**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 35352 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**

To process the data, we first removed data from players who have played for shorter than two months. These players were often new to the game and thus were unstable and can be unpredictable when it comes to why they quitted the game. To better model the population, we removed data points from these players from the data set.

We then calculate the performance dependent variables – one set of variables for each player. Specifically, the performance dependent variables include average kills, average deaths, average assists, average game duration, average winning percentage, hero selection diversity (i.e., we counted the number of unrepeated heroes selected by that user), average game finish rate, game frequency (i.e., number of games per week), game history (i.e., total number of weeks playing DotA2), number of games (i.e., total number of games played), number of weeks not played since the last game.

We also want to pay attention to that player’s last period of game play before she quitted the game. For example, we calculated Z-scores of performance characteristics (e.g., kills/deaths/assists) of the player against her average performance throughout her entire gaming history. We also calculated the average performance characteristics of the player’s last period of game play such as winning percentage, game finish rate, and game frequency. In our data processing, we defined the last week and the last month as the last period of game play. The above characteristics are calculated for the last week and the last month respectively.

Table 1. shows all dependent variables we calculated:

|  |  |
| --- | --- |
| **Symbol** | **Meaning** |
| AverageKills | Average kill count in a game |
| AverageDeaths | Average death count in a game |
| AverageAssists | Average assist count in a game |
| AverageDuration | Average duration for a game (second) |
| AverageWinningPercentage | Average winning percentage |
| HeroSelectionDiversity | The total number of unrepeated heroes selected |
| GameFinishRate | Average game finish rate |
| GameHistory | The total number of weeks of game play |
| NumberOfGame | The total number of games played |
| NotPlaySinceLastGame | Number of weeks the player has not played since the last one |
| STDKills | The standard deviation of kill count in a game |
| STDDeaths | The standard deviation of death count in a game |
| STDAssists | The standard deviation of assist count in a game |
| STDDuration | The standard deviation of duration count in a game |
| LWZKills | Average z-scores of kill count in the last week |
| LWZDeaths | Average z-scores of death count in the last week |
| LWZAssists | Average z-scores of assist count in the last week |
| LWZDuration | Average z-scores of duration count in the last week |
| LMZKills | Average z-scores of kill count in the last month |
| LMZDeaths | Average z-scores of death count in the last month |
| LMZAssists | Average z-scores of assist count in the last month |
| LMZDuration | Average z-scores of duration count in the last month |
| LWGameFrequency | Number of games played in the last week |
| LWWinningPercentage | Winning percentage in the last week |
| LWGameFinishRate | Game finish rate in the last week |
| LMGameFrequency | Number of games played in the last month |
| LMWinningPercentage | Winning percentage in the last month |
| LMGameFinishRate | Game finish rate in the last month |

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

We will investigate the effects of players’ previous performance on the decision to quit the game, by building a logistic regression model to model the likelihood of quitting a game. We suspect that the performance characteristics especially the average performance of the last time period will have a stronger effect on the decision to quit the game. We will also add the dynamics between the performance of the last time period and the average performance of a player as dependent variables. These dynamics are normalized using Z-scores and subtraction by the averages of each player.

[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/#](https://docs.opendota.com/)