is a different perspective shooter than first person perspective. We expect the system to use these terms to steer towards recommend more Shooter games.

Finally we have look at the terms involved in the top two themes, Sci-Fi and Fantasy:

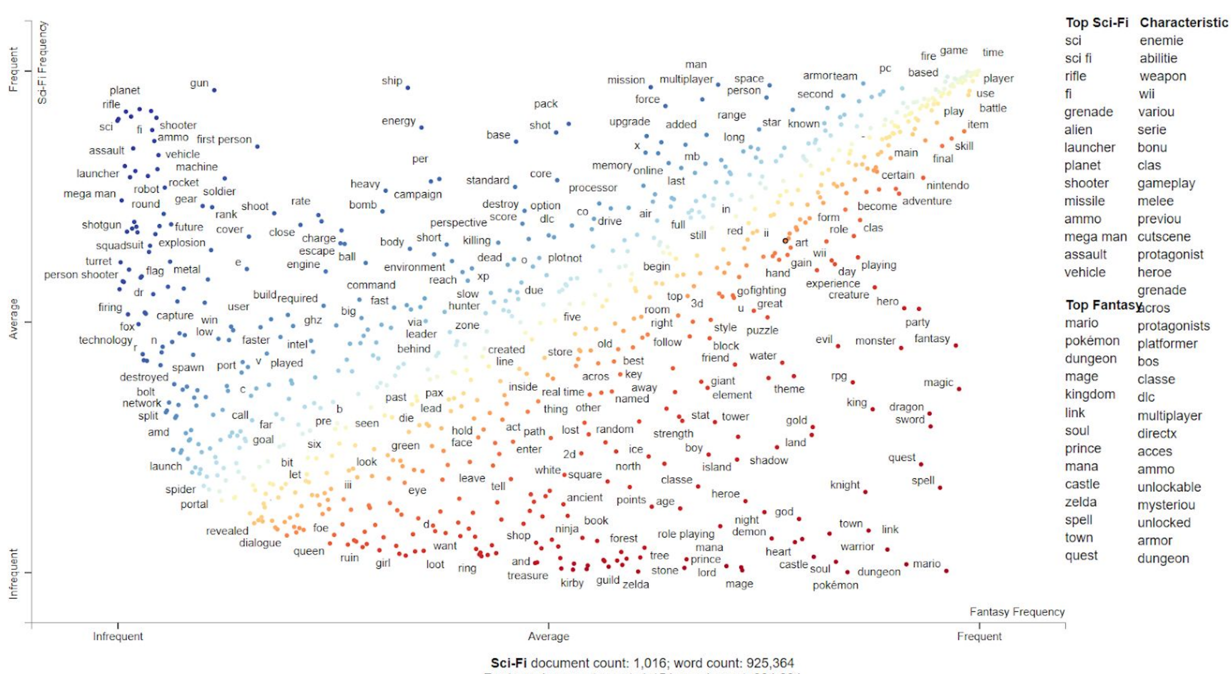


Figure 18. Scattertext for Action vs. Action-Adventure games.

- Terms used frequently among both themes include commonly used game words such as: games, player, enemy, playstation, soundtrack, etc. Reference to violence (combat, damage, kill, attack, etc.) and weapons (grenade, ammo, armor, guns, etc.) seem to be dominant among both themes. We expect the system to use these terms to recommend both Sci-Fi and Fantasy games.
- Terms more frequently used for Sci-Fi games, but not Fantasy, include science fiction terms, such as 'planet', 'plasma', 'alien', etc. Popular references to Sci-Fi games such as Halo and Mass Effect are made. We expect these terms to steer the system to recommend more Sci-Fi games.
- Terms more frequently used for Fantasy, but not Sci-Fi, include terms such as 'dungeon', 'dragon', 'demon', 'knight', and many other fantasy themed terms. Many references to popular Fantasy games such as Legend of Zelda, Uncharted, Pokemon, and Final Fantasy. We expect the system to use these terms to steer recommendations towards more Fantasy games.

5. Modeling

5.1 Pre-Processing

The original dataset is split into two different datasets, one for the content-based filtering implementation and one for the collaborative-based filtering implementation.

