Proceedings of the International Conference on Intelligent Computing and Control Systems (ICICCS 2019)
IEEE Xplore Part Number: CFP19K34-ART; ISBN: 978-1-5386-8113-8

Victory prediction in League of Legends using Feature Selection and Ensemble methods

Ani R [1], Vishnu Harikumar [1], Arjun K Devan [1], O.S. Deepa [2]
[1] Department of Computer Science and Applications,
Amrita Vishwa Vidyapeetham, Amritapuri, India
anir@am.amrita.edu, vishnu harikumar12@gmail.com, arjunkdevan1995@gmail.com
[2] Amrita School of Engineering, Coimbatore, Amrita Vishwa Vidyapeetham
os_deepa@cb.amrita.edu

***Abstract*— Prediction of winners in the online video games has become an important application for machine learning based prediction models. The main goal of the present study is to achieve a good prediction rate for a popular Electronic sport called League of Legends. League of Legends is a Multiplayer Online Battle Arena (MOBA) game that combines intensity of a Real-time strategy with various Role-playing elements. Feature selection is done and only relevant features that affect the match outcomes are considered. Prediction is done by using ensemble models of classification algorithms and the performance was evaluated. The important performance metrics and their influence on each game model were also analyzed. The results show that the reliable match result prediction is possible in the League of Legends game.**

Keywords—League of legends, esports, feature selection, ensemble methods, accuracy.

I. INTRODUCTION

Victory prediction gains an increasing interest from researchers and Electronic sports (Esports) industry. The term Esports refers to digital sports that are played on a competitive level with millions of spectators. In the year 2019, it is estimated that more than 400 million people will be watching some sort of Esports. This is mainly due to the increase in online streaming media, particularly Twitch and YouTube gaming. Apart from the people watching it, the global Esports market generated 905 million dollars worldwide in the year 2018 which was a 35 percent increase from 2017. At the current pace, it is said it will generate an amount of 1.5 billion dollars in revenue by the year 2020.

The main aim of victory prediction is to determine which team of players or player will emerge victorious in a particular match. Victory prediction can be studied in human vs. human, human vs. Artificial Intelligence and Artificial intelligence vs. Artificial intelligence situations. Prediction can be done using pre-match, within-match, or post-game features. The present study analyzes the prediction rate for human vs. human matches using pre-game and within-game features and also a combined model of both for League of Legends.

League of Legends provides vast testing grounds for machine learning algorithms due to the availability of a voluminous amount of data. For instance, Isaac da Silva Beserra and Lucas Camara [1] analyzed keystroke and mouse dynamics for user identification to find out whether the person using the account is the actual account holder or not. League of Legends back-end servers contains more than 800 features per player across millions of players. In League of Legends, players can participate in player versus player as well as player versus environment matches. Furthermore, the performance metrics in League of Legends are similar to other competitive MOBAs such as Dota2 and Smite, and to a degree also to First Person Shooters (FPS) such as Fortnite and Player Unknown Battlegrounds (PUBG), making League of Legends a broadly applicable case.

II. GAME DESCRIPTION

League of Legends is a MOBA game. In essence, MOBA games typically consists of five players on each team with each player controlling a single character. Contrary to MMORPG (Massively Multiplayer Online Role-Playing Game) games, there is no unit construction nor there is a massive number of players on a map at a time in a MOBA game, much of the strategy revolves around cooperative team play and individual character development.

A professional League of Legends match consists of ten players, five members on each team on a map called Summoners Rift. Before the match begins, each team member chooses a champion to play, which is a character with unique designs and abilities that already exist within the game. Furthermore, they are allowed to customize their champion by selecting 2 powerful abilities (Summoner Spells) as well as special stones called runes which gives extra enhancements to a champion. Each Champion has four activated abilities and a passive ability which can be used to kill enemy Champions or monsters. In addition to champions, each team member is assigned a role/position (Jungle, Top Laner, Mid Laner, ADC, Support) which he/she must fulfill until the match finishes. League of Legends is a team game and all five roles are essential for the team's success.

Once the match begins, each player has to go their respective lanes after buying their starting Items using Gold. Gold is the in-game currency for League of Legends match and Items are objects which are used for in-game enhancements on Champions. The goal of the game is to crush the enemy base

(Nexus) while protecting your own. But before destroying enemy Nexus, there are a bunch of obstacles a team should overcome to reach that point. Each match on an average last around 20-40 minutes. Finally, whichever side destroys the Nexus first gets the victory.

III. Related Work

Hao Yi Ong, Sunil Deolalikar, and Mark Peng [2] used K-means for clustering player behavior and used Logistic Regression, Gaussian Discriminant Analysis and Support Vector Machines for determining match outcomes in League of Legends.

Johansson and Wikstrom [3] have made a real-time prediction in their paper, they attempted to create a model with different parameterized versions of the Random Forest algorithm to predict the winning team of Dota2 (MOBA) matches using partial game-state data.

Weiqi Wang [4] used multi-layer feed forward neural networks to predict the game outcome for Dota2 on hero-draft data.

