DOTA 2 Win Prediction

Nicholas Kinkade
University of California, San Diego
9500 Gilman Dr
La Jolla, CA 92093
nkinkade@eng.ucsd.edu

Kyung yul Kevin Lim
University of California, San Diego
9500 Gilman Dr
La Jolla, CA 92093
kyl063@eng.ucsd.edu

ABSTRACT

In this paper, we present two win predictors for the popular online game Dota 2. The first predictor uses full post-match data and the second predictor uses only hero selection data. We will explore and build upon existing work on the topic as well as detail the specifics of both algorithms including data collection, exploratory analysis, feature selection, modeling, and results.

1. INTRODUCTION

DotA 2, or Defense of the Ancients 2, is an online multiplayer game created by Valve. It has a user-base of over 8 million players and competitions with prize pools of up to $10 million.

A DotA 2 match consists of a fight between two teams of five players. Both sides, Radiant and Dire, attempt to destroy one another's fortress, the Ancient. Each player controls a single character, known as a hero. Each hero has an array of unique powers which are accessed by leveling up. In order to level up, heroes must gain experience (XP) by killing enemy heroes and minions. By landing killing blows, heroes also gain gold. This gold can be used to purchase helpful items.

Currently, there are 110 different heroes to choose from. Once one hero is chosen, other players may not choose that same hero. The choice of heroes plays a large role in determining the match outcome. Every hero has different strengths and weaknesses. These strengths and weaknesses can serve to complement other heroes on the friendly team or counter heroes on the opposing team. In general, hero roles can be divided into the following:

- Carry - Very item and experience dependent hero. Becomes strongest when game reaches long duration.

- Support - Hero strength is less dependent on items or experience. Helps protect the carry.

- Initiator - Hero that has an ability with which to begin strong team fights.

- Ganker - Hero which has the ability to single-handedly kill enemy heroes.

A team missing any of the above roles will generally struggle in some important aspects of the match.

There appears to be two interesting machine learning problems associated with a Dota 2 match. The first is hero recommendation, recommending a hero at each step of the picking process to maximize the probability of victory. The second is win prediction, predicting which team will win based on some given match data. The first problem appears to be heavily dependent on the second. A hero recommendation system can only be successful if it depends on an accurate win predictor. Because of this, we chose win prediction as the primary interest of this paper.

2. RELATED WORK

Dota Picker[1] is an application which uses specified metrics for hero recommendation. The two available metrics are hero advantages (matchup between two heroes) and win rates. Unfortunately, the creators have not divulged specific details of their approach nor their accuracy.

DotaBuff[2] is a website that provides various DotA 2 statistics. Currently, it does not provide any form of win prediction or hero recommendation, but their various data visualizations are useful.

The paper `How Does He Saw Me?`[3] describes a DotA 2 hero recommendation system which depends on a win predictor. It specifies two possible models for win prediction. The first uses logistic regression with a binary hero feature vector. The second uses K-nearest neighbor classification using a custom weight to specify distance between teams. Accuracy of first model is 69% and accuracy of the second model is 68%.

The paper `To Win or Not to Win`[4] describes a DotA 2 win predictor. It specifies two possible models for win prediction. The first uses logistic regression with a binary hero feature vector. The second combines this predictor with a genetic fitness metric, weighting each predictor equally in the final prediction. Accuracy of the first model is 69% and accuracy of the second model is 74%.

3. DATA SET

Data for 62,000 matches was collected using the Steam Web API[5] over the period of 11/20/2015 to 11/22/2015. The following information was collected for each match:

- Winning Side

- Duration

- For Every Player:
  - XP Per Minute
  - Gold Per Minute
  - Kills
  - Assists
  - Deaths

Only matches with the following requirements were included:

- The skill level is 'very high'. This allows us to avoid player skill skewing expected hero dynamics, and also ensures the players all have a basic understanding of the game already, improving our prediction performance.

- The game mode is 5v5. Other modes such as 1v1 do not represent the intended game style of DotA 2.

- Players are present throughout the full duration of the match. Otherwise, the match is heavily favored against the team with leaver(s).

- The game lasts at least 10 minutes. Very short games are likely the result of one team purposely losing.

The data came in sorted by time of match, and was shuffled so that we don't overfit to a particular timezone or region of the world.

Then we separate the data into 52,000 training matches, 5,000 validation matches, and 5,000 test matches.

