Academic Journal of Science and Technology
ISSN: 2771-3032 | Vol. 12, No. 1, 2024

Research on Prediction of Tennis Competition Fluctuations Based on CRITIC Weight Method and CNN-LSTM Model

Rong Wang [1,#], Yujie Sun [2,#]

[1] College of Textiles, Donghua University, Shanghai, 201620, China
[2] College of Information Science and Technology, Donghua University, Shanghai, 201620, China
#These authors contributed equally.

Abstract: This paper mainly studies the correlation between the "momentum" index and match outcome in tennis matches, and establishes a model to predict competition fluctuations. First, the index weights associated with "momentum" were determined by the CRITIC weight method, and the Spearman correlation coefficient between "momentum" and "performance coefficient" was 0.479, which indicated that there was a significant correlation between these two. Then, this paper analyzed the correlation between other indicators and the "performance coefficient", and found that the successful first serve, unforced error, running distance, hitting number and other indicators had an important influence on the fluctuation of the game. In order to predict competition fluctuations, CNN and LSTM were combined to form a prediction model, with good results in the 2023 Wimbledon competition data (RMSE <1). In addition, the prediction model was used to prove that competition fluctuations and player wins are not random, and using at least one game data to explore the factors associated with the change of competition situation, laying the foundation for developing predictive models. The index weight allocation was conducted by CRITIC method, which explored the relationship between "momentum" and the result of competition using Spearman correlation analysis, established the CNN-LSTM deep learning model for prediction, and achieved high prediction accuracy.

Keywords: CRITIC Weight Method; CNN-LSTM; Tennis Competition Fluctuations.

1. Introduction

Tennis is elegant and intense, known as the world's second largest ball sport. The thrilling and stimulating competition process makes it highly ornamental. Both sides on the tennis court are doing their best to become eternal winners. Seemingly lonely confrontation exudes endless glory and confidence, which makes it welcomed by people all over the world. The extraordinary performance of one of the greatest Grand Slam players ever ended when veteran Novak Djokovic was defeated in the 2023 Wimbledon final by rising star Carlos Alcaras. It was his first Wimbledon loss since 2013.These kinds of incredible swings, always be boiled down to momentum’s impact.

In sports events, a team or player may feel they have some kind of power which make them play better or worst during a match. It seems self-evident that it is difficult to measure such a phenomenon. Nowadays, fans have become more and more enthusiastic about predicting the outcome of games while coaches will adjust their strategy before the games. Therefore, our team decided to build a model to evaluate how well a sportsman performs during the competition and to assess the role of momentum in it. Moreover, predicting the results.

To study the correlation between “momentum” index and game results in tennis competitions and predict the fluctuations, flow chart is shown in figure 1.

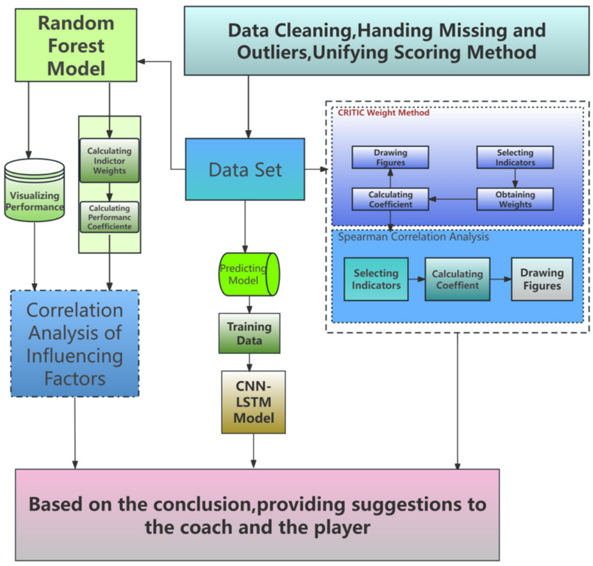


Figure 1. flow chart

To achiebe this goal, we have made the following assumptions and justifications: Ignoring the sudden physiological factors such as the physical discomfort of the athletes. Neglecting the influence of the field, weather and the atmosphere on the site performance. Ignoring the impacts of hard factors such as equipment on the forecast results. Athletes' performance can be evaluated and predicted by the model.

What’s more, the primary notations used in this paper are listed in Table 1.