Ravari, Spronck, Sifa and Drachen [5] used Gradient boosted and Random Forest for victory prediction in game Destiny for several game modes. They also analyzed the different performance metrics and their influence on each mode.

Ravari, Bakkes and Spronck [6] employed Gradient Boosting and Random Forests to predict the match winner for various game modes in Star-craft and obtained optimal results.

IV. Dataset Description

Our League of Legends dataset includes performances of players in professional Esport matches. In total, the dataset includes 1500 matches and each match consists of 10 professional players. Hence a total of 15000 instances were collected. The dataset consists of 97 features which comprises of pre-match and within-match features. The data was then separated based on pre-match and within-match features and also a combined model for both was created. A general explanation of some of the features used in the game is explained in the sub sections below.

A. Pre-match Features

- Champion: it’s an avatar or character which the player assumes. There is a total of 143 Champions in League of Legends and each champion has unique designs and abilities.
- Bans: each player is allowed to Ban a Champion which he/she does not wish to play against.
- Summoner Spells: abilities or Spells which can be used to upgrade a Champion. A player can select 2 summoner spells from a list of 10.
- Role: each team member is assigned a position which he/she must fulfil until the match finishes. All five roles are essential for the team’s success.

B. Within-match Features

- First Blood: it is the first kill of the game and gives bonus Gold to the player who gets it.
- Gold Spent Percentage Difference: GSPD is the average measurement of the Gold difference between both teams towards the end of a match.
- KD: Kill (K) is an event of reducing the enemies hit points to zero. Once the enemies hit points have been reduced to zero it results in a Death (D).
- CSat10: Creep Score at 10-minute mark. Creeps are small minions which players kill in the game for Gold in lanes.
- CSdat10: Creep Score difference between both teams at 10-minute mark.
- Expat10: Experience earned at 10-minute mark.
- KPM: Kills acquired by each player Per Minute.

V. Methodology

For our models, 60 percent of the data was taken for training and the rest 40 percent for testing purposes. As part of our pre-processing we used Feature Selection algorithms for finding the most significant features. Prediction is done using ensemble models of classification algorithms and the performance was evaluated. Accuracy will be used as the standard for measuring the performance

A. Feature Selection

Datasets today are very rich in information with lots of data. So, finding relevant information from a huge dataset is going to be a challenging task because there will be lots of features. Feature Selection [7] will help select a subset of relevant features without much loss of the total information. Some of the merits of Feature Selection is as follows:

- Accuracy of the resulting subset is increased, if the right subset is chosen.
- Reduces complexity of the model making it easier to interpret.
- Enables the algorithm to train faster

We used Recursive Feature Elimination with cross validation [8], [9], [10] for selecting the relevant features. In Recursive Feature Elimination (RFE) the features are selected by recursively considering smaller and smaller sets of features. RFE requires a specified number of features to keep, however it is often not known how many features are correct or valid. For finding the optimal set of features we use cross-validation with RFE to score different feature subsets and then select the best scoring subset of features. We also ranked the top-3 features which affects the match outcome most in the order of their Gini importance. Gini importance is a measure of variable importance based on the Gini impurity index used for the calculation of splits during training. In the next subsection we explain the various ensemble methods adopted for calculating our predictive performance.

B. Ensemble methods

1) Random Forest: Random Forest (RF) [11], [12], [13], [14] is an Ensemble learning method which uses decision trees for prediction i.e. it builds multiple decision trees and combines them together to get a more accurate and stable prediction. RF has been used successfully for prediction problems in online games such as in [3], [5], [6].

2) AdaBoost: Boosting has always been a very successful technique for solving the two-class classification problem. AdaBoost [15] is a popular boosting approach which helps you combine multiple weak classifiers into a single strong classifier. A weak classifier is one which performs poorly, but performs better than random guessing. AdaBoost can be sensitive to noisy data and outliers at times which makes it less reliable.

3) Gradient Boosting: Gradient Boosting (GB) [16] is also used for converting a set of weak learners into a single strong learner. The key difference between GB and AdaBoost is how the two methods identify the limitations of weak learners. GB uses an ensemble of weak learners, such as regression trees, and optimizes a loss function (measure indicating how good a model's coefficients are at fitting the underlying data) to generalize them whereas AdaBoost uses high weight data points to overcome the limitation of weak learners. GB has also been successfully used for video game prediction as shown in [5], [6].

4) Extreme Gradient Boosting: Extreme Gradient Boosting (XGBoost) is an advanced version of GB method. It's a powerful algorithm which can deal with all sorts of data. The key difference between both is that XGBoost uses a more regularized model formalization to control over-fitting, which gives it better performance. It implements parallel processing and is much faster compared to GB and also has a built-in mechanism to handle missing values.