For the content-based filtering dataset, the original dataset is reduced to include only one record for each unique game (duplicate game records are filtered out), for a total of 4223 records. Each unique game has its own similar list of similar games, which is used to evaluate the content based filtering recommendations. All of the game's metadata is combined into one feature, the bag of words.

For the collaborative-based filtering dataset, in order to ensure similarities between users, the original dataset is reduced to include only games that have been reviewed at least 2 times, and users that have made at least 2 reviews. This reduced the dataset from 24023 reviews to 15809 reviews, which is now ready to be utilized by the matrix factorization and clustering models to predict user ratings for unrated games. Because matrix factorization assigns a value of 0 to missing values in the user-item rating matrix, each record's score is increased by a value of 1, so that the ratings range from 1.0 to 6.0, to prevent reviews with a score of 0 to change to a predicted value.

5.2 Content-Based Filtering

The approach that the content-based filtering will recommend games based on the initial game's similarity to other games, will rely on the information from the game's metadata (genre, theme, description. etc.). Here, the TF-IDF algorithm will be implemented to weigh words in each game's bag of words, which is an amalgamation of all of the game's metadata data into a single feature, and assign importance to those words based on the number of times they appear in the bag of words. Each game is stored as a vector of the game's attributes, where the angle between each vector, in an n-dimensional space, is calculated to determine the similarity between each game's vector. The measure of similarity that is used is called cosine similarity, because when the angle between two vectors decreases, the value of cosine will increase, signifying more similarity between the vectors, or games.

Because the 'similar games' feature will evaluate how accurate the TF-IDF method recommendations, we want to determine which combination of features will create the most accurate recommender system. The metrics measured will be precision, recall, and F1 score, in which each game's precision, recall, and f1 score is summed and averaged over the total number of unique games. Ideally we don't want to recommend too few or too many games for the user to choose from, so the baseline number of recommendations will be 10, for which:

1. Precision = # of games from the recommendations that are in the similar games list / 10
2. Recall = # of games from the recommendations that are in the similar games list / # of games in the similar games list
3. F1 = 2 * $\left(\frac{precision * recall}{precision + recall}\right)$

We start with all features combined to make the bag of words, calculate the metrics and accuracy of the recommendations, then systematically remove one feature from the bag of words. If the removed feature creates more accurate recommendations, it stays removed, otherwise, it remains in the bag of words. The order of which feature is removed was determined using the previous analysis of the composition of the similar games list (Fig 6,7,8,9,10,12,13,14). The results of the TF-IDF recommendations are shown below:

Table 1. Results of TF-IDF recommendations for each bag of words combination

Bag of Words includes:	Avg. Precision	Avg. Recall	Avg. F1
Every feature	0.087	0.191	0.103
Every feature except characters	0.088	0.193	0.104
Every feature except	0.109	0.206	0.122

characters and game body*			
Every feature except characters, game body, franchises	0.110	0.209	0.123
Every feature except characters, game body, franchises, publishers	0.111	0.211	0.125

* "Game body" refers to full length description of the game

The features used (combined for the bag of words), for most accurate recommendations, were genres, themes, concepts, developers, platforms, game_deck (short summary of game). Any feature, or features, removed from this bag of words will lower the accuracy metrics, so we stop the feature selection here. The best precision, 0.111, indicates that the system, on average, can recommend at least 1 game in each game's similar games list, and the best recall, 0.211, indicates that the system, on average, can recommend at least 20% of the games in each game's similar games list. Again, this does not mean that the recommendations system is not accurate, with the inclusion of all features, as accuracy, for the sake of this dataset, is judged by the games within the similar games feature list.

To display an example of what the system recommends for the game Pac Man:

Table 2. Example recommendations for Pac Man. Left table: Similar games list for Pac-Man. Right table: Recommended list of games from the system.