Table 1. Notations

Symbol	Description
X_i	The value of the raw data
mean(x)	mean of the variable
sd(x)	the standard deviation of the variable
R_j	Index conflict
ρ	Spearman correlation coefficient
C_j	amount of information
W_j	weight coefficient
S_j	the standard deviation of the JTH index
k	k dimensionality
X	matrix

2. Model Establishment and Solution

2.1. CRITIC Method

CRITIC Method is an objective weight assignment method, whose basic idea is to determine the objective weight of the index.[1] It is based on two basic concepts. The one is to contrast the intensity. It represents the size of the value gap of each evaluation scheme of the same index in the form of standard deviation. The other is to evaluate the conflict between the indicators. The conflict between indicators is based on the correlation between indicators. For example, if there is a strong positive correlation between the two indicators, then we can say that the conflict between the two indicators is low [2].

The steps are as follows:

First is the indicator variability:

$$\begin{cases} \overline{X_j} = \dfrac{1}{n}\sum_{i=1}^{n} X_{ij} \\ S_j = \sqrt{\dfrac{\sum_{i=1}^{n}(X_{ij}-\overline{X_j})^2}{n-1}} \end{cases} \tag{1}$$

Second is the indicator conflict:

$$R_{\mathrm{j}} = \sum_{i=1}^{p}(1-r_{ij}) \tag{2}$$

Third is the amount of information:

$$C_{\mathrm{j}} = S_j\sum_{i=1}^{p}(1-r_{ij}) = S_j \times R_j \tag{3}$$

Last is the weight:

$$W_{\mathrm{j}} = \frac{C_j}{\sum_{j=1}^{p} C_j} \tag{4}$$

This method is suitable for judging the stability of data and analyzing data with certain correlation between indicators or factors.

In an effort to evaluate the correlation between momentum and the outcome, and to prove that the game swings and victories are not random, this study used the CRITIC method to calculate the weight of the players' each performance on the outcome. First, analyze the weight of each index according to the calculation results, and then obtain the weight analysis matrix through the weight calculation results. The calculation results are shown in Table 2:

Table 2. The weight calculation results

index	Index variability	Index conflict	Quantity of information	Weight (%)
unf_err	0.332	5.841	1.94	24.466
ace	0.213	6.118	1.301	16.411
break_pt_won	0.117	5.989	0.7	8.83
double_fault	0.128	5.708	0.733	9.237
net_pt_won	0.263	6.209	1.634	20.598
break_pt_missed	0.154	5.921	0.909	11.465
p1-p2	0.13	5.488	0.713	8.993

Specially, the index variability is the standard deviation, the larger the standard deviation, the greater the weight; the conflict is the correlation coefficient, the stronger the correlation between the indicators, the lower is the conflict and the weight; the information amount is the index variability * conflict index, and the weight is the normalization of the information. By comprehensively comparing the size of each data in the table, the importance of the indicators is further obtained, as shown in Figure 2:

2.2. Spearman Correlation Analysis

Spearman Correlation analysis is a method to calculate the non-linear relationship between two variables. At the same time, it does not assume the distribution of the data. The basic idea of this method is to rank the values of two (or multiple) variables and calculate the rank correlation (Spearman correlation coefficient) between them [3].

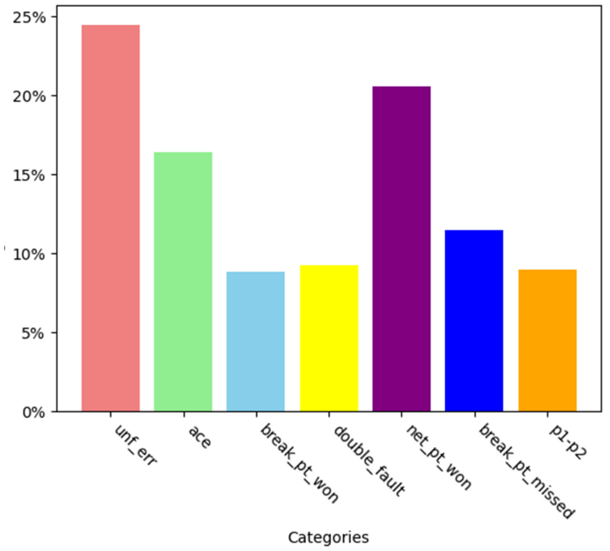


Figure 2. Weight Coefficient

$$\rho = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}} \tag{5}$$

Spearman correlation coefficient ranges from-1 to 1. A value of -1 indicates a complete negative correlation, a value of 1 indicates a complete positive correlation, and a value of 0 indicates no correlation between the two variables.

