Spearman Correlation and Entropy Weight Method Based Best-of-n Competition Momentum Prediction

Yayelin Mo [1, *], Yong Huang [2]

[1] School of Electronics and Information Engineering, Beijing Jiaotong University, Beijing, China, 100044

[2] School of Mechanical Engineering and Mechanics, Xiangtan University, Xiangtan, China, 411105

* Corresponding Author Email: 22725060@bjtu.edu.cn

Abstract. This paper builds a novel model for predicting momentum in competitive settings, particularly in Best-of-n competitions, using a feature extraction method. The study focuses on quantifying various momentum indicators using relevant match data to capture the dynamic nature of momentum shifts. Key indicators include fundamental momentum and psychological factors such as player confidence, opponent pressure, and unforced errors, as well as offensive momentum. These elements are integrated into a comprehensive assessment framework that allows for a holistic understanding of competitive dynamics. The model employs advanced statistical techniques and algorithms to enhance prediction accuracy, ensuring reliable outcomes. Through further evaluation, the approach is visually demonstrated in predicting momentum trends, making it applicable across diverse competitive environments. By leveraging these insights, stakeholders could make informed decisions, improve strategic planning, and enhance performance in various competitive arenas. Ultimately, it contributes to a deeper understanding of the intricacies of momentum in sports and other competitive fields.

Keywords: Spearman Correlation, Entropy Weight, Momentum Prediction Model.

1. Introduction

Statistical techniques have evolved significantly in recent years, finding applications across a range of fields, include sport analytics [1]. In tennis, a sport famous for its blend of elegance and technical complexity, data-driven insights are transforming traditional approaches to player performance and match outcome predictions [2]. Predictive models are largely based on historical data and betting odds, have achieved notable success [3]. But their accuracy remains limited, with typical prediction rates reaching only about 70% [4]. This limitation suggests that key information influencing match outcome is yet to be fully captured, especially concerning the psychological and momentum-driven aspects of competition.

Momentum, a phenomenon where a player's success in one phase of a match influences performance in subsequent phases, identified as a critical yet underutilized factor in match prediction [5]. Studies indicate that win a close first set, for instance, could psychologically impact both players, often give the winner an edge in the following set. While previous research has acknowledged the importance of momentum, few models incorporate it effectively due to its complex, non-linear nature [6]. Traditional linear approaches fall short in represents such dynamics, where the associations are always monotonic rather than strictly linear.

To address this gap, this paper proposes a momentum prediction model specifically designed for the "best-of-n" competition format common in tennis [7]. This model based on Spearman's correlation coefficient, which capture monotonic relationships in ranked data and is particularly effective for non-linear analysis of momentum trends [8]. Additionally, the Entropy Weighting Method (EWM) is applied to assign objective weights to each key feature, allowing the model to actively adjust the focus derived from performance indicators relevant to momentum [9]. By combining these two methods, this study develops a comprehensive approach on predicting momentum shifts and the impact on match outcomes.

This research contribution lies in the integration of Spearman's correlation and EWM within a best-of-n framework, offer a much more nuanced model that not only enhances prediction accuracy but also reflects the psychological and sequential flow of a tennis match. This integrated approach provides valuable insights for analysts, coaches, and players, advancing the understanding of momentum as a powerful predictor in competitive sports.

2. Methodologies

2.1. Spearman Correlation Coefficient

Considering that the model going to implemented involves complex data processing and analysis. Particularly, the selection of data revolves around the 2023 Wimbledon Men's Singles final, contain all of the data from 31 matches [10].

Spearman Correlation Coefficient is a non-parametric statistical measure, which is calculated by ranking the data for each variable and computing the correlation based on the rankings, it is robust while dataset doesn't follow a normal distribution or contains outliers, making it suit for measure the correlation in our cases.

The formula for the coefficient is as follows:

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2-1)}, \tag{1}$$

Where ρ represents the coefficient, d_i denotes the rank differences, and n be the number of data points.

