ISSN 2616-5775 Vol. 7, Issue 12: 96-103, DOI: 10.25236/AJCIS.2024.071214

A momentum synthesis model of the tennis ball based on the TERFAS

Boya Zhan

College of Computer, Jiangxi University of Traditional Chinese Medicine, Nanchang, 330004, China

Abstract: *The 2023 Wimbledon final showcased momentum shifts in Carlos Alcaraz's victory over Novak Djokovic. This study investigates momentum in tennis to inform tactical strategies. Firstly, we developed the Tennis Elite Rating and Factor Analysis System (TERFAS). Specifically, the factor analysis and combined weighting were used to quantify the impact of serve, break points and player skill. Then, 10 key metrics were identified and their impact on match performance as measured by real-time momentum insights was assessed using non-linear regression models. Additionally, we confirmed that momentum influences match outcomes using a logistic regression model validated by Pearson correlation. Furthermore, the ARIMA and Markov models were integrated into the MarkoRima Tennis Forecaster (MRTF) to provide strategy recommendations in order to improve accuracy. Meanwhile, we verified the validity and generalisability of the models using entropy analysis. Finally, simulation experiments show that the model designed in the article has good robustness and sensitivity.*

Keywords: *TERFAS model, factor analysis method, Markov model, Arima model, relative entropy*

1. Introduction

In the 2023 Wimbledon final, Carlos Alcaraz triumphed over Novak Djokovic, ending Djokovic's decade-long dominance at the tournament [1]. The match was characterized by pronounced momentum fluctuations, with Djokovic commanding the first set 6-1, followed by Alcaraz narrowly winning the second set in a tie-break, and both players alternating control in the subsequent sets. Alcaraz ultimately secured a 6-4 victory in the decisive set, underscoring the significant role of momentum in tennis. Given the clear impact of these momentum shifts on match outcomes, this study aims to explore the dynamics of momentum in tennis and its influence on player strategies.

Momentum is often described as the accumulated force or intensity of an action or sequence of events, and it plays a key role in influencing the outcome of a game [2]. Athletes and teams frequently perceive momentum shifts during competition. Yet how to capture and quantify this abstraction is a huge challenge[3, 4]. Moreover, identifying the specific match events that trigger or shift momentum is not always straightforward. In response to these challenges, we present a series of models designed to analyze momentum and evaluate its impact on match progression.

Building on the analysis of momentum in tennis, we developed the Tennis Elites Rating and Factor Analysis System (TERFAS) using factor analysis and combinatorial entropy weighting to quantify the effects of serve performance, break points and individual skills. A nonlinear regression model was created to visually represent the impact of these indicators on match outcomes. We then introduced a logistic regression model using maximum likelihood estimation, with the Pearson correlation confirming that momentum influences match results. To improve predictive accuracy, we incorporated a Markov model and time-series analysis to track momentum shifts and provide strategic recommendations, identifying second-serve performance and pressure-handling as key factors[5, 6]. The model was validated with women's tennis data using entropy analysis, confirming its robustness, though further adjustments may be needed for broader applications.

2. Model Assumptions

Based on the common practices and principles commonly used to analyze sports competitions and build mathematical models, we propose some conventional assumptions to simplify the problem and build a feasible model while ensuring that each assumption is reasonable.

(1) Suppose that the results are influenced by differences in skill level and performance. This is a

ISSN 2616-5775 Vol. 7, Issue 12: 96-103, DOI: 10.25236/AJCIS.2024.071214

basic assumption, as a player's skill and performance levels usually affect the outcome of a game.

(2) Suppose that the score and the situation in the game can reflect the dynamics of the game. This means that scoring changes and situation shifts during matches can be used to analyze momentum.

(3) Since the server is much more likely to win in tennis, consider reducing the entropy weight to balance the winner against the receiver.

3. Model Preparation

3.1 Definitions and Notations

In this work, we use the nomenclature in Table 1 in the model construction. Other used symbols will be introduced once they are used.