Therefore, to further explore the correlation between momentum and competition results, this study used Spearman Correlation Method to improve the accuracy and credibility of the results. First, test whether there is a statistically significant relationship ($P < 0.05$) between the momentum and competition results, and then analyze the positive and negative correlation coefficient and the degree of correlation. Finally, summarize the analysis results to obtain the correlation coefficient relationship as shown in Table 3:

Table 3. Table of correlation coefficients

	Q20-p1per	power4
Q20-p1per	1	0.49
power4	0.49	1

2.3. Thermal Map of the Correlation Coefficient

Draw the correlation coefficient thermal map for the data in figure 3:

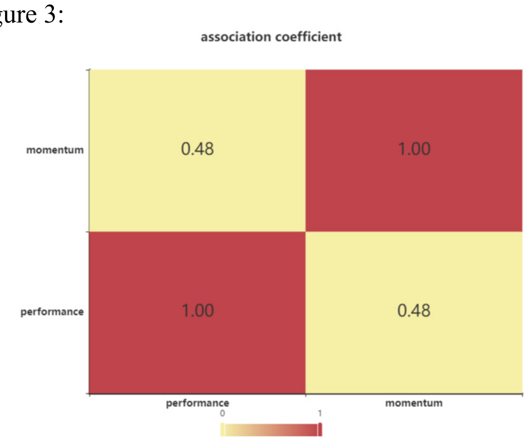


Figure 3. Orrelation Coefficient Thermal Map

As can be seen from the figure 3, there is a great correlation between the momentum and the results of the game, so it can be further explained that the trained data results have high confidence.

3. Model Establishment and Solution

3.1. CNN-LSTM Model

CNN and LSTM are two common used deep learning models that can be used for classification and prediction. CNN is a neural network model for image processing, while LSTM is a neural network model for sequential data processing.[4]

This paper used the combination of the two, which can not only improve the robustness and convergence of the model, but also reduce the prediction error and improve the prediction accuracy [5].

First, construct the data as a feature matrix containing the spatiotemporal correlation of the traffic flow:

$$X=\begin{bmatrix} x_1^1 & \cdots & x_{t-m}^1 \\ \vdots & \ddots & \vdots \\ x_1^i & \cdots & x_{t-m}^i \end{bmatrix} \tag{6}$$

Second, put he feature matrix X into the multilayer CNN and exxtract the spatial features of the traffic flow data to obtain the sequence feature vector. And put the feature vector into the multi-layer LSTM network to extract time-dependent features.

In order to prevent data overfitting and long training time, Droput layer is introduced to the training process of the model.

After the model is established, the training data set is divided, trained, and the test set is predicted, and the results are shown in the figure: The blue line represents the actual value and the orange line represents the predicted value. Comparison between the real value and the predicted value is shown in figure 4.

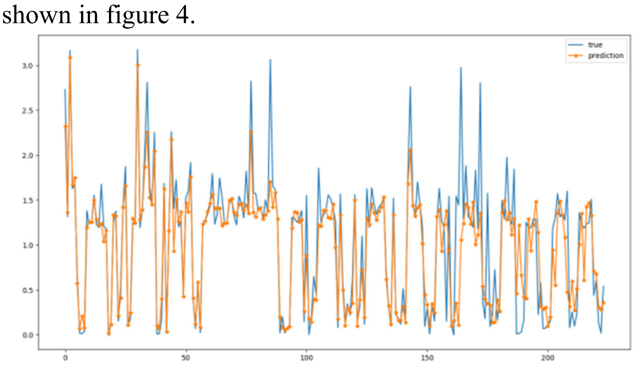
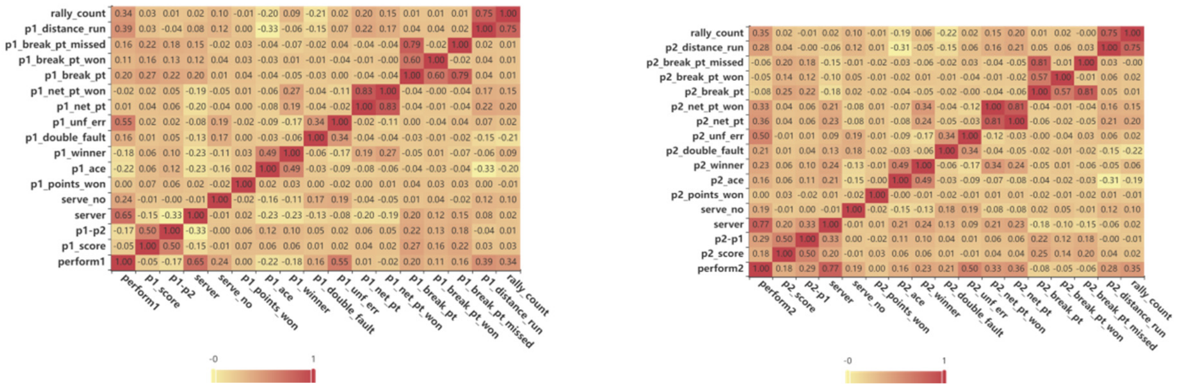