2.2. Entropy Weighting Method

In competitive match, the information from all sides changes constantly, more than that, they are all scattered. More and more information will be derived if the degree of dispersion continues to increase, and the entropy weight method (EWM) is a significant information weight method. Entropy weighting provides a more responsive approach to model the "momentum", particularly in the case of complex datasets.

As an objective weighting method which work pleasingly allocate objective weights in different machine operation. Suppose event *X* represent the score of two players, *x* represents the possible scenarios in event *X*, *p(x)* is the possibility of *x*. Then define:

$$I(x) = -ln((p(x)), 0 \leq p(x) \leq 1 \tag{2}$$

If event *X* occurs when: x_1, x_2, ..., x_n, then define the entropy of event X is:

$$H(X) = \sum_{i=1}^{n} p(x_i) I(x_i) = -\sum_{i=1}^{n} p(x_i) \cdot ln(p(x_i)) \tag{3}$$

Then form the positive definite matrix:

$$X = \begin{bmatrix} x_{11} & \cdots & x_{1m} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nm} \end{bmatrix} \tag{4}$$

Suppose the normalized matrix *Z*, the *i* th indicator for the *j* th:

$$Z_{ij} = \frac{x_{ij}}{\sqrt{\sum_{i=1}^{n} x_{ij}^2}} \tag{5}$$

Then get:

$$Z = \begin{bmatrix} z_{11} & \cdots & z_{1m} \\ \vdots & \ddots & \vdots \\ z_{n1} & \cdots & z_{nm} \end{bmatrix} \tag{6}$$

Calculate the probability matrix *P*:

$$P_{ij} = \frac{z_{ij}}{\sum_{i=1}^{n} z_{ij}} \tag{7}$$

Calculate the entropy and the wight of indicators with calculating the information value. For the j -th indicator, to calculate its information entropy:

$$e_j = -\frac{1}{ln\,10}\sum_{i=1}^{n} p_{ij}\, ln\left(p_{ij}\right),\ j = 1,2,\ldots,m \tag{8}$$

The definition of the information value is:

$$d_j = 1 - e_j \tag{9}$$

Accordingly, the grater the d_j, the greater the relative information and furthermore, we could have the entropy weight for each indicator:

$$\omega_j = \frac{d_j}{\sum_{j=1}^{m} d_j},\ j = 1,2,\ldots,m \tag{10}$$

Therefore, by combining the feature extraction and entropy weighting, with a specified data selection, the weights for each indicator of momentum could be obtained.

2.3. Best-of-n model

Meanwhile, players who are faced in a bunch of ns one shot opposition stages, we call it a "best-of-n" competition. A basic framework is considered in Figure 1.

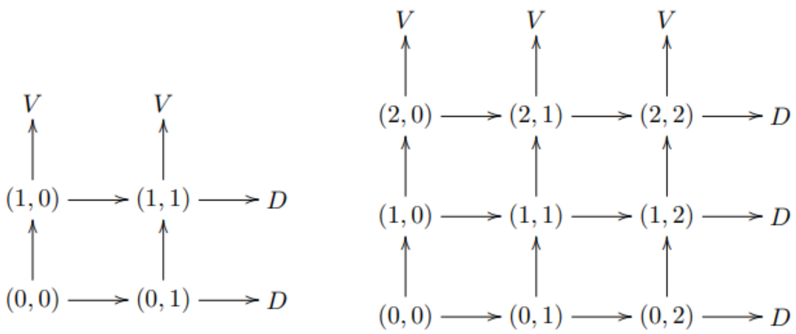

Figure 1. Best-of-3 and best-of-5 competitions tree

From above, it could witness that a competitor will win this match until he achieves the success of $(n+1)/2$ encounters.

