Research on winning rate prediction of e-sport league of legends based on machine learning

Haiting Cui*[a,b]
[a]School of Sport Communication and Information Technology, Shandong Sport University, Jinan 250102, Shandong, China; [b]Faculty of Education, Shandong Normal University, Jinan 250014, Shandong, China

ABSTRACT

E-sport project League of Legends is a top event with a complex mechanism, in order to improve accuracy win rate prediction of League of Legends, the logarithmic energy entropy generates expansion vectors as grouping strategy to enhance contextual information between use of heroes with stronger relationship features, using regulating force mechanism to assign heroes feature weights, extracts heroes lineup selection and role relationship position signal features in turn, establishes a win rate prediction model to combine convolution neural network as a classifier to fully extract the deeper features. Simulation results show that the proposed grouping strategy and feature combination strategy can obtain an overall prediction accuracy 75.32%, which can effectively improve the prediction model accuracy.

Keywords: League of Legends, machine learning, convolution neural network, win rate prediction

1. INTRODUCTION

1.1 League of legends e-sport competition

League of Legends is a classic hero-versus-hero MOBA group competition item developed by Riot Games, and operated by Tencent Games as an agent on the mainland of China. The game possesses 166 individual heroes, and League of Legends has more than 70 million active users per month, which is loved by young people. On the big data analysis of a large number of e-sport events generated by the League of Legends, the pre-match data analysis of various match lineups and intelligent mining of data to assist in the pre-match lineup hero selection, prediction of the match outcome and the corresponding tactical and strategic adjustments play a crucial role.

1.2 Machine learning

Machine learning is a multi-disciplinary cross-discipline, involving probability theory, statistics, algorithmic complexity theory and other disciplines, dedicated to the study of how computers simulate or implement human learning behaviour, in order to acquire new knowledge or skills, and to re-organize the existing knowledge structure so that it continuously improves its own performance. Common classifiers used in machine learning are Naive Bayesian (NB), XGBoost model, Support Vector Machine (SVM) mode[1].

convolution Neural Network (CNN) is a class of Feed-forward Neural Networks that contains convolution computation and has a deep structure, which is one of the representative algorithms of deep learning. The sharing of convolution kernel parameters within the hidden layers and the sparseness inter-layer connections enables convolution neural networks to grid-like topology features with a small amount of computation, and is therefore well suited for modelling applications of natural language processing for e-sport League of Legends tournaments[2]. Therefore, this paper starts from the big data analysis of the actual League of Legends tournament, and uses machine learning related technologies such as convolution neural network to predict the winning rate of the League of Legends tournament, in order to be able to improve the accuracy of the tournament's prediction, and to provide a reference basis for the coaching team's decision-making and the technical team's design.

*361873903@qq.com; phone 17705311529

International Conference on Optics, Electronics, and Communication Engineering (OECE 2024), edited by Yang Yue, Proc. of SPIE Vol. 13395, 133953E · © 2024 SPIE · 0277-786X · Published under a Creative Commons Attribution CC-BY 3.0 License · doi: 10.1117/12.3049149

2. PROCESSING OF BIG DATA AND GROUPING STRATEGIES FOR EXPANDING VECTORS

2.1 Processing of league of legends match data

E-sport tournament theme website can provide a huge amount of e-sport tournament data, such as player e-sport https://www.wanplus.cn/lol/teamstats, the hero alliance tournament data comes from here.

For the hero matching relationship of the League of Legends tournament data extracted and intercepted from the website is screened, and 53530 field data are obtained, and the pre-processing of the data tournament mainly includes abnormal data detection, data normalization constraints, data feature processing and data typing. election conditions and victory results, the choice of heroes and roles are divided into the upper single, fielding, single, auxiliary, shooter five roles, a unified identity, feature extraction and then coding, the data typing is A1-A5, respectively, represent the heroes chosen by A side in the five positions of upper single, playing field, middle single, auxiliary and shooter, and B1-B5 represent the five heroes chosen by B side, and finally 42320 pieces of retained data are obtained showed as Table 1.

Table 1. Matrix of results after data pre-processing.

