Academic Journal of Computing & Information Science
ISSN 2616-5775 Vol. 3, Issue 5: 17-22, DOI: 10.25236/AJCIS.030503

Game Predication of FIFA Football World Cup Based on Support Vector Machine

Shengtai Ding

Suzhou Science and Technology Town Foreign Language School
SSFLS@szlunhua.com

ABSTRACT. *The purpose of this paper is to evaluate the state of the team through the player's post-match scores and game data. the correlation of these indicators is very strong. Then SVM, a classic classification algorithm, is used to predict the winner of FIFA Football World and it has good classification performance. The results showed that SVM with gauss kernel suppressed that with linear kernel. The average score of all players in a team reflects the strength and state of the team. Using player ratings to predict game results is more accurate than using historical records and the relationship between team wins and losses.*

KEYWORDS: *Game Predication, FIFA Football World Cup, Support Vector Machine*

1. Introduction

FIFA World Cup is a global football festival held every four years, founded by FIFA [1-2]. The participating teams are the top 48 national teams from all over the world, which symbolizes the highest honor in the football world and the most famous and influential football match. The World Cup schedule is divided into two stages, one is the qualifier stage, and the other is the final stage. The qualifiers are divided into six major competitions, namely Europe, South America, Asia, Africa, North America and Oceania. Each FIFA member country (region) team that has signed up for the World Cup needs Conduct qualifiers in your region and compete for places in the World Cup finals. There are only 48 places to enter the finals, and the host can directly qualify for the finals. The final stage of the group stage is divided into 12 groups, each with four teams, determined by drawing lots. The top two teams in each group of points qualify for the 1/8 finals, and then the winners go to the 1/4 finals, and finally the finals for the "Hercules Cup".

The media and the team can use these analyzed data to determine whether the players or lineups in the team are suitable for this week, whether further adjustments are needed, whether young blood needs to be introduced to the team, and what will happen to the game after the team has adjusted and so on. These are also some of the decision-making factors and issues that the team management needs very much.

ISSN 2616-5775 Vol. 3, Issue 5: 17-22, DOI: 10.25236/AJCIS.030503

Through the data of each FIFA World Cup, the support vector machine method is used to predict the situation of the next World Cup.

There are many factors that affect the performance of sports competitions [3-5], including the training level of the athletes, the competitive state, the performance of the opponents, the attention of the society, the coach and the person, and the weather conditions, etc., and these factors show a non-linear relationship. For real sports enthusiasts, the four-year FIFA World Cup is absolutely as influential as the Olympic Games. To host the World Cup, in addition to the level of football development as the foundation, it must also be guaranteed by a certain degree of economic strength. Taking the past ten years as an example, the host countries have invested huge amounts in stadium construction, and many commendable sports architecture boutiques have emerged.

2. Model Construction

Before introducing support vector regression, we first introduce the support vector machine (SVM). When using the perceptron to classify, it is necessary to find multiple hyperplanes that can be classified to classify the data, and try to ensure that other points are based on hyperplanes during optimization. However, the actual situation is that there are many points very close to the hyperplane, and these points are often difficult to correctly classify, and the misclassification rate is high. In order to make it easy to classify points that are originally very close to the hyperplane, we need to adopt some measures, and the idea of SVM is derived from keeping the points very close to the hyperplane away from the hyperplane. The definition of SVM is shown in Figure 2-4. Set the classification hyperplane as $ax + b=0$. It can be seen from the figure that all samples can not only be well separated by the hyperplane, but also maintain a certain amount with the hyperplane. Such a classification hyperplane is better than the classification hyperplane of the perceptron. It can be proved by mathematical methods that there is only one such hyperplane. As shown in the Figure 1, the hyperplane represented by the dotted line is parallel to the classification hyperplane and maintains a certain distance. The vector corresponding to the dotted hyperplane is defined as a support vector. Through mathematical derivation, the distance from the SVM to the classification hyperplane is 1/2|w|, and the distance between the two planes is 2/|w|.

