

Research on the prediction of sports competition results based on momentum modeling

Kan Hu*, Mingyi Ke

School of Physics and Electrical Engineering, Hanjiang Normal University, ShiYan, China, 442000

* Corresponding Author Email: 17371103187@163.com

Abstract. This study examines the role of “momentum” in sports competitions. Using a trained model to predict all samples, we analyzed the correlation and effect of “momentum” with athletes' scores through Pearson's correlation coefficient and one-way linear regression. The results showed that “momentum” was moderately correlated with athletes' scores and had a significant effect on the outcome of the game. We also built a judgmental descriptive model, aggregated and normalized the data, and performed binary logistic regression analysis. The results show that the model has good accuracy in predicting whether an athlete will be able to win a race. In addition, we validated it using machine learning algorithms and plotted ROC curves and AUC values. The results show that traditional machine learning models perform well in predicting the outcome of a single match. Finally, suggested strategies are proposed for different types of players, including studying the opponent's game video in depth, emphasizing on psychological preparation, making specific tactical plans, and maintaining flexibility and adaptation to improve performance in the game.

Keywords: Pearson's correlation coefficient, One-way linear regression, Momentum.

1. Introduction

In sports, victory often depends on a combination of many factors, one of which is the concept of “momentum”, which has received much attention[1-2]. During a match, the ups and downs of athletes' performances and changes in the game can affect the flow of this kinetic energy, which can have a significant impact on the outcome of the match. The purpose of this study is to investigate the mechanism of kinetic energy in sports competitions by analyzing and modeling the data using the kinetic energy model. We will provide insights into the associations between athlete performance and game outcomes, with a view to revealing the potential impact of “momentum” on game outcomes. Through this study, we expect to provide new theoretical support and practical guidance for outcome prediction and tactical strategies in the field of sports competition[3-4].

2. Momentum Working or Not

All we need to do is to use the trained model to make a prediction for all the samples, and this prediction is actually called " momentum ", based on which we only need to prove that the correlation between this momentum and the actual scoring of the players is significant and can significantly affect the outcome of the game, in order to show that the momentum is playing a role in the game[5-6].

2.1. Model Establishment:

We used Pearson's correlation coefficient to analyze the correlation and one-way linear regression to analyze the influence

The Pearson correlation coefficient[7] is a statistical indicator of the degree of linear correlation between two variables, and its value is between -1 and 1. The closer the Pearson's correlation coefficient is to -1 or 1, the stronger the linear correlation between the two variables; if the Pearson's correlation coefficient is close to 0, it means that there is almost no linear correlation between the two variables. Its calculation formula is:

$$r = \frac{\Sigma(X - \bar{X})(Y - \bar{Y})}{\sqrt{\Sigma(X - \bar{X})^2 \Sigma(Y - \bar{Y})^2}} \quad (1)$$

Two variables X (momentum) and Y (whether the player scores or not), which contain n observations each. where $\bar{X}$ and $\bar{Y}$ are the average of X (momentum) and Y (whether the player scores or not), respectively.

The p -value is used to measure the probability of the observed data occurring under the null hypothesis (i.e., there is no correlation between the two variables). The t -statistic is calculated based on the sample size and correlation coefficient with the formula:

$$t = r \sqrt{\frac{n - 2}{1 - r^2}} \quad (2)$$

where n is the number of observations and r is the Pearson correlation coefficient.

Using the t -statistic and the degrees of freedom here as $(n - 2)$, calculate the p -value.

Finding a linear equation to best fit a set of data points allows this linear equation to predict, with the highest possible accuracy, the value of a dependent variable (whether or not to score), based on the value of an independent variable (momentum). This relationship is considered linear, meaning that changing one or more of the independent variables will cause a proportional change in the dependent variable[8].

2.2. Solving and Evaluation of results

Where the Pearson's correlation coefficient is 0.482, there is a moderate correlation with whether the player scored or not, while the p -value is small and there is a significant linear correlation between the two.

