

Quantitative Analysis and Prediction Methods for Sports Competition Results

Junyi Liu*, Xiaozhe Zhu#, Shiying Wang#

School of Information and Communication Engineering, Communication University of China,
Beijing, China, 100024

* Corresponding Author Email: L1968851627@163.com

#These authors contributed equally.

Abstract. In the era of big data, sports performance prediction becomes critical by analyzing accumulated competition data, especially the influence of momentum factors on competition results, although its definition and quantification are controversial in academia. This study explores the application of momentum in predicting the results of sports competitions and proposes a comprehensive research framework combining sports science and sports psychology. By collecting and analyzing the competition data, the feasibility of momentum as an important index in predicting the results of the competition is confirmed. The results show that momentum not only reflects the immediate performance of athletes in the game but also predicts their future results. This finding is of great significance for coaches and athletes to formulate strategies and adjust tactics in the game. Through continuous accumulation and analysis of data, we can more accurately predict the results of the game, and improve the level and fairness of sports competition.

Keywords: Sports psychology, Entropy Weight-Topsis, Momentum, PSO-BP neural network.

1. Introduction

In the context of today's big data era, the collection and analysis of sports data have become increasingly important ^[1]. With the continuous accumulation and in-depth excavation of sports competition data, the performance prediction of athletes has become the focus of researchers. Accurate prediction of sports performance not only helps coaches and athletes to better formulate training and competition strategies but also has important value and significance for sports scientific research. Among them, momentum is widely considered to be one of the key factors affecting the results of the competition ^[2]. However, although the influence of momentum is widely accepted in practice, the definition and quantification of the concept of momentum and its specific impact on athletes' performance and competition results are still controversial in academia.

In the era of big data, there have been studies on big data analysis of tennis. Koronas V evaluated tennis players' beliefs and positions through big data analysis in 2021 ^[3]. In the field of sports psychology, Scamardella F applied psychology to the field of sports, which proved the applicability of professional psychology in the field of sports ^[4]. Mahammed R F conducted a study on the confident behavior of tennis and table tennis players ^[5]. In this study, we used the entropy weight-TOPSIS method to evaluate the performance of the players and used the PSO-BP neural network to predict the results of the game. For sports physique evaluation, Zhuo C proposed a new multi-dimensional sports physique evaluation and ranking method based on TOPSIS theory ^[6]. Huang W applied the BP neural network to study the prediction of the results of table tennis technical and tactical diagnosis ^[7]. Zhao Z applied the PSO-BP neural network algorithm to airport passenger prediction classification based on risk ^[8]. Yan L used computer vision to realize tennis recognition in 2021, which proved the feasibility of the application of intelligent algorithms in the field of sports ^[9].

In this work, this study used the Entropy Weight-TOPSIS model, derived the comprehensive performance of each player at each scoring point, and visualized the game flow for comparison. To determine whether "momentum" plays a role in the match, we constructed a momentum model based on positive and negative indexes and conducted a Spearman correlation analysis on whether the point was scored or not. We selected features such as scoring rate, return score rate, and wonder shot for

quantification. Using the PSO-BP neural network model, we predicted the outcome of each set for the players.

2. Exploring a Dynamic Scoring System for Athlete Performance

2.1. Data Processing

Through observing the provided data of www.contest.comap.com, we identified some issues and took corresponding measures. Upon referencing the official match website, we discovered inaccuracies in the attached data. These errors include discrepancies in the number of sets played and missing information on running distances.

For instance, in the match labeled "2023-wimbledon-1403", only two sets were played, and further investigation revealed that it was due to a player withdrawal. Additionally, in some match instances, there was no record of the running distances for both players.

To address these issues, we conducted data cleaning to minimize errors and enhance the quality and accuracy of the data.

Next, we select the indexes that are conducive to the results of the game, such as whether to serve, continuous score, which are recorded as positive indexes; and unfavorable indexes of the results of the game, such as double fault, break_pt_missed, are recorded as negative indexes. The specific classification is shown in Table 1.

