Mid or Feed
A Role Classification Model for Dota 2 Players

Angelo Vincent Delos Santos, Luis Gabriel del Rosario, Edrene Bryze Miranda
Department of Computer Science
College of Engineering
University of the Philippines, Diliman

***Abstract*—Dota 2 is a well-developed computer game loved by millions of players. The game has evolved quickly through time from being a casual time-killer to a systematic and cross-country competition. In the game, players get to choose a hero to pit against the opponent. These heroes, later on, has been classified to Core, Support, Offlane, etc. The goal of this study was to develop a model that could assess a players skill set and predict what type of hero one should train more. Given this, the team used different machine learning techniques to determine a players recommended role. After a few test runs on different techniques, the team has found that artificial neural networks with hyperbolic tangent as its activation function produced the best results compared among others. Future applications and development of this project is to extend its features to determine a players weakness and widening the scope not just the two main roles but for the other minor ones, too.**

I. Introduction

Defense of the Ancients is a long-running MOBA game that has been played by millions of players all across the globe. The objective of the game is to destroy the opposing teams ancient. At the start of the game, a 5-man team is organized and each player of the team chooses their respective heroes to help achieve the games goal. These heroes are then, through time, classified according to their capabilities.

The issue of choosing a hero based on their role has been bothering the games players through the years. Some players tend to develop an inclination towards certain heroes and certain roles, to the point that they decline to select heroes vital to the teams lineup just because they dont want to - a phenomenon known to many as cancerous. Oftentimes, they tend to get too comfortable with these options that they refuse to broaden the pool of heroes they know how to play. The team, with two of its members actively playing DOTA 2, decided to tackle this issue.

This project aims to help players decide which class of heroes to play, whether it be for training or, if they feel like theyve trained enough, for actual competitions. This will hopefully ensure that each player will select the best hero for the lineup. The project also aims to gauge a players capacity to play certain classes of heroes and provide information in which they are most effective at. Through this, a players potential and capability will be utilized and maximized for the players enjoyment and their teams victory.

II. Short of Review of Related Studies

Egger et. al. used logistic regression to classify the role of the player within a specific game, given the replay data of the game itself[1]. They observed the players style in the game in order to determine what role they were taking up at that moment. They were able to get a dataset of 708 labeled games from the online DOTA community. Based on the attributes of the game, which includes kills, deaths, and assists, their logistic regression model yielded an accuracy of 96.15%.

Gao, et. al., used a similar method to predict either the hero itself or the heros role, based on game replay data[2]. Their logistic regression model yielded an accuracy result of 89% for professional matches, and 84% for public matches.

These methods made use of game replay data, which has much more information as compared with basic match data. These two groups also showcased a deep understanding of the games mechanics. Moreover, the groups focused on predicting the role of a player in the context of a match, and not as a binding classification for the player themself. Our team aims to create a model that makes use of the most recent matches of a player in order to determine the role the player most likely will fill.

III. Methodology and Results

The dataset used was the set of professional players from the OpenDota API. The data fetched consisted of a total of 1198 players with their fantasy role, the class that they are classified into, and the heroes that they picked for their 20 most recent games.

The dataset was prepared in such a way that the roles are represented as integers instead of strings: 0 for versatile, 1 for core, and 2 for support. These roles will be the labels for classification. The feature vector is taken from the 20 most recent games of the player; it includes the ID (integer) of the hero used for that game, as well as an indicator whether the players team won or lost.

First, a list of account IDs was taken from the list of professional players in the website OpenDotas database, as well as their role. Most of the players had roles incorrectly labeled as 0, and so the team had to manually cross-check the actual role online. From there, a script was made to get the 20 most recent games of the given account ID, take the necessary information (namely, the hero ID and whether the player won or not) and store it in a local file.