4. EXPLORATORY ANALYSIS

4.1 Hero Analysis
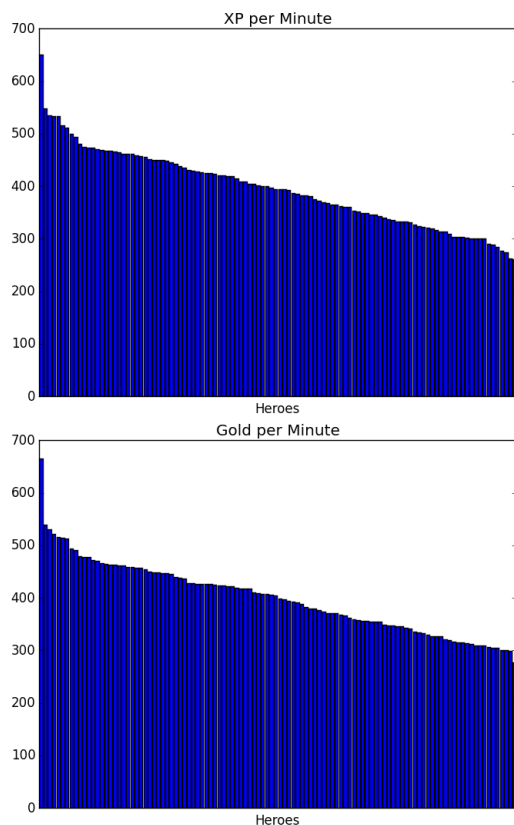
4.1.1 Hero XPM and GPM

Figure 1: XP, Gold Per Minute Bar Graphs

XP per minute and gold per minute appears to vary widely between different heroes. It is generally the case that a hero with high XPM will also have high GPM, and a hero with low XPM will also have low GPM. At the top of the lists, we have heroes which generally fill the role of carry: Meepo, Alchemist, Templar Assassin, and Shadowfiend. At the bottom of the lists, we have heroes which generally fill the role of support: Io, Lion, Techies, and Ancient Apparition. Since Dota 2 is a game highly dependent on team synergy, XPM and GPM are not a direct indicator of hero strength but rather hero role. It may, however, be the case that XPM and GPM can give us some indication of a hero's relative strength to other heroes within its role.
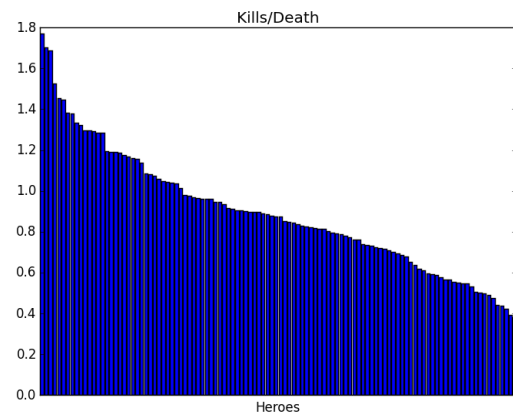
4.1.2 Hero Kills per Death

Figure 2: Kills Per Death Bar Graph

Kills per death also appears to vary widely between different heroes. At the top of the list, we have heroes which generally fill the role of ganker: Ursa, Templar Assassin, and Riki. At the bottom of the lists, we have heroes which generally fill the role of hard support: Keeper of the Light, Io, and Dazzle. Kills per death is not a direct indicator of hero strength but rather hero role. Heroes which have lower XPM, GPM, and KD are generally hard support which aim to boost teammates XPM, GPM, and KD through strong buffs and heals. A good team composition is one which has a spread of different roles so that each hero may feed the strengths of the others.

4.1.3 Hero Game Duration

With a few exceptions, average game duration does not vary significantly across different heroes. Similarly, average won game duration also does not vary significantly. Hero outliers include Nature's Prophet and Lycan. Their game durations generally fall shorter than that of the average hero since their spells allow for strong early game gold advantage via pushes. This metric does not appear to be extremely relevant to our goal of win prediction since it only applies to a small subset of heroes.
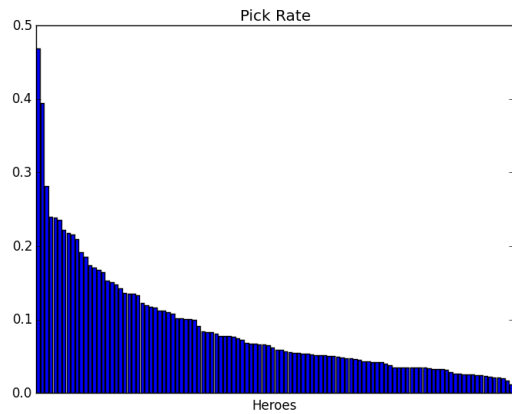
4.1.4 Hero Pick Rate

