Transactions on Computer Science and Intelligent Systems Research
ISSN: 2960-1800, eISSN: 2960-2238 | Volume 5
AIDML 2024

Research on Momentum Evaluation and Turning Point Prediction in Tennis Matches Based on RSR and GSRF

Bowei Dong, Kaizhi Dong *

School of Computer Science and Technology, Ocean University of China, Qingdao, China, 266100

* Corresponding author: kaifengdong10086@gmail.com

#These authors contributed equally.

Abstract. In the men's singles final of the 2023 Wimbledon Open, rising star Carlos Alcaraz defeated veteran Novak Djokovic, an astonishing turnaround and result that sparked an in-depth exploration of the sports phenomenon of "momentum". When one side of a match gains momentum, its technical level and adaptability will be greatly utilized, which will directly affect the victory or defeat of the match. Most of the past researches focus on verifying the authenticity of momentum in sports games and the factors affecting the change of momentum. In this paper, taking the above match as the research object, we use the rank and ratio comprehensive evaluation method (RSR) to mine the match data, and propose a new prediction model GSRF by combining the random forest model and the lattice search algorithm, and based on the independently established momentum model and the momentum difference model, we predicted the turning point of the momentum in the match and the result of the match. The difference effect size of Cohen's d-value of 0.035 and accuracy R^2 of 0.826 was obtained, which verified the higher accuracy and generalizability of the model. The study helps coaches and instructors optimize training programs, develop game strategies, and provide new ideas and approaches for coaches and instructors.

Keywords: Momentum assessment; RSR; SHAP; GSRF; Turning Point Forecast.

1. Introduction

Momentum can be observed in sports such as tennis, rugby, table tennis, and basketball, and several past studies have focused on the authenticity of momentum [1], the factors that influence changes in momentum [2-3], the evaluation of players through momentum [4], and the causal relationship between momentum and the outcome of the game [5].

For example, Den Hartigh and van der Sluis et al. found that scoring patterns during a match have an impact on a player's PM by examining the development of psychological momentum (PM) in table tennis [6]. Dietl et al. found that by studying the psychological and behavioral states of sports teams during positive and negative momentum that when the team has negative momentum, subjects experience negative psychological changes and the team puts in a rapidly diminishing amount of effort [7], whereas during positive team momentum, the order in which the team exerts effort is more variable and adaptive. On this basis, Den Hartigh et al. also found that negative changes in teams were stronger than positive changes in teams with positive momentum, i.e., there were asymmetries [8] and gender differences [9]. It has also been shown that athletes show significant individual variability [10] in the face of momentum changes due to differences in psychological quality [11] and physical condition. In addition to this, Yue et al. used Exploratory Data Analysis (EDA) to explore variables related to match outcomes and proposed the Glicko tennis match outcome prediction model with an accuracy of about 71% [12]. Zeitvogel et al. defined a stress index and created a dynamic in-match prediction model for match-winning percentage [13]. Šarčević et al. proposed a mathematical model based on conditional probability and empirical Bayesian estimation and finally combined them through a unified hybrid approach based on Monte-Carlo simulation, with an average of 25% smaller simulation errors [14].

In past studies, although momentum has been adequately characterized qualitatively, there is a lack of quantitative research on momentum during matches, which does not fully exploit the value of momentum in sporting events. We have analyzed some data for all the matches in the men's singles final of the 2023 Wimbledon Tennis Championships in detail and established a momentum assessment model and a momentum turning point prediction model. This is to more fully explore the value of momentum in shaping athletes' abilities [2-3], orienting the trend of the match, deciding the winner, and formulating tactical strategies in the matches [15].

2. Model Establishment

2.1. Establishment of a Momentum Assessment Model

The article begins by using the Rank and Ratio Synthesis (RSR) method of evaluation to file and categorize all data labels recorded during the course of the race according to the degree of their impact on the outcome.We also reviewed relevant literature to conduct a comprehensive evaluation of feature importance. After determining the final data labels, the labels were ranked according to the feature importance, and the input parameters of the momentum model were set in this order to establish the momentum assessment model, and then the accuracy of the momentum assessment model was verified using the paired-sample Wilcoxon signed rank test.

Rank-sum ratio (RSR) refers to the average or weighted average of the rank sum of rows (or columns) in the table, which is a non-parametric measure of the composite index with the characteristics of continuous variables in the interval of 0~1. RSR is a non-parametric statistical analysis, with no special requirements for the selection of indicators, it applies to a variety of evaluation objects, and it can eliminate outliers because the values used for the calculation are ranked. Interference. In addition, it also integrates the method of parametric analysis, and the result is more accurate than the non-parametric method alone, it can be used in a wide range of applications, both in direct ordering and in graded ordering.