Table 1. Classification of positive and negative indexes

Positive	Negative
Break_pt_won	Break_pt_missed
Net_pt_won	Net_pt_missed
Server	Unf_err
Continuous score	Double_fault

Taking " 2023-Wimbledon-1301 " as an example, we use the indexes of player 1 as a benchmark. Recalculate the original data, the formula is as follows:

$$IV = IV_1 - IV_2 \tag{1}$$

Except for continuous scores, other indices have original values in the table. The calculation formula for continuous score is as follows:

$$CS = PW_n - PW_{n-1} \tag{2}$$

2.2. Solution

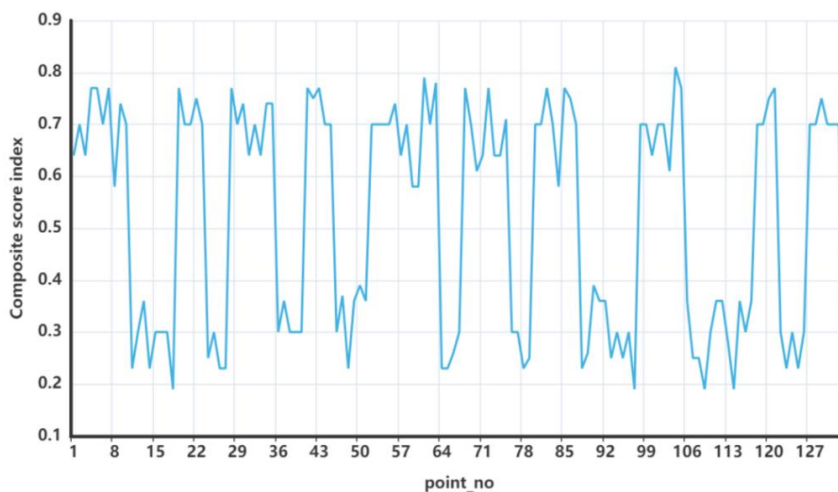


Figure 1. The fluctuation of the composite score index

As shown in Figure 1, the comprehensive score exhibits regular fluctuations over time. When the comprehensive score is greater than 0.5, it indicates that at that score point, P1 performs better, while if it is less than 0.5, P2 performs better. The specific degree is reflected in the numerical values.

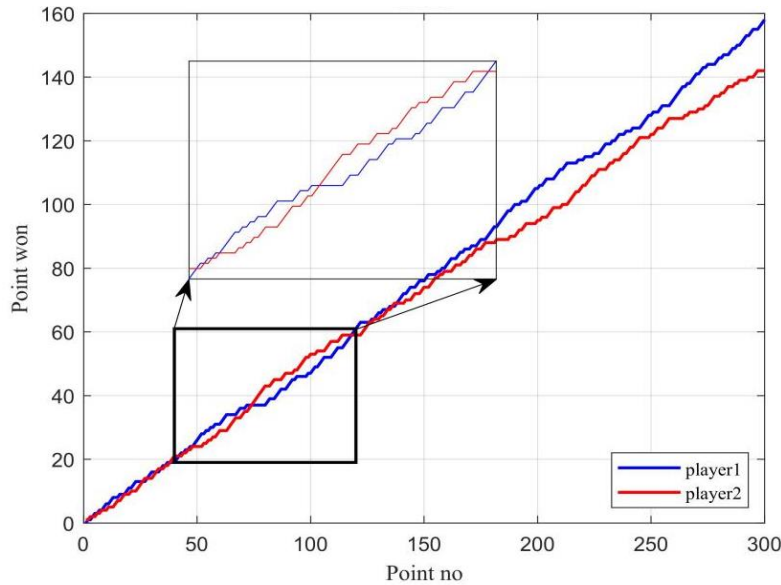


Figure 2. The fluctuation of points_won

In addition, we also conducted data visualization of the game scoring process. It can be seen from Figure 2 that the scores of the two sides in the first half of the game were very close. However, in the second half of the game, the score gap between the two sides gradually expanded, and one side began to occupy a clear advantage.

3. Exploring Momentum's Impact on Match Outcomes

3.1. Establishment of the Model

To address the momentum of players during the match, we propose the Momentum Model.

$$m = \prod_{i=1} P_i - \sum_{j=1} N_j \tag{3}$$

This model is primarily composed of positive indexes, negative indexes, and the momentum at the previous scoring point. The positive and negative indexes are manually selected, and their respective weights have been determined using the entropy weight- TOPSIS model used in the previous section. The formula is as follows:

$$P_i = \begin{cases} 1 & P_i = 0 \\ 1 \times (1 + W_i) & P_i = 1 \end{cases} \tag{4}$$

$$N_j = \begin{cases} 0 & N_j = 1 \\ 1 \times (1 + W_j) & N_j = 0 \end{cases} \tag{5}$$

At the same time, we get the weight of each index, as shown in Figure 3. It can be seen that the weight value of double_fault is the smallest, 0.184, and the weight value of break_pt_won is the largest, 31.515.