The two were then analyzed by one-way linear regression. Where the coefficient of determination is 0.233, which indicates that the model explains 23.3% of the variability in the data, implying that the model has limited explanatory strength and most of the variability in the data is not captured, the adjusted coefficient of determination is close to the coefficient of determination, which is slightly adjusted to the explanatory strength of the model after taking into account the effect of the number of independent variables[9-10].

F-statistic & Prob (F-statistic), they denote the F-statistic and its test of significance, reflecting whether the whole model is significant or not, here it is 376.3 and probability value close to 0, which indicates that the statistical significance of the model is very high and momentum has a significant effect on the direction of the game. t -value is 19.406 and p -value close to 0 (< 0.0001), which indicates that the the effect of momentum on the direction of the game is highly significant and positively correlated.

All things considered, momentum plays a more significant role in the game.

3. Model Testing Optimization

Based on Evaluation of predictive models, we can know that the prediction for each score is very difficult, if we can in general (i.e., for each game) can judge whether to get the victory, is also very valuable, therefore, we in Judgmental Descriptive Model based on the aggregation of the data, each sample represents the content for a game of the information, label that is, for the current game game whether to get the victory.

The samples with abnormal scores were excluded, the samples with missing speeds were deleted, aggregated according to the mean value, and similarly. Similarly, in order to verify whether the proposed metrics have a significant effect on whether a player gets a GAME victory or not, we subjected the data to a binary logist regression analysis to determine the significance of these metrics. It can be seen that the accuracy of binary logistic regression is at 71.4, which is already very good for

predicting whether or not a player will get a single GAME in an actual match, and relative to the results of Evaluation of predictive models, the existing results have many variables that were not significant that are also starting to be significant.

Similarly, we take machine learning algorithm for comparison and apply 5-fold cross-validation method for validation and the results are as table 1.

Table 1. 5-fold cross-validation method

	Accuracy	Recall	Precision	F1	Auc
LGBM	0.66	0.67	0.66	0.67	0.72
XGB	0.65	0.67	0.65	0.66	0.71
SVC	0.71	0.75	0.7	0.72	0.75
MLP	0.7	0.73	0.7	0.71	0.75
LR	0.7	0.73	0.7	0.71	0.76

The ROC curve was plotted and AUC values were obtained as figure 1.

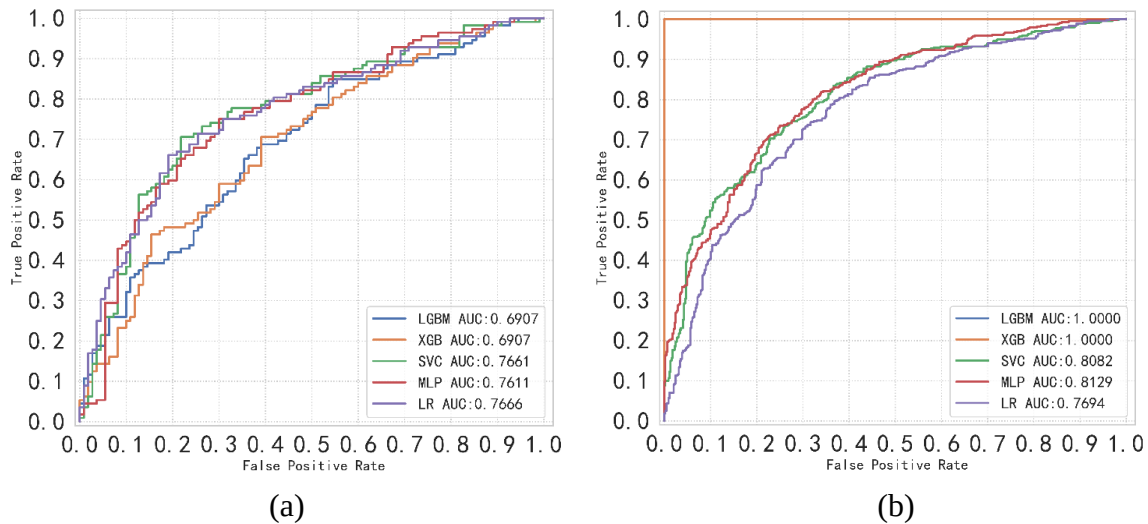


Figure 1. (a) shows the ROC curves for the test set,(b) shows the ROC curves for the train set

For the prediction of a single GAME, we can find that the traditional machine learning models have better performance relative to the tree model, among them, the best performance should be SVC because it has the highest f1 score, which means that the prediction of all the samples can take into account the samples in each category, and we get the graph of feature importance scores of the algorithmic models with the best results this time as figure 2.

