Names: Niehao Chen, Junlin Wu, Irene Jiang

Title: DOTA2 Pro Match Prediction

Abstract

In this project, we use machine learning techniques to predict the result of pro DOTA2 matches based on the heroes selected by each team after the pick and ban phase. We employed various models, including neural networks, decision trees, and random forest, to analyze historical data from DOTA2. After comparing the accuracy of each model, we found that the random forest model gave the best prediction results with an accuracy of 51.4%. We also investigated the counter relationship between heroes and found that considering this factor could improve the accuracy of the model to 58.3%. Our model's prediction is consistent and robust, and it can be used to help the DOTA2 community and players to have a more interactive game.

Keywords: DOTA2, Decision Tree, Random Forest

Introduction

The combination of heroes played in a DOTA2 match significantly influences the team's success. Various companies have tried fitting a predictive model on the winning rate based on hero selections; however, the robustness of their prediction is unknown and relationships between heroes that would influence the winning result is still undiscovered. Therefore, in this project, we attempt to predict pro match results based on the selected heroes after the pick and ban phase using machine learning techniques and find relationships between certain heroes that would affect the result, and then test the robustness of our result at last. We analyze historical DOTA2 data by using API and train our model based on the heroes' features and stats. The model can help players make informed decisions during the pick and ban phase and improve their chances of winning. For the model, neural networks, decision trees, random forest, and

XGBoost were employed to predict the winning team and we found that the best-performing model was the random forest with an accuracy of 51.4%.

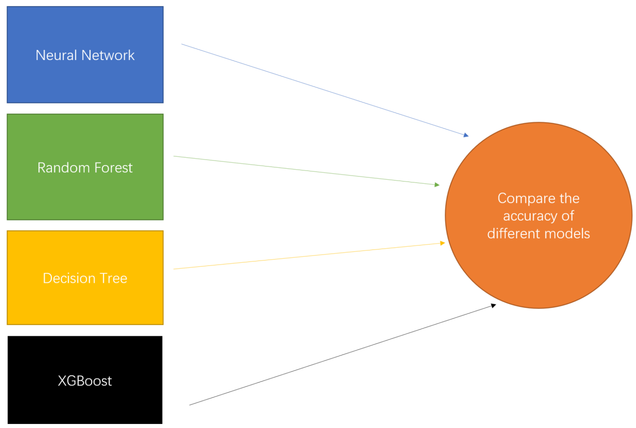
	match_id	radiant_win	region	patch	first_pick_team	selection_0_id	selection_0_base_health	selection_0_base_health_regen	selection_0_base_mana	selection_0_base_mana_regen	...
0	6974098635	False	3	51	0	100	200	1.00	75	0.5	
1	6992915881	True	3	51	1	14	200	2.00	75	0.0	
2	6085918133	True	10	48	1	47	200	0.25	75	0.0	
3	6619687384	False	13	50	1	92	200	0.25	75	0.0	
4	6092247026	True	5	48	1	47	200	0.25	75	0.0	

Methods

In order to predict the winning rate of each team based on their heroes selections, we employed various types of machine learning models to analyze historical data from DOTA2 and we want to identify the hero features that have the most influence on the outcome of a match. We experimented with neural networks, which involves converting the data into tensors and creating data loaders. Since the accuracy of predictions using neural networks is insufficient, our team also attempted to build a predictive model using decision trees, random forest, and XGBboost. We then evaluate and determine which model best suits our dataset for prediction by looking at cross validation and accuracy of each model.

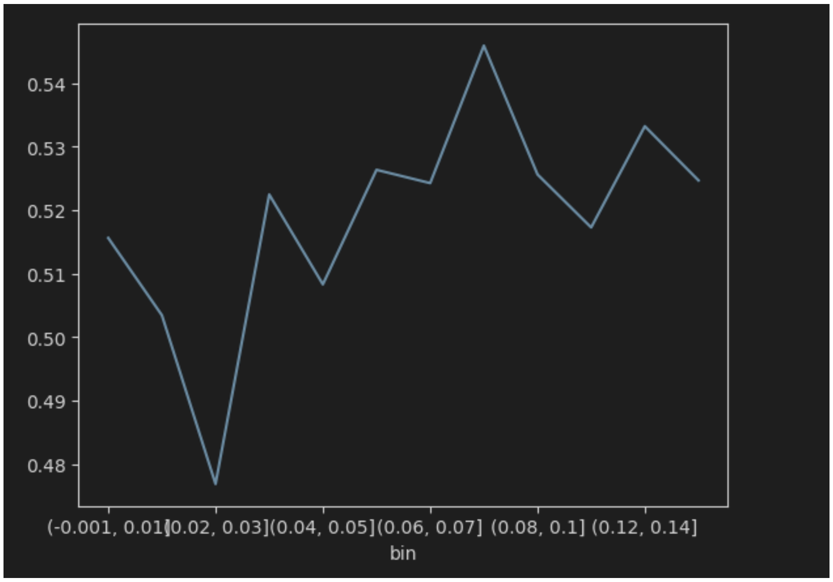
Once a basic predictive model has been established, our team will investigate whether the relationship between heroes' and it can influence the accuracy of the model. This involves exploring how the model captures the Anti-Mage VS Medusa counter relationship, a well-known counter in the DOTA2 community, using the data available.

Lastly, our team evaluates the confidence and accuracy of the decision trees used in the model. This will be accomplished by testing the tree's ability to identify the outcomes of matches accurately while simultaneously determining its confidence level in the predictions made. This approach enables us to assess the reliability and robustness of the model, ensuring that it can be used to make accurate predictions about the outcome of matches in DOTA2.



Results

After employing different types of models (neural networks, decision trees, random forest, XGBoost), we found that the model that uses the method of random forest gives us the best prediction result, which is about 51.4% accuracy. Therefore, by using the random forest model, we can predict the winning team 51.4% the time correctly after each team chooses their heroes in those pro matches. At the same time, we also analyzed the potential winning strategies, such as choosing the hero based on some counter relationship inside the game. We found out that if the model learns that Anti-Mage counters Medusa, the accuracy of our best model's (Using Random Forest) accuracy will be improved to 58.3%. Lastly, we examined the prediction of each tree in the Random Forest model and found out that as each tree predicts a winning probability that is further away from 50% chance of winning, the more accurate the prediction result is. Therefore, when our algorithm is more confident of the winning result by giving a prediction that is not close to a tie, the accuracy in predicting the winning team is higher. So, this shows our model prediction is robust and could be used to help the DOTA2 community and the players.



Here is a confusion matrix for our random forest prediction:

```
confusion_matrix(filtered_y, rf.predict(filtered_X), normalize="all")

array([[0.34466019, 0.14563107],
       [0.27184466, 0.23786408]])
```

Discussion

We acknowledge that our model's accuracy is not perfect, but it is a good start in predicting the results of DOTA2 matches based on hero selections. We recognize that there are many other relationships between heroes that our model has yet to learn. If we want our model to have better prediction accuracy, we need to train it on more data and let it learn about more hero counter relationships.

Additionally, it is important to note that the characteristics of the pro players and team corporations are also essential in determining the outcome of a match. As such, we could try to predict non-pro DOTA2 matches since the match results may be more dependent on hero selections. Our model will help the DOTA2 community and players make more informed decisions during the pick and ban phase, improving their chances of winning.

The strength of our model is that it currently gives us a prediction that is robust and we have discovered some relationships between heroes selections that could influence the winning rate, such as the example between Anti-Mage and Medusa, which are useful for players to consider when selecting heroes.