Dissecting Dota

A statistical analysis of Dota 2

Madhukara S Holla, Sathkrith
1st year MSCS, Northeastern University

October 26, 2023

Problem description

Dota 2 is a popular online multiplayer game where strategy, individual player skill, and resources play a pivotal role in the outcome of each match. While the objective of the game is to destroy the enemy team's base, achieving this goal is a complex endeavor with strategic decisions and team dynamics.

Each player controls a hero, a unit with unique abilities and attributes that evolve as the game progresses. A team consists of five players, each assuming a distinct role in the game. "Carry" is responsible for dealing damage, "Support" facilitates the growth and provides vision, "Mid/Off-laner" tackles various objectives depending on the game's situation. The efficiency and success of each role are tied to the in-game resources they acquire. These resources, primarily gold and experience, dictate a player's progression, item acquisition, and overall strength relative to their opponents.

This project seeks to answer a crucial question:

> Does a specific distribution of in-game resources among different team roles correlate with a higher likelihood of victory?

Furthermore, can we determine an optimal strategy for resource allocation to maximize the chances of winning?

Summary of the Dataset

The dataset contains several metrics that are essential for understanding the dynamics of each game:

Experimental Unit: Match

- Player roles: Identification of each player's role
- Heroes picked: Specific heroes chosen by players
- Player rank: Indicative of player skill level
- Resource metrics: Timestamps associated with gold, experience, net-worth, objectives destroyed, etc.
- Match duration and outcome: Context about the pace and result of the match

Sampling method

1. Consider a population of games from the same skill bracket, and the same patch.
2. Randomly sample 2500 games and pick the winning team, and then randomly sample 2500 games and pick the losing team. (Do not consider both teams from the same match).
3. Consider the resource distribution at the 10 minute mark for each game.
   (First 10 minutes of the game is the most balanced and has the most impact).

Proposed methods

The primary methodological approach proposed is Multiple Logistic Regression with the match outcome as the response variable.

Predictors

- Player roles
- Resources acquired by each role (Gold, Exp)
- Interaction terms - Role x Resources

Additional predictors to be considered: Objectives captured by the team, Heroes picked by a role.

Methodological difficulties

- Multicollinearity between predictors due to player interactions and synergies.
- Random effects in the predictor - heroes picked, items acquired, etc.
- High variance in the predictor - The game features 124 distinct heroes and over 100 items. Neglecting to include specific variables for each hero or item can introduce significant variability into the predictors.

These problems can be resolved by restricting the analysis to a specific role such as "Support" or "Carry". We plan on finalizing the number of predictors after exploratory analysis of the dataset.

Timeline

- 03.Nov.23 - Finish exploratory analysis of the dataset and verify model assumptions.
- 10.Nov.23 - Verify initial results and finalize the predictors.
- 17.Nov.23 - Finish statistical analysis and explore the results.
- 01.Dec.23 - Finish hypothesis testing and validation.
- 11.Dec.23 - Report submission

References

Dataset sourced from `https://stratz.com/`