March-April 2020
ISSN: 0193-4120 Page No. 1300 - 1304

DOTA2 Winner Predictor using Machine Learning

R. Mythili[1], M. Sangeetha[2], G. Parimala[3]
[1,2,3]Department of Information Technology, SRMIST, Kattankulathur, Kancheepuram
[1]mythilir1@srmist.edu.in, [2]sangeetm@srmist.edu.in, [3]parimalg@srmist.edu.in

Article Info
Volume 83
Page Number: 1300 - 1304
Publication Issue:
March - April 2020

Article History
Article Received*: 24 July 2019*
Revised*: 12 September 2019*
Accepted*: 15 February 2020*
Publication*: 14 March 2020*

Abstract
DotA2 (Defense of the Ancient 2) is an online multiplayer game. It's a real time strategy game played between two teams of five players. DotA2 rules are extremely mind-boggling. We have created a model to predict the winner of Dota2 by using Logistic Regression. By obtaining the information of heroes and the average MMR of all the ten players we are predicting the outcome of the match. The win percentage is dependent on hero selection and MMR of players. We also determine the optimum 10th pick by invoking a brute force execution.

***Keywords*:** Logistic Regression, DotA2, Heroes, MMR, Radiant, Dire.

1. Introduction

DotA2 is a multiplayer online fight arena and a real time strategy game played between two teams of five players. It's developing rapidly. The largest tournaments boast prizes larger than the ICC World Cup, and viewership is more than MLB World Series and NBA Finals. There are two sides in this game (Radiant and Dire). Each match takes place over two distinct phases: the draft phase and game play phase. Players should play as a team to win DotA2. In this game, each hero has one of the three primary attributes, "Strength", "Agility", "Intelligence". Some heroes are designed in such a way that they are meant to support others. When a hero levels up, so does their attributes.

In DotA2 matches, interaction between the team members is the key to create a good team. In DotA2 there are 110 heroes. One player can choose only one hero and if a player chooses hero A, no other player in that team can choose that hero. The hero who is being selected by the player plays a major role in the outcome of the match. DotA2 heroes are classified into four categories: Carry, Support, Ganker, Initiator.

A team Figure 1 and Figure 2 should have all these four categories of heroes, if not they will struggle during the match. There are 2 major concern associated with this machine learning model. One is hero recommendation, which will maximize the victory percentage. And second is the win prediction, i.e., predicting which team is going to win based on hero selection and average MMR (match maker rating) of the players. MMR is the value which determines the skill level of each player.

Figure 1: Radiant team

Figure 2: Dire Team

2. Literature Review

There are ample amount of machine learning projects in Dota2. Atishagarwala and Michael Pearce [1] did a project on Dota2 coupled with machine learning. Their model is based on predicting the winner by the selection of heroes and interaction between the heroes. But prediction percentage is more when interaction between the heroes is not considered. Nicholas Kinkade and Kyung yul Kevin Lim [2] did a paper on DOTA 2 Win Prediction. They proposed two win predictors for DotA2.

In 2013, Kevin and Daniel presented a paper on a survey which they did on machine learning algorithm applied to DotA2. And also a recommendation engine which helps the players to choose the heroes to maximize the winning percentage in DotA2.

3. Data-Set Collection

Valve cooperation is an American based video game developing company. They are the developers of the game DotA2. They store the records of the matches, end-game results. These data can be accessed by using their Web API. This Web API is developed by Valve developers to retrieve the match data and player data. There is python

March-April 2020
ISSN: 0193-4120 Page No. 1300 - 1304

library dota2api, which can be used to collect the required data.

Data for 60,000 matches was collected using the Steam Web API over the period of 18/02/2019 to 20/02/2019. The following information was collected for each match:

- Winner
- Game Duration
- **For Each Player:**
  - Experience Gained Per Minute
  - Gold Gained Per Minute
  - Kills Per Game
  - Assists Per Game
  - Deaths Per Game

A. Match Data

DotA2 has different mode of games, in which some of are not regular DotA2 modes, we need to filter out those games that are not useful in our prediction. In our predictor, we only consider those games that is ”All Pick”, ”Ranked All Pick” and ”Random Draft”. And players with zero records are filtered out. Invalid matches were filtered out manually. These match data are stored in the form of Json files for further developments.

B. Player Data

The player data for each single game is collected, we use the API to retrieve all the players' data and use it for the ten players who are involved for the match. We retrieve the match history for a single player. A set of match history is taken under collection. The match history allows us to retrieve the match ID for each game. These data are also stored in the form of Json files for further developments.