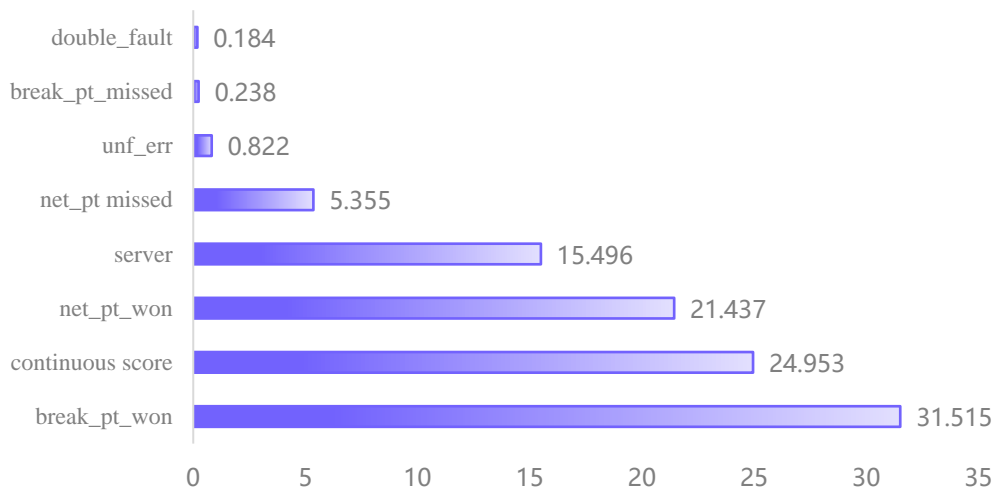


Figure 3. The weight of each index

3.2. Solution

Jaworski J et al. used Spearman correlation analysis to identify the causal relationship between the position in the sports ranking and the analytical variables of postural stability and proved the correlation [10].

After data processing, we used Spearman correlation analysis between the momentum model score and the "winner of the point". The result was a correlation coefficient of 0.884, indicating a strong correlation. This proves that "momentum" has a significant impact during the game.

4. Competition Trend Prediction and Momentum Analysis

4.1. Data Processing

To determine the change of flow direction in a game, it is necessary to predict the flow direction of the game according to the results of each set in the game and the performance of the players. We use the original data to calculate more performance of the players, such as the first-serve scoring rate, the second-serve scoring rate, and so on. The calculation process of some indicators is shown in Figure 4.