Table 1: Notations used in this paper

Symbol	Description	Unit
β0	The baseline log-odds predicted by the model	/
BIC	A measure of the goodness of fit of statistical models	/
F1	Harmonized average of precision and recall	/
AVE	Assess the internal consistency and reliability of each construct in the measured model	/
CR	Consistency and reliability of the constructs in the measurement model	/

3.2 Data preprocessing

Before modeling, we focused on data preprocessing to address potential noise, outliers, and missing values. The data, sourced from http://www.mcm.edu.cn, was standardized, renumbered for 31 matches, and unified in time format for consistency. Outliers were identified using box plots, and missing values were handled with the K-nearest neighbor method (K=5) to improve robustness and prevent overfitting.

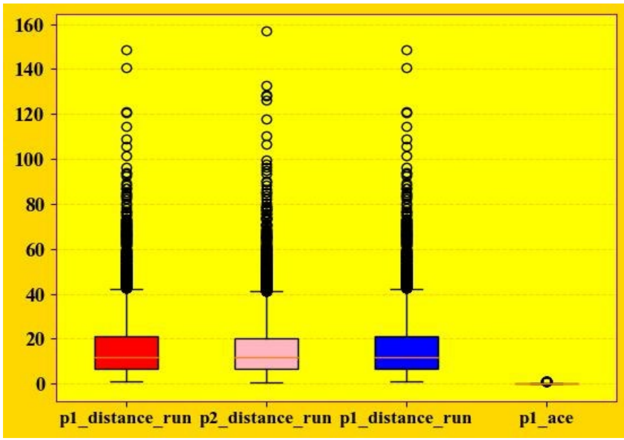

Figure 1: Outlier value display

Figure 1 illustrates the presence of outliers in the "distance_run" data, whereas no outliers were detected for the ACE variable.

4. Establishment of Tennis Elites Rating and Factor Analysis System

4.1 Extraction of the raw variables

In analyzing modern tennis matches, ten key technical indicators were selected as independent variables, including ACE per game, success rate, scoring rate, second serve success rate and breakpoint save rate, based on data from the ATP. Additionally, mental state, physical fitness and technical skills were considered, with error rate, skill score, and endurance value included as variables. A nonlinear regression model was developed to quantify the impact of these factors on match outcomes. Endurance was simplified as the distance moved per minute, and correlation analysis was conducted on error rate

ISSN 2616-5775 Vol. 7, Issue 12: 96-103, DOI: 10.25236/AJCIS.2024.071214

and scoring skills.

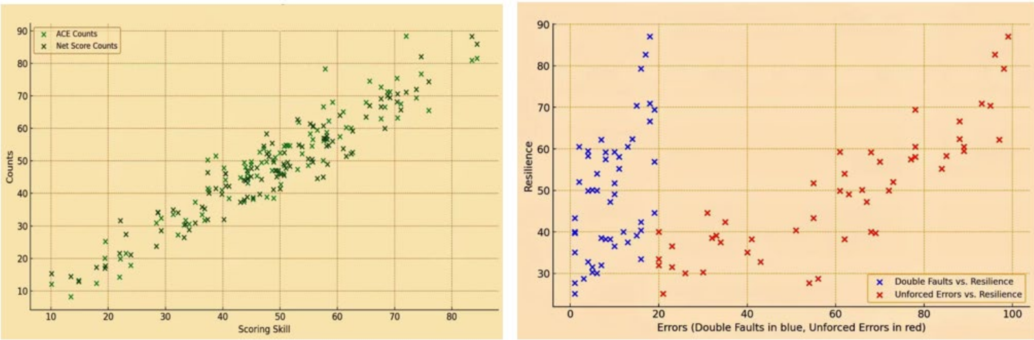

Figure 2: Correlation analysis of error rate and scoring skills

The figure 2 shows a positive correlation between scoring skill and successful plays (ACEs and net scores), while the right plot indicates that resilience decreases more significantly with increasing double faults than with unforced errors.

4.2 Model preparation

Before conducting multivariate statistical analysis like factor analysis, a thorough preliminary examination of the dataset is essential to handle outliers and ensure accuracy. We previously completed data preprocessing. To ensure reliability, 15 matches were selected as the sample size. Factor analysis requires variables to follow a normal or approximate normal distribution for effective correlation extraction. The Kaiser-Meyer-Olkin (KMO) test yielded a value of 0.772, indicating moderate collinearity, while Bartlett's test produced a P-value of 0.000, rejecting the null hypothesis and confirming the suitability of the factor analysis. These results demonstrate a correlation between variables, validating the method's feasibility.

