Wining Rate Prediction of Game League of Legends

Jianxun Zhao

Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China
1811020207@stu.hrbust.edu.cn

Abstract. Currently, machine learning is widely used to predict the winning rate of various games. As one of the most popular games among young people, predicting League of Legends is significant for both the players and the game. In order to identify the factors that affect game results, and predict the wining rate of each team, this study introduced Particle Swarm Optimization method to the Random Forest algorithm. This method is named PSO-Random Forest algorithm, which allows automatic and precise parameters optimizing. Besides, five other algorithms were deployed in the experiment as the baseline. For the prediction of the wining rate, experiment shows that the proposed method is effective. PSO-Random Forest algorithm achieved the highest accuracy, reached an accuracy of 0.7479, exceeding other baseline algorithms. This study revealed the main factor that affects the wining of game LOL and introduced PSO method to Random Forest algorithm to find the optimal parameters.

Keywords: Deep Neural Network; Wining Rate Prediction; Video Game; Particle Swarm Optimization;

1 Introduction

Nowadays, deep learning technology is widely applied in the fields of data mining and analysis, helping people extract helpful information and knowledge from massive data. For the area of video games, this area is one of the domains with huge data amount. It is possible to train an efficient win-rate prediction model by introducing data mining into the wining rate prediction. Traditional game AI generally refers to non-player-controlled characters (NPCs) in games, those characters can only act as the preset way. In the modern gaming industry, AI is no longer the pre-programmed NPC; they can now learn the previous dataset of the game and act with intelligence. However, due to the complexity of the games, and the dynamic real-time changes of the battle situation, making research on games is more complex, thus predicting winners has become a highly relevant topic in artificial intelligence research.

While the MOBA (multiplayer online battle arena) games began to popular in the last decade, much effort has been put upon the wining rate prediction of it. Researchers have released much work to the AI for games to better fit players' entertainment needs. For instance, S. M. Lucas [1] introduced an interchangeable actuator model for the automated game AI tuning in game Planet Wars. R. De Schaetzen and A. Sestini [2] introduced an efficient path-following solution using

Y. Wang (ed.), *Proceedings of the 2024 International Conference on Artificial Intelligence and Communication (ICAIC 2024)*, Advances in Intelligent Systems Research 185,
https://doi.org/10.2991/978-94-6463-512-6_75

quadratic Bézier curves to estimate the path curvature for ground vehicles controlled by game AI. Ye et al [3] used a supervised learning model achieved the human-level performance of game AI. P. Tian [4] divided heroes into different group with each hero's feature, combined with reinforce learning to form a hero-featured model for the wining rate predict of game honor of kings. Y. Li [5] introduced an algorithm based on decision tree algorithm named ID3. Their model can adapt to the player's operation automatically. Y. Takano [6] combined hybrid reward architecture with the gaming AI.

This study summarizes the work of previous researchers and analyses a dataset on the wining rate of game League of Legends. Proposed PSO-Random Forest algorithm, and reveals what factor affects the wining rate most by comparing the result with different models for wining rate prediction.

2 Dataset and Method

2.1 Data And Processing

As dataset is the core of game review and wining rate prediction, this study chooses only high rank games among the recorded games (https://www.kaggle.com/datasets/bobbyscience/league-of-legends-diamond-ranked-games-10-min). This article is based on Kaggle's publicly available League of Legends (LOL) game data. In the game "League of Legends," players embody "champions" endowed with distinctive powers, engaging in combat with a squad of adversaries controlled either by fellow players. Players attempt to destroy the enemy's base for the final victory [7-8].

The involved dataset contains the basic numeral record of 9880 games (solo queue) from high ELO (diamond I to master) in game LOL. To examine the balance of the dataset, this study counted wins per side and shows the result in Fig. 1. It is clear that the dataset is well balanced.