Figure 3: Pick Rate Bar Graph

Pick rate appears to vary widely between different heroes. The most picked heroes are Shadow Fiend, Windranger, and Juggernaut with pick rates of 47%, 40%, and 28% respectively. The least picked heroes are Naga Siren, Elder Titan, and Chen each with a 2% pick rate. Pick rate should roughly correlate to players' opinion on hero strength, however, players may be biased towards easy to play heroes, fun heroes, or heroes with a specific role. Pick rate is therefore not an extremely reliable indication of hero strength. As will be discussed in future sections, it does appear that heroes at the bottom of pick rate are generally weak within their designated roles. Pick rate also gives us an idea of the spread of hero data. Since the least picked heroes have a pick rate of 2% and we have a data set of 62,000 matches, we still have a reasonable amount of data to base our assesment of these heroes (1200+ games).
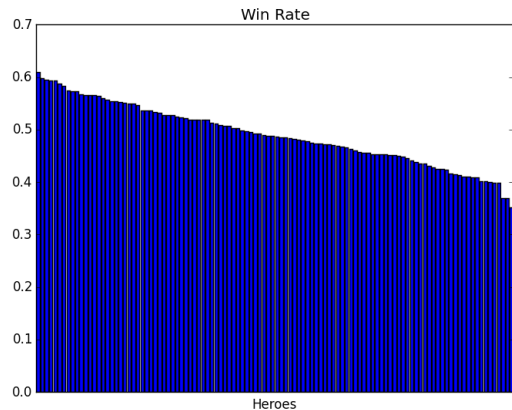
4.1.5 Hero Win Rate

Figure 4: Win Rate Bar Graph

Win rate appears to vary widely between different heroes. Win rate varies evenly between 36% and 60%. At the upper end of the spectrum, we have Lycan, Omniknight, and Undying with 60% win rates. At the lower end of the spectrum we have Storm Spirit, Naga Siren, and Enchantress with 36% win rates. From this analysis, we can conclude that individual hero selection has a strong impact on win likelihood, though some heroes have a stronger correlation than others. In the next section we will explore how hero pairings affect win rate.

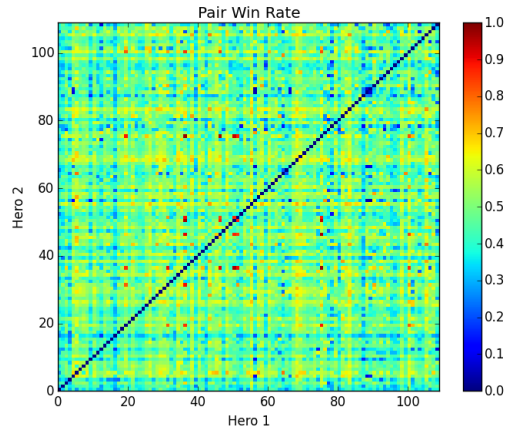### 4.1.6 Hero Pair Win Rate

Figure 5: Pair Win Rate Heat Map

The above heat map displays the win rate when two heroes are on the same team. While much of the graph lies close to 50%, there are a significant amount of outliers in both the positive and negative direction. This indicates that while individual hero picks can have an affect on overall win rate, hero pick combinations are also important. The pairings with high win rates indicate strong hero synergy. As an example, the top three pairings are (Beastmaster, Lycan), (Luna, Lycan), and (Beastmaster, Luna) with win rates of over 90%. These three heroes have damage auras which can stack to give all team members high damage. This is just one example of hero synergy, other forms can be seen when looking at other high win rate pairs. At the opposite end of the spectrum, we have hero pairings which indicate anti-synergy. As an example, the three bottom pairings are (Keeper of the Light, Io), (Enchantress, Keeper of the Light), and (Enchantress, Io) with win rates of below 10%. Each of these heroes are considered hard support. Having more than one hard support on one team results having very few heroes which can actively contribute to fights (other than just providing heals and buffs).