No.	A1	A2	A3	A4	A5	B1	B2	B3	B4	B5	Result
1	Deep sea	Titan arrow	Bombardier	Mackerel	Gunner ancient	Lords huntre	Nerzumang	Goddess	Alien rock	Sparrow	Win
…	…	…	…	…	…	…	…	…	…	…	…
42320	Ice sharps	Chantor death	Mechanical	Trickster	Master cards	Flamshannen	Blind Monk	Trick siren	Soul clean	King break	Loss

2.2 Generation of data expansion vectors

2.2.1 Acquisition of word vectors. Skills of each hero is a balance point of reciprocal coordination, the result of data analysis is to obtain the coordination of the hero of the party, inhibit the opponent's hero's output strength[3]. Therefore, the selection of hero word vectors is divided into two levels, one is to obtain the collocation relationship of the model, and the other is to expand the competitive relationship of the vectors showed as Figure 1.

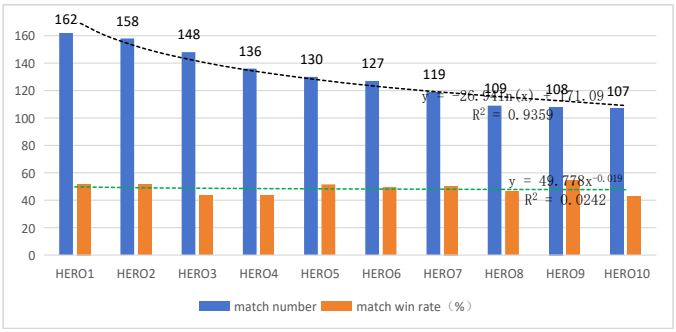

Figure 1. Heroes (time assassin) best ten matching heroes match count and match win rate.

2.2.2 Expansion of word vectors. From the acquisition of the hero word vector to establish a training corpus, expand the feature indicator location marking, repeated training and iterative screening[4], so as to calculate the priority of the relevance of this vector and the expansion of the space, the number of growth levels, the correlation coefficient between the semantic vocabulary is the degree of the relationship between the heroes with the degree of cooperation, the energy entropy objective function of the training model is:

$$\mathrm{LogE} = \sum n_{i=1}\ \log 2(p_i^n)\cdot N \tag{1}$$

where, p_i represents the regression component reveal rate of the i hero attribute selection, for a specific hero appearance, the attribute play should be stable, n is the length of the corresponding energy entropy, N is the coefficient of the expansion dimension of the word vector, so as to derive the value of the similarity power between the heroes[5].

3. CONSTRUCTION OF WIN RATE PREDICTION MODEL

3.1 Hero lineup selection

3.1.1 Establishment of collocation relationship. The order of hero lineup selection affects the matching relationship of heroes, and the heroes that get higher performance are chosen and prevented from being chosen by the opponent, and have a higher chance of winning. The winning rate of the match result is affected by the three aspects of the hero's appearance rate, the hero's appearance order, and the hero's banned rate showed as Figure 2.

Figure 2. Graph of some top heroes and auxiliary support heroes.

3.1.2 Correlation of competitive relationship. The establishment of the competitive relationship by the players certain hero selection experience and familiarity, according to the judgement for the prejudgment, and even cold heroes[6], heroes of the competition and restraint relationship is often associated with the win rate of the game, heroes using the difference between the expected win rate and the actual win rate of the existence of the competition to quantify the possibility of the competitive correlation coefficient is calculated as follows:

$$E_{ij} = \frac{P_i - P_i \times P_j}{P_i + P_j - \partial \times P_i \times P_j} \tag{2}$$

E_{ij} indicates the expected and probable rate of competitive association between heroes and hero hand-offs, and the competitive relationship coefficient takes values in the range of 0 to 1, indicating the degree of comprehensive restraint pairing relationship.

3.2 Characteristics of role-relative position signaling

The matching degree of hero position can be used to represent the correlation coefficient between the frequency of appearance of heroes in the position and the winning rate, and it is necessary to find out the degree of fit and the combination order of the data set and the hero position, and transform the order of the hero position into the problem of arranging the hero features, calculate the logical order and the possibility of appearance of the contextual matching, and then solve the bias of the judgement of the front and back position in the machine model[7].

3.3 Regulatory power mechanism distribution

3.3.1 Hero win rate contribution rate. Data features and further loop iterations to match the hero weights before entering the classification period. In the ranking of LOL hero contribution, the hero's matching mode and position, the operation and development of the front, middle and late stage, the equipment and skill attribute synthesis and enhancement and the timing of the phasing, will give you a vital role in the late group battle, the calculation and selection of the hero's win rate contribution needs to take this into account, that is, the ability to have the ability to open the group and bear the damage.