C. Data collector

- Match_id - Official match ID stored in valve servers
- Radiant_win - True if radiant won, false if dire won.
- Radiant_team - String containing 5 hero IDs corresponding to the heroes in the radiant team.
- Dire_team - True if dire won, false if radiant won
- avg_mmr - Average MMR of people with public MMR in the game.
- num_mmr - Number of people with public MMR in the game

These are the data which is being collected and saved in a file.

4. Requirement Analysis

A. Software Requirements

- Python
- Scikit-learn for Python
- PostgreSQL
- Flask

B. Hardware Requirements

- Minimum system requirements - Pentium IV 2.6 GHz
- Hard Disk - Minimum 20 GB

5. System Design

A. Data farming

Collection of updated and normalized data to our model

B. Storage of Data

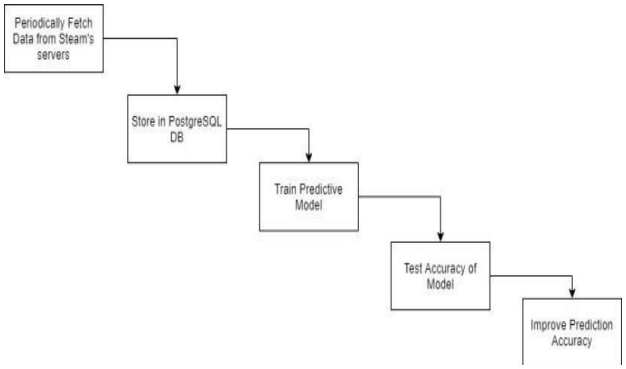
Design and implement a DB Schema to store and retrieve our large repository of match data.

C. Training and prediction

Implementing different ML characterization systems to decide the ideal model for our motivations. The algorithm utilized for preparing is Logistic Regression and the assessment is done through cross approval. The trained model is saved to a pickle file for later use.

D. Analysis of results

Plotting and analyzing the results should reveal interesting information about the nature of the system.



E. Logistic Regression

The appropriate regression analysis to conduct when the dependent variable is binary. Like all regression analyses, the logistic regression is a predictive analysis. Logistic regression is used to describe data and to explain the relationship between one dependent binary variable and one or more independent variables. Binary logistic regression major assumptions:

The dependent variable should be binary. There should be no duplicate in the data, which can be assessed by converting the continuous predictors to standardized scores, and removing values below -3.29 or greater than 3.29. There should be no high correlations among the predictors. This can be assessed by a correlation matrix among the predictors. Tabachnick and Fidell (2013) suggest that as long correlation coefficients among independent variables are less than 0.90 the assumption is met.

March-April 2020
ISSN: 0193-4120 Page No. 1300 - 1304

F. Reporting R2

Numerous pseudo-R2 values have been developed for binary logistic regression. These should be interpreted with extreme caution as they have many computational issues which cause them to be artificially high or low. A better approach is to present any of the goodness of fit tests available; Hosmer-Lemeshow is a commonly used measure of goodness of fit based on the Chi-square test.

G. Overfitting

When selecting the model for the logistic regression analysis, another important consideration is the model fit. Adding independent variables to a logistic regression model will always increase the amount of variance. However, adding more and more variables to the model can result in overfitting, which reduces the generalizability of the model beyond the data on which the model is fit.

6. Proposed System

A. Issues in Existing Methodology

Without a doubt DotA2 game is complex and unusual and impacted by such huge numbers of variables separated from the group methods, for example - player expertise, gold and experience development, the capacity of the team members to facilitate among each other, any abrupt and basic group battles that could change the diversion and so on. There are an excessive number of elements to count here and we focus on hero selection by the players. The proposed framework will take in the significant considers alone request to give an outcome like the last result. The system will have the capability to suggest heroes also in-order to analyze the game and to increase the fan base of DotA2 gaming.

B. Describing need for new Technology

We present techniques and results for win prediction in professional games using extremely high skill publicly available game data to help us with professional level training data and ensure the training data provides most reliable prediction models. We thoroughly evaluate common prediction algorithms and their configuration to identify the best performing algorithm and configuration on various aspects of MOBA data. This not only indicates which algorithm and configuration to use and when, but also, how much optimization is required for highest prediction accuracy. The new system includes utilizing a probability provider prepared with a great pre-coordinate information and hero information. Logistic Regression Algorithm is the premise of this mind-boggling framework. With Logistic Regression calculation, you can accomplish likelihood rates. This calculation quickly forms results in a split of seconds. It is the fundamental idea of practically all the constant systems. Logistic Regression Algorithm gives high exact outcomes exact qualities. The framework includes information gathering and extraction.

