A study of trends in tennis matches

Authors: Ma Yihan, Liu Wenyu, Li Junjie

Date: 2024-02-20T23:58:07+00:00

Abstract

This study aims to accurately predict shifts in match trends and trajectories by analyzing match flow. To capture match flow, we first defined an A-value and developed a decision tree model. Furthermore, we constructed a nonlinear autoregressive neural network to implement predictive functionality. During the model refinement process, we computed the Pearson correlation coefficient to quantify the degree of influence. The results demonstrate that the model achieved relative success in implementing predictive functionality. The number of aces, double faults, and unforced errors are key influencing factors.

Full Text

Preamble

A Study of Trends in Tennis Matches

Yihan Ma, Wenyu Liu, Junjie Li
Jinan University and Birmingham University Joint Institute, No. 855, East Xingye Avenue, Panyu District, Guangzhou, China

Keywords: nonlinear autoregressive neural network, flow of play, model prediction, Wimble-

Abstract: This study aims to accurately anticipate shifts in tennis matches by analyzing the flow of play. To achieve this, we enhanced our model by scrutinizing factors that impact its forecasting capabilities. We defined an A-value to capture the flow of play and developed a decision tree model to predict it. Additionally, we constructed a Nonlinear Autoregressive Neural Network for forecasting. During model improvement, we calculated Pearson correlation coefficients to gauge factor impact. Results indicate the model performs its predictive function with relative success, with aces, double faults, and unforced errors identified as key influencing factors.

1. Background

In contemporary society, individuals increasingly pursue higher quality of life through sports, with tennis emerging as a popular choice. This demanding sport requires both physical prowess and mental acuity, as players must react swiftly to maintain advantageous positions. Momentum represents a critical factor in predicting set outcomes, influenced by various elements including service advantage, scoreline, and relative match strength. Through data analysis, we can gain valuable insights into how momentum impacts match trends and directionality.

2. Construction of the Model

Our analysis focuses on the 2023 Wimbledon-1301 match, using Player 1 as our exemplar case study.

2.1. Quantification of the Flow of Play

2.1.1. Construction of the Decision Tree Model To quantify momentum, we created a parameter called A and constructed a decision tree model to predict its value. We examined various factors and categorized different scenarios, each exerting unique influence on A. We defined the relative A-value as the difference between Player 1 and Player 2's A-values, enabling us to track game flow through the sign and magnitude of this relative metric.

We utilized consecutive points scored and serving side as primary and secondary decision criteria, assigning A-value significance to overall player performance. Analyzing both players' A-values enables comprehensive performance assessment.

A-value calculation follows specific rules: it initializes at 0 and remains unchanged absent a scoring streak. During a streak, A increases by 1, 1.5, or 2 depending on whether two serves, one serve, or no serve occur. Figure 1 illustrates this algorithm through a flowchart.

We further explore the relative A-value's implications: positive and negative signs indicate Player 1 or Player 2 dominance, respectively, while the absolute value quantifies the performance differential. Graphs 1 and 2 visualize A-value fluctuations for both players and corresponding changes in the relative A-value.

2.2. Forecast for the Swings in a Game

We constructed a nonlinear autoregressive neural network to accurately forecast match outcomes. The model incorporates relevant A-values and divides data into three sets: 70% for training, 15% for validation, and 15% for testing. Figure 2 depicts the neural network architecture. Following final training, validation, and testing, we focus on the Mean Squared Error (MSE) indicator, which demonstrates the model's reliability in predicting match fluctuations.

3.1. Test for Other Matches in the Same Tournament

To test model generalizability, we applied it to the 2023 Wimbledon-1302 match as test data. We calculated relative A-values and input them into the neural network. Results show that fewer than five examples exhibited errors as low as 2.294 units. Supplementary tests revealed that over 75% of samples had minimal errors compared to actual values, with approximately 25 samples showing negligible deviation. The neural network achieved an MSE of 0.68, demonstrating accurate relative A-value prediction and effective gameplay forecasting.

3.2. An Exploration of Factors Affecting Model Accuracy

Yinman Zhang (2009) suggests that aces, double faults, and unforced errors may play pivotal roles in determining outcomes of men's hard court tennis matches. To investigate further, we utilized Pearson correlation coefficients. We counted aces accumulated after each point as an independent variable, with relative A-value at each moment as the dependent variable, enabling clear visualization of their relationship. We applied the same methodology to double faults and unforced errors.

Calculating Pearson correlation coefficients for each relationship yielded values of approximately 0.80 for all three variables (double faults, unforced errors, and aces), indicating strong associations and suggesting they are primary determinants of match outcomes.

4. Strengths and Weaknesses of the Model

Our model offers valuable versatility, applicable across various competitions and scenarios. By consolidating multiple factors into a unified framework, it enables accurate prediction of game flow and momentum. However, certain cases suffer from missing essential data, requiring informed assumptions during model development. Accessing more comprehensive data resources would undoubtedly improve results. Notably, striking a balance between realism and elegance presents a challenge, and our model prioritizes realism.

5. References

[1] COMAP Mathematics Competitions. (n.d.). Retrieved from https://www.comapmath.com/MCMICM/index.html
[2] Zhang, Yinman. (2009). Winning factors in hard court matches of the world's best men's tennis singles players. *Journal of Beijing Sport University*, 32(10), 135-137.

Note: Figure translations are in progress. See original paper for figures.