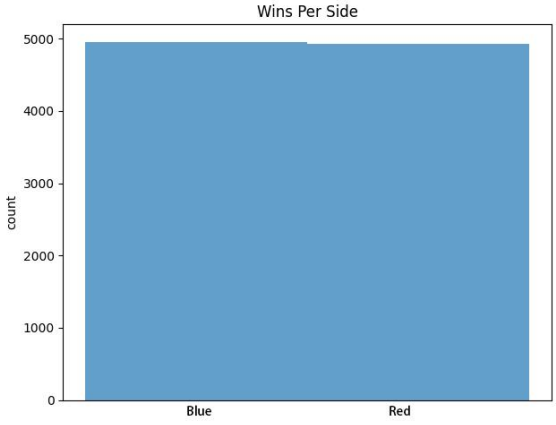

Fig. 1. Schematic diagram of wins per side.(Photo/Picture credit : Original)

2.2 PSO-Random Forest Algorithm

Particle Swarm Optimization (PSO) is a method based on the observation of ani-mal clusters, utilizing the sharing of information among individuals in the animal group, the movement of the entire group undergoes an evolutionary process from disorder to order in the problem-solving space, thereby obtaining the optimal solution.

The phenomenon of the prediction accuracy of the random forest algorithm changing with the variation of parameters is very obvious. Incorrect parameters setting may result in a decrease of more than 10% in accuracy. In order to find the optimal parameters, this study introduced PSO method to Random Forest algorithm. Using PSO optimization, parameters in the random forest model, including n_estimators(the number of trees) and max_depth(the maximum depth of each decision tree) can be iterated to find the set of values that make the prediction accuracy the highest. Based on this optimization, the classical Random Forest algorithm is improved and called PSO-Random Forest algorithm. This algorithm can automatically determine the optimal values of the parameters through PSO method and prevent overfitting [9-10].

In this theory, all the regression machine learning or deep learning models can be used for predicting the game result. This study selected Deep Neural Network, K Nearest Neighbors, Random Forest, Logistic Regression and Naive Bayes algorithm as the baseline algorithms [11].

2.3 Methodology

This section introduced the methodology and the related algorithms we used, as well as our way of dealing with the issue in detail. We assembled our selected data of different algorithms. Using computer simulation, we acquired the performance of different machine learning models on this issue [12-14]. Fig. 2 below shows the procedure we took for the simulation on each model

To test the accuracy of wining-rate predict of PSO-Random Forest algorithm, we compared the performance of different m algorithms. All the algorithms mentioned bellow can achieve the predicting function, but because of the nonlinear model, different algorithm resulted in different performance [15].

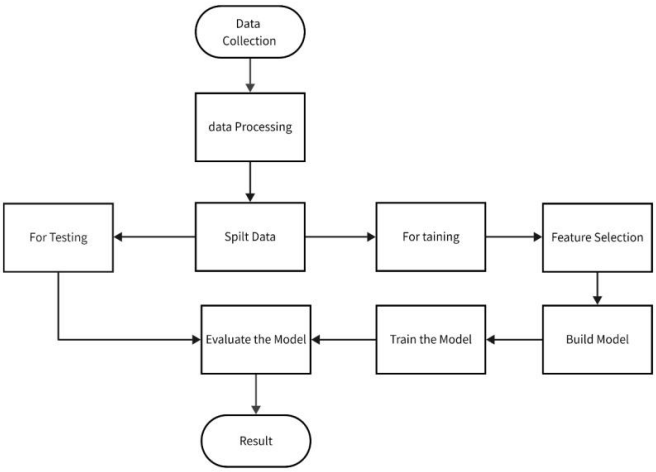

Fig.2. Proposed system flow of wining rate prediction(Photo/Picture credit : Original)

The following part is a brief introduction of the baseline algorithms this study used.

Deep Neural Network is a multi-layer network with massive neural cells in each layer (Fig.3). Deep neural networks can achieve specific advantages by simulating the process in which biological neurons receive signals and process them with the strength of the signal and make response. When processing huge data input DNN model's performance would not reach a limit like older algorithms.

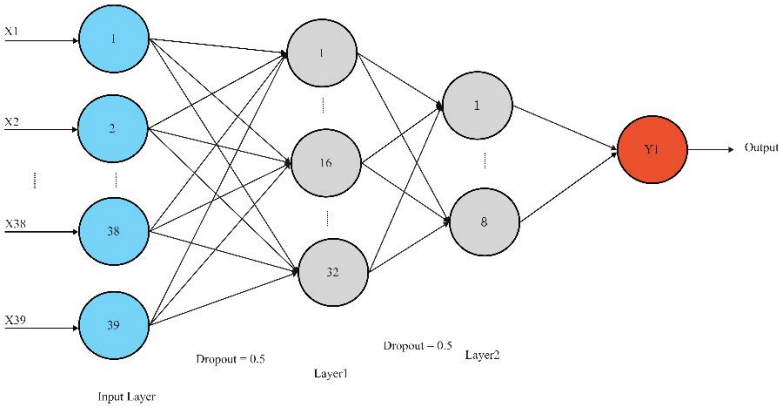

Fig.3. Schematic diagram of common Deep neural networks(Photo/Picture credit : Original)