Ranking of indicator values.

For positive indicators:

$$R_{ij} = 1 + (n-1)\frac{X_{ij} - min(X_{1j}, X_{nj}, \cdots, X_{nj})}{max(X_{1j}, X_{nj}, \cdots, X_{nj}) - min(X_{1j}, X_{nj}, \cdots, X_{nj})} \tag{1}$$

For negative indicators:

$$R_{ij} = 1 + (n-1)\frac{max(X_{1j}, X_{nj}, \cdots, X_{nj}) - X_{ij}}{max(X_{1j}, X_{nj}, \cdots, X_{nj}) - min(X_{1j}, X_{nj}, \cdots, X_{nj})} \tag{2}$$

Calculate the rank sum ratio RSR value.

When the weights are equal:

$$RSR_i = \frac{1}{n \times m}\sum_{j=1}^{m} R_{ij} \tag{3}$$

When the weights are different:

$$WRSR_i = \frac{1}{n}\sum_{j=1}^{m} W_j R_{ij} \tag{4}$$

Where $\mathrm{W_j}$ is the weight of the j-th evaluation index, $\sum_{j=1}^{m} \mathrm{W}_j = 1$.

Use the entropy weight method to get the weight calculation results, according to each specific evaluation index according to the size of its index value to rank, to get the rank R, with the rank R to replace the original evaluation index value, according to the results of the ranking to establish the rank data matrix of each index.

Calculate the RSR value and RSR value ranking, list the distribution table of RSR, and get the Probit value.

Using the probit value (the probability unit corresponding to the cumulative frequency) as the independent variable and the RSR value as the dependent variable, calculate the linear regression equation and fit the corresponding RSR estimate.

$$\mathrm{RSR(WRSR)} = \mathrm{a} + \mathrm{b} \times \mathrm{Probit} \tag{5}$$

Sorted based on the fitted RSR values and graded hierarchically.

Table 1. Momentum Parameter Table

Variables	Sign	Player ω gains points
Player ω is the serving side	a	+10
Player ω is serving the second serve	b	-1
α = (Player ω's winning score - Player ν's winning score)	c	+(α/100)
Player ω serves an unreachable winning serve	d	+2
Player ν serves an unreachable winning serve and Player ω wins	e	+5
Player ω hits an unreachable winning shot	f	+2
Player ν hits an unreachable winning shot and Player ω wins	g	+5
The unreachable shot is a backhand	h	+3
Player ω loses both serve opportunities and loses the point	i	-2
Player ω commits an unforced error	j	-1
Player ω approaches the net	k	+1
Player ω wins the point at the net	l	+3
Player ω wins the opponent's serve	m	+5
Player ω misses the opportunity to win the opponent's serve	n	0
β = Player ω's running distance in the previous game	o	-（β/100）
Player ω scores consecutively twice	p	+1
Player ω scores consecutively three or more times	q	+3

Where ν is the opponent of ω in the match.

After obtaining the graded sorted data labels, we directly define momentum on this basis as a sports game process, a team or a player in the game to occupy the upper points, full of confidence in the victory, high morale, or a series of successive successful performance, mastered the rhythm of the game and has a sustained impact on the subsequent game, we call the size of this momentum for the P-value, we set the initial P-value as 10, and then the P-value as the weight of the data labels on the outcome of the game. To visualize more intuitive, we set the initial value of the P-value as 10, according to the data labels that have an impact on the outcome of the game the weight size of the input data of the momentum evaluation model to adjust the parameters, to get the momentum evaluation model shown in Table 1.

$$\mathrm{P} = 10\mathrm{a} - \mathrm{b} + \left(\frac{\alpha\mathrm{c}}{100}\right) + 2\mathrm{d} + 5\mathrm{e} + 2\mathrm{f} + 5\mathrm{g} + 3\mathrm{h} - 2\mathrm{i} - \mathrm{j} + \mathrm{k} + 3\mathrm{l} + 5\mathrm{m} - \left(\frac{\beta\mathrm{o}}{100}\right) + \mathrm{p} + 3\mathrm{q} \tag{6}$$

2.2. Definition of Momentum Turning Point

Turning points are categorized in this model as turning points where player1's momentum decreases and turning points where player2's momentum decreases by extracting the difference in momentum between player1 and player2, which is defined as a momentum-decreasing turning point when it exceeds a predetermined threshold. The model accurately captures this transition through absolute and relative momentum changes. The specific implementation is as follows:

$$\text{momentum\_change}(t)\ is \begin{cases} \text{True}, & (|\text{m\_diff}(t)| > \text{threshold}) \\ \text{False}, & \text{otherwise} \end{cases} \tag{7}$$

Where m_diff(t) is the difference in momentum between Player 1 and Player 2 at a particular time point at time t, the threshold is a pre-defined threshold. If the absolute change in momentum difference between consecutive time steps exceeds this threshold, it is considered a significant change and is noted as a turning point.

