Unveiling the Deep Connection Between Athlete Momentum and Match Scoring through Statistics Analysis

Boya Zhang, Yifan He, Yiyu Shi*

College of Science, Minzu University of China, Beijing, China, 100081

* Corresponding Author Email: syy13469014537@126.com

Abstract. The concept of momentum in ball sports is characterized by significant changes in-game performance and mental states, which play a key role in determining the outcome of a match. This study takes the 2023 Wimbledon men's final as an example and constructs a model for assessing athletes' performance during tennis matches based on binary logistic regression. It also constructs a tennis match momentum assessment algorithm to explore the relationship between the momentum change among match players and the performance change (score difference) of both players. It was found that the momentum change of match players was closely related to the score difference. This study is not only conducive to improving the athletes' competitive level but also has a profound impact on the scientific and professional development of tennis.

Keywords: Momentum, Athlete Performance Assessment, Binary logistic regression, Momentum state transfer algorithm.

1. Introduction

In competitive sports, the concept of momentum is often cited as a key determinant of performance outcomes. This elusive phenomenon is frequently discussed by athletes, coaches, and analysts. Momentum means that a player benefits from a psychological and/or physiological boost[1]. For tennis, there is a strong link between male athletes' mastery of potential energy and their ability to score points[2].

Multivariate logistic regression is widely used in sports technology and athlete-level assessment research and its special form is binary logistic regression[3-5]. Existing studies on assessing the performance and momentum of tennis players mainly focus on evaluating players' performance with factor analysis and fuzzy comprehensive analysis, but few studies have quantitatively analyzed the relationship between players' performance trends and momentum trends [6-7].

Based on binary logistic regression, this paper constructs a model for assessing players' performance during a tennis match from the perspective of the time when the match takes place. It also constructs a momentum assessment algorithm for tennis matches to investigate the relationship between players' performance and momentum during the matches.

2. Construction of the Athlete Performance Assessment Model

2.1. Variable Selection and Data Description

2.1.1 Variable Selection

The data for the 2023 Wimbledon men's singles tournament were sourced from the official Wimbledon tennis tournament website. Variables reflecting factors pertinent to player performance and scoring occurrences were chosen based on the research findings of Miguel Crespo and Rafael Martinez-Gallego[8] et al. The relevant variables are summarized in Table 1.

Table 1. Description of relevant variables

Variant	Hint	Variant	Hint
y	Score	x_7	Fail to break the bonds
x_1	First serve situation	x_8	Average distance traveled
x_2	Hitting untouchable balls	x_9	Serve speed of the golfer
x_3	Use standard posture when scoring	x_{10}	Change of serve angle
x_4	Own fault	x_{11}	Depth of serve
x_5	Score at the net	x_{12}	Returner Return Depth
x_6	Successful breakout	/	/

Two notes are made regarding the data: firstly, the variables' values need to be segregated between serve and return players; secondly, outliers such as data errors and records that defy conventional logic due to statistical errors have been addressed to ensure the accuracy of the remaining data.

2.1.2 Descriptive Statistical Analysis of Players' Performance

Based on the data presented in Table 2 for all players in various indicators, it can be analyzed that most of the players are proficient in making effective use of scoring opportunities at the net and minimizing individual errors. However, their ability to capitalize on break points was noticeably lacking, and the likelihood of serving or returning quality shots was relatively low.

Table 2. Descriptive statistical analysis of match data for all players

Variable	Average	Standard Deviation	Variance	Skewness	Kurtosis	Frequency(=1)
y	0.5	0.5	0.25	0	-2	0.5
x_1	0.18	0.384	0.148	1.664	0.769	0.18
x_2	0.168	0.374	0.14	1.779	1.164	0.168
x_3	0.114	0.318	0.101	2.432	3.915	0.114
x_4	0.128	0.334	0.111	2.23	2.975	0.128
x_5	0.082	0.274	0.075	3.046	7.281	0.082
x_6	0.012	0.107	0.011	9.171	82.117	0.012
x_7	0.022	0.147	0.022	6.49	40.125	0.022
x_8	0.054	0.073	0.005	2.95	18.721	/
x_9	0.374	0.403	0.162	0.194	-1.881	/
x_{10}	0.278	0.448	0.201	0.993	-1.014	0.278
x_{11}	0.161	0.367	0.135	1.847	1.412	0.161
x_{12}	0.159	0.366	0.134	1.861	1.464	0.159

Individually, the performance of Alejandro Davidovich Fokina (player1) and Holger Rune (player2) in "2023-wimbledon-1304" is analyzed. Examining the results of 3-6, 6-4, 6-3, 4-6, 6-7[8-10] and scrutinizing the scorecards reveals a consistent presence of deuce in nearly every round. This suggests that player1 and player2 exhibited comparable performance and strength.