K Nearest Neighbors, refers to a network whose cells use their closest K neighboring values to represent itself. Here K refers to the top k most similar data in the sample dataset. In a binary classification problem, to determine which category a sample belongs to, this algorithm examines the sample's K closest sample points and

calculates which category's value is higher. And the higher one is that sample's classification.

The Random Forest algorithm evolved from the decision tree algorithm, which constructs multiple decision trees by randomly sampling data and features. This algorithm effectively reduced prediction errors and improved the accuracy of classifiers. For each training set, the algorithm constructs a decision tree, randomly extracts a portion of the features during splitting, finds the optimal solution among the extract-ed features, applies it to nodes, and then performs splitting.

Logistic Regression is a method of using sample data to obtain a logical model and applying it to classify new data. By using a logistic regression model, the output results are compressed into an S-shaped curve of 0-1, representing the probability of a positive sample. If a sample's probability is higher than 50% it will be classified as 1, otherwise it will be classified to 0.

Naive Bayes model uses the Bayesian formula to predict the sample's classification. This algorithm first calculates the prior probabilities with labels in the training set and then calculates the conditional probabilities of each feature separately.

To test the models, this paper use the accuracy of test set, precision, recall and f1-score to judge the model. Table 1 shows the definitions of the error metrics the study used.

Table 1. Metrics and definytions

Metrics	Definition
Accuracy	Correct predicts among all predicts
Precision	Positive samples correctly predicted / Total number of samples predicted positive
Recall	The proportion of correctly determined positive instances to the total positive instances
F1-score	Weighted harmonic average of precision and recall

3 Result and Analyze

Firstly, this study examined the correlations of the data features. As there's total 39 features, to simplify the learning process and make the result more precise, this study dropped some of them with weak correlation ship with the result.

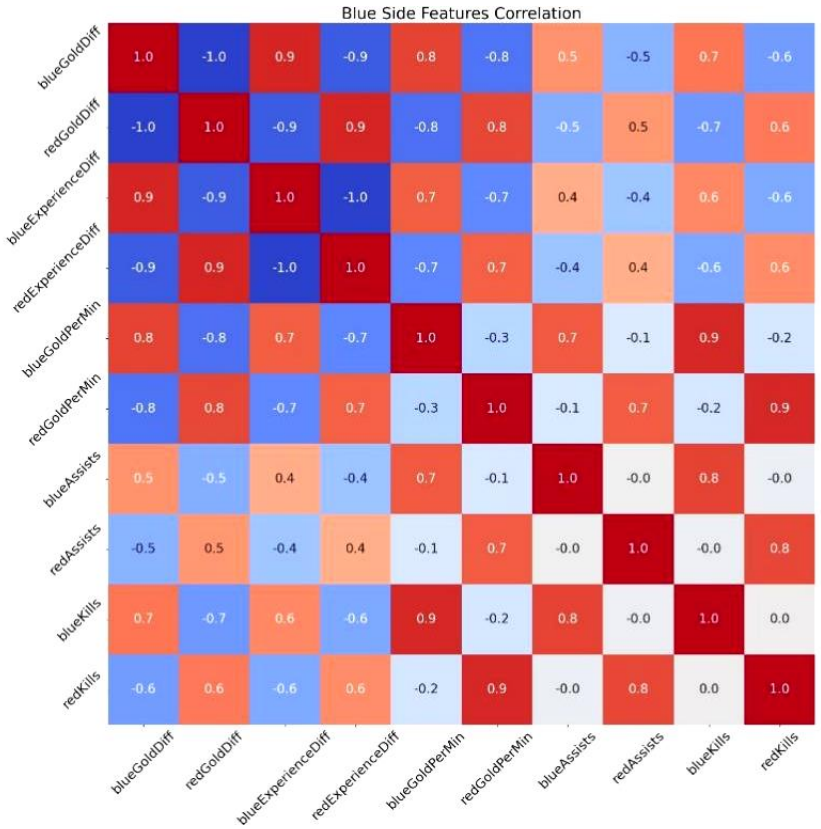

Fig. 4. correlation of each feature(Photo/Picture credit : Original)

Fig. 4 shows the correlation between features and the blue team's victory. According to Fig. 4, the most relevant features that lead to the wining of blue team are concluded as table 2.

