An Integrated Assessment Model of Momentum and Performance in Tennis Based on PCA And TOPSIS

Zhiming Liang, Xiukang Shen *, Yuxing He

School of Resource Environment and Safety Engineering, University of South China, Hengyang, China, 421001

* Corresponding Author Email: 18607452216@163.com

Abstract. This study conducted a complex analysis of the performance of athletes in tennis matches, taking into account that the performance of athletes is not only affected by the direction of the serve, but also by a variety of factors. Therefore, this study used a combination of Principal Component Analysis (PCA) and TOPSIS to solve this problem. Specifically, this study first extracted three indicators based on the original data and determined the weights of these indicators using the PCA method. Then, the momentum is reflected by weighting and summing these indicators. Next, this study used the TOPSIS model to combine momentum with other metrics in order to assess, visualise and analyse the performance of the players' scores at each point in the match. With this approach, this study was able to capture the flow of the game as scoring occurred during the match. Finally, this paper tested our model on several games and demonstrated that the model better captures the flow of the game as it is scored.

Keywords: Tennis game, principal component analysis, TOPSIS modeling.

1. Introduction

Tennis is a highly competitive and complex sport, and the performance of players is affected by a variety of factors, including technical level, psychological state, and match situation. Therefore, it is crucial to accurately analyze and predict athlete performance for coaches and players alike. In recent years, with the development of data science and machine learning technology, the use of these technologies to analyze and predict the results of sports competitions and athletes' performance has become one of the hot spots in research [1-2].

In tennis, a player's performance depends not only on their technical level but also on various factors such as the direction of the serve and double faults. Therefore, traditional analytical methods may not comprehensively capture these complex relationships [3]. The aim of this study is to explore a methodology that combines Principal Component Analysis (PCA) and TOPSIS modeling to more accurately assess athletes' performance. PCA can be used to determine the weights of different indicators, while TOPSIS modeling can combine these indicators to comprehensively assess athletes' performance [4].

Through analysis of tennis game data and model validation, our aim is to provide athletes with a reliable and accurate method for assessing and predicting game results, as well as developing more effective training and game strategies. This study's results are not only crucial for improving tennis players' performance levels but also provide valuable references and lessons for analyzing other sports competitions using data science techniques [5]. The data source for this paper is https://www.wimbledon.com/. We have selected the 2023 Wimbledon Gentlemen's Singles match data to analyse and have coded the number of matches. For example, 2023-wimbledon- CAND701, where "2023" denotes the year, "CAND" denotes the two players Carlos Alcaraz and Novak Djokovic, "7" denotes round 7, and "01" denotes the first match listed.

2. PCA-based TOPSIS modeling

The player's performance is quite complex, not only limited to whether he/she is on the serve side or not, but also affected by a variety of factors. Therefore, we used a combination of Principal Component Analysis (PCA) and TOPSIS to solve this problem.

The study extracted three indicators, namely score, point spread, and advantage serve (AD), from the raw data. The weights of these indicators were determined using the PCA method. The momentum was then reflected by weighting and summing these indicators. The TOPSIS model was used to combine momentum with other metrics to calculate the players' scores at each time point, assessing their scoring performance during the game and visualizing and analyzing it. This method allowed us to capture the game's progression as points were scored.

Finally, we tested multiple matches and demonstrated that our model can more accurately capture the flow of the match during scoring.

2.1. Data pre-processing

Prior to modelling, the data needed to be initially checked and processed. Following the scoring rules of tennis matches, the legitimate scores should be 0, 15, 30, 40, and AD. However, 18 abnormal data points were found in the data for the 2023-wimbledon-CANJ304 matches in the p1_score and p2_score columns, which included scores between 0 and 9. Therefore, we eliminated these anomalous samples, ultimately leaving 319 samples in the 2023-wimbledon-ADFHR304 tournament.

Meanwhile, we did not find any missing values in the data. The data preprocessing process is shown in Figure 1.