C. Advantages of Proposed System

The most significant advantage of this proposed system is the minimum time complexity. The ability to most accurately predict the winner in seconds time. The system also provides another feature were it suggests heroes who can be used to fill up the remaining slots of the team in-order to provide a high win probability by choosing the recommended heroes. The Logistic Regression Algorithm used in this project is known worldwide which produces high level of accuracy and precision with rapid computation speeds. The system proposed can also be further enhanced by fine tuning the factors considered for prediction and the feature extraction system to improve the accuracy and also to decrease the failure results. The idea can also be incorporated with deep learning concepts to find a pattern between different types of players and using hardware technology to increase the collection of data. This can be really effective for the DotA2 fans and also can be implemented for other arcade games of the gaming world.

7. Methodology

A. Feature Engineering

Hero Selection

Every player has to select a different hero for each game. DotA 2 heroes do not just look different, they also serve different purposes. Certain hero compositions will help players win a game. There are 112 heroes in total.

Net worth at the start mark

The higher the net worth, the greater advantage a hero has. Thus, heroes with higher net worth will be more likely to win the game.

Other Features

Other features include net death counts from each team, team composition of hero roles and the interaction between each hero's role and its net worth. Having a bad composition of roles may put a team at a disadvantage.

B. Two-Queries

We implement two different queries which enable us to exact maximum value from our model.

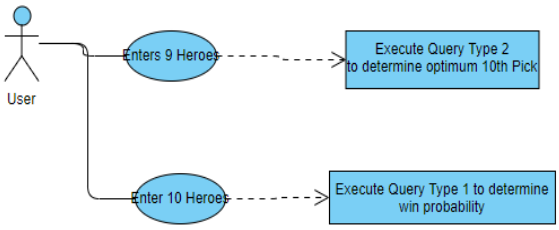
Query Type 1

This query is the standard query which invokes our program to determine the win probability for each side of the given 10 hero lineup.

Query Type 2

This query invokes a brute force execution of query type 1s to determine the hero pick to give the maximum probability of winning. Only a 9 hero input is required for this query type.

March-April 2020
ISSN: 0193-4120 Page No. 1300 - 1304



Full query

Insert all 10 heroes in a game and predict the winner. The program returns the probability of Radiant Team winning.

```
from training.query import query

full_result = query(3000,
                    [59, 56, 54, 48, 31],
                    [40, 41, 52, 68, 61])
```

C. Implementation

DotA2 has a player matching system which will allow the game to bring in same level of players for each game, this is called “matchmaking ranking”(MMR). Therefore we provide a slot for the user to enter the MMR level. This is done in order fully utilize the available 110 playable heroes in DotA2 and sometimes it is possible that a player is not familiar with the hero that he used in a particular game which lead to the hero not completely be utilized. Moreover, the player’s experience level might be different. This helps us to simply the problem to consider the hero selection in a game. We use the hero selection data to predict which team will win the game and analyse how hero selection contributes to the final game result.

We define the implementing vector as SET X containing 224 heroes starting from

{ X1 X2 . . . X112 X113 X114 . . . X224 }

As each DotA2 game has two sides - one is ”Radiant” and the other one is ”Dire”.

The first half of the implementation vector(X1 ...X112) represents the “Radiant” side. The second half of implementation vector (X113 ...X224) represents the ”Dire” side.

For the first half of the implementation vector:

1, a player of radiant side plays hero i

0, no player of radiant side plays hero i

For the second half of the implementation vector:

1,a player of dire side plays hero i -112

0, no player of dire side plays hero i - 112

We note down each match as

Y = 1, if radiant team wins

0, if dire team wins

Logistic Regression is used to determine the probability of which team is winning. The formula which helped to figure out the regression is:

$$min_{w,c}\frac{1}{2}w^T w + C\sum_{i=1}^{m} log(exp(-Y_i(\phi(x)_i^T w + c)) + 1)$$

The 60000 match data which we collected and divided them to as many folds possible. So for each fold, we trained our model using 2500 matches. Then we validated using 500 matches. When predicting the winner of the game, we get the probability of who wins among the Radiant or Dire team for a given match and then choosing the team with higher probability as the winner, this is simply how the predictor works.

Scikit-learn for Python helped us to compute the values for the logistic regression.

Example - Logistic Regression for 3000 matches of 5 folds

Fold	1	2	3	4	5
Radiant	0.48	0.49	0.54	0.58	0.55
Dire	0.52	0.51	0.46	0.42	0.45

The table shows us that each team won at different folds.