4.3 Exploratory Factor Analysis Method (EFA)

We conducted an exploratory factor analysis and obtained a matrix table of component score coefficients, reflecting the load of each variable on common factors. The factor scoring coefficients were used to calculate the total score of influencing factors in tennis matches. The following formula represents the principal component calculation:

$$F = \frac{0.412}{0.758} \cdot F_1 + \frac{0.216}{0.758} \cdot F_2 + \frac{0.129}{0.758} \cdot F_3 \tag{1}$$

This formula is derived from the component matrix table and is used to quantify the contribution of each factor.

4.4 Visualization and Performance Analysis

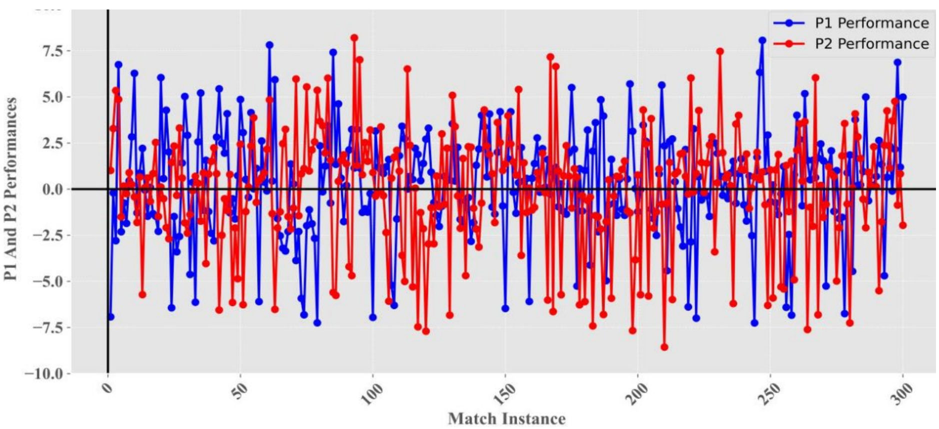

Figure 3: P1 And P2 Performances for match

ISSN 2616-5775 Vol. 7, Issue 12: 96-103, DOI: 10.25236/AJCIS.2024.071214

In this problem, we build a TERFAS model to capture the game process when the score occurs and explain the player momentum changes during the game. Based on this, we build a nonlinear regression model to finally form a visual matching process. By establishing the above model, the model obtained the following results:

Figure 3 illustrates the momentum shifts between P1 and P2. Specifically, P1 led in the first 50 minutes, followed by P2 gaining the advantage between 50 and 100 minutes. P2 outperformed P1 from 100-150 minutes, while P1 regained the lead early in the 150-200 minute range, but P2 later surpassed. In the final period (250-300 minutes), both players were evenly matched.

4.5 Construct a logistic regression model

4.5.1 Model Construction and Fitting

A logistic regression model was constructed using Player 1's momentum score as the independent variable and Player 1's win outcome as the dependent variable. The model parameters were estimated using the maximum likelihood method, where a positive coefficient for Player 1's momentum score indicates a positive correlation with winning probability, while a negative coefficient for Player 2's momentum score shows an inverse relationship with Player 1's winning probability. Pearson's correlation coefficient was also calculated, showing a moderately positive correlation of 0.587 between Player 2's momentum score and match outcome, confirming that momentum significantly influences game results.

The model parameters are fitted by maximum likelihood estimation to maximize the probability of the observed data at a given feature.

4.5.2 Prediction of the probability of winning

The fitted logistic regression model is used to predict the winning probability of player 1 based on the momentum score and start to build a logistic regression model. The output of the logistic regression model is the probability of the event, and the calculation formula is:

$$P(Y=1)=\frac{1}{1+e^{-(\beta_0+\beta_1 X_1+\beta_2 X_2)}} \tag{2}$$

4.6 Analyze the influence of the momentum score

By fitting a logistic regression model, we can analyze the influence of the momentum score (β_1) on the outcome of the game. If β_1 is significantly greater than 0, it means that the momentum score is positively correlated with the winning probability; otherwise, if β_1 is significantly less than 0, it indicates a negative correlation. The results are as follows:

Table 2: Parameter estimation results of the logistic regression model

Intercept (β_0)	Coefficient of the player-1 momentum score (β_1)
−0.0142	0.949

As can be seen from the table 2 above, the momentum score of player 1 has a positive effect on the outcome, that is, for each unit increase in the momentum score, the log odds (log odds) of player 1 will increase by 0.949 units.