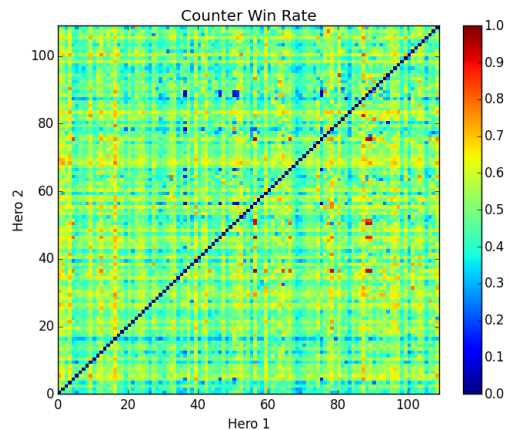### 4.1.7 Hero Counter Win Rate

Figure 6: Counter Win Rate Heat Map

The above heat map displays the win rate when two heroes are on opposing teams. While much of the graph lies close to 50%, there are a significant amount of outliers in both the

positive and negative direction. This indicates hero counter picking is an important aspect in win rate. The pairings with high win rates indicate strong hero countering. As an example, the top three pairings are (Lycan over Io), (Centaur Warrunner over Meepo), and (Leshrac over Shadow Demon) with win rates of over 78%. In the first case, Lycan is very fast and does significant damage while Io is slow and fragile. In the second case, Centaur Warrunner does damage in large areas causing him to destroy Meepo since Meepo consists of multiple characters. In the third case, one of Shadow Demon's spells which can be cast on friendly heroes causes Leshrac to be set up to do immense damage. Further examples of hero countering can be seen at higher counter win rates.

4.2 Match Analysis

To get a more holistic understanding of the game we analyzed the Gold/min, Exp/min, and Kills/min of each side as well as their differences on each side.

Of the 62000 matches, 2792 were pre-filtered to remove non-relevant matches. These were matches that had duration lower than 10 minutes, or was not played on the standard modes(5v5 on All Pick, All Draft, All Random, Captain's Mode, Captain's Draft, Single Draft, Random Draft).

Before exploring specific per-minute data, it should be mentioned that the average duration of a match was 30 minutes and 53 seconds, with a standard deviation of 6 minutes and 42 seconds, and the radiant side's win rate was 56.5%.

4.2.1 Gold-Per-Minute Analysis

Radiant's average GPM was 2141.42 with a standard deviation of 527.39. On the Dire side, the average was 2014.27 and standard deviation was 544.36. The Radiant's average being higher is consistent with the higher Radiant win rate, and the hypothesis that the side with the higher GPM at the end of the game would have won, simply by strength in resources.

We can count exactly how many games each side won with a lower GPM than the other side. Out of the 59208 matches, Radiant won 248 matches(0.42% of total games) with less GPM than the Dire, and Dire won 262 matches(0.44% of total games) with less GPM than the Radiant, for a total of 0.86% of matches in which a side won in spite of being down on GPM.

Figure 7 is a scatter plot and a Logistic Regression Fit on how GPM differences creates a clear divide on which side won.

4.2.2 Exp-Per-Minute Analysis

Much like the GPM analysis, the hypothesis is that the side with the higher XPM at the end of the game would have won. Radiant's average XPM was 2101.33 with standard deviation 516.52, and Dire's was 2016.65, with standard deviation 544.65. Again, on average Radiant—the side with higher average win rate—had the higher average XPM.

The abnormal cases where a side with lower XPM than the other side won was 355, or a negligible 0.60% of total games. Radiant won 221(0.37% of total games) of those games, whereas Dire won 134(0.23% of total games) of those games. Curiously, this number is lower than the GPM, perhaps suggesting experience differences are harder to surmount than gold differences, though we would have to gather more data to prove that hypothesis.
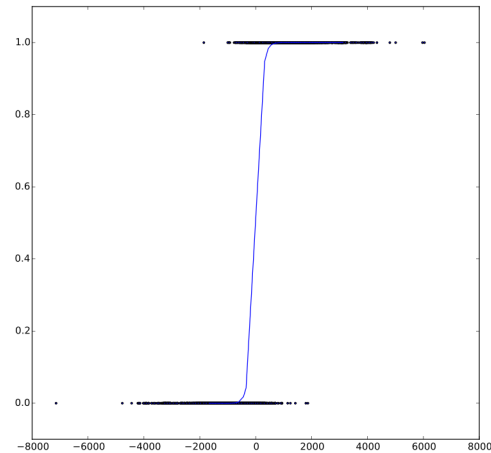

Figure 7: Radiant GPM - Dire GPM plotted with Radiant Win(1) or Loss(0) with Logistic Regression Fit

Figure 8 is a scatter plot and a Logistic Regression Fit on how XPM differences creates a clear divide on which side won.