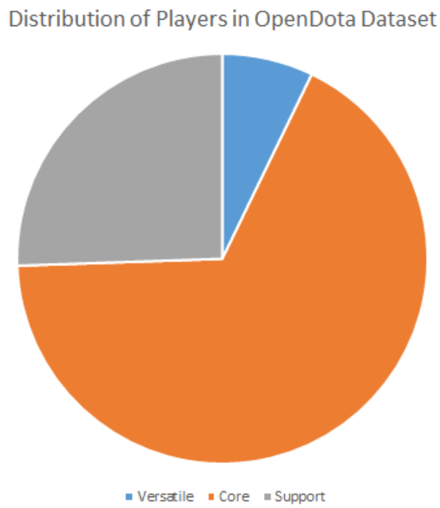

Fig. 1. Distribution of players in the OpenDota dataset

From here, different machine learning algorithms were tested, to see which one yielded the most favorable result. The team wanted to use that which produced the highest and most balanced accuracy for each different classification. Hence, the confusion matrix was used, in order to easily compute for the accuracy of each classification.

The team tried Naive Bayes, Support Vector Machines (SVM), and Artificial Neural Networks (ANN), and messed around with different activation functions (tanh, logistic) and kernels (poly, rbf). It turns out that using an ANN with tanh as its activation function yielded the most favorable result, as is shown below:

$$M_{train} = \begin{bmatrix} 17 & 0 & 0 \\ 0 & 465 & 0 \\ 0 & 0 & 237 \end{bmatrix} M_{test} = \begin{bmatrix} 0 & 48 & 21 \\ 2 & 281 & 58 \\ 0 & 54 & 16 \end{bmatrix}$$

Note that the rows of the matrix represent the actual label (0, 1 and 2, respectively), and the columns represent the predicted label. The other models tried are as follows:

ANN with logistic activation function:

$$M_{train} = \begin{bmatrix} 16 & 1 & 0 \\ 0 & 465 & 0 \\ 0 & 0 & 237 \end{bmatrix} M_{test} = \begin{bmatrix} 3 & 47 & 1 \\ 5 & 269 & 67 \\ 0 & 50 & 20 \end{bmatrix}$$

Multinomial Naive Bayes:

$$M_{train} = \begin{bmatrix} 16 & 1 & 0 \\ 0 & 465 & 0 \\ 0 & 0 & 237 \end{bmatrix} M_{test} = \begin{bmatrix} 1 & 52 & 16 \\ 8 & 271 & 62 \\ 2 & 44 & 24 \end{bmatrix}$$

SVM with Polynomial Kernel:

$$M_{train} = \begin{bmatrix} 17 & 0 & 0 \\ 0 & 465 & 0 \\ 0 & 0 & 237 \end{bmatrix} M_{test} = \begin{bmatrix} 0 & 68 & 1 \\ 0 & 338 & 3 \\ 0 & 64 & 6 \end{bmatrix}$$

Default SVM:

$$M_{train} = \begin{bmatrix} 17 & 0 & 0 \\ 0 & 465 & 0 \\ 0 & 0 & 237 \end{bmatrix} M_{test} = \begin{bmatrix} 0 & 69 & 0 \\ 0 & 341 & 0 \\ 0 & 70 & 0 \end{bmatrix}$$

In most cases, versatile players were classified wrongly for the testing datasets, with most of them being classified as ”Core”. This goes to show that the representation within the dataset for versatile players isnt as distinct as with other classifications. This may stem from the fact that this part of the dataset was manually encoded.

Moreover, as seen in Figure 1, the distribution of the dataset was completely unbalanced; the core players greatly outnumbered the support and versatile players. This leads to the relatively low accuracy of the model when classifying support players, especially for the SVMs, which are known to do well even with uneven datasets. This might be a sign that the role of a player cannot be easily determined by the games they have played. However, statistical tests involving correlation and causation are within the scope of this paper.

IV. Conclusion

At the moment, the accuracy of the tool isnt high enough to be deployed. Players will more often than not be classified as ”Core” than as ”Support” or ”Versatile.” Most Dota players wouldnt find this tool useful due to the possible misinformation it might give. Future research may involve more advanced models and a larger dataset for tighter training.

It is possible that the features selected for the model are too few, and that there could be more information about the match or the player that are involved in the classification of the players role. In line with this, further analysis on the matches of the game, as well as comparison with other games (and other studies done on said games), may be done.

References

[1] C. Eggert, M. Herrlich, J. Smeddinck, and R. Malaka. (2015). *Classification of Player Roles in the Team-based Multi-player Game Dota 2.*

[2] Gao, L., Judd, J., Wong, D., and Lowder, J. (2013). *Classifying Dota 2 Hero Characters Based on Play Style and Performance.*