2.3. Establishment of the SHAP Model

SHAP (Shapley Additive exPlanations) is an algorithm for interpreting the predictions of a machine learning model that interprets the output of a particular prediction based on the Shapley value in cooperative game theory, which takes into account the contribution of each feature to this prediction.

Three characteristics of the SHAP algorithm are: first, it has excellent global interpretability, which effectively reveals the average impact of each feature on the overall model output. Its strong local interpretability allows it to drill down to each prediction to clarify the specific impact of each feature on the model output. Second, SHAP values satisfy the consistency property of Shapley values, ensuring that the sum of each feature's contribution to the overall model is equal to the difference between the model output and the benchmark. Third, SHAP not only provides an explanation of the positive impact of features, but also explicitly states the negative impact on model output, making its explanation more comprehensive.

Analysis using the SHAP algorithm helps coaches and players understand the role of key factors in the game to develop more effective game strategies.

$$\varphi_{\mathrm{i}}(f) = \sum_{S\subseteq N\setminus\{i\}} \frac{|S|!(|N|-|S|-1)!}{|N|!}[f(S\cup\{i\}) - f(S)] \tag{8}$$

Where $\phi i(f)$ is the Shapley value of the selected feature; f is the prediction of the change in player momentum that is not explicitly stated but exists in the model; N is the total number of features used in the model; S is the set of features that does not include the ith feature in each iteration; $|S|$is the number of features in the subset S; $\mathrm{f(S\cup\{i\})}$ is the output of the model when including the ith feature in the subset S; $\mathrm{f(S)}$ is the output of the model when excluding the ith feature from the subset S; and $\mathrm{f(S)}$ is the output of the model when excluding the ith feature from the subset S. S; $\mathrm{f(S\cup\{i\})}$ is the output of the model when the ith feature is included in the subset S; and $\mathrm{f(S)}$ is the output of the model when the ith feature is excluded from the subset S.

3. Results

3.1. Analysis of the Momentum Evaluation Model

The RSR modeling method is a non-integer rank method, the number of grades is 3, and the variable weights are the entropy weight method. Choose at random one of the two opponents of a match as player p1, and the other as player p2, and compare the number of points won by p1 in a match, whether p1 hit a winning serve, whether p1 hit a winning shot, whether p1 missed two serves and lost a point, whether p1 made an unforced error, whether p1 reached the net, and whether p1 won the point at the net, Whether p1 won the p2 serve, whether p1 had a chance to win the p2 serve, whether p1 missed a chance to win the p2 serve, and p1's running distance as a positive indicator, and similarly each of p2's data as a negative indicator, with the winner of that matchup as the index value.

In the 2023 Wimbledon Championships Men's Singles Final match between Carlos Alcaraz and Novak Djokovic, for example, the momentum assessment model calculated the final P-value of the player Carlos Alcaraz to be 20.12233, while Novak Djokovic's final P-value was 20.09482, with Carlos Alcaraz leading by a very small margin. Novak Djokovic for the victory. The real-time p-values of the two players were calculated based on the time series of the match and visualized and analyzed to make a line graph as shown in Figure 1.

