Predicting the Next Gloomhaven

Using clustering and topic modeling to determine winning game mechanics

Tiffany Green
March 2022

Business problem:

Is it possible to determine what board game mechanics (e.g. worker placement, bluffing, set collection) as described by players will predict the popularity of a game?

Anticipated Data Science Approach:

The boardgamegeek.com (BGG) website collects a huge quantity of data. The dataset found at the link below captures the reviews of 350,000 unique users. BGG tracks a user review score, which is the mean of all reviews on a scale of 1 to 10. They also track a board game score using a proprietary algorithm. This score ranks every game against every other game on the site, with number 1 being considered the best game. This game is currently Gloomhaven, which has held the top spot for some time. Lastly, they track a geek rating, which is a Bayesian averaged version of the average user rating, to prevent games with relatively few reviews from skewing results.

I had planned to use user reviews with those games that had fewer reviews weighted lower. At the time, I thought the geek rating was made up of ratings from competition gamers. Instead, I'll use the geek rating in place of user review score as a dependent variable, since they've already done the Bayesian averaging to remove the influence of games with fewer reviews.

Reviews often mention the mechanics in the game that appeal to reviewers, so an NLP approach will be used. The following models hold promise: unsupervised sentiment analysis, bag of words analysis, and clustering/topic modeling on phrases and their frequencies. The desired output is the words from the reviews that are found in the reviews of popular games.

Datasets:

https://www.kaggle.com/jvanelteren/boardgamegeek-reviews

Deliverables:

Slide deck and report
Github repo with code from Jupyter notebook