Drawing insights from the data table and radar chart, it can be inferred that player1 should concentrate on refining skills and techniques. On the other hand, player2 should prioritize strengthening basic training, maintaining a stable mindset, and minimizing personal errors.

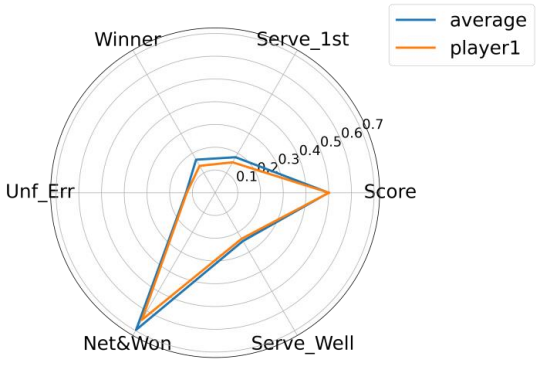


Figure 1. player1 Radar chart of performance

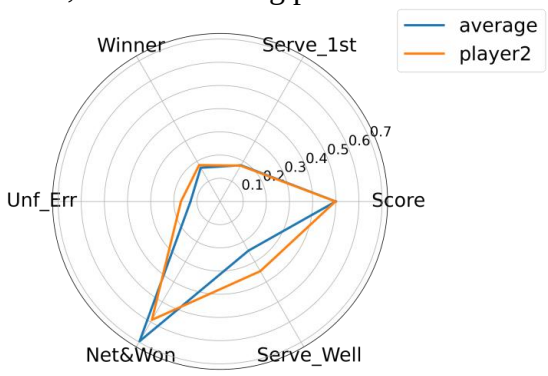


Figure 2. player2 Radar chart of performance

2.2. Empirical Studies

2.2.1 Model Derivation

Based on the extracted factors influencing athletes' scores, a binary logistic regression model is formulated to analyze athletes' performance. The model is represented as:

$$\ln\left(\frac{p(y=1)}{1-p(y=1)}\right)=\alpha+\sum_{i=1}^{k}\beta_i x_i \tag{1}$$

Where: y denotes the athlete's score, x_i represents the ith explanatory variable influencing the athlete's score, k is the number of explanatory variables, β_i signifies the partial regression coefficient of the explanatory variables, indicating their correlation with whether the athlete scores or not, and α is the intercept term.

Subsequently, the probability of the athlete scoring in each round can be expressed as:

$$p(y=1\,|x)=\frac{\exp\left(\alpha+\sum_{i=1}^{k}\beta_i x_i\right)}{1+\exp\left(\alpha+\sum_{i=1}^{k}\beta_i x_i\right)} \tag{2}$$

2.2.2 Model Building

Upon computing the partial regression coefficients β_i of the model, the indicators influencing the athlete's score at moment t are incorporated into the model, yielding the athlete's value at moment t, denoted as Z_t. This is expressed as:

$$Z_t=\alpha+\sum_{i=1}^{k}\beta_i x_i \tag{3}$$

After performing a log transformation on Z_t and conducting additional calculations to obtain the value of $p(y = 1|x)$, this computation is further translated into the performance score formula, denoted as Q_t. This formula represents the athlete's performance at moment t by scoring Q_t. Subsequently, the discrepancies in performance among different athletes at moment t are delineated through Q_t, reflecting variations in performance situations. This process of quantification embodies the essence of capturing performance differentials across athletes at a given moment.

2.2.3 Multicollinearity Analysis

To ensure the accuracy of the binary logistic regression analysis, a multicollinearity analysis of the independent variables was conducted. Following testing, the VIF values for each independent variable ranged from 1.025 to 3.398, all below the threshold of 10. This suggests that multicollinearity was not a significant issue among the independent variables and confirms their substantial impact on athletes' scores and performances at a 95% confidence level.