Similar Games list for Pac-Man
Lock-n-Chase
Lady Bug
Katamari Damacy
Spore
Mario Bros.
Dig Dug
Bit Boy!!
Bit Trip Void
The Legend of Zelda: Spirit Tracks
Jungler
Gubble
Monaco: What's Yours Is Mine
Grand Theft Auto
Wolfenstein-3D
Metal Gear
Galaxian

Recommended Games	In Similar Games list?
Dig Dug	Yes
Ms. Pac Man	No
Galaxian	Yes
Galaga	No
Centipede	No
Mario Bros.	Yes
Frogger	No
Pac Man Championship Edition	No
Tetris	No
Space Invaders	Yes

.

For the example of Pac-Man, the system managed to recommend 4 games, out of the 10 recommended games, that were in the similar games list, for a precision of 0.4 and recall of 0.25. This is a relatively high precision and recall, compared to the average metrics. However, with a closer look at the similar games list for Pac-Man, while many games can be argued as similar, games such as Wolfenstein-3D, Metal Gear, and Grand Theft Auto do not seem to be very similar to a classic arcade game, Pac-Man. When looking at the list of recommended games, each game represents classic arcade games that would have been on the same, or next to, the machine with Pac-Man. Therefore, here is an example of why not to trust the similar games list as the evaluating feature for the content based system, as many would agree that the recommended list of games for Pac-Man, with sequels/prequels/spin-offs and same style of games are truly similar to Pac-Man.

5.2 Collaborative-Based Filtering

Collaborative-based filtering involves the prediction of unknown ratings through the similarities between users. Here, model-based algorithms, such as clustering (KNN) and matrix factorization (SVD, NMF) are implemented with the package: Surprise. 11 algorithms are run, using cross validation, on the dataset to determine which algorithms are most accurate with predicting ratings, using RMSE and MAE as metrics. The results are shown below:

Algorithm	test_rmse	test_mae	fit_time	test_time
SVDpp	0.876658	0.664882	7.524015	0.250665
SVD	0.885243	0.671107	1.715744	0.075928
BaselineOnly	0.891370	0.673919	0.101063	0.024599
KNNBaseline	0.935266	0.693460	0.165082	0.315176
CoClustering	0.984642	0.714413	1.319195	0.056860
KNNBasic	0.995987	0.712726	0.131267	0.310606
KNNWithMeans	0.998343	0.729940	0.228719	0.340424
KNNWithZScore	1.022443	0.745451	0.382309	0.289892
SlopeOne	1.075540	0.790935	0.152590	0.161901
NMF	1.149029	0.905406	1.807499	0.040558
NormalPredictor	1.321550	1.029843	0.019945	0.043218

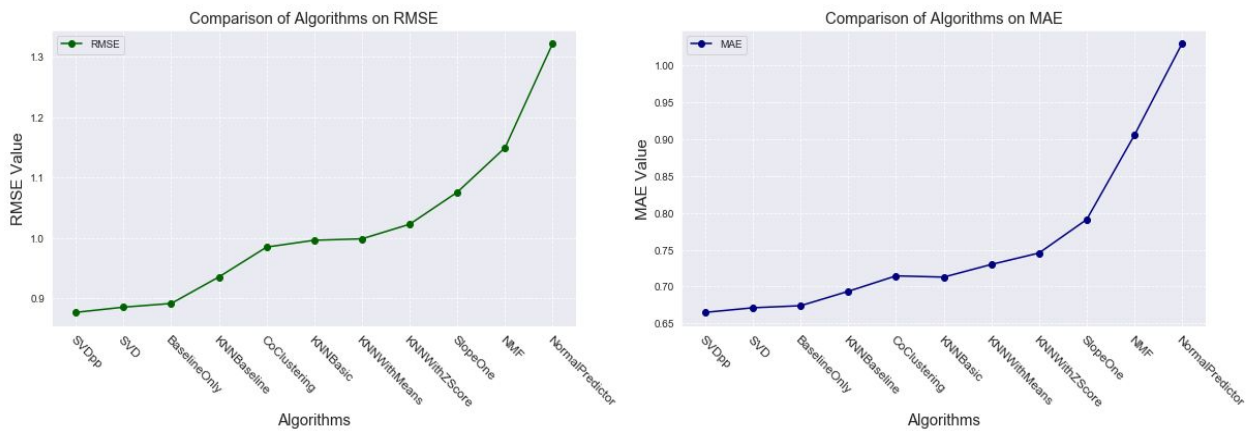


Figure 19. Results and visualization of each algorithm's cross validation on dataset