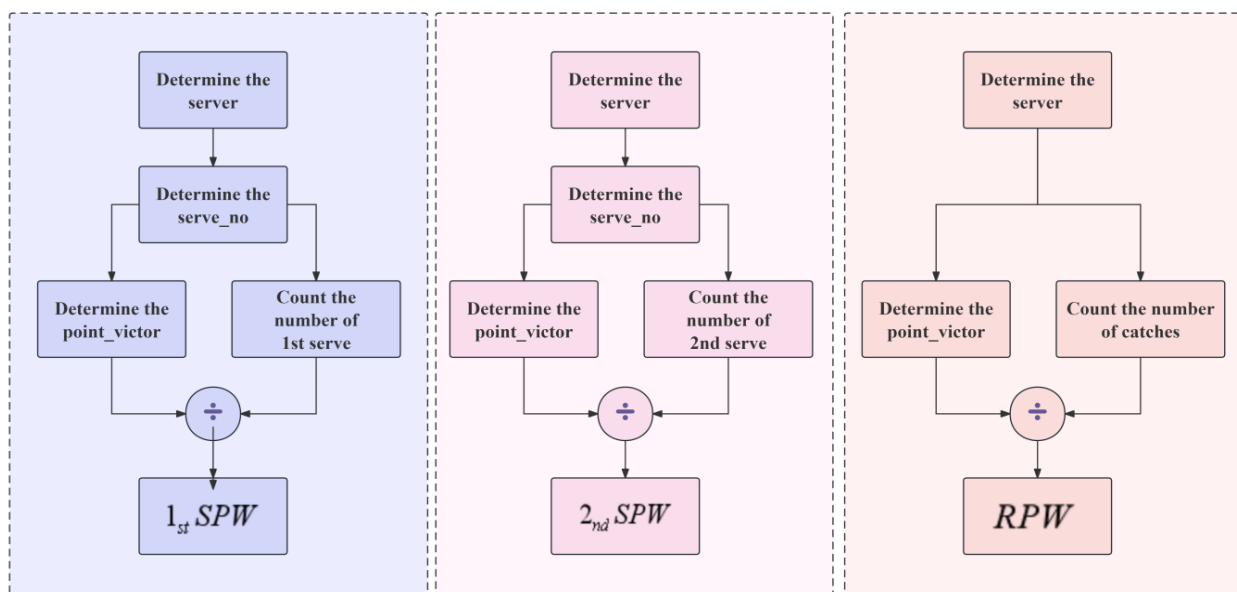


Figure 4. The calculation process of some indexes

For indicators that do not have a given value, such as serve depth, we simply score different situations, as shown in the Table 2, Table 3, and Table 4.

Table 2. The rating of serve_width

Serve_width	Score
<i>B</i>	1
<i>BC</i>	2
<i>BW</i>	3
<i>C</i>	4
<i>W</i>	5

Table 3. The rating of serve_depth

Serve_depth	Score
<i>NCTL</i>	1
<i>CTL</i>	3

Table 4. The rating of return_depth

Return_depth	Score
<i>ND</i>	1
<i>D</i>	3

Table 5 is part of the processed data of Carlos Alcaraz in "2023-wimbledon-1301". After the above data processing of the 31 games of this Wimbledon, the feature data are normalized by min-max. The normalized feature variables are used as the input layer of the BP neural network, and the results of each disc are used as the output layer. The victory is 1, and the failure is 0. The PSO-BP neural network is used for binary classification prediction. The data set is divided into training sets, verification sets, and test sets, and the ratio is 70: 20: 10.

Table 5. Processed data of Carlos Alcaraz's 2023-wimbledon-1301

Set	1stSPW	RPW	ACE	Winner	Serve depth	Serve width
1	0.786	0.348	3	9	60	115
2	0.778	0.277	4	10	63	136
3	0.857	0.452	4	11	50	94
4	0.708	0.438	1	11	65	119

4.2. Solution

In the construction of the model, the Trainlm function in the MATLAB toolbox is used to realize the binary classification prediction of each game result based on the PSO-BP neural network. We only need to set the parameters of the model to test the prediction effect of the model.

To predict the fluctuation of the game and judge which player can win the game, the input data of the two are the selected characteristics, and the output is the result of their respective wins and losses. For example, the input data of player 1 is the service scoring rate, receiving scoring rate, total scoring rate, wonder shot, physical fitness, and skills. If the output data is 1, the failure is 0.

To avoid overfitting, the sample data is randomly divided for training, and the ratio of training, testing, and verification data is 70: 20: 10. The learning rate is set to 0.01, the minimum initialization error is 0.00001, and the mean square error is used to measure the network performance. When the number of training times reaches 1000, the iteration stops.

4.2.1 Model Solving

The test of PSO-BP neural network prediction model. Figure 5 is the mean square error analysis of

the training set, the validation set, the test set, and the full set. The mean square error of the training set is 0.85762, and the mean square error of the test set is about 0.86251, which is maintained at a high value. It can be shown that our prediction results have a certain accuracy. However, because our model is a binary classification model, only the MSE of the original data is not comprehensive, so we chose the calculation accuracy to comprehensively evaluate the model.