2.3. Results Of The Experiment

The coefficients of the established model have been computed to derive the final performance scenario formula:

$$\begin{aligned}Q_t=&\,0.122x_1+4.66x_2+1.866x_3-5.922x_4+4.429x_5+4.238x_6\\&-3.412x_7+0.965x_8+1.942x_9+0.252x_{10}+0.3x_{11}+0.251x_{12}\end{aligned} \tag{4}$$

Upon analyzing these coefficients, it's observed that, except for factors related to losing points due to personal errors and missing breakpoints, all other variables positively influence athletes' scoring performance, aligning with real-world observations. Notably, the magnitudes of the coefficients

suggest that x_2, x_4 through x_7 are pivotal factors impacting athletes' scoring outcomes. Furthermore, it indicates that being on the serving side in a tennis match enhances the likelihood of winning.

By the established model, the performance data of both Player1 and Player2 in a match have been graphically represented. To enhance the clarity of performance score fluctuations, the data has been magnified by a factor of 100. The presented results are shown in Figures 3 and 4:

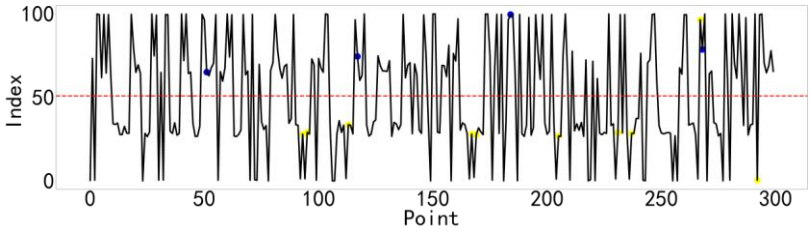


Figure 3. Trend of the game Player 1

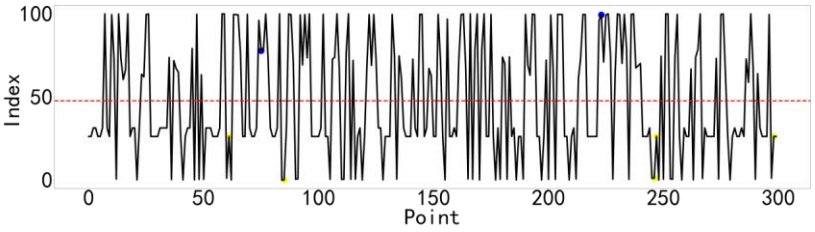


Figure 4. Trend of the game Player 2

In Figures 3 and 4, the blue points denote winning moments at critical junctures, while the yellow points represent losses at such pivotal instances. The red line signifies the average performance level. Observing the graph, it's evident that both players' performances fluctuated throughout the match. Notably, during critical moments, there's a heightened likelihood for players to capitalize on opportunities, potentially leading to victory. Furthermore, the points garnered during these critical phases surpass the average value, whereas points lost tend to fall below it. This underscores the model's precision and rationality in evaluating player performance.

2.4. Model Verification

Upon scrutinizing the model's training accuracy, the training error stands at 0.19. Additionally, P-R and ROC curves were generated.

The ROC curve exhibits an area under the curve (AUC) of approximately 0.886, surpassing the 0.5 threshold, indicative of superior generalization performance. Concurrently, the P-R curve showcases an overall favorable model performance, with a total correct rate reaching 0.81. These results affirm the model's validity and accuracy.

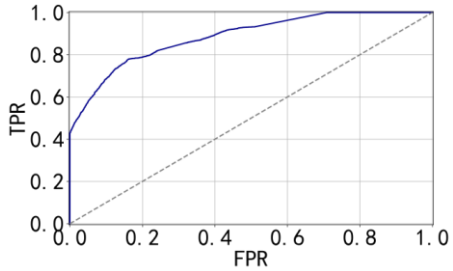


Figure 5. ROC curve

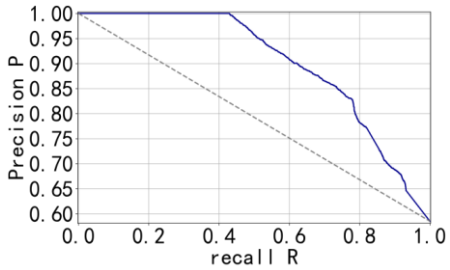


Figure 6. P-R curve

3. Momentum's Impact on Tennis Scoring

3.1. The process of calculating M_t

This paper defines a formula for the calculation of momentum based on the factors affecting momentum on the court and the rules of the tennis game called “Momentum state transfer algorithm”:

$$M_t = M_{t-1} + \alpha_t * \beta_t * N_t + Q_t \quad (5)$$

Where M_t denotes the "momentum" of an athlete at time t, and similarly M_{t-1} denotes the "momentum" of an athlete at time t-1. And αt denotes the characteristic variable of whether the athlete scores or not at moment t (i.e., scoring at moment t is assigned a value of 2 and not scoring is assigned a value of -2); β_t denotes the serving weight at moment t (1.5 for the serving side and 1 for the receiving side). Q_t is the performance score at moment t, and the formula is shown in Equation (4). N_t denotes the weight of consecutive wins added to the momentum score at moment t (base score is 1, each additional win adds 0.2 to the base score)

3.2. Mt of different players

Data from the men's final of the 2023 Wimbledon Tennis Championships were selected to make a graph of the momentum scores of the opposing teams, using the number of balls scored (point) as the horizontal coordinate and the momentum of the players as the vertical coordinate.

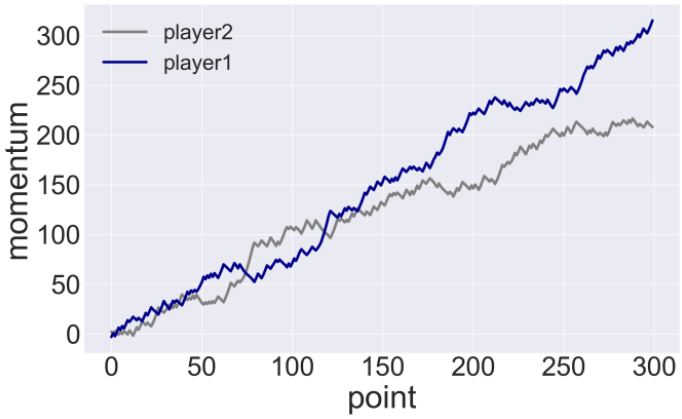


Figure 7. Momentum change

When analysed in the context of the match, the first set was more intense at the 30th to 40th scoring points (in the graph, the two curves are entwined at the 30th to 40th scoring points), and in the end, the set was won by player 1, Carlos Alcaraz (the momentum of player 1 was higher than that of player 2 at the end of the first set in Figure 7), which is the same as the direction of the lines under the corresponding scoring points in Figure 7.

3.3. Exploring the relationship between momentum and changes in scoring

In order to investigate whether the so-called "momentum" is related to the change of the score on the field (situation), the data of the difference in momentum between two players in the same game and the cumulative score difference between the two players will be visualised, as shown in Figure8 and Figure9: (PS: the difference is player1-player2)

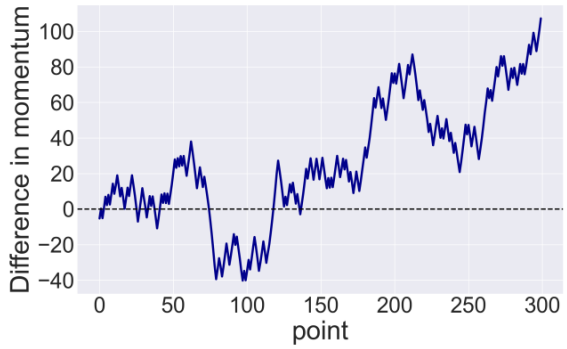


Figure 8. Difference in momentum

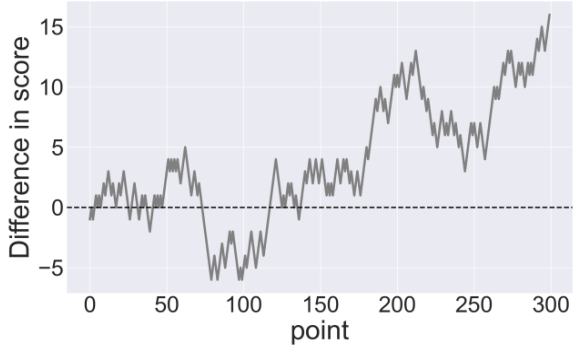


Figure 9. Difference in score

As can be seen in the above graph, the two curves have nearly identical shapes. Since the first athlete Carlos Alcaraz won the tie-break in this match with 3:1, and the opponent won in the second set (i.e., 80th~110th points), and both graphs are negative in this interval, which is in line with the actual situation, then it proves that the momentum model is established reasonably, and the process of the match "momentum " has a strong correlation with the change of on-court scoring (situation).