Table 2. Features correlate with blue wins

Statistics Index	Correlation
blueGoldDiff	0.51
redGoldDiff	0.51
blueExperienceDiff	0.49
redExperienceDiff	0.49
blueGoldPerMin	0.41
redGoldPerMin	0.42

To improve computational efficiency and reduce noise, this study selected the most impactful features from the dataset. We introduced XGBoost model into our study, using variance analysis to measure the extent of correlation between features and

target variables. Fig. 5 shows the result of inputting our dataset into XGBoost model. It is clear that the blue team's gold difference affects the blue team's wining rate the most. As our label data is the blue team's wining rate, red team's gold difference only has much less effect on the result.

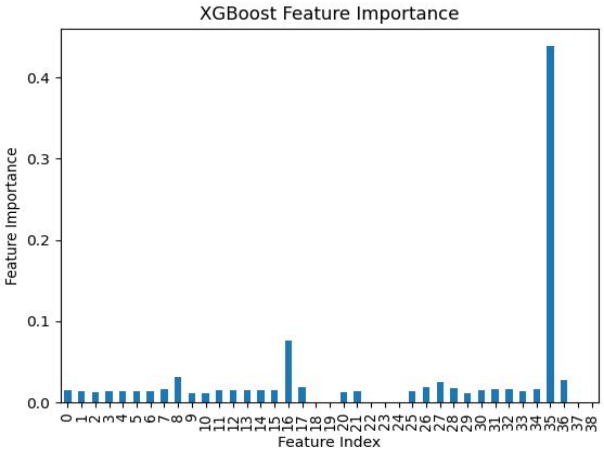

Fig.5. XGB Feature Importance Result(Photo/Picture credit : Original)

After figuring out the importance of features, we dropped out all the features that have little correlation or low in feature importance with the label column--blueWins.

Using the method described in Chapter 2 for simulation, the results of different algorithms were obtained, as shown in Table 3.

Table 3. Simulation results

Metrics		DNN	Logistic Regression	Random Forest	KNN	Naive Bayes	PSO-Random Forest
Accuracy		0.727	0.740	0.747	0.679	0.732	0.751
precision	0	0.75	0.74	0.74	0.68	0.73	0.76
	1	0.71	0.74	0.75	0.68	0.73	0.76
recall	0	0.69	0.74	0.76	0.68	0.74	0.77
	1	0.77	0.74	0.74	0.68	0.73	0.74
f1-score	0	0.72	0.74	0.75	0.68	0.73	0.76
	1	0.74	0.74	0.74	0.68	0.73	0.77

From the discussion and simulation above, this study found that the PSO-Random Forest algorithm fits the game most. The outcome yielded the highest accuracy of 0.751, along with nearly the highest precision, recall, and f1-score. Accuracy is the main factors determining the superiority or inferiority of algorithms, and other evaluating indicators including precision, recall, and f1-score, reveals the accuracy of algorithms from different perspectives. Thus, the experiment result could easily lead us to the concussion that the algorithm this paper introduced has a better performance compared to the baseline algorithms. On the contact, other models achieved the accuracy range from 0.679 to 0.747, lower than PSO-Random Forest algorithm.

4 Conclusion

This study proposes a new algorithm called PSO-Random Forest algorithm, and compared its performance with different baseline algorithms through experiment.

The playing conditions are fair between the two teams. The blue side benefits from more golds and kills and the red side benefits from more neutral objectives and experience. From the result, it is clear that the more gold a team has, the more items they can buy, the easier they win the game. The analysis reveals a significant impact of in-game gold differences on win rates, which is of great significance for understanding game dynamics and formulating strategies. Also, the fast pace and the high randomness of the game means that statistics like gold difference or kills do not determine the final victory. It is the charming place of the game, and also the reason why we can't reach higher predicting accuracy.

Traditional random forest algorithm has some shortcomings in parameter selection and automatic tuning, which may lead to poor model training results. Deciding the parameters could be a hard task when optimizing the algorithm. Introducing PSO optimization method could figure out the fitness parameters, avoid such drawbacks and improve the prediction performance, generalization ability and computational efficiency of the model.