The algorithm with the lowest RMSE and MAE, or with most accurate predictions, is the matrix factorization algorithm SVDpp, a derivative to the second best algorithm, SVD. SVDpp achieves better predictive accuracy than normal SVD, due to the addition of implicit feedback information. This means that a user exhibits bias or preference to games that they have reviewed, and these preferences are taken into consideration within the algorithm. SVDpp and SVD will be used to train the dataset and predict ratings for a test set to see how the algorithm can handle unseen data.

Next, GridSearch cross validation is implemented to determine the best parameters for both SVDpp and SVD, which were determined as follows:

1. n_epochs: 25
2. lr_all: 0.01
3. reg_all: 0.4

With this information, the dataset is split into a training and testset, and both SVDpp and SVD, with best parameters, are used to fit the training set, and used to predict for the testset. The results are shown below:

Table 3. SVDpp and SVD evaluation of test data

Algorithm	RMSE	MAE
SVDpp	0.8682	0.6662
SVD	0.8708	0.6679

Looking at the residual plot:

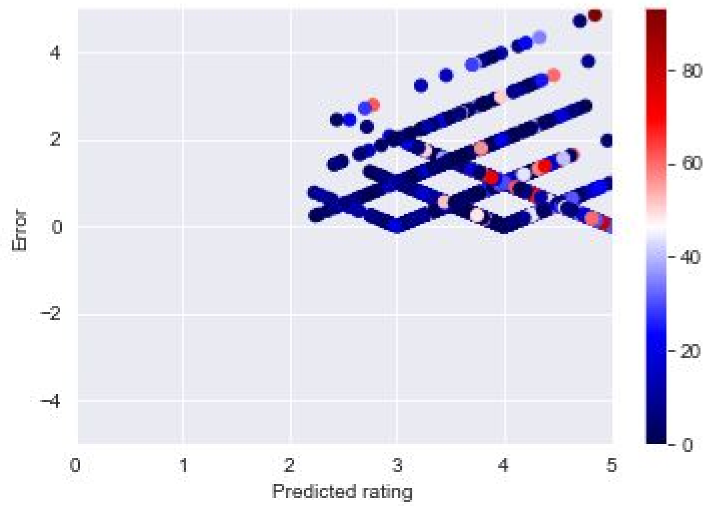


Figure 20. Residual plot for SVDpp test set

The system overpredicts, especially for 0 and 1 ratings, which can be attributed to the highly positive skewed distribution of ratings. The residuals would be a lot more balanced, across an error of 0, if more negatively rated games were included in the dataset, and help alleviate the overprediction problem, by balancing the distribution of ratings. The implementation of sentiment analysis could help with this issue by using the text of the user reviews to provide more weight towards negative reviews, and in turn, not cause the system to overpredict for ratings. Because of this issue, the content based filtering method in recommending games is preferable to use for this dataset, as it does not rely on other users to make recommendations.

The way that this system would recommend games, for each user, is to sort the games, all 4243 unique games, based on the highest predicted rating to lowest. Then we would print the first ten games of the sorted list, as the recommended games for that user.

6. Further Research and Recommendations

6.1 Future Work

For the content-based filtering method, while the game features provide substantial information in describing each of the 4223 unique games, the recommendations can always expand with added new games and their game information. A new evaluator list for the recommendations should also be determined, possibly through surveying gamers to determine what they feel are the most similar games to the games they have played, and strengthening how accurate the system will recommend games. Multiple surveys of similar games should provide enough evidence to determine a consensus list of similar games that can be used to evaluate the system.

For the collaborative-based filtering method, the dataset lacks negatively rated games, causing the system to overpredict ratings, so adding more reviews for these games may help alleviate this issue, but if not, will increase the amount of reviews for each game to make better rating predictions. Another idea that could be implemented is a hybrid system that combines collaborative based filtering with content based filtering, and add sentiment analysis, which would use the user review feature to recommend games tailored towards the individual querying the system.