3.3.2 Allocation of attention regulation. Attention regulation makes heroes with larger weights of hero win rate contribution have higher parameter appearance rate appearance rate and, thus focusing more advantageous resources to optimize the iterative global win rate prediction, avoiding falling into the local optimum[8]. The allocation mechanism of attention regulation can be viewed as the weight allocation standard of attention, and the scoring mechanism is defined as:

$$\theta_i = \int （q, X） \cdot \sum_{i=1}^{N} \partial_i X_i \tag{3}$$

3.4 Structure of the win rate prediction model based on convolution neural network

3.4.1 Model building environment. The model language is Python environment, using the model of machine learning for the win rate prediction environment construction, the process of model construction is showed as Figure 3.

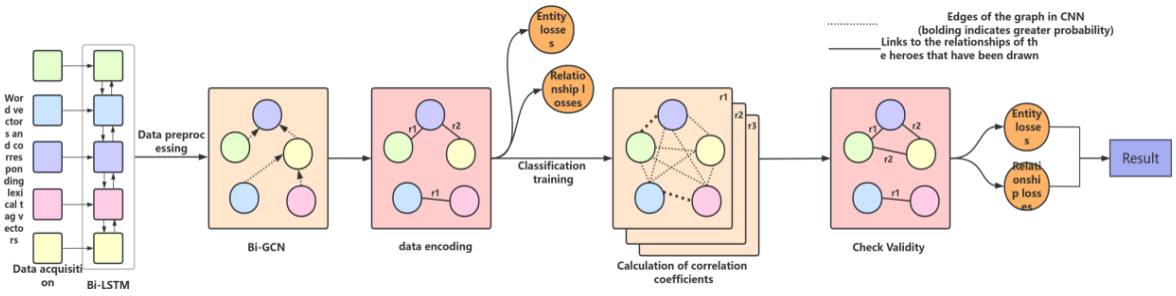

Figure 3. Flow of building the model.

Model win rate prediction needs to build assessment indicators environment[9], the closer the prediction accuracy rate to 1 indicates the model is more accurate, will be expected to assess the accuracy rate samples obtained at the full accuracy rate and prediction accuracy rate, the sum of the actual accuracy rate to get the model's accuracy rate formula is:

$$\mathrm{Ar} = \frac{\mathrm{TP+TF}}{\mathrm{TP+TF+FP+FF}} \tag{4}$$

3.4.2 Model structure of convolution neural network. Assuming that the hero lineup matrix is G, then the rows of G represent the hero vectors in the lineup, and there are 10 heroes in each input lineup, and the word vector dimensions are composed of collocation relationship and restraint relationship together, which is now set to R. Then G is a 10R relationship dimension matrix, and when the lineup information is inputted into the convolution neural network classifier as follows:

$$G=f\begin{pmatrix} A1 & \cdots & A5 \\ \vdots & \ddots & \vdots \\ B1 & \cdots & B5 \end{pmatrix} \tag{5}$$

Using 10 filters, it is possible to generate 10 hero features, using convolution neural networks to mine the data for local feature models for data fitting, obtaining curves as shown in Figure 4. Hero radar relationship attributes as shown in Figure 4.

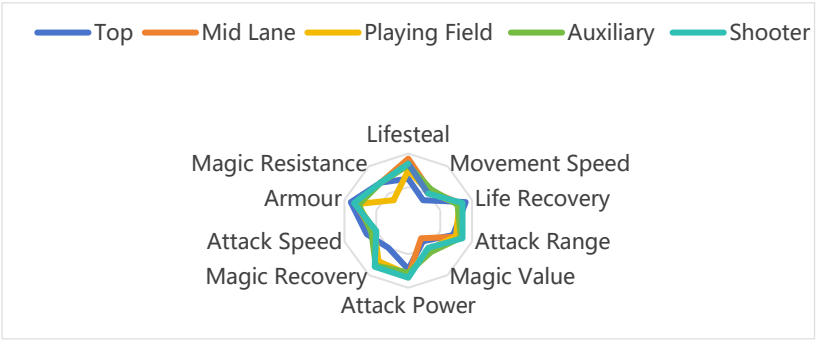

Figure 4. Radar analysis of hero screening data.

