Data Science Analysis of Strategies in Counter-Strike: Global Offensive

Michael Bejma
Satwik Chodisetti
Matthew Fleming

Summary

Counter Strike: Global Offensive (CS:GO), developed by Valve Corporation is a massively-popular competitive first person shooter online video game wherein teams of 5 play against each other in an attack/defend scenario (terrorists vs. counter-terrorists). CS:GO is designed to have few to no random elements, meaning talented players can leverage their understanding of fundamental mechanics and teamwork in order to defeat the other team over (usually) 30 rounds of gameplay, wherein the first to 16 round wins will be the victors of the match.

We will be closely investigating factors such as the item economy, team strategy, and skill of players in order to determine what contributes to a victory on a per round, and ultimately per match basis. What we intend to include will be outlined in the below section on our goals.

Goals

Our goals are to model several aspects of the game, and determine how they affect round wins. We will also be examining how the player's rank (determined by their performance in previous games) affects things such as strategies used, as well as general effectiveness in combat.

Our goals include in no particular order:

- Heat maps of action across the various maps provided in the dataset, and a *k-means* analysis to determine the 'hottest' areas.
- A *regression* and *RFE* analysis and correlation matrices of factors such as rank, economy, and use of certain weapons (for example, whether players with a higher rank make more effective shots). The best of these factors will be collected and used for the *k nearest neighbour analysis*.
- Creating algorithms to identify strategies such as bomb site rotation, grenade throws, attacker pushes intended as diversions, and economy management strategies. Then, determining what skill levels more properly utilise these strategies to their advantage.
- Performing a *k nearest neighbour* analysis of factors derived from the earlier *regression* and *RFE* analyses in order to determine what extent these factors contribute to the outcome of the round. This will require creating a new dataset, derived from the other datasets.

Summary of The Datasets

Given the digital medium of CS:GO, obtaining accurate datasets is an automatic process. Our data was obtained via Kaggle. The author of the datasets says the data was extracted through official competitive replays found on the website https://csgo-stats.net/ after the September 2018 game update. The dataset includes a wealth of information taken from some 1400 matches (some 410,000 rounds).

Included in the Kaggle are .csv datasets for damage, grenades thrown, as well as map data. The datasets for damage and grenades thrown include entries which provide information about game state, and the players who performed the action. These datasets will be used in conjunction with the map dataset in order to create visualisations of the strategies employed by the teams. We intend to do a *k nearest neighbour* analysis derived from the results of our *regression* analysis. All of this will require some data preparation, and interaction between datasets.

Techniques

We intend to use *regression*, *RFE, k-means*, and *k nearest neighbour* techniques to achieve what has been outlined in the goals section above. *Data visualisation* through graphics is also very important to the analysis given the three-dimensional nature of the game. *Regression* and *RFE* will be used to determine what factors go towards the *k nearest neighbour* analysis.

Project Plan

Our project plan is as follows:

- Perform analysis on player rank vs. game factors. This would entail the initial *regression* and *RFE* analyses. This will be done early in order to prepare for the *k nearest neighbour analysis*.
- Create heat maps/visualisation of action. This is where the *k-means clustering* and visual representation of action/grenade throws would be performed.
- Determine strategies and model their effectiveness on outcome of the round/match. Algorithms to identify strategies will be implemented. This and the analyses outlined above will contribute to a *k nearest neighbour* analysis, in order to draw a conclusion as to what factors most strongly contribute to success.

Preliminary Exploration of The Data

Preliminary exploration of the datasets have shown that the datasets complement one another rather in a tidy manner. Similar analyses of the same datasets (with different goals) are available on Kaggle, which will also serve as a useful resource if problems are to arise.