An exploratory study in survival rates of online game players

Yang Zhang (yangz3)

Introduction

With the advances in computer and internet, gaming online becomes a common part of everyday life. People play online games together with friends, or in most cases with strangers for entertainment. In fact, NBC News reported that there were mover 1 billion hours of League of Legends played in a month in 2012 [1]. That was only one game among many others with even more popularity.

As a result, we are interested in what makes people want to play the game. Or in other words, what makes a player decide to stay in community and keep playing. And what would make the player quit the game. One of the most intuitive approaches is to compare between games with different designs, and compare the user playing time. We can then investigate the effect of design factors and the interactions between them on a user's playing time. However, this would require reliable and uniformed sources from multiple games, which is beyond our reach. Instead, we focus on one game – Defense of the Ancients (DotA) 2, and compare between players. Specifically, we want to investigate what happened in the games that made a DotA player decided to not play the game anymore. Equivalently, it will also inform us what will make a player keep playing the game.

Data structure

Our data comes from the OpenDotA website, where we can get access to over a million DotA players data ever since their first game. We downloaded 5758 players' data (https://www.dropbox.com/s/cbraofsl18oavyr/download.tar.gz?dl=0). Each player has one json file which contains all games that player was ever involved. Each game has a data point in the json file, formatted as follows:

{"match_id":3835021371,"player_slot":4,"radiant_win":false,"duration":2986,"game_mode":22,"lobby_type":7,"hero_id":96,"start_time":1523822948,"version":null,"kills":0,"deaths":9,"assists":19,"skill":null,"leaver_status":0,"party_size":null}

Where we have the match id, the player's side, game result, game duration, game mode, lobby type, hero id, game start time, version, number of kills, number of deaths, number of assists, skill, leaver status, and party size. A full document of these data can be found here [2].

Analysis plan

We will first apply 20/60/20 rules to the data and analyze the distribution of our datapoints by creating histograms and tables. Specifically we are interested in players who has not played for more than three months. We will consider these players as ones that have already quitted the game.

We will investigate the effects of players' previous performance on the decision to quit the game, by building regression models to model the likelihood of quitting a game. We will also investigate the interactions and the causality of factors.

[1] NBC News. https://www.nbcnews.com/tech/tech-news/league-legends-players-log-1-billion-hours-month-flna1C6423906

[2] OpenDotA API Documentation https://docs.opendota.com/#