ISSN 2616-5775 Vol. 3, Issue 5: 17-22, DOI: 10.25236/AJCIS.030503

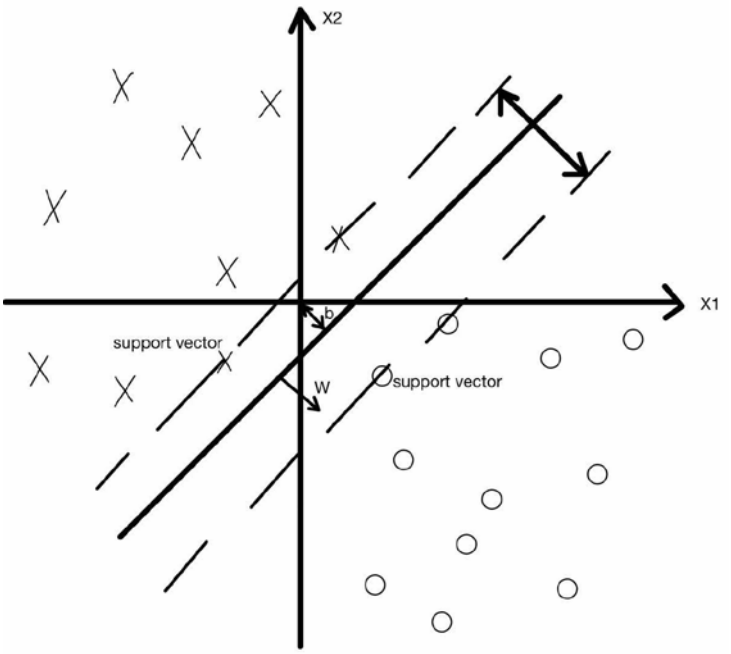

Figure. 1 Introduction of SVM

3. Data research

Research data is obtained from the website www.kaggle.com, which contains a large amount of data in different industries. What I check is from the database "FIFA World Cup" collected and created by Andre Becklas. All these data are from the real football World Cup every four years from 1930 to 2014. All data indicators collected by this database are as follows:

'Name', 'Age', 'Photo', 'Nationality', 'Flag', 'Overall', 'Potential', 'Club', 'Club Logo', 'Value', 'Wage', 'Special', 'Preferred Foot', 'International Reputation', 'Weak Foot', 'Skill Moves', 'Work Rate', 'Body Type', 'Real Face', 'Position', 'Jersey Number', 'Joined', 'Loaned From', 'Contract Valid Until', 'Height', 'Weight', 'LS', 'ST', 'RS', 'LW', 'LF', 'CF', 'RF',' RW', 'LAM', 'CAM', 'RAM', 'LM', 'LCM', 'CM', 'RCM', 'RM', 'LWB', 'LDM', 'CDM', 'RDM', 'RWB', 'LB', 'LCB', 'CB', 'RCB', 'RB', 'Crossing', 'Finishing', 'Heading Accuracy', 'Short Passing', 'Volleys', 'Dribbling', 'Curve', 'FK Accuracy', 'Long Passing', 'Ball Control', 'Acceleration', 'Sprint Speed', 'Agility', 'Reactions', 'Balance', 'Shot Power', 'Jumping', 'Stamina', 'Strength', 'Long Shots', 'Aggression', 'Interceptions', 'Positioning', 'Vision', 'Penalties', 'Composure', 'Marking', 'Standing Tackle', 'Sliding Tackle', 'GK Diving', 'GK Handling', 'GK Kicking', 'GK Positioning', 'GK Reflexes', 'Release Clause'.