4.7 Model results and evaluation

The table 3 presents the predictive evaluation metrics for the cross-validation, training, and test sets, used to assess the performance of the logistic regression model. Hyperparameters are adjusted via cross-validation to ensure a reliable and stable model.

Table 3: Model results

	Precision	Recall	Accuracy rate	F1
Training Set	0.871	0.871	0.874	0.872
Test Set	0.956	0.956	0.956	0.956

The model performs better on the test set than the training set, showing strong generalization and no overfitting. High accuracy, recall, precision, and F1 scores indicate the model reliably predicts match outcomes. Consistently high performance on unseen data demonstrates its predictive reliability. These

ISSN 2616-5775 Vol. 7, Issue 12: 96-103, DOI: 10.25236/AJCIS.2024.071214

results confirm that the momentum score, particularly for Player 1, is a strong predictor of match outcomes. In conclusion, momentum is a significant factor in tennis, and using statistical models, it can be effectively quantified to support strategic decisions for coaches and players.

5. Establishment of prediction model

In a tennis match, the changing dynamics of the game and fluctuating player momentum are critical to predicting the outcome of the match. However, traditional prediction models often fail to fully capture these complex and variable dynamics. To address this problem, we developed a comprehensive prediction model combining the Markov and time series models to capture game dynamics from multiple perspectives. This approach improves prediction accuracy, generalization and provides a more robust system for understanding and predicting momentum fluctuations in tennis, as shown in the figure 4 below:

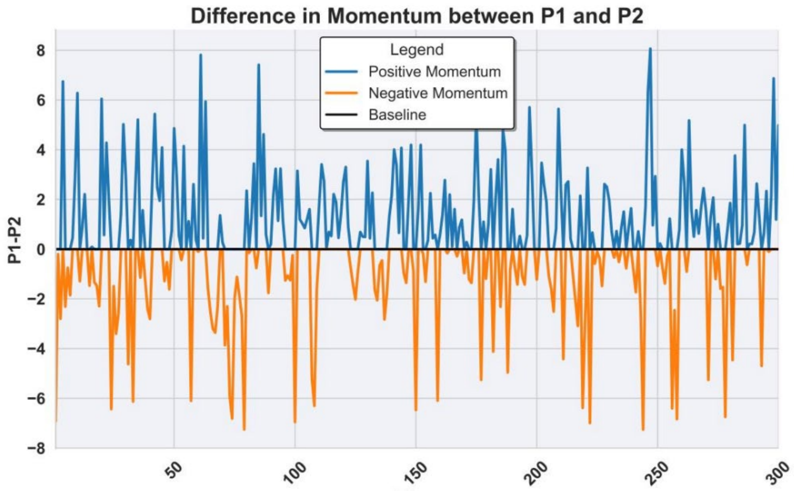

Figure 4: Visual preview of the momentum score

5.1 Define the status

In this model, we define four possible states for players during the game: Player A performs well if they score, and poorly if they lose points; similarly, Player B performs well if they score, and poorly if they lose points. The initial state is determined by the outcome of the first score in the game.

5.2 Data processing and transfer matrix construction

To calculate Player A's gains and losses over time, we summarize their scores and lost points. The probability of moving from one state to another is determined by computing the state transition frequency and normalizing it. This results in a transition probability matrix, which shows the probabilities of transitioning between all states. The matrix size remains constant, and the state transition matrix formula is:

$$P_{ij}^{(n+m)} = \sum_{K=1}^{I} P_{ik}^{(n)} P_{kj}^{(m)} \tag{3}$$

5.3 Model prediction results and evaluation

According to the transition matrix and the initial state distribution, the current state and the transition matrix are used to predict the next state, with the time as the X axis, and the score as the Y axis, resulting in the following figure.