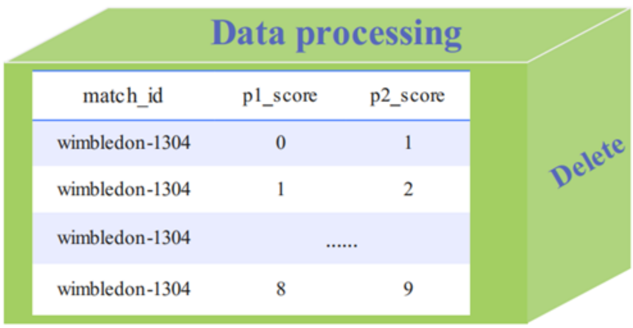

Figure 1. Diagram of the data pre-processing process

2.2. Quantification of Momentum

Momentum is a challenging concept to quantify directly. However, it is possible to measure the characteristics that indicate changes in momentum. In this text, we will use the term 'momentum' to refer to these characteristics. Momentum is defined as:

$$M_i = \alpha S_i + \beta(S_{1i} - S_{2i}) + \gamma Ad_i \quad (1)$$

Where M_i, S_i, S_{1i}, S_{2i}, Ad_i denote the momentum of Athlete 1 at the ith time point, whether or not he scored (if he scored the value is 1, and vice versa), the cumulative score of Athlete 1 at the ith time point, the cumulative score of Athlete 2 at the ith time point, and whether or not Athlete 1 scored Ad at the ith time point (if he scored Ad the value is 1, and vice versa), respectively, and α, β, and γ denote weights.

Principal Component Analysis (PCA) is used to determine the weights of each indicator, considering their possible correlation. The degree of explanation of the variance of the principal components on the three indicators indicates the strength of the correlation, the greater the degree of explanation, the stronger the correlation, the weight should be greater, and finally the final weight of each feature is calculated by normalization. The flowchart is shown in Figure 2.

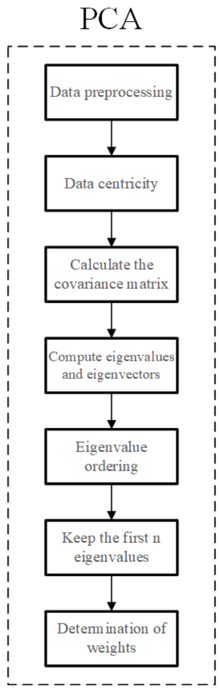

Figure 2. PCA flowchart

The momentum of player1 was calculated for each of the four games with IDs 2023-wimbledon-ADFHR304, 2023-wimbledon-DMJL403, 2023-wimbledon-ARND504, and 2023-wimbledon-CAND701, as shown in Figure 3.