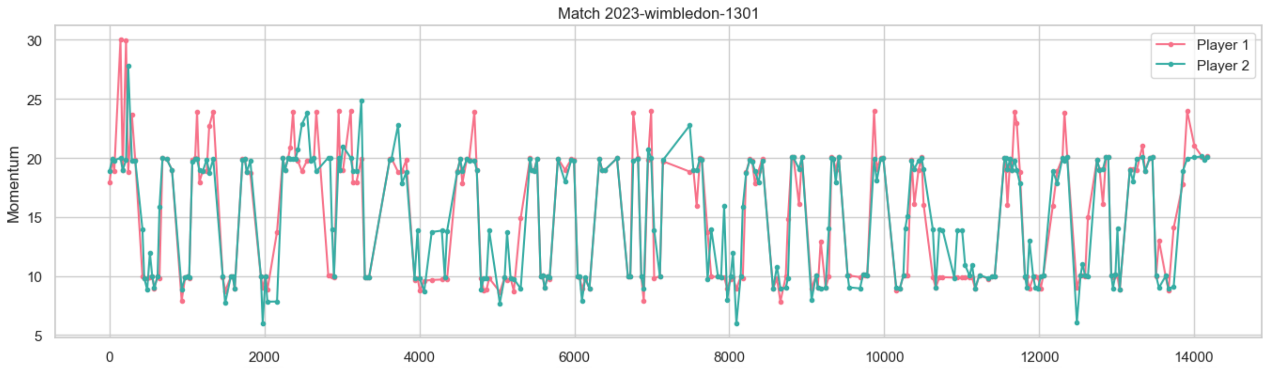

Figure 1. Comparison Chart of P-Values for this Competition

From Figure 1, it can be seen that the momentum of Player 1, i.e. Carlos Alcaraz, had an advantage over Player 2, i.e. Novak Djokovic, throughout the match. Carlos Alcaraz has 15 momentum maxima in the whole match, while Novak Djokovic has only 5 momentum maxima, and the maxima corresponding to the maxima are lower than the maxima corresponding to the maxima of Carlos Alcaraz.

Subsequently, we use the Shapiro-Wilk test to get the significance P-value of 0.000001, which presents significance at the level and rejects the original hypothesis, therefore the data does not satisfy the normal distribution and can be paired with samples Wilcoxon signed rank test. Based on the variable actual match results paired with the match results shown in the momentum model, a significance P-value of 0.162 was obtained, which does not present significance at the level, and the original hypothesis could not be rejected, so there is no significant difference between the actual match results paired with the match results shown in the momentum model. The magnitude of difference in Cohen's d value is 0.035, the magnitude of difference is very small. Therefore, it can be judged that momentum plays a role in determining the winner and has a very high accuracy.

3.2. Analysis of Momentum Turning Points

Take 2023-wimbledon-1701 as an example to analyze the relationship between the turning point and the winners and losers, in the figure we can easily find that the turning point has a close correlation with the winner of the game, the results are shown in Figure 2.

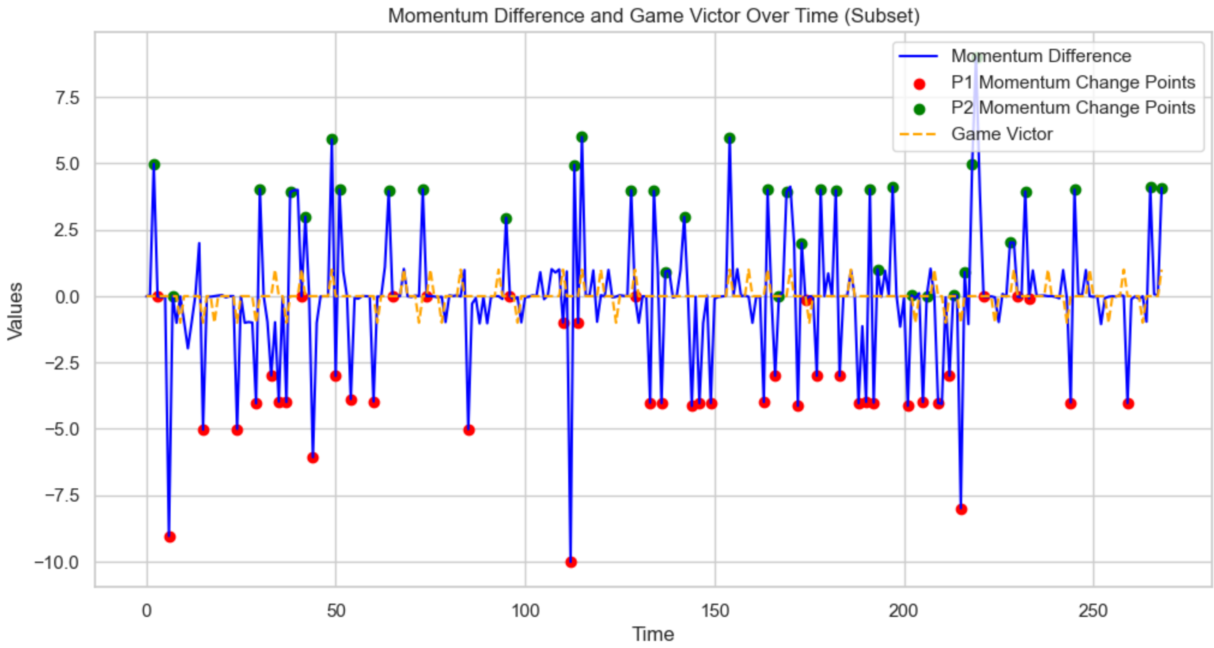

Figure 2. Relationship graph between the halfway turning point of Player 1 and the outcome of the match