3. Experiments

3.1. Feature Extraction

Initially conduct a thorough examination of the dataset with the aim of identifying key features. This process involved calculating the Spearman correlation coefficient to measure the monotonic relationship between variables. Subsequently, the features are meticulously extracted, resulting in the figure presented below:

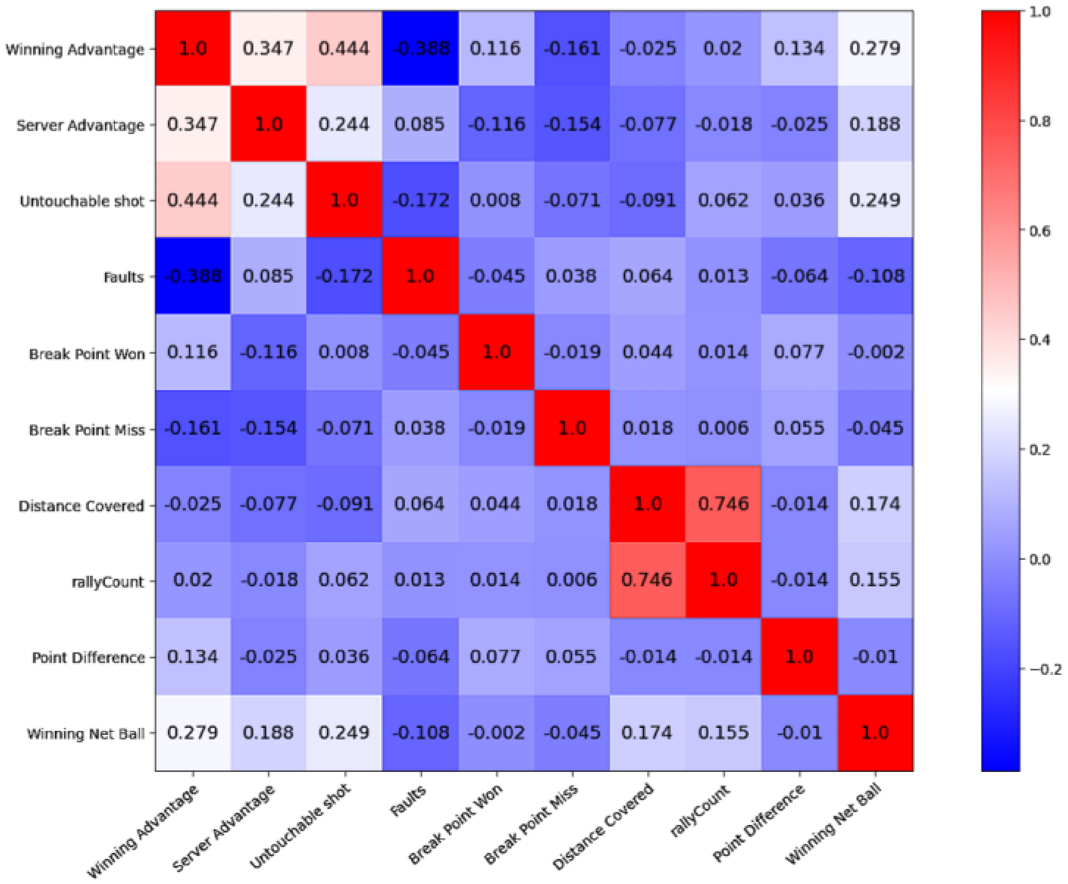

Figure 2. Spearman Correlation Coefficient Plot of "Momentum"

The figure summarizes key features: server advantage shows higher win probability for servers; point difference advantage indicates leads; untouchable shots are aces and winners; faults are errors that hurt momentum; winning advantage of net ball covers net points; break point winning advantage highlights breaking serve.