ISSN 2616-5775 Vol. 7, Issue 12: 96-103, DOI: 10.25236/AJCIS.2024.071214

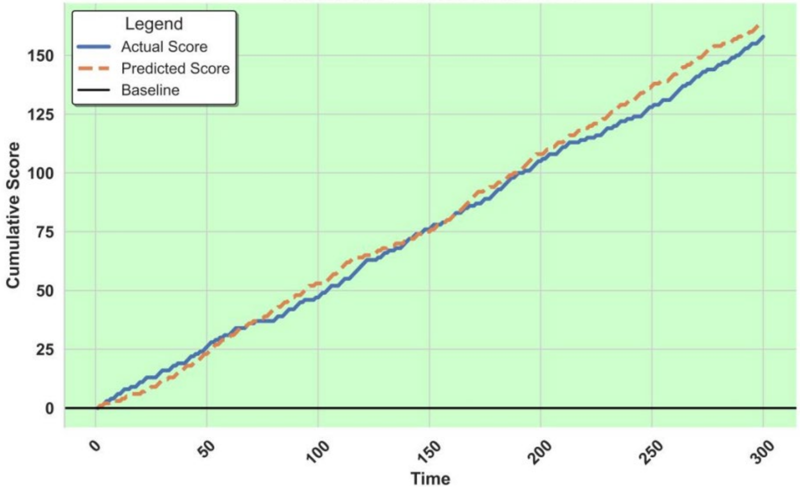

Figure 5: Prediction results

From figure 5, through the predicted score and the real match score fitting, the results show a very high degree of fit between our model and the actual data. This not only validates the effectiveness of the model, but also highlights its excellent performance in predicting scores in tennis matches.

5.4 Establishment of ARIMA

We used the ARIMA model to predict player momentum in tennis, consisting of autoregressive (AR), integration (I), and moving average (MA) components. A stability test (ADF) ensured the data's stationarity, and BIC was used to optimize model parameters. A white noise test confirmed uncorrelated residuals, indicating a good model fit.

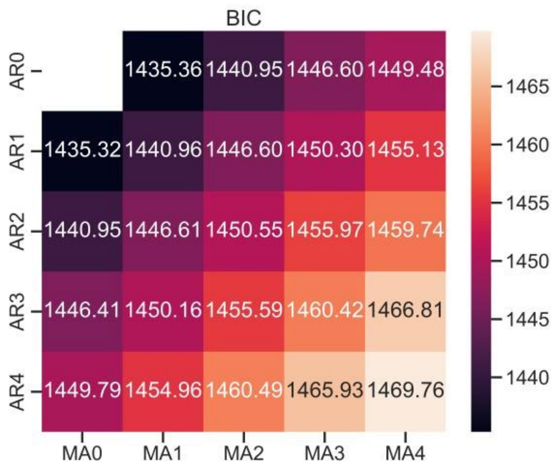

Figure 6: ARIMA model test table

From the figure 6, the BIC values suggest that the AR0-MA0 model provides the best prediction for momentum shifts in tennis matches, offering accuracy without added complexity. This model minimizes overfitting, leading to more reliable and generalizable predictions.

5.5 Prediction of model results

We used the ARIMA (1, 0, 1) model to predict future data based on historical trends, achieving a good fit and strong performance. The results are shown below.

ISSN 2616-5775 Vol. 7, Issue 12: 96-103, DOI: 10.25236/AJCIS.2024.071214

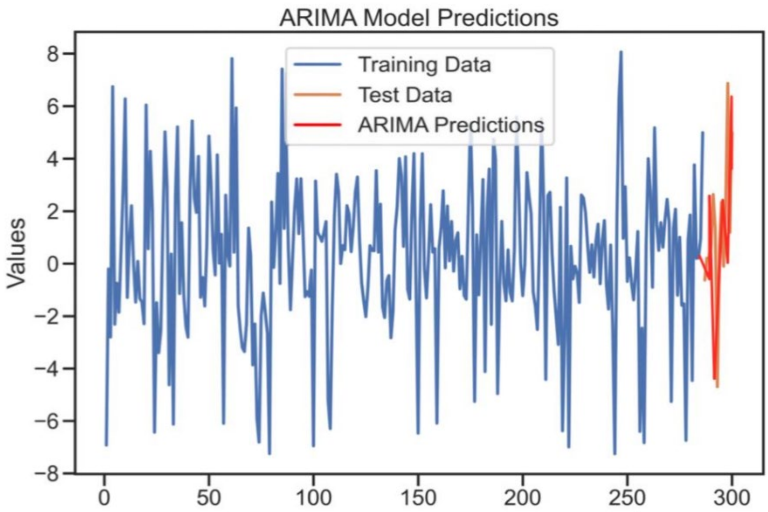

Figure 7: Result of prediction

From the figure 7, the prediction results closely follow the actual data points of the test set, forming a good match.