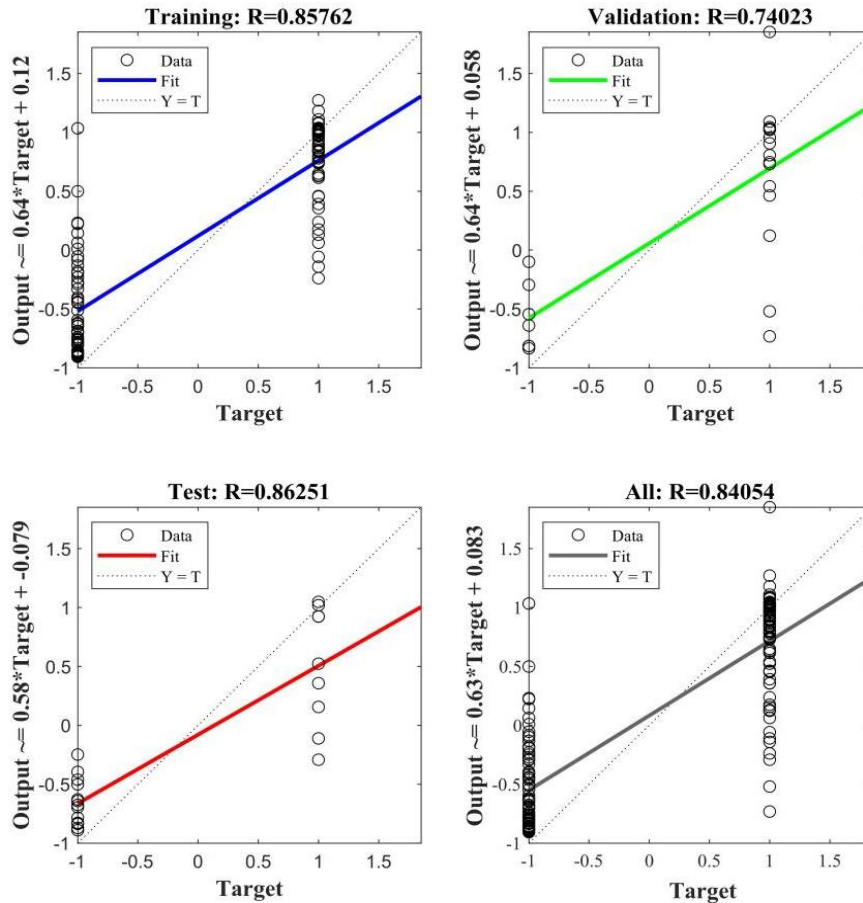


Figure 5. Root mean square error results

To quantify the prediction effect of the BP neural network, the original data is compared for visualization, and the results are shown in Figure 6. The accuracy rate, recall rate, and precision rate of the prediction results are calculated, and the accuracy rate is as high as 81.8%.

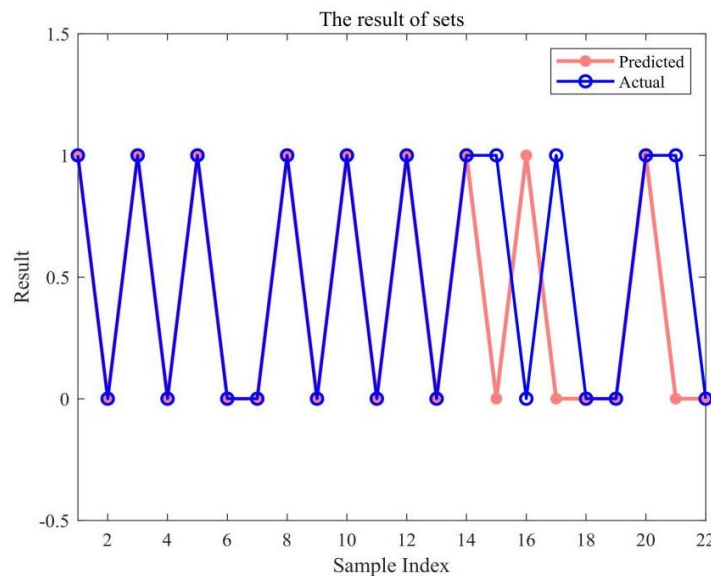


Figure 6. Comparison of the predicted results with the real results

In the classification problem, to avoid the interaction between the precision rate and the recall rate, we calculated the F1 value of about 0.8, indicating that the PSO-BP neural network model we established can identify the predicted correct samples as accurately as possible. The statistical results of the test set are shown in Table 6, the confusion matrix heat map of the predicted results is in Figure 7.