3.2. Information entropy and value

For the purpose of define a model of momentum to track the flow in the match, the specific 2023 Wimbledon Men's Singles final is selected, more factor that affect momentum according to the time series have been applied and explored, including Server Advantage (SA), Point Differ- ence Advantage (PDA), Untouchable Shot (ACE), Faults (DUF), Winning Advantage of Net Ball (NWA), Break Point Winning Advantage (BA), Break Point Missing (BM). Table 1 shows the information entropy and the information value obtained from calculation:

Table 1. Information entropy and value

Item	Information entropy e	Information value d
Server Advantage	0.92	0.08
Point Difference Advantage	0.996	0.004
Untouchable Shot	0.796	0.204
Winning Advantage of Net Ball	0.71	0.29
Break Point Winning Advantage	0.527	0.473
Faults	0.996	0.004
Break Point Missing	0.997	0.004

The weights of different indicators by Entropy Weighting Method (EWM) mentioned above as figure below:

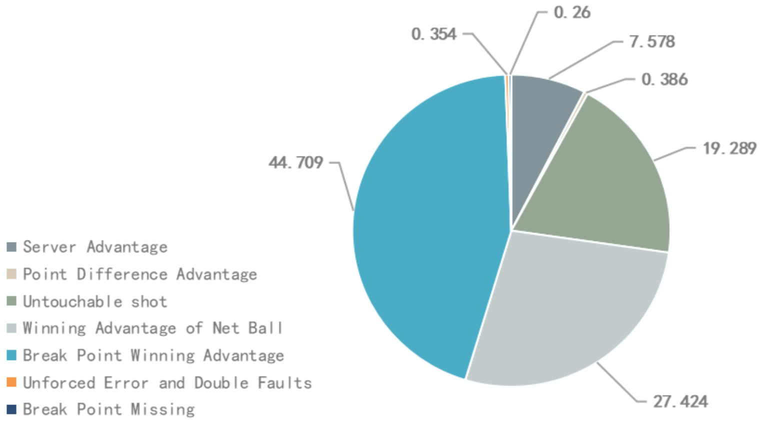

Figure 3. Weights of different indicator

Then, the data has been proceeded. Handling the abnormal and redundant data, the first few rows of the updated dataset are displayed. As Table 2 shows:

Table 2. Data including Momentum on Time Series

Match ID	P1	P2	ET	P1M	P2M
2023-Wimbledon-1701	Carlos Alcaraz	Novak Djokovic	00:00:00	0.10304	0.30945
2023-Wimbledon-1701	Carlos Alcaraz	Novak Djokovic	00:00:25	0.10472	0.38877
2023-Wimbledon-1701	Carlos Alcaraz	Novak Djokovic	00:01:29	0.10936	0.46613
2023-Wimbledon-1701	Carlos Alcaraz	Novak Djokovic	00:02:17	0.11095	0.9140
2023-Wimbledon-1701	Carlos Alcaraz	Novak Djokovic	00:03:02	0.111	0.914
2023-Wimbledon-1701	Carlos Alcaraz	Novak Djokovic	00:03:38	0.11295	0.91404

This optimized data set now incorporates the momentum metrics for two players, which have been calculated to provide a more nuanced understanding of their performance dynamics.

3.3. Visualisation Depiction

For understanding the momentum changes of players and track the match flow more comprehensive, combining the dataset and conducted two different visualizations of momentum trend on time series. Figures 4 to 5 presents it inline chart and scatter plot respectively.

Firstly, in Figure 4, the green dashed line represents the simulation of momentum changes of Alcaraz, and the orange represents to Djokovic, provides a clear observation of momentum fluctuation throughout the match.