4.2.3 Kills-Per-Minute Anaylsis

KPM analysis is slightly different as the number scales are different, and it doesn't directly reflect resource differences between two teams, but the results look alike with GPM or XPM. Radiant had a higher than Dire KPM of 0.9501 and standard deviation of 0.4602, whereas Dire had an average 0.8690 KPM, and a standard deviation of 0.4430.

The times a team won with a lower KPM than the other side was 1111, much higher than the abnormal GPM or XPM cases. This accounts for 1.88% of total games, a small but non-trivial amount of games. These games are divided between Radiant and Dire as 679(1.15%) and 432(0.73%) respectively.

The higher number of anomalies for kills per minutes can be accredited to the fact that kills don't directly reflect the resource gained and lost on each side. Even though kills are a great way of gaining and denying resources, it is not the only way to do so. For example, there is a common type of strategy that sets up a situation in which even if the team bleeds away kills, they gain a net positive in resource elsewhere on the map and eventually win the game, referred to as "ratting" by the dota community.

Figure 9 is a scatter plot and a Logistic Regression Fit on how KPM differences creates a clear divide on which side won.

These three stats will be used in our first predictor of trying to classify who won given gpm, xpm, and kpm data.

5. PREDICTIVE TASKS

We created two different win predictors. The first win predictor uses full post-match data and the second predictor uses only hero selection data. The first win predictor
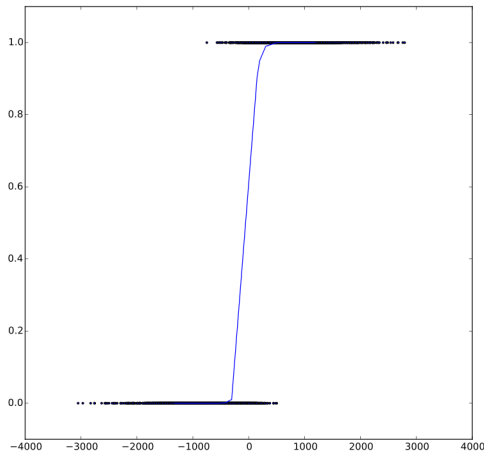
Figure 8: Radiant XPM - Dire XPM plotted with Radiant Win(1) or Loss(0) with Logistic Regression Fit

Figure 9: Radiant KPM - Dire KPM plotted with Radiant Win(1) or Loss(0) with Logistic Regression Fit

doesn't appear to hold any real world use, but we were interested in seeing to what extent a win is dependent only on provided post-match stats. If successful, the second win predictor should allow for the creation of a robust hero recommendation algorithm.

6. WIN PREDICTION (POST GAME DATA)

Since the outcome of the match is already known after a match has concluded, a predictor using post game data doesn't immediately have a use. This prediction was done to serve as the base of the options for the various models we have as well as validation of our exploration.

6.1 Feature Selection

Given the exploratory data from above, it seems all GPM, XPM, and KPM at the end of the match are clear indicators of who won, and therefore they should be all great features to utilize. They are all real number values that can be directly plugged into a classifier model. The reason we want to use the normalized per-minute value as opposed to the total gold or experience gained in a game is because longer games will naturally have a large number and this adds a confounding factor to our models where longer games will be predicted differently than shorter games.

However, duration of the match still matters as in Dota strategy discussions, there is often said to be a factor that the Dire side gains an advantage in the later portions of the game, as they have an easier approach to the boss non-playable character Roshan. Therefore we include the duration of the game as a feature as well.