VI. Result and Discussion

In this section, we present the performance of our models. Table 1 shows the accuracy of our models tested on different ensemble methods. For pre-match data, an accuracy of 95.52 percent was achieved using Random Forest algorithm. The relatively high accuracy was mainly due to the feature" Ban". Each team can Ban a total of 5 Champions and since each Champion has unique skills and abilities, it has a huge impact in a professional match. It can also be seen that it has the highest Importance rate for pre-match data as shown in Table 2. Banning strong Champions which are over-powered tremendously increases the chance of player or team of players winning a match. As for within-match data, we got an accuracy of 97.77 percent which was also achieved using Random Forest algorithm. In within-match data the features" Tower" and" KD" were most important performance metrics as seen in table 2. Towers are structures which the players need to destroy in order to get to the enemy Nexus (main base). Each lane has 3 towers and also 2 towers at the enemy nexus. The enemy nexus will not be open until these towers are destroyed." KD" (Kills (K) and Deaths (D)) being in the top-3 was an expected result before we even ran our algorithm. Kills and Deaths are important performance metrics in not only MOBA games but also in FPS (First Person Shooter) and MMORPG based games and also in several other genres. Finally, for our combined model which contains features from both pre-match and within-data, we were able to increase the performance even further by getting 99.75 percent. Furthermore, for all our models Random Forest algorithm achieved the highest possible accuracy compared to all other algorithms that were used.

TABLE 1. Accuracy of our models evaluated using different ensemble methods

Methods	Pre-match	Within-match	Combined
Random Forest	95.52%	98.18%	99.75%
Adaboost	57.22%	96.31%	96.25%
Gradient Boosting	65.67%	96.82%	97.01%
Extreme Gradient Boosting	65.12%	96.83%	97.21%

TABLE 2. Top-3 features with their importance

Pre-match	Within-match
Ban - 0.18	Turret – 0.22
Champion - 0.08	KD – 0.19
Patch no – 0.07	GSPD – 0.06

VII. Conclusion

In this work, we tried to get a good prediction rate for the MOBA, League of Legends. Our models achieved a performance between 95 and 99 percent in terms of accuracy using Random Forest algorithm. This indicates that reliable prediction of match results is possible in professional League of Legends matches using both pre-match and within-match data.

Acknowledgment

We are extremely thankful to Mrs. Ani R, Assistant professor and Chair person, Department of Computer Science and Applications, Amrita School of Engineering, Amritapuri for her primary support throughout the project We also thank our colleague P.S Narayanan for his help in the development aspects of our project.

References

[1] I. da Silva Beserra, L. Camara, and M. Da Costa-Abreu, "Using keystroke and mouse dynamics for user identification in the online collaborative game League of Legends," p. 8 (6 .)-8 (6 .), 2017.

[2] H. Y. Ong, S. Deolalikar, and M. V Peng, "Online Multiplayer Games," pp. 1–5.

[3] F. Johansson, J. Wikström, and F. Johansson, "Result Prediction by Mining Replays in Dota 2," 2015.

[4] R. Bierig, "Predicting Multiplayer Online Battle Arena (MOBA) Game Outcome Based on Hero Draft Data Weiqi Wang Supervisor :"

[5] Y. N. Ravari, P. Spronck, R. Sifa, and A. Drachen, "This is a repository copy of Predicting Victory in a Hybrid Online Competitive Game : The Case of Destiny . Proceedings of the Thirteenth Artificial Intelligence and Interactive Digital Entertainment Predicting Victory in a Hybrid Online Competitive Game ," 2017.

[6] Y. N. Ravari, S. Bakkes, and P. Spronck, "StarCraft Winner Prediction," pp. 2–8, 2014.

[7] I. Guyon, "An Introduction to Variable and Feature Selection 1 Introduction," vol. 3, pp. 1157–1182, 2003.

[8] P. M. Granitto, C. Furlanello, F. Biasioli, and F. Gasperi, "Recursive feature elimination with random forest for PTR-MS analysis of agroindustrial products," vol. 83, pp. 83–90, 2006.

[9] R. Kohavi, "A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection," no. March 2001, 2013.

[10] A. V. Vidyapeetham and A. V. Vidyapeetham, "A Correlation Based SVM-Recursive Multiple Feature Elimination Classifier for Breast Cancer Disease using Microarray," pp. 2677–2683, 2016.

[11] L. E. O. Breiman, "Random Forests," pp. 5–32, 2001.

[12] R. Ani, A. Augustine, N. C. Akhil, O. S. Deepa, Á. Classi, and Á. Random, "Random Forest Ensemble Classi fi er to Predict the Coronary Heart Disease Using Risk Factors," pp. 701–710, 2016.

[13] R. Ani, J. Jose, M. Wilson, and O. S. Deepa, "Modi fi ed Rotation Forest Ensemble Classi fi er for Medical Diagnosis in Decision Support Systems," 2018.

[14] F. Algorithm, S. Sreejith, S. Rahul, and R. C. Jisha, "A Real Time Patient Monitoring System for Heart Disease Prediction Using Random."

[15] Y. Freund, R. E. Schapire, and M. Hill, "Experiments with a New Boosting Algorithm Rooms f 2B-428 , 2A-424 g," 1996.

[16] J. H. Friedman, "Stochastic gradient boosting," vol. 38, pp. 367–378, 2002.