This work is mainly limited by the dataset. On the one hand, by collecting more data, the result may change for DNN model performs better when there is more data input. For another, our dataset does not include the information of the heroes each team chose. In the game, there is usually a counterbalance relationship between different hero characters. This means that some heroes have an advantage over others, while being at a disadvantage against others.

In the future work, the study can collect more data and form a more systematic about game LOL's wining rate predict. With the continuous advancement of technology and the increasing richness of data, the PSO Random Forest algorithm will play a greater role in the field of game AI, providing players with deeper insights and valuable feedback for game developers..

References

1. Lucas, S.M.:Game AI Research with Fast Planet Wars Variants, 2018 IEEE Conference on Computational Intelligence and Games (CIG), Maastricht, Netherlands, 1-4 (2018)
2. Schaetzen, R. De and Sestini, A.: Efficient Ground Vehicle Path Following in Game AI,2023 IEEE Conference on Games (CoG), Boston, MA, USA, 1-4 (2023)
3. Ye, D. et al.: Supervised Learning Achieves Human-Level Performance in MOBA Games: A Case Study of Honor of Kings, in IEEE Transactions on Neural Networks and Learning Systems, 33(3), 908-918 (2022)
4. Tian, P., Lan, W. and Zhang, X.: Hero featured learning algorithm for winning rate prediction of Honor of Kings, 2022 IEEE Conference on Games (CoG), Beijing, China, 322-329 (2022).
5. Li, Y. and Xu, D. -W.: A game AI based on ID3 algorithm, 2016 2nd International Conference on Contemporary Computing and Informatics (IC3I), Greater Noida, India, 681-687, (2016)
6. Takano, Y., Ouyang, W., Ito, S., Harada, T. and Thawonmas, R.: Applying Hybrid Reward Architecture to a Fighting Game AI, 2018 IEEE Conference on Computational Intelligence and Games (CIG), Maastricht, Netherlands, 1-4, (2018)
7. Janusz, A., Grad, Ł. and Grzegorowski, M.: Clash Royale Challenge: How to Select Training Decks for Win-rate Prediction, 2019 Federated Conference on Computer Science and Information Systems (FedCSIS), Leipzig, Germany, (2019).
8. Zhou, Y. and Li, W.: Discovering of Game AIs' Characters Using a Neural Network based AI Imitator for AI Clustering, 2020 IEEE Conference on Games (CoG), Osaka, Japan, 198-205, (2020).
9. Cao, H., Li, S., Gao, M. and Zheng, L.: Research on the end-game of DouDizhu: A method to improve AI of imperfect information game, 2022 International Conference on Cloud Computing, Big Data and Internet of Things (3CBIT), Wuhan, China, 208-212, (2022).
10. Logofatu, D., Leon, F. and Muharemi, F.: General Video Game AI Controller - Integrating Three Algorithms to Bring a New Solution, 2019 23rd International Conference on System Theory, Control and Computing (ICSTCC), Sinaia, Romania, 856-859, (2019)
11. Frans, K.: AI Charades: Language Models as Interactive Game Environments, 2021 IEEE Conference on Games (CoG), Copenhagen, Denmark, 1-2, (2021)
12. Jusoh, S. and Al Fawareh, H.: AI Games and Algorithms: An Overview of Categories, 2023 15th International Conference on Electronics, Computers and Artificial Intelligence (ECAI), Bucharest, Romania, 1-6, (2023)
13. Zhang, Z., Bai, F., Quan, H. -B., Yin, R. -J. and Tao, W. -Q.: PEMFC Output Voltage Prediction Based on Different Machine Learning Regression Models, 2022 5th International Conference on Energy, Electrical and Power Engineering (CEEPE), Chongqing, China, 401-406, (2022)
14. He, H.: Research and Application of Different Machine Learning Algorithms in ILPD Risk Prediction Model, 2023 IEEE 3rd International Conference on Electronic Technology, Communication and Information (ICETCI), Changchun, China, 1330-1334, (2023)
15. Vu, Q. H., Ruta, D., Ruta, A. and Cen, L.: Predicting Win-Rates of Hearthstone Decks: Models and Features that Won AAIA'2018 Data Mining Challenge, 2018 Federated Conference on Computer Science and Information Systems (FedCSIS), Poznan, Poland, 197-200, (2018)