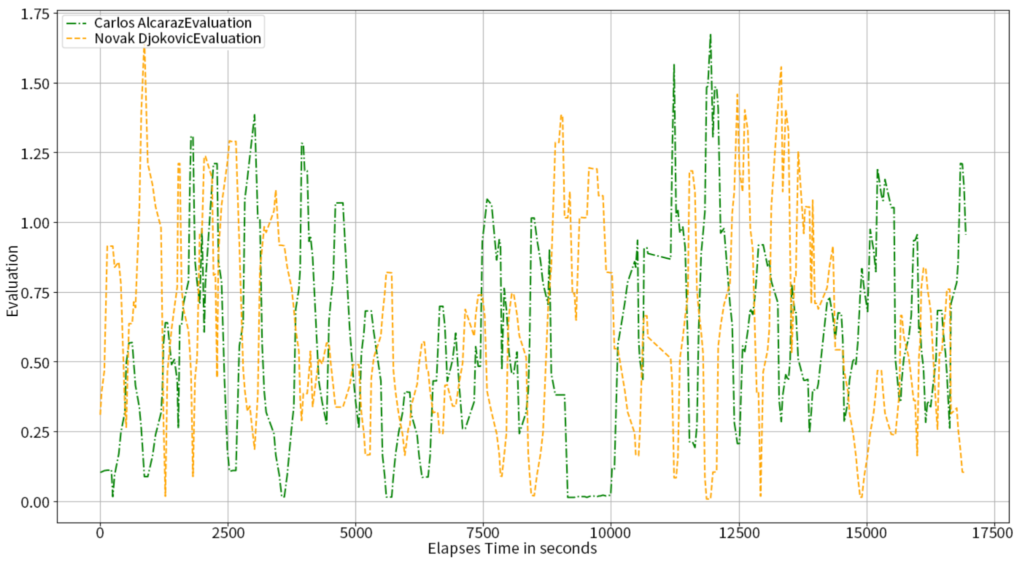

Figure 4. Momentum Flow

Secondly in Figure 5, is the plot where the colour intensity indicates the degree of the momentum. Blue represent Alcaraz and vice versa, revels that at different stages of the match, there are different variations in momentum, offering a benchmark for understanding the momentum.