Table 6. Model evaluation results

	Accuracy	Recall	Precision	F1
Testing set	0.818	0.727	0.889	0.800

Note : $F1 = 2 * (precision * recall) / (precision + recall)$

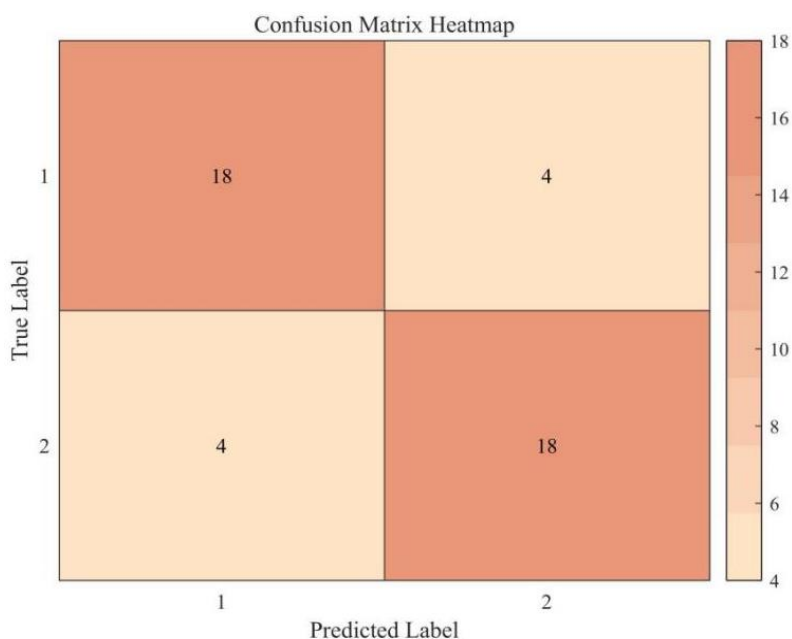


Figure 7. Confusion matrix heat map of prediction results

4.2.2 Correlation Analysis

According to the competition results predicted by the neural network, the accuracy rate reaches 81.8%, which shows that the change of the selected features has a certain influence on the fluctuation of the competition, which leads to the change of the competition results.

To find out which factors are most relevant to the fluctuation of the competition, some characteristics, and competition result data are selected to carry out correlation analysis and calculate the correlation coefficient. Some results are shown in Table 7.

Table 7. Correlation analysis of various factors and results

	Return depth	Serve depth	Winner	ACE	Distance run	Unf_err	Return of serve success rate	Total score rate	Result
Result	0.382 (0.000***)	-0.023 (0.733)	0.274 (0.000***)	0.199 (0.003***)	0.009 (0.891)	-0.224 (0.001***)	0.543 (0.000***)	0.721 (0.000***)	1 (0.000***)

Note : * * * represents the significance level of 1 % respectively

By observing the significant level of the correlation coefficient, it can be found that the total score rate and return of serve success rate are the most relevant to the results, and the correlation size also lays the foundation for the selection of player performance indexes in the future.

4.3. Advise

Each player will have their own relatively strong or weak aspects. According to the different fluctuations of the "momentum" in the historical competition, we can better give the athlete the

suggestion of the competition, so that he can play his advantage in the field. To compare the ability between athletes more clearly and intuitively, we cluster the same type of indexes, and finally summarize them into five characteristics: Ingenuity, Resilience, Efficiency, Skill and Stamina, and then score the players. The characteristics are shown in Figure 8.

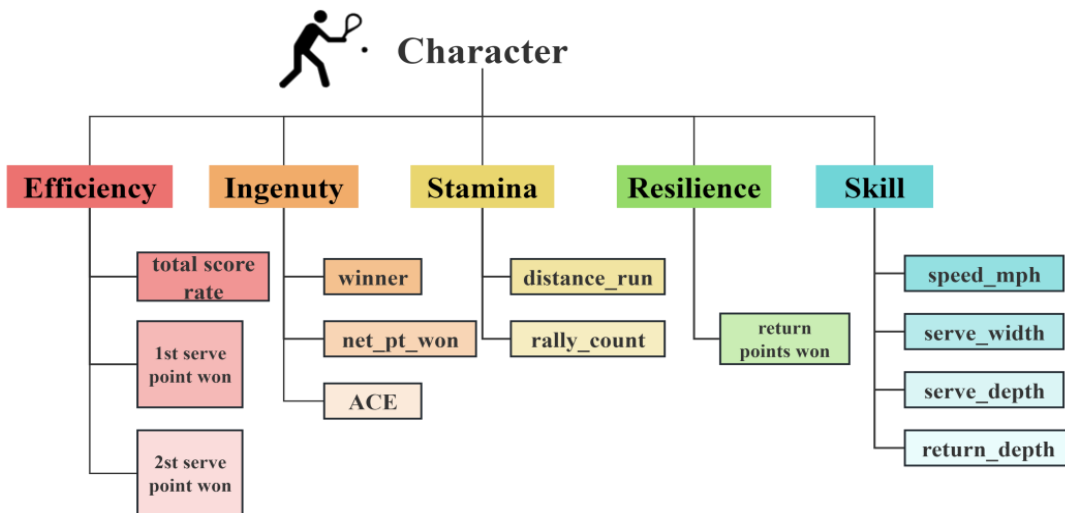


Figure 8. Composition of characteristics

Taking the top four players of Wimbledon in 