There is the danger of double counting since KPM affect GPM and XPM directly, and also all actions that gain experience points(such as killing creeps or heroes) always give gold. This is accounted by using a model that avoids double counting, as detailed in the next section.
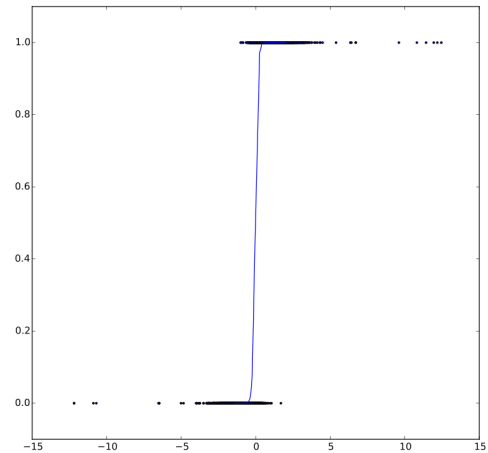
6.2 Model

Since GPM, XPM, and KPM are in fact all dependent with each other, we cannot directly use a Linear Regression model. Because the end goal is to classify a binary win(1) or loss(0) state, we can use a Logistic Regression model that will classify the Radiant side winning as 1, and the Dire side winning as 0.

Another model to consider is the Random Forest Classifier. The Random Forest Classifier will pick from an ensemble of trees that randomly pick between the different features to use. GPM and XPM are very similar, but there is a slight concern that the KPM is actually not a direct representation of a team's resource. For this reason, a Random Forest Classifier that has estimators without KPM might improve our accuracy.

6.3 Results

As can be seen from Table 1, our predictor reaches mostly perfect accuracy. Due to this, there wasn't much validation tuning required; the Logistic Regression used $\lambda = 1$, and the Random Forest Classifier used 50 estimators. Such a high accuracy can be attributed to the fact that the features are numbers after a match has concluded and it should be very easy to find out who won after someone has won the match. Further, since DotA is in many ways a resource game, these resource numbers directly factor into the game's outcome.

Perhaps a more interesting predictor will use data either in the start or middle of the game.

7. WIN PREDICTION (GIVEN PICKS ONLY)

At the start of a match, only the Dire/Radiant side and the heroes on each side is known. Therefore the prediction becomes a harder problem and we explore this in this section.

7.1 Feature Selection

Features	Logistic Regression	Random Forest Classifier
GPM	99.15%	99.08%
XPM	99.45%	99.40%
KPM	97.83%	97.38%
GPM + XPM	99.58%	99.78%
GPM + KPM	99.15%	99.65%
XPM + KPM	99.45%	99.37%
GPM + XPM + KPM	99.58%	99.81%

Table 1: Varying features and their effect on both of the models' accuracies

Our initial features are nearly identical to [3]. We start with an offset feature:

$$X_0 = 1$$

This feature should allow the model to consider the general advantage of Radiant over Dire. There are currently 110 heroes. We represent a matchup via binary features corresponding to which heroes are on the Radiant and Dire side as follows:

$$X_{1+i} = \begin{cases} 1 & \text{if hero } i \text{ is on Radiant side} \\ 0 & \text{otherwise} \end{cases}$$

$$X_{111+i} = \begin{cases} 1 & \text{if hero } i \text{ is on Dire side} \\ 0 & \text{otherwise} \end{cases}$$

These features should allow the model to consider the individual impact of each Radiant and Dire hero on a match outcome. In order to take into account hero synergies, the following feature was constructed. Let R represent the heroes on the radiant side, D represent the heroes on the dire side, and S represent synergy. S_{ij} can be defined as the win rate when hero i and hero j are on the same team (see Figure 5). Hero synergy on the radiant side can be defined as

$$S_R = \sum_{i \in R} \sum_{j \in R, i \neq j} S_{ij}$$

Similarly, hero synergy on the dire side can be defined as

$$S_D = \sum_{i \in D} \sum_{j \in D, i \neq j} S_{ij}$$

We can then construct a single feature which represents the difference between hero synergy on the radiant and dire side:

$$X_{221} = S_R - S_D$$

Our hope is that this synergy feature will also capture the notion of role distribution within a team. Generally, support and carry heroes will have high synergy since one benefits the other. By using team synergies as a feature, good team role distribution should be inherently rewarded while poor team role distribution should be punished. In order to take into account hero countering, the following feature was constructed. Let C represent countering. C_{ij} can be defined as the win rate when hero i is playing against hero j (see Figure 6). Hero countering of the radiant side over the dire side can be defined as

$$C_R = \sum_{i \in R} \sum_{j \in D} C_{ij}$$

It is unnecessary to calculate C_D as the information would be redundant with C_R since $C_{ji} = 1 - C_{ij}$. We can now construct a single feature which represents the hero countering of the radiant side over the dire side:

$$X_{222} = C_R$$

7.2 Model

We considered two basic models to use as predictors: logistic regression and random forest classification. After significant experimentation, we determined that logistic regression was the most appropriate model. Overfitting was not an issue with logistic regression. Training accuracy and validation accuracy are nearly identical at 73.2% and 72.9%. This is likely due to the somewhat random nature of match outcome which can be attributed to varying player skill within a match and the linear separability nature of logistic regression. For random forest classification, however, overfitting was a large issue. While we were able to get 99% accuracy on the training set, this would result in validation accuracy of 55%. Varying several parameters of random forest classification, we were able to achieve a maximum of 67% accuracy on the validation set. As logistic regression achieves a significantly higher accuracy, it follows it is the more appropriate of the two models.