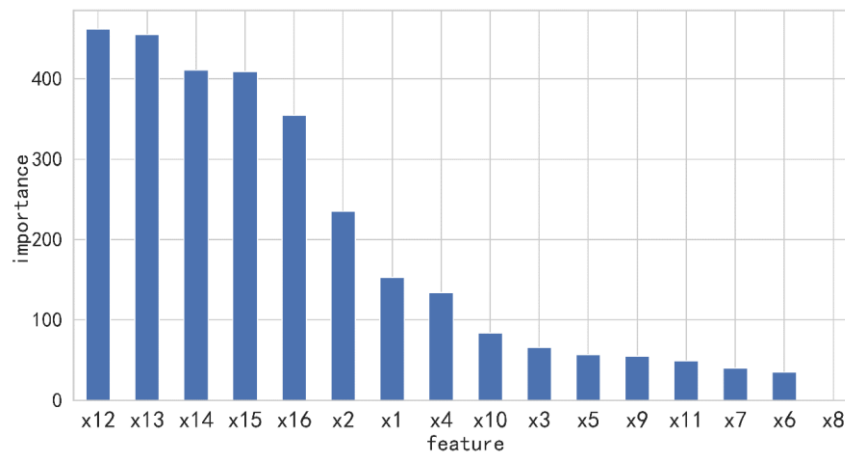


Figure 2. Histogram of Feature Importance Scores for the SVC Model

It is not difficult to find that the feature importance scores in this model test changed compared to

the first test in Evaluation of predictive models, the highest feature importance score for whether a single GAME was won was X_{12} , followed by X_{13} , X_{14} , on this basis, we re-trained the model, again using the best performing SVC model, selecting the classic match in the 2023 Wimbledon men's singles final in which 20-year-old Spanish star Carlos Alcaraz defeated 36-year-old Novak Djokovic for real-time performance visualization.

Since SVC cannot compute feature importance scores with a Gaussian kernel, we used the information gain algorithm for feature importance score computation as table 2.

Table 2. feature importance score computation

	feature	score
3	x4	0.10108
1	x2	0.095181
6	x7	0.063331
10	x11	0.016805
9	x10	0.009771
15	x16	0.001026
5	x6	0.0004
0	x1	0
2	x3	0
4	x5	0
7	x8	0
8	x9	0
11	x12	0
12	x13	0
13	x14	0
14	x15	0

Based on this result, it can be known that the game's score leading progress is most conducive to improving a player's performance, followed by whether the last point was scored (average), which means that this player has a hot hand, and also the no-touch return score, which came third, shows the player's hand.

Differences in "momentum" fluctuations mean that different players have different abilities to maintain or regain momentum in the face of pressure, loss or adversity. Some players may perform at a consistently high level throughout the game, while others may show more ups and downs.

Suggested strategies for players facing new matches with different players:

- study the opponent's game footage in depth: find out how their momentum has changed in previous matches, paying special attention to those key moments that can change the opponent's momentum.
- Emphasize mental preparation: adjust your mental expectations and tactical play according to the momentum fluctuation characteristics of your opponent. For those opponents who are prone to fluctuations in momentum, develop strategies to capitalize on this, such as trying to establish a lead early in the game.
- Develop tactical plans: Develop specific tactics based on your opponent's weaknesses. If opponents are prone to losing momentum in certain specific situations (e.g. after conceding a goal), this can be exploited by stepping up the offense or changing tactics.
- Staying adaptable: Players need to be prepared to adapt to changes in the game, especially when the opponent tries to change the game situation to regain momentum. Keeping tactics flexible and adaptable is key.
- Enhance your own momentum management: As well as analyzing the opposition, players need to train themselves on how to build and maintain positive momentum during the game and how to recover quickly when faced with adversity.

