CAP5610 – Introduction to Machine Learning

Match Prediction Accuracy in *Dota 2*

Final Project Proposal

Nickolas Barran

Nestor Maysonet

Abstract

In this proposal, we outline a project that trains machine learning classifiers and compares performance in predicting match results in *Dota 2*, a popular online game. We will use a dataset of 50,000 ranked competitive matches to predict which team will win based primarily on character selection. A possible direction is to also train a classifier on the full game state 5 minutes into the game and predict which team wins or predicting match outcomes solely using individual player statistics to determine which side is favored. The project will involve preparing the *Dota 2 Matches* dataset and selecting features to train an efficient and accurate classifier.

1 Introduction and Problem Statement

Dota 2 is an incredibly prominent and lucrative Multiplayer Online Battle Arena (MOBA) game in the vein of *League of Legends* and *Smite*. Players control characters dubbed 'heroes' and compete in two teams of five to destroy each other's structures. Each team has a stream of computer-controlled 'creeps' that automatically charge the enemy and attack opposing units. To win, players kill enemy heroes and creeps to gain experience, which levels their heroes, and increases their gold, which is used to purchase items. Items have a wide-ranging variation of effects; their purchase order and selection can have great impact on a match. It is no exaggeration to say that matches worth millions of USD have been lost based on hero selection alone.

Valve, the proprietor of *Dota 2*, already has a machine learning system in *Dota Plus*, a subscription service which provides match analysis, hero and item recommendations, and detailed statistics on player performance. In order to recommend hero choice and items, a match predicting algorithm is required. By training classifiers and evaluating their performance, we can discover the likely methods that Valve uses to build their statistics service.

In other words, we want to compare classifier performance to see if we can build a relatively efficient and accurate classifier like what Valve or other research uses.

2 Motivation

Dota is a highly complex and difficult game that has attracted much attention from statisticians and computer scientists. *OpenAI* created the *OpenAI Five* by having their bots play over 10,000 years of games via Deep Reinforcement Learning. Two-time world tournament *The International* champions OG members, Sebastien Debs and Johan Sundstein directly credited analyzing OpenAI games for inspiring changes in their teamplay that led to their second championship win at *The International 9*[1]. Even Bill Gates tweeted that *OpenAI*'s achievements were a milestone in artificial intelligence[2]. Observations while training machine learning classifiers on *Dota* can probably be generalized to other MOBAs or even other team-based competitive games such as *Counter-Strike*. In a world where esports are raking in ever-increasing revenues and professional teams turning more and more to statistical analysis, match outcome prediction is highly desirable.

3 Hypothesis

It is hypothesized that using Deep Learning, like *OpenAI*, or TensorFlow, which is likely utilized by *Dota Plus*, will result in a trained classifier with better accuracy than the Logistic Regression, PCA, KNN, and Random Forest classifiers used in the research as noted by Semenov et al.[3]

4 Methodology

Implementing a neural network and TensorFlow then comparing performance on the Dota 2 Matches dataset. The features should be limited to hero selection and/or individual player metrics to determine match outcome before it is played. By tuning different hyper-parameters and comparing performance with accuracy as the primary metric it is possible to evaluate which approach results in a better classifier.

5 Expected Results

We expect to find an accurate model that can predict the outcome of a match based on primarily character selection prior to the start of a match. By observing matchups, team composition, and other features, the model should predict whether a team is going to win based on past matchups and game results.

6 Related Works

Kinkade and Lim explored two predictors, one using the full post-match data and the other only using hero selection data. They found that the first predictor had no real application as

[1] https://openai.com/projects/five/
[2] https://twitter.com/BillGates/status/1011752221376036864
[3] http://ceur-ws.org/Vol-1842/paper_05.pdf

using data about who already won would contradict its use. Their second predictor resulted in a 73% accuracy by using an additional factor of hero countering.

A paper by Berner et al. utilized thousands of GPUs over a span of months to build a distributed training system known as OpenAI Five. This AI system was the first one capable of defeating the world champions of Dota 2. They opened OpenAI Five to the Dota 2 community for competitive play and it won 99.4% of over 7000 games.

Jack Scott explored item prediction using a fully connected neural network with two hidden layers. The features used involved the current game time and the players current items, hero selection and skill rating to determine which item they would purchase next. Using a ROC curve he detected about a 60% true positive rate and a 10% false positive rate.

7 Project Timeline

Week 1- Clean and prepare dataset

Week 2 – 5 – Develop training models and compare results

Week 6 – Complete write up and submit

8 References

Kinkade, Lim (2015) Dota 2 Win Prediction
http://jmcauley.ucsd.edu/cse258/projects/fa15/018.pdf

Berner, et al. (2019) https://arxiv.org/abs/1912.06680

Jack Scott (2018) https://cs230.stanford.edu/projects_spring_2018/reports/8290559.pdf