This paper further explores the relationship between the momentum difference data and the cumulative score difference data through the k-s test as well as the Pearson correlation test. The former was standardized and tested to obtain p-value = 0.4218, which means that the original hypothesis was rejected at the 99.9% confidence level that the two sets of data are not significantly different and may come from the same distribution; the latter test obtained a Pearson correlation coefficient of about 0.998 for the two sets of data, which proves that the two sets of data are strongly correlated.

In summary, it can be concluded that there is a strong positive correlation between the momentum of the players in the match and the score situation of the match.

4. Conclusions

This paper establishes the assessment model and momentum algorithm of athlete performance. First of all, the relevant indexes of athletes' performance are visualized, and the assessment model of athletes' performance is constructed based on binary logistic regression, which leads to the calculation formula of the performance situation. By analyzing the formula, it is found that except for the factors of losing points to their own mistakes and missing breaking points, all other factors have a positive influence on the occurrence and performance of athletes' scoring points. The data from one game was selected to visualize the performance scores of both opponents, and it was found that the modeling was more accurate and effective. For the athlete momentum algorithm, data from the men's final of the 2023 Wimbledon Tennis Championships was used to visualize and compare the difference between the opposing sides' momentum scores and points scored, respectively, and it was found that the two profiles were approximately the same. Then, Pearson correlation analysis was done on the two sets of data, standardized by the k-s test. The Pearson correlation coefficient was about 0.998, and the k-s test rejected the original hypothesis, concluding that the player momentum during the match had a strong positive correlation with the match score situation.

This study reveals algorithms for assessing athletes' whole game performance and momentum in the field of sports science and psychology, which contribute to solving the problems related to the improvement of athletes' game strategy assessment and prove the feasibility of the methodology.

References

[1] Dietl, H.M. and C. Nesseler, Momentum in tennis: Controlling the match. INTERNATIONAL JOURNAL OF SPORT PSYCHOLOGY, 2017. 48(4): p. 459-471.

[2] Gauriot, R. and L. Page, Does Success Breed Success? A Quasi-Experiment on Strategic Momentum in Dynamic Contests. The Economic Journal (London), 2019. 129(624): p. 3107-3136.

[3] Zou, X., Comprehensive power evaluation and regression model prediction analysis of world women's hardcourt tennis singles. Bulletin of Sports Science and Technology Literature, 2022. 30(01): pp. 67-70.

[4] Ting Jiang and Qing Li, A study on the factors influencing the winners and losers of men's singles matches in professional tennis - based on the four Grand Slam tournaments from 2014-2018. China Sports Science and Technology, 2021. 57(07): pp. 62-68.

[5] Ma Jiacai et al, Modeling and research analysis of match-winning factors of professional tennis players. Bulletin of Sports Science and Technology Literature, 2023. 31(04): pp. 66-68+118.

[6] Den Hartigh, R.J.R., J.K. Van der Sluis and F.T.J.M. Zaal, Perceiving affordances in sports through a momentum lens. Human Movement Science, 2018. 62: p. 124-133.

[7] Meier, P., et al., Separating psychological momentum from strategic momentum: Evidence from men's professional tennis. Journal of Economic Psychology, 2020. 78: p. 102269.

[8] Crespo, M., & Martinez-Gallego, R. (2023). Editorial: Tennis: Testing and performance. Frontiers in Sports and Active Living, 5. Retrieved from https://www.frontiersin.org/articles/10.3389/fspor.2023.1190917

[9] Liu, H., J. Lu and A. Salvo, Willingness to fight on: Environmental quality in dynamic contests. RAND JOURNAL OF ECONOMICS, 2023. 54(2): p. 189-239.

[10] Uzomba, G.C., et al., The Regression Modelling of Kinanthropometric and Kinematic Variables in Relation to Ball Velocity of Nigerian Female Tennis Players. Annals of Applied Sport Science, 2020. 8(1): p. 1-8.