Since we can see that the model provided almost same predictions for both the sides. To overcome this we started training the model with other data which involved:

- Hero Pair Win Rate
- Team Radiant Hero’s synergy and counter energy towards Team Dire
- Team Dire Hero’s synergy and counter energy towards Team Radiant

After adding extra features to the model it showed improvisation in the results

Fold	1	2	3	4	5
Radiant	0.45	0.41	0.56	0.60	0.57
Dire	0.55	0.59	0.44	0.40	0.43

The table shows us that there is a bigger difference in the probability of the winning and losing side.

8. Result

The DotA2 project has been successfully completed with complete analysis and performance check of all the modules. Accuracy of the system has been found by computing the ratio of number of correct decisions made to the total number of decisions. The accuracy was found to be more than 75% every time. Finding the best win percentage was possible. The project has its own hero suggestions which paved way to an efficient way to help the players get more wins and also increasing the scope of

the project where it can be further enhanced coupled with suitable algorithms.

9. Conclusion

Our predictor can accurately predict the win rate of a DotA2 match and we have successfully completed creating a predictor with 75 percentage accuracy at the beginning of the match with the information on the hero picks and other features. To achieve our aim, we analyzed the data set from the heroes of the game and also the additional features available in the game. The data collected on heroes containing their relationship with other heroes and their counters helped us to provide a more precise result. This shows us that hero taken has an important role in deciding the winner. The results showed us that we can introduce more features in order to increase our efficiency of the predictor as the game has more in-game factors. Our predictor is not a full efficient system as in DotA2 a match's result can be changed according to the happenings in that match and with respect to the player moves as we humans can never be read or understood completely. Therefore, more information on what happens during the game will greatly add to our predictor and improve performance.

10. Future work

As DotA2 is a complex game to work on and to predict a winner accurately is a difficult task to achieve with only the help of few features which we tried in our project. Therefore in near future we would be looking forward to work and analyze on all the available in-game aspects with which we can train our model in-order to provide a fully efficient predictor that will help the DotA2 game and it's viewers.

References

[1] Classification of player roles in the team-based multi-player game dota 2. In International Conference on Entertainment Computing, 112–125. Springer. [Hall 1999] Hall, M. A. 1999.

[2] Result prediction by mining replays in dota ¨ 2. Master's thesis, Blekinge Institute of Technology, Karlskrona, Sweden. [Kalyanaraman 2015] Kalyanaraman, K. 2015.

[3] To win or not to win? a prediction model to determine the outcome of a dota2 match. [Kinkade, Jolla, and Lim 2015] Kinkade, N.; Jolla, L.; and Lim, K. 2015.

[4] Dota 2 win prediction. Technical report, University of California San Diego. [Kohavi and John 1997] Kohavi, R., and John, G. H. 1997.

[5] Wrappers for feature subset selection. Artificial intelligence 97(1-2):273–324. [Pobiedina et al. 2013] Pobiedina, N.; Neidhardt, J.; Calatrava Moreno, M. d. C.; and Werthner, H. 2013.

[6] Ranking factors of team success. In Proceedings of the 22nd International Conference on World Wide Web, 1185–1194. ACM. [Rioult et al. 2014] Rioult, F.; Metivier, J.-P.; Helleu, B.; ´ Scelles, N.; and Durand, C. 2014.

[7] Performance of Machine Learning Algorithms in Predicting Game Outcome from Drafts in Dota 2. Cham: Springer International Publishing. 26–37. [Song, Zhang, and Ma 2015] Song, K.; Zhang, T.; and Ma, C. 2015.

[8] Predicting the winning side of dota2. Technical report, tech. rep., Stanford University. [Sun, et al. 2016] Sun, et al., Y. 2016. Internet of things and big data analytics for smart and connected communities. [Witten and Frank 2000] Witten, I. H., and Frank, E. 2000.

[9] Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations. San Francisco: Morgan Kaufmann. [Yang, Harrison, and Roberts 2014] Yang, P.; Harrison, B. E.; and Roberts, D. L. 2014.

[10] Identifying patterns in combat that are predictive of success in moba games. In Proc. of Foundations of Digital Games (FDG'14). [Yang, Qin, and Lei 2016] Yang, Y.; Qin, T.; and Lei, Y.-H. 2016. Real-time eSports Match Result Prediction. ArXiv e-prints.

[11] Min Feature Replacement Algorithm Based Multiple Imputation Based Gap Analysis for Clinical Data. International Journal of Innovative Technology and Exploring Engineering (IJITEE). ISSN: 2278-3075, Volume-8 Issue-5 March, 2019.