Among them, the coefficients of hero pairing and competitive relationship expressed by the word vectors described in the previous section are used as inputs to the model, and the top, field, mid, auxiliary, and shooter variations corresponding to the heroes are labeled respectively to extract the before-and-after memory relationship of the transmission position[10].

4. ANALYSIS OF SIMULATION RESULTS

4.1 Analysis of data results

Using XGBoost model and PSO-SVM model, which are commonly used machine learning models in data mining field

for comparison, the win rate prediction model CNN based on convolution neural network machine learning maintains good restraint relationship and echo relationship, and the accuracy rate in win rate prediction is better. It can be seen that the first two commonly used learning methods are much lower than the convolution neural network model. Among the traditional methods, the XGBoost model achieved the best prediction effect, with a prediction accuracy of 65.23%, but none of the League of Legends win rate prediction models based on the traditional methods reached an accuracy of 70%, which is far from the neural network in terms of the model's expressive power, which shows that the traditional algorithms can't predict the match results very well.

4.2 Testing the effectiveness of the winning percentage model

Lineup data with five known labels were used to determine whether the classification results of the predictive model matched the actual labels. Through the validation experiment of this win rate prediction model, it can be seen that the CNN win rate prediction model already has some practical value, and the research method of this paper on the League of Legends win rate prediction problem is also more reasonable. A CNN win rate prediction model without using the expansion vector as input is constructed, and the model can achieve an accuracy of 75.32%, which also verifies the role of the restraint relationship in the win rate prediction, and it can be learn through the model effect validation experiments that the CNN win rate prediction model has a better model application effect.

5. CONCLUSION

A CNN win rate prediction model without using the expansion vector as input is constructed, and the model can achieve an accuracy of 75.32%, which also verifies role of the restraint relationship in the win rate prediction, and it can be learn through the model effect validation experiments that the CNN win rate prediction model has a better model application effect. As it is difficult to quantify the characteristics of objective and subjective factors such as the environment and process of the player's game, the prediction model constructed in this paper excludes the influence of the above factors, and it is correct and feasible to expand the vector to quantify the hero's restraining relationship, and it is possible to use the knowledge of the neural network to solve the problem of the interrelationships between the heroes and the problems related to the selection of the lineup to effectively improve the accuracy of the prediction model, and the use will effectively improve the accuracy of the original model. The CNN model can capture the local features of the input lineups, and at the same time, through the feature extraction of the CNN layer, it reduces the number of neural network weights, which makes the fitting speed of the prediction model faster.

REFERENCES

[1] Daniel, C., "The impact of ball recovery type, location of ball recovery and duration of possession on the outcomes of possessions in the English," Science and Medicine in Football 4(3), 196-202 (2020).

[2] Johansson, F. and Wikström, J., "Result prediction by mining replays in Dota2," Faculty of Computing 2, 371 (2015).

[3] Rajwinder, S. and Harshita, P., "An efficient language-independent acoustic emotion classification system," Arabian Journal for Science and Engineering 45(1), 3111-3121 (2020).

[4] Daniel, K., "The theory of the quantum kernel-based binary classifier," Physics Letters A 384(21), 401-405 (2020).

[5] Dai, S., Li, Z., Li, L., Cao, D., Dai, X. and Lin, Y., "Investigating the dynamic memory effect of human drivers via ON-LSTM," Science China (Information Sciences) 63(09), 83-93 (2020).

[6] Lee, J. H., Lee, M. and Kim, J. W., "A study on Korean language processing using TF-IDF," The Journal of Information Systems 28(3), 105-121 (2019).

[7] Intan, N. Y., "Bi-directional long short-term memory using quantized data of deep belief networks for sleep stage classification," Procedia Computer Science, 116, 530-538 (2017).

[8] Sriramakrishnan, "Deep learning approach for customer records prediction for future products issues," Scientific Research Publishing, 7(3), 44-54 (2019).

[9] Hao, J., "Machine learning made easy: a review of scikit-learn package in python programming language," Journal of Educational and Behavioral Statistics, 44(3), 348-361 (2019).

[10] Kumar, V., "Problem formulations and solvers in linear SVM: a review," Artificial Intelligence Review, 52(2), 803-855 (2019).