From Figure 2, we can see that turning points are closely related to winning or losing this game. After SHAP analysis, we derive the impact of each feature on the turning point of the race and their specific contribution to different samples, as shown in Figure 3.

Analysing Figure 3, we can see that points won by the opponent at the net, and unforced errors by ourselves often lead to a turnaround in the match, where our momentum is weaker than the opponent's, which in turn affects the win or loss of the match.

3.3. Analysis and Recommendations for Turning Points

We used the influencing factors derived from the SHAP model analysis as input parameters for the GSRF model to predict the changes in the fluctuations of the turning point, after tuning the parameters by the particle swarm algorithm (PSO).

In this study, the accuracy of each model was verified using the cross-validation method, and the following table shows the results of each indicator within the training set, cross-validation set, and validation set after parameter optimization as shown in Table 2, with an R^2 of 0.826, which shows that the model fits well and meets expectations.

Table 2. Model Evaluation

	MSE	RMSE	MAE	R^2
Training set	0.006	0.004	0.039	0.959
Cross-validation set	0.058	0.081	0.074	0.826
Test set	0.048	0.041	0.051	0.873

4. Conclusion

We captured momentum and momentum difference in tennis matches and quantitatively analyzed them to make the overall direction of the match and the real-time status of the players more intuitive and to provide data support for coaches to further develop match strategies. We used Rank and Ratio Comprehensive Evaluation (RSR) to mine the data, analyze the weights, and establish a momentum assessment model to visualize and analyze the momentum direction of the game. We also proposed a GSRF turning point prediction model by combining the random forest model and the grid search algorithm to visualize and analyze the change of the momentum turning point during the game. From this study, it can be concluded that the winners and losers of matches in tennis are not random; opponents winning points at the net, players committing unforced errors, losing two service chances and losing points, running distances, the current point difference, the difference in sets, and the difference in the rankings of the two players all affect the momentum of the players. The turning points of the match are centered at the beginning and end of each set, and when there are 40-40 and AD-40 points in each set, which are crucial moments in the match and have a decisive impact on the outcome of the match. This study shows that our model is able to predict the variation of turning points in tennis matches better with an R^2 of 0.826, which indicates that the model has good rationality. This study can help coaches discover players' strengths and weaknesses and develop training and match strategies for players.