5.6 Search for the most relevant factors

Factor analysis identified the most relevant factors through total variance interpretation and a scree plot. The interpretation showed that three principal components explained 75.77% of the variance, with minimal improvement when adding a fourth (84.77%). Thus, three principal components were selected for further analysis.

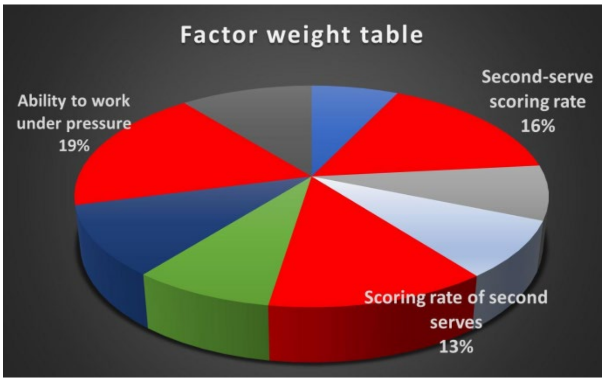

Figure 8: Weight diagram

According to the figure 8, draw the above figure. Finally, we identified the scoring rate of second serves, second-serve scoring rate, and ability to work under pressure, as the most relevant factor, namely the momentum effect on the player.

6. Conclusions

In this article, we develop three distinct models to assess the momentum of each player and predict the fluctuation time of that momentum. The models are divided into two categories: an evaluation model and a prediction model. In the evaluation model, we assign corresponding weights to various momentum-related factors to quantify momentum. This allows us to provide a numerical representation of a player's momentum at any given point, which can then feed into the prediction model. Using this setup, we are able to predict the timing of future momentum shifts with a high degree of accuracy.

Our results demonstrate that momentum is a real and measurable phenomenon that significantly influences the outcome of games. By quantifying momentum and predicting its fluctuations, we give coaches valuable information that allows them to make timely and informed strategic decisions. These insights have the potential to improve in-game adjustments, optimizing both player performance and

ISSN 2616-5775 Vol. 7, Issue 12: 96-103, DOI: 10.25236/AJCIS.2024.071214

overall team strategy by addressing shifts in momentum as they occur.

In addition, we carefully analyzed each model to ensure their effectiveness in capturing and predicting momentum changes. Our evaluation shows that the models have strong predictive capabilities and can reliably forecast future momentum shifts. This predictive power offers valuable insights that can help coaches make timely strategic decisions and enhance performance management in competitive scenarios.

References

[1] Sun Y, Xiong R, Zhu Q, et al. Balance motion generation for a humanoid robot playing table tennis[C]//IEEE-RAS International Conference on Humanoid Robots. IEEE, 2011:19-25. DOI:10.1109/Humanoids.2011.6100826.
[2] Ravesloot M J L, Frank M H, Van Maanen J P, et al. Positional OSA part 2: retrospective cohort analysis with a new classification system (APOC)[J].Sleep And Breathing, 2016, 20(2). DOI:10. 1007/s11325-015-1206-y.
[3] Chen Y F, Fang T H, Huang C C, et al. Research on the Flows of Backspin Movement in Table Tennis[J].Advanced Materials Research, 2011, 201-203:2263-2266. DOI: 10.4028/ www.scientific. net/AMR.201-203.2263.
[4] A. A. M , G. I. D .Justification Of The Method For Developing Accuracy Of Throwing A Tennis Ball At The Target In III – Class Students[J].Journal of Advanced Zoology, 2024, 45.
[5] Bartz S P. Delayed rebounds in the two-ball bounce problem [J].European journal of physics: A journal of the European Physical Society, 2022(3):43.
[6] S, J, Haake, et al. Oblique Impact of Thick Walled Pressurized Spheres as Used in Tennis[J]. Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science, 2016.DOI:10.1243/095440605x32084.