ISSN 2616-5775 Vol. 3, Issue 5: 17-22, DOI: 10.25236/AJCIS.030503

Figure. 2 Countries of FIFA Football World Cup

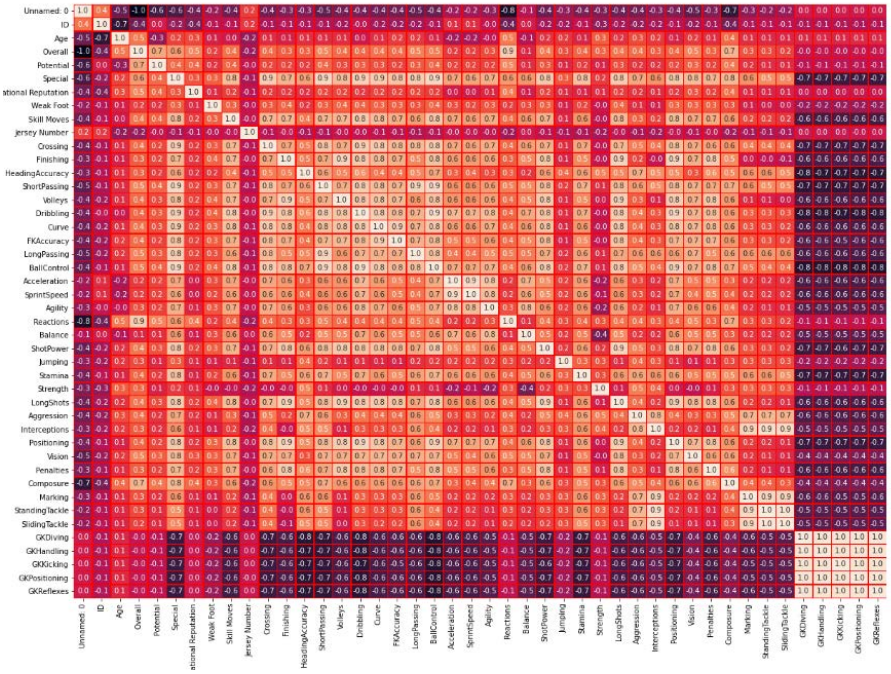

Figure. 3 correlation result of each indicator

The figure above is the correlation result of each indicator. We can see that the correlation of these indicators is very strong.

ISSN 2616-5775 Vol. 3, Issue 5: 17-22, DOI: 10.25236/AJCIS.030503

4. Results

SVM model has two very important parameters c and gamma [7]. Where C is the penalty coefficient, that is, the tolerance to error. The higher the C, the less tolerance of error and easy over fitting. The smaller C, the less fitting. If C is too large or too small, the generalization ability becomes poor. Gamma is a parameter of RBF function after it is selected as kernel [8]. The larger the gamma, the less the support vector, and the smaller the gamma value, the more support vectors. The number of support vectors affects the speed of training and prediction. he comparison of the experimental results is shown in Table 1.

Table 1 evaluation of SVM with different kernels

	F_1 score	ROC
Linear kernel	0.66	0.63
Gauss kernel	0.71	0.68

The experimental results of the various models used in this article. SVM with Gauss kernel has the higher F_1 score and ROC on the test set, which is 0.71 and 0.68 respectively. The SVM model works very well on the training set, and the test set is average. Since the data set used in this article is not large enough, the complex model may not be well trained and prone to overfitting. Under the improved data set, SVM is a very promising model.

5. Conclusions

This thesis mainly introduces the application of player ratings in the prediction of match results. Player ratings are closely related to the performance of players and teams. After getting accurate player ratings, they can be used in many areas of football. For example, a player's rating reflects a player's abilities and status to a certain extent, and analyzing a player's rating changes over a period of time can reflect the fluctuation of the player's status or the rise or fall of abilities. It can be used for player value prediction. A team's victory or defeat does not depend on its strength. The better-performing team loses because of luck or refereeing. The victory of the game does not fully reflect the strength of the team.

References

[1] Dvorak, Jiri, et al. "Medical report from the 2006 FIFA world cup Germany." British journal of sports medicine 41.9 (2007): 578-581.

[2] Cornelissen, Scarlett. "More than a sporting chance? Appraising the sport for development legacy of the 2010 FIFA World Cup." Third World Quarterly 32.3 (2011): 503-529.

ISSN 2616-5775 Vol. 3, Issue 5: 17-22, DOI: 10.25236/AJCIS.030503

[3] Ljdokova, Gallia Mikhaelovna, Oleg Anatolevich Razzhivin, and Kadriya Rafikovna Volkova. "Powerlifters' ways to overcome confounding factors at competitions." Life Sci J 11 (2014): 481-484.
[4] Javadi Khasraghi, Hanieh, and Abdollah Aghaie. "Crowdsourcing contests: understanding the effect of competitors' participation history on their performance." Behaviour & Information Technology 33.12 (2014): 1383-1395.
[5] Yustres, I., et al. "Swimming championship finalist positions on success in international swimming competitions." PloS one 12.11 (2017): e0187462.
[6] Wu, Lyndia C., et al. "Detection of American football head impacts using biomechanical features and support vector machine classification." Scientific reports 8.1 (2017): 1-14.
[7] Igiri, Chinwe Peace. "Support Vector Machine–Based Prediction System for a Football Match Result." IOSR Journal of Computer Engineering (IOSR-JCE) 17.3 (2015): 21-26.
[8] Chen, Rung-Ching, and Chung-Hsun Hsieh. "Web page classification based on a support vector machine using a weighted vote schema." Expert Systems with Applications 31.2 (2006): 427-435.