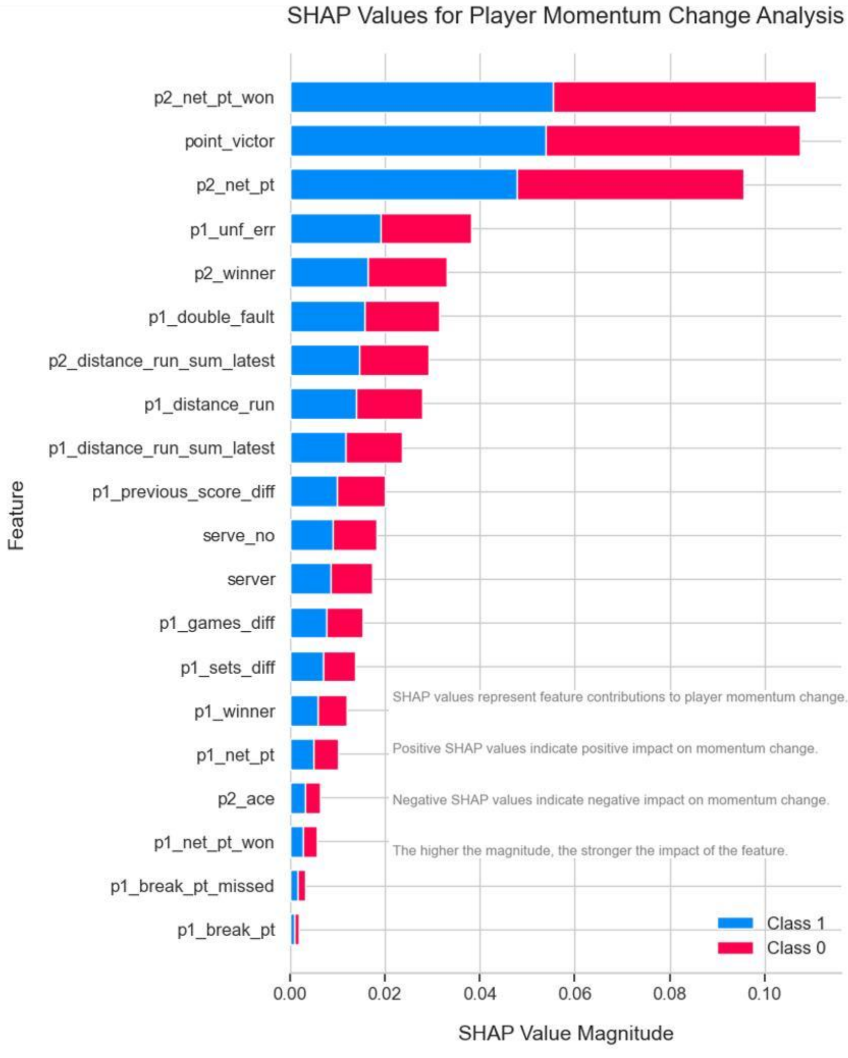

Figure 3. Feature contribution ranking

The factors during the actual game are more complex, and further research can consider other relevant factors and introduce more features and indicators to enrich the inputs and outputs of the model in order to improve the complexity and depth of our model. In the future, the model can also be combined with other models, such as the match strategy recommendation model.

References

[1] Gauriot R, Page L. Does success breed success? A quasi-experiment on strategic momentum in dynamic contests [J]. The Economic Journal, 2019, 129 (624): 3107 - 3136.

[2] Giles B, Peeling P, Reid M. Quantifying change of direction movement demands in professional tennis matchplay: An analysis from the Australian Open Grand Slam [J]. The Journal of Strength & Conditioning Research, 2022.

[3] Thornton H R, Armstrong C R, Gamble T, et al. Quantifying the movement characteristics of Australian Football League Women's competition [J]. The Journal of Strength & Conditioning Research, 2022, 36 (12): 3415 - 3421.

[4] Angelini G, Candila V, De Angelis L. Weighted Elo rating for tennis match predictions [J]. European Journal of Operational Research, 2022, 297 (1): 120 - 132.

[5] Meffert D, Breuer J, Ohlendorf L, et al. Towards an understanding of big points in tennis: Perspectives of coaches, professional players, and junior players [J]. Journal of Physical Education and Sport (JPES), 2021, 21 (2): 728 - 735.

[6] van der Sluis J, Huijgen B, Zaal F, et al. The Effect of Psychological Momentum on Affordances in Tennis [C]//FEPSAC 2022: Sport, exercise and performance psychology: challenges and opportunities in a changing world. 2022.

[7] Dietl H, Nesseler C. Momentum in tennis: Controlling the match[J]. UZH Business Working Paper Series, 2017 (365).

[8] Den Hartigh R J R, Gernigon C, Van Yperen N W, et al. How psychological and behavioral team states change during positive and negative momentum [J]. PloS one, 2014, 9 (5): e97887.

[9] Milekhina A, Breznik K, Restaino M. Gender comparison of in-match psychological traits of tennis players: dynamic network analysis [J]. Statistica Applicata-Italian Journal of Applied Statistics, 2023 (2).

[10] Richardson P A, Adler W, Hankes D. Game, set, match: Psychological momentum in tennis [J]. The Sport Psychologist, 1988, 2 (1): 69 - 76.

[11] Depken C A, Gandar J M, Shapiro D A. Set-level strategic and psychological momentum in best-of-five matches in professional tennis [J]. Applied Economics Letters, 2023, 30 (6): 735 - 739.

[12] Yue J C, Chou E P, Hsieh M H, et al. A study of forecasting tennis matches via the Glicko model [J]. Plos one, 2022, 17 (4): e0266838.

[13] Zeitvogel T, Lucas T. Analyzing and Predicting the Influence of In-Match Dynamics on Tennis Outcomes [J]. 2022.

[14] Šarčević A, Pintar D, Vranić M, et al. Modeling in-match sports dynamics using the evolving probability method [J]. Applied Sciences, 2021, 11 (10): 4429.

[15] Prieto-Lage I, Paramés-González A, Torres-Santos D, et al. Match analysis and probability of winning a point in elite men’s singles tennis [J]. Plos one, 2023, 18(9): e0286076.