6.2 Recommendation to Clients

Based on the findings from the data exploration and modeling, it is recommended that:

1. Gamers should use the content based filtering method of recommending games over the collaborative based method, as the dataset is highly skewed towards positive ratings, and games rated negatively will be overpredicted upon.
2. If using the content based filtering method, gamers should be aware that features considered are genres, themes, concepts, platforms, developers, and short game summary, as it produced the best metrics when evaluated by the similar games list. If the gamer enjoys their game due to other features, such as their favorite character, or franchise, the system is unlikely to recommend a sequel/prequel/spin-off to their game.
3. Even though a few features are considered by the content-based filtering system, this does not mean that the recommendations are correct, as the features were determined through evaluating the system with a man-made list of similar-games to the target game. The gamer should feel that these are not the correct similar games to the target game, as different lists of games should be used to evaluate the system, in terms of how well the system can recommend similar games.

7. Conclusion

7.1 Data Exploration Conclusion

To summarize the findings from the exploratory data analysis of the review/game data:

1. **Ratings and Reviews:** The dataset is heavily skewed towards positively rated games (many 4's and 5's). Many users have only rated one game, and many games have only been reviewed upon one time.
2. **Genres, Themes, and Concept:** Majority of games reviewed upon were of genres: Action and Action-Adventure, themes: Fantasy and Sci-Fi, concepts: Achievements and Polygonal-3D. Highly associated popular genres include: Shooter with Action, popular themes: Post-Apocalyptic with Sci-Fi. The percentage of games, in the similar games list, that are of the same genre, theme, or concept to the target game is relatively high, compared to the others features, indicating these features will be highly considered when recommending games evaluated by the similar games list. This does not mean that the other features, in general, should not be used to recommend games.
3. **Characters and Franchises:** The percentage of games, in the similar games list, that are of the same character or franchise is extremely low, indicating that these features should not be included in recommending game.by the content-based system. Once again, this does not mean these features, in general, should not be included, but for the purposes of evaluating the system with the similar-games list, they may not be.
4. **Platforms, Publishers, and Developers:** Strong association between most consoles and PC exists, as most games are available digitally on PC, and games overlap between Playstation and Xbox (few exclusives). The percentage of games, in the similar games list, that are of the same platform as the target game is relatively high, indicating that it will be an important feature to include when recommending games. However, the percentage of games, in the similar games list, that are of the same publisher, or developer, is extremely low.
5. **Term Frequency:** Looking at the frequency of terms, the popular terms all pertain to general gaming language, such as: 'game', 'player', 'character', 'xbox 360', etc. Action terms dominate: 'weapon', 'enemies', 'first person shooter', 'real time combat', etc.

7.2 Modeling Conclusion

To summarize the findings from content based filtering and collaborative based filtering modeling:

1. **Content based filtering:** The optimal combination (bag of words) of features, when evaluated by the similar games list, were genres, themes, concepts, platforms, developer, and short game summary. While this produced the best metrics, the similar games list seems to have unusual game choices, that doesn't seem similar to the target game at hand. I would recommend the user to do some prior research on the games recommended

and determine whether they deem those games similar to their favorite games, before deciding to play them, because the target variable similar games list has limits and flaws.

2. **Collaborative based filtering:** Out of all CF algorithms ran on dataset, SVDpp and SVD produced the lowest error, RMSE and MAE, indicating they are the optimal algorithms to run for this dataset. According to the residual plot, the system overpredicts as a result of the uneven distribution of the ratings (heavily skewed to positive ratings). A balanced dataset with negative and positive reviews would prevent this issue and produce better ratings predictions than this skewed dataset. In this case, content-based filtering is the preferable method to recommending the games.

Overall, we've only scratched the surface with what the recommendations system can produce, given the limited data we have used. Hopefully, gamers will continue to voice their opinions towards newer and newer games, writing reviews, assigning ratings, and providing a list of games they felt were similar to their most enjoyed games, on GiantBomb or other gaming databases, expanding the potential that the recommendation system can have to accurately recommend the most enjoyable games to all gamers.