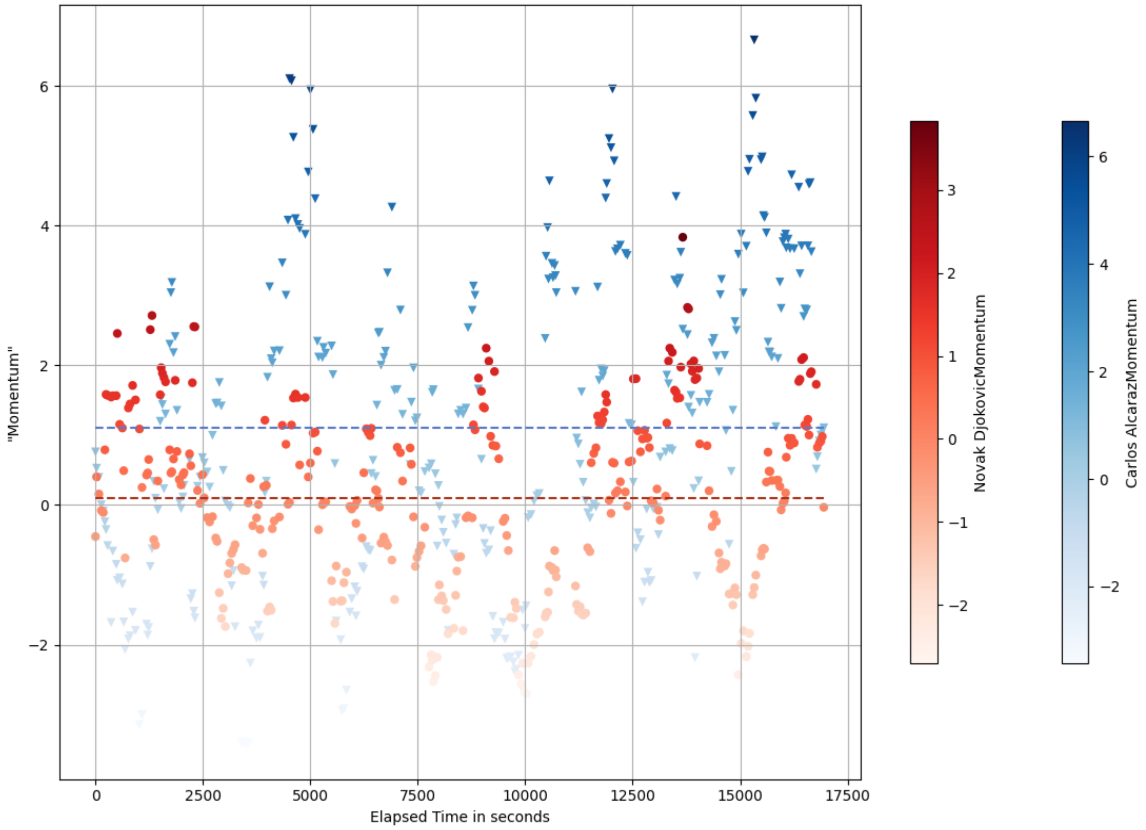

Figure 5. Final performance

This highly anticipated final ended with Alcarez lifted the trophy, overview of the whole momentum changes based on the model we built, the winner is much even more on his state. In addition to the predictions, the post-match statistics also indicates that the flow of the match are might concerned with a subjective state, momentum.

4. Conclusion

This paper mainly focus on the momentum of athletes in tennis and includes extensive, detailed analysis. With considerable effort to data processing and feature extracting. A robust model has been successfully developed to predict momentum shifts in competitive tennis, with a specific focus on Best-of-n competitions. Through comprehensive feature extraction and integration of advanced statistical methods such as Spearman's Correlation Coefficient and the EWM, the model effectively quantifies various momentum indicators, capture both psychological and performance-related factors. By applying these indicators to the 2023 Wimbledon Men's Singles final, the model demonstrates its predictive capacity, providing a detailed visualization of momentum trends across the match. The outcomes suggest that momentum is not only influenced by point differentials and physical performance but also by psychological elements like player confidence and opponent pressure. This model serves as a tool for a deeper understanding of competitive dynamics, offering insights valuable for strategic planning and performance enhancement in sports and other high-stakes fields. The integration of objective statistical methods and the visual depiction of momentum shifts provides a comprehensive framework with potential applications across various competitive environments, and it will contribute to a broader exploration of momentum's role in performance outcomes.

References

[1] Tao Chen, Qingliang Fan, Kai Liu, Lingshan Le. Identifying Key Factors in Momentum in Basketball Games [J]. Journal of Applied Statistics, 2020, 48 (16), 3116 - 3129.

[2] Benjamin S. Baumer, Gregory J. Matthews, Quang Nguyen. Big ideas in sports analytics and statistical tools for their investigation [J]. WIREs Computational Statistics, 2023, 15 (6): e1612.

[3] Junyan He, Shashank Kushwaha, Jaewan Park, Seid Koric, Diab Abueidda, Iwona Jasiuk. Sequential Deep Operator Networks (S-DeepONet) for predicting full-field solutions under time-dependent loads [J]. Engineering Applications of Artificial Intelligence, 2024, 127 (A), 107258.

[4] Wilkens Sascha. Sports prediction and betting models in the machine learning age: The case of tennis [J]. Journal of Sports Analytics, 2021, 7 (2): 99 - 117.

[5] Andrea Giovanni Reina, Thierry Njougouo, Elio Tuci, Timoteo Carletti. Speed-accuracy trade-offs in best-of-n collective decision making through heterogeneous mean-field modeling [J]. Phys. Rev. E, 2024, 109 (5), 054307.

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

[7] Craig A. Depken, John M. Gandar, Dmitry A. Shapiro. Set-level Strategic and Psychological Momentum in Best-of-three-set Professional Tennis Matches [J]. Journal of Sports Economics, 2022, 23 (5): 598 - 623.

[8] Han Yu, Alan D. Huston. A robust Spearman correlation coefficient permutation test [J]. Communications in Statistics - Theory and Methods, 2022, 53 (6): 2141 - 2153.

[9] Raman Kumar, Sehijpal Singh, Sehijpal Singh et al. Revealing the benefits of entropy weights method for multi-objective optimization in machining operations: A critical review [J]. Journal of Materials Research and Technology, 2021, (10): 1471 - 1492.

[10] Nathan Eugster, Christian Ewerhart. Rules in International Tennis Tournaments: A Game Theoretic Perspective [J]. Journal of Sports and Games, 2024, 6 (1), 33 - 42.