Figure 4. Comparison between the real value and the predicted value

By calculation, the RMSE of the predicted result is 0.3596. According to regulations, the closer the RMSE is to 0, the better the prediction effect is, so this model has higher accuracy. Therefore, it can be shown that the model can effectively predict the fluctuation in the game.

3.2. Spearman Correlation Analysis

The correlation among performance of the player, who is the server, unforced error, the running distance between each socre, and the number of swings is more prominent than that in other indicators. In this paper, it is believed that the running distance and the number of swings is reflections of the physical exertion situation of the players. And who serves means having greater advantages in the game, which will affect the performance of the players naturally. Non-forced errors mean losing points for nothing. The influence on players' mentality has also played an important role in previous studies of momentum. So this paper attributes "momentum" to an indicator of the impact of various situations on the pitch on the players' mentality. This indicator has a considerable effect on the performance of the players. Therefore, in order to determine which indicators are important in predicting the outcome of the competition, in this part, Spearman correlation analysis method is used again.[6] The heat map of the correlation coefficient is obtained as shown in Figure 5a and Figure 5b:



(a) (b)

Figure 5a and 5b. Thermal diagram of the correlation coefficient

4. Conclusion

In this paper, the data of each point of Wimbledon 2023 is cleaned to dealt with the missing values and outliers and unify the two different scoring methods.

Firstly, in order to reflect the players' performance from the game data intuitively, this paper establishes a Random Forest Classification Model. Through the training of the model, the weights of indicators which influence on players in the match are obtained to calculate the "performance coefficients" of the players. So that players' performances be evaluated visually.

Secondly, according to the relevant literature, indicators related to "momentum" are obtained the weight by the CRITIC Weight Method in order to quantitative 'momentum'. The Spearman coefficient between "momentum" and "performance coefficient" is 0.479 with significant correlation. This paper also analyzes the correlation between other indicators and "performance coefficient". The results indicate that successful first serve, unforced errors, running distance, number of shots and other indicators have great impact on the swings in matches.

Thirdly, in order to predict the swings in matches in view of the amount of indicators, this paper combines CNN and LSTM to form a prediction model. Model works well (RMSE<1) for matches data of Wimbledon,2023.

We have found that, the evaluation model we established in this paper can accurately explore the factors that have a great influence on the competition results. Furthermore, the prediction model can predict the competition results well.

In a word, the models we established provided us with a platform to quantify the performance indicators and momentum indicators.

References

[1] Zhang Yu, Wei Huabo. Multiple attribute decision making based on CRITIC combination empowerment method [J]. Journal of statistics and decision, 2012, (16): 75-77. The DOI: 10.13546 / j. carol carroll nki tjyjc. 2012.16.009.

[2] Liu W, Liu A, Qin H, et al. Application of Hybrid Multi-criteria Decision-making Approach to Analyze Wastewater Microalgae Culture Systems for Bioenergy Production. [J]. Environmental research,2024,256:119234-119234.

[3] Li Chao, Yang Hui, Bao Bowen, et al. Spearman Correlation Coefficient Abnormal Behavior Monitoring Technology Based on RNN in 5G Network for Smart City[A].16th International Wireless Communications and Mobile Computing Conference, 2020: 1440-1442.

[4] Li Mei, Ning Dejun, Guo Jiacheng. CNN-LSTM model based on Attention mechanism and its Application [J]. Computer Engineering and Applications,2019,55(13):20-27.

[5] Zhu X, Chen G ,Ni C , et al.Hybrid CNN-LSTM model driven image segmentation and roughness prediction for tool condition assessment with heterogeneous data[J].Robotics and Computer-Integrated Manufacturing,2024(90):90102796-.

[6] Han Y, D. A H. A robust Spearman correlation coefficient permutation test [J].Communications in Statistics - Theory and Methods, 2024,53(6):2141-2153.