

Project Proposal

Recommendation System for Video Games

This would be a recommendation system that utilizes both collaborative filtering and content based filtering, in classifying games into elements that will be recommended or not recommended for a user based on user reviews and ratings, as well as, using a user-item feature space. These recommendations would be valuable to avid gamers. Luckily, game content and other users/game critics who have a rich history in reviewing many games, across many genres, provide information that can help recommend different titles for a user who will have a high chance of enjoying those games.

Data is extracted using Gamespot.com API queries, for which there exists 14862 critic reviews, for 10355 games, ranging from games released in 1995 to games released presently. User features include their review of the game, good points and bad points, which provide rich information as to why they chose a rating, ranging from 1-10. Game features include genre, description of the game, platform, release-date, franchise, distribution type, provides content based information that can be used to find highly similar types of games, for which the user will enjoy. The user review dataset will be extracted initially, in which a unique id for the game is provided as a feature, which will be used to query for the respective game's features, and merged to the user reviews and ratings data. Additional features will include region of release (U.S., Europe, etc.), overall region rating of the game, publisher information, and developer information, that provide additional support for recommending games.

Two primary methods will be implemented: collaborative filtering and content-based filtering. Collaborative filtering will make use of the behavior of other users for the games they enjoy, or their preferences, to make recommendations. Non-negative matrix factorization (NMF) will be implemented, as we will be dealing with sparsity, which utilizes dimensionality reduction, to assign recommendations based on filling in unknown ratings for games the user has not played yet, based on other user's reviews and ratings, and recommending the highest rated games. 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 me and you 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.

Deliverables for this project include Python code of the implemented algorithms for the data wrangling, data storytelling (identifying unique trends between the predictors and the target variable), regression modeling and visualization processes. A paper documenting the methodology, displaying in depth analysis of the results and visuals from each process, and discussion of the results will be available. A powerpoint slide deck presentation will also display the results and any unique trends in the data from the data storytelling process, as well as, a succinct discussion of the results of the project. All files and documents will be stored within a Github repository.