The following table shows the test accuracies achieved with various combinations of the features described in the previous section:

Offset	Matchup	Synergy	Countering	Accuracy
✓				56%
	✓			64%
✓	✓			64%
		✓		66%
			✓	66%
		✓	✓	67%
✓	✓	✓		68%
✓	✓	✓	✓	73%

From this table, we can see that each feature significantly improves accuracy, with the exception of the offset feature which only appears to make fractions of a percent difference. We can see that hero individual performance (matchup feature), hero team synergy (synergy feature), and hero countering (countering feature) all have a strong impact on win rate with little overlap between the three.

7.3 Results

We use two baseline predictors for comparison. The first baseline predictor is a random predictor. As would be expected, the random predictor results in 50.1% accuracy. The second baseline predictor chooses the team which has heroes

with the highest combined individual win rate. This predictor performs relatively well with 63% accuracy. Our final predictor, described in the model section, gives an accuracy of 73%. This is a significant improvement on both baseline predictors. This can be attributed to the fact that it considers the same factors as the second baseline predictor but also considers two further dimensions of synergy and countering. Our final predictor has 4% higher accuracy than the predictor proposed in `How Does He Saw Me?`[3] and 1% lower accuracy than the predictor proposed in `To Win or Not to Win`[4]. The main difference between our algorithm and the algorithm from these two papers is the fact that we additionally consider hero countering when making a prediction. It appears that the genetic algorithm used to represent hero synergy in `To Win or Not to Win`[4] may be a more appropriate synergy metric than our proposed metric, resulting in the 1% accuracy difference.

8. CONCLUSION

We started out asking the question, can we accurately predict the win rate of a DotA match, and we've successfully answered the question by creating a predictor with 73% accuracy at the beginning of the match with only information on the hero picks.

To achieve that goal, we explored the data set from the perspective of the heroes that are in the game as well as the post-game resource stats and found interesting relationships between hero picks, pairings, and counters as well as a clear indication that DotA at its core can be seen as a resource game where having more resources lead to a win most of the time.

Our predictor was not perfect and that's because a match of DotA is not decided at the pick stage, but what happens within a game affect the outcome heavily. Therefore, more information on what happens during the game will greatly add to our predictor and improve performance.

9. FUTURE WORK

Based on the work we've done so far, we believe it is not hard to extend this into a hero recommender at the beginning of the match. Given the heroes allies have picked and the heroes the opponent have picked, we can answer the question "What hero should I pick to maximize my win rate?" This can be of great use to individuals, as well as at the full team level.

Further, in Captain's Mode where there are Bans as well as selections, the recommender can recommend bans that will maximize the opponent team's win rate, making this a useful tool for the captain who decides on the bans and picks.

The GPM, XPM, KPM predictor in this case was trivial, but we could also train on those features during a match as well. Since the Dota Web API allows gathering the match details that happened during a match, with a larger data pipeline, it can be possible to ask and answer the question "Given a certain GPM, XPM, KPM at 15 minutes into the game, what is the win rate?" This can also be asked as a resource prioritization question, in which a player might wonder "Should our team be focusing on experience at the moment, or kills?"

10. REFERENCES

[1] Dota Picker. http://dotapicker.com/
[2] Dota Buff. http://www.dotabuff.com/
[3] Conley, Perry (2014). How Does He Saw Me? A Recommendation Engine for Picking Heroes in Dota 2. http://cs229.stanford.edu/proj2013/PerryConley-HowDoesHeSawMeARecommendationEngineForPickingHeroesInDota2
[4] Kalyanaraman (2014). To win or not to win? A prediction model to determine the outcome of a DotA2 match. http://cseweb.ucsd.edu/ jm-cauley/cse255/projects/Kaushik_Kalyanaraman.pdf
[5] Steam Web API. https://developer.valvesoftware.com/wiki/Steam_Web_API