4. Conclusions

In this study, we used the momentum model to predict and analyze the results of sports competitions, and found that momentum has an important impact on the results of competitions. Through data analysis and modeling, we reveal the association between athlete performance and game outcomes and verify the validity of the momentum model. In practical tests, we found that the model was stable in assessing athletes' performance, but had some limitations in generalization ability. Nonetheless, this study provides a new theoretical framework and methodology in the field of sports competition, which helps to guide the tactical strategy and outcome prediction of the game. In the future, we will further optimize the model to improve its generalization ability and explore more factors on match outcomes to enhance the accuracy and usefulness of the model.

References

- [1] YANG Longfei,LIU Qingyao,ZHANG Jiangnan. Research and analysis on analyzing and predicting the results of men's singles badminton matches in China and the structure of winning and losing[C]// China Wisdom Engineering Research Association,China Bandi Association,Guangdong Physical Fitness Association. Proceedings of the 10th China Physical Training Science Conference (in Chinese). North Central University;, 2023: 5.
- [2] Qiu-Shi Wang,Garin Gahangnur,Lan Ding et al. Research progress of big data analysis based on open source information of professional tennis[C]// Chinese Society of Sports Science. Abstracts Collection of the 13th National Sports Science Conference - Wall Paper Communication (Athletic Training Section) (II). Beijing Sport University;, 2023: 3.
- [3] Dong Meiyin, Lian Bizhen. Advantage analysis and revisit study of the "four factors" statistical index in CBA league[C]// Chinese Society of Sports Science. Abstracts of the Thirteenth National Sports Science Conference--Special Report (Athletic Training Section). Beijing Sport University; Shijiazhuang College;, 2023: 3.
- [4] Shao-Liang Zhang,Half-Mate Yang,Lei Chen. Research on integrated learning-driven prediction model for Chinese soccer super league[C]// Chinese Society of Sports Science. Abstracts Collection of the Thirteenth National Conference on Sports Science - Presentation (Sports Statistics Division). Tsinghua University;, 2023: 3.
- [5] Yang T, Cheng Z, Feng LJ et al. Correlation analysis between osmotic behavior and match winning and losing in 2022 FIFA World Cup matches[C]// Chinese Society of Sports Science. Abstracts Collection of the 13th National Sports Science Conference - Special Presentation (Sports Statistics Division). Northern Nationalities University; South China Normal University; Weinan Normal College;, 2023: 3.
- [6] TAN Wentao, LONG Qiusheng, LI Zhilan. A review of research on badminton match result prediction model based on support vector machine[C]// Chinese Society of Sports Science. Abstracts of the Thirteenth National Conference on Sports Science - Special Report (Sports Statistics Division). Guangzhou Institute of Physical Education; Guangdong Second Normal College;, 2023: 3.
- [7] Zhang L F, Zhu Y, Wu Xuan et al. Data-driven analysis of the winning factors of Chinese elite male singles tennis players--The case of Zhang Zhizhen[C]// Chinese Society of Sports Science. Abstracts Collection of the 13th National Sports Science Conference--Wall Paper Exchange (Athletic Training Section) (III). School of Physical Education, Southwest University; School of Physical Education and Health, Dalian University of Technology;, 2023: 3.
- [8] Ban Yue. Research on soccer match result prediction based on multiple machine learning models[C]// Chinese Society of Sports Science. Abstracts Collection of the 13th National Sports Science Conference - Wall Communication (Sports Engineering Division). Capital Sports Institute;, 2023: 3.
- [9] Jiang HF,Yu HL,Wei L. Research on the analysis of influencing factors of soccer game based on fuzzy gray correlation analysis [J]. Computer and Digital Engineering, 2023, 51 (03): 555-560.
- [10] Xiu Tingche, Yu Hongyan, Huang Wenyan. Construction and simulation application of table tennis excellent women's singles match result prediction model based on BP neural network[C]// Athletic Training Branch of Chinese Society of Sports Science. Abstract collection of 2022 National Sports Training Symposium (II). Department of Physical Education, Shanghai Jiao Tong University;, 2022: 1.