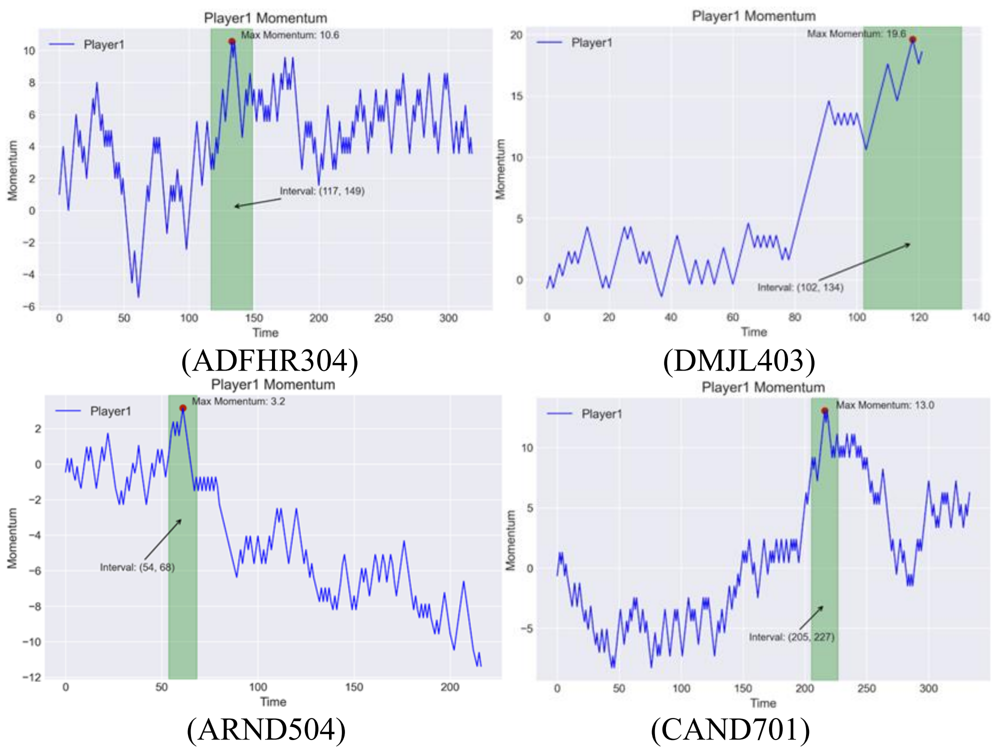

Figure 3. Momentum of player1 for each of the four games

2.3. Quantification of performance

Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) is also known as the distance method of superior and inferior solutions. It is a commonly used comprehensive evaluation method, by detecting the distance between the evaluation object and the optimal solution and the worst solution to rank, if the evaluation object is closest to the optimal solution and furthest away from the worst solution at the same time, then it is the best; otherwise, it is not optimal [6]. The flowchart of the TOPSIS algorithm is shown in Figure 4:

Figure 4. TOPSIS flowchart

Momentum in tennis has been discussed in detail by Philippe Meier [7], Ben Moss [8], and other scholars [9-10]. According to their studies, momentum may indicate a player's mental state, while a player's scoring performance during a match is influenced by both mental state and strategic technique. Therefore, we have analyzed the possible scoring influences of tennis players and finally identified that the factors [11] affecting momentum are shown in Table 1:

Table 1. Indicators for calculating the significance of characteristics

Factors	Explaination
Momentum	The momentum of the players
p1_sets	sets won by player 1
p1_games	games won by player 1 in current set
server	server of the point
p1_ace	player 1 hit an untouchable winning serve
p1_winner	player 1 hit an untouchable winning shot
p1_double_fault	player 1 missed both serves and lost the point
p1_unf_err	player 1 made an unforced error
p1_break_pt_won	player 1 won the game player 2 is serving
p1_break_pt_missed	player 1 missed an opportunity to win a game player 2 is serving
p1_distance_run	player 1's distance ran during point (meters)
rally_count	number of shots during the point

We added momentum, server, and other metrics to the TOPSIS model for very large metrics. According to the information in the game, the value of 'serve' can be 1 or 2. For player1, a value of 1 is more favorable because it means that the player serves first, i.e., a smaller value is more favorable. Therefore, in the TOPSIS data normalization step for player1, we convert the 'serve' metric into an extremely large metric. For player 2 with a value of 2, a larger value is more favorable, i.e., the indicator is very large at this point, so no data normalization is performed. Similarly, the other very small indicators were uniformly converted to very large indicators. Subsequently, data visualisation was used to analyse the players' momentum and scores over the course of the game.

2.4. Analysis of results

Figure 5 shows that Carlos Alcaraz began the match with low momentum, while Novak Djokovic took the lead. However, over time, Alcaraz gradually increased his momentum, and the scores of the two matches were closer together, and this pattern coincided with the actual match result.

This pattern ultimately coincided with the actual match result, as Carlos Alcaraz initially fell behind but then reversed to beat Djokovic 3-2.

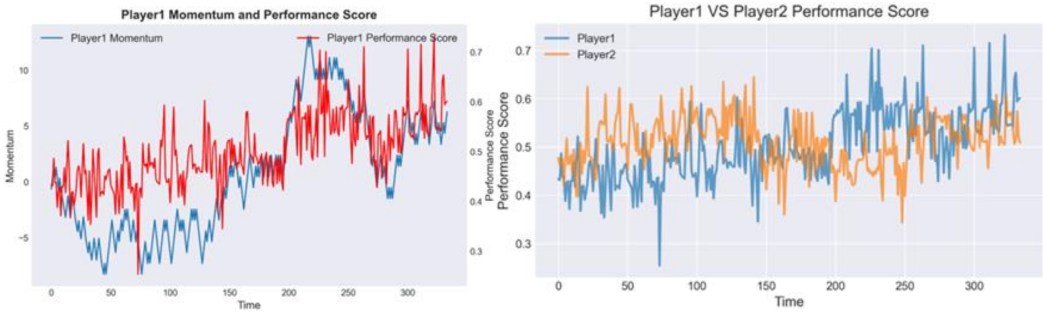

Figure 5. Carlos Alcaraz and Novak Djokovic performance score

To verify the model's applicability, we selected three matches with the most, median, and least amount of match time point data (match IDs ADFHR304, DMJL403, and ARND504). We used the model to predict the outcomes of these matches, as shown in Figure 6, which are plotted in the following figure. After comparing the scores with the actual match scores, we conclude that the TOPSIS model better reflects the relationship between player momentum and performance at the time of the match.

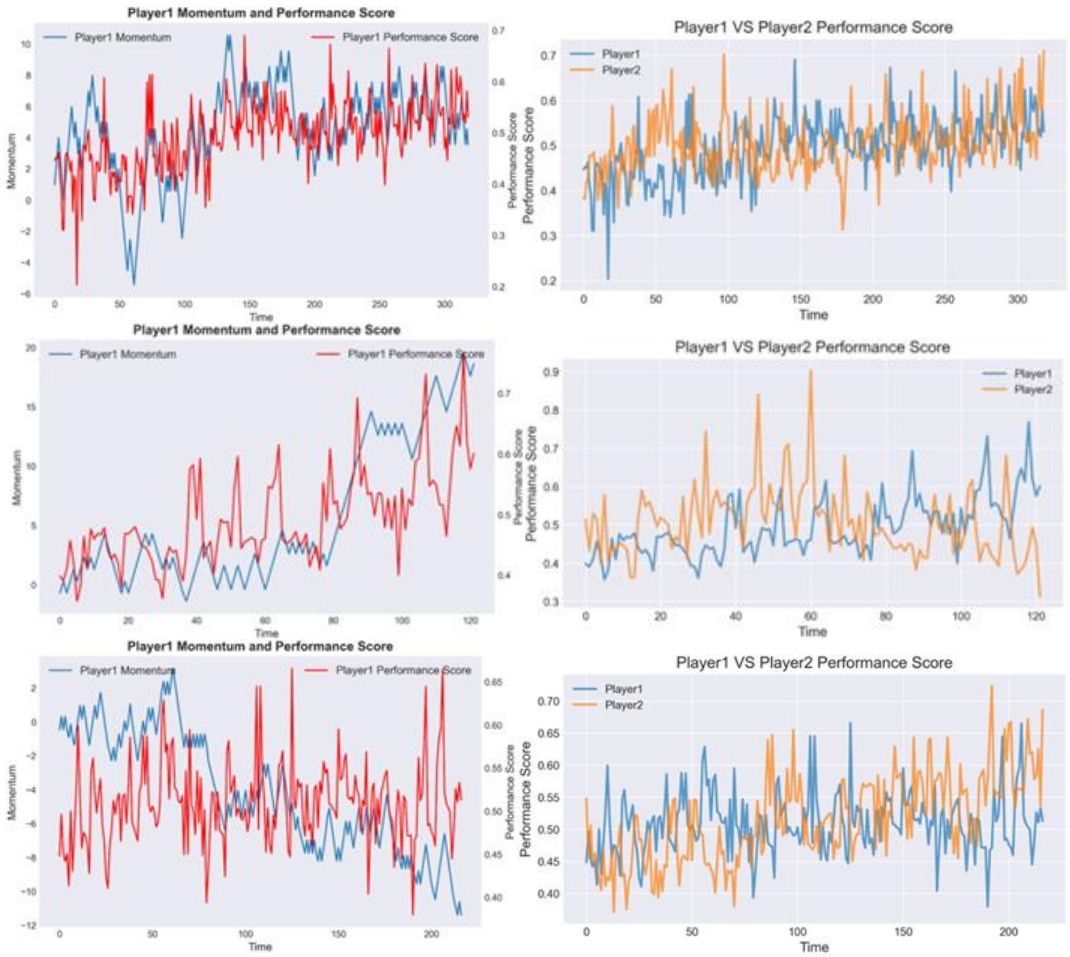

Figure 6. Predicted vs. actual scores for the three games

3. Conclusion

This study analyses athletes' performance in tennis matches using Principal Component Analysis (PCA) and TOPSIS models. Through the processing of match data and model validation, we successfully quantified the athletes' momentum and performance and proposed the relevant affecting momentum. The findings suggest that the TOPSIS model can accurately reflect the relationship between athletes' momentum and performance. This provides players with a reliable method for assessing and predicting match outcomes. The study is significant for improving the performance level of tennis players and provides a useful reference for analyzing other sports competitions using data science techniques.

References

[1] WANG Qiusi, GALIN GAHENNUR, DING Lan, et al. Research progress based on big data analysis of open source information of professional tennis [C] // Chinese Society of Sports Science. Abstracts Collection of the 13th National Sports Science Conference - Wall Paper Communication (Athletic Training Section) (II). Beijing Sport University; 2023: 3.

[2] C.Y. Huang. Research on tennis match performance prediction model based on BP neural network [D]. Fujian Normal University, 2022.

[3] Rong Zhang. Result prediction and player analysis of tennis matches [D]. Yunnan University, 2022.

[4] ZHOU Wu, ZHANG Chi, CAO Yingying. Analysis of the competitive performance of men's professional tennis singles players in the 2019 season [J]. Hubei Sports Science and Technology, 2021, 40 (06): 549-556.

[5] LIN Long, CHEN Tao, RONG Miaofeng. Comprehensive evaluation of kill quality of excellent female badminton singles players based on TOPSIS+RSR method [J]. Journal of Guangzhou Sports Institute, 2021, 41 (06): 105-110.

[6] Zhang Limin. Comprehensive evaluation of key points serve and receive performance of excellent male tennis singles players in the past ten years [D]. Guangzhou Sports Institute, 2021.

[7] Meier P, Flepp R, Ruedisser M, et al. Separating psychological momentum from strategic momentum: Evidence from men’s professional tennis [J]. Journal of Economic Psychology, 2020, 78: 102269.

[8] Moss B, O’Donoghue P. Momentum in US Open men’s singles tennis [J]. International Journal of Performance Analysis in Sport, 2015, 15 (3): 884–896.

[9] Richardson P A, Adler W, Hankes D.Game, Set, Match: Psychological Momentum in Tennis [J]. Sport Psychologist, 1988, 2 (1): 69-76.

[10] M. H D, Cornel N. Momentum in tennis: Controlling the match [J]. INTERNATIONAL JOURNAL OF SPORT PSYCHOLOGY, 2017, 48 (4): 459-471.

[11] Xie Roshi, Wang Sanbao. Training interpretation of winning factors in competitive tennis [J]. Journal of Guangzhou Sports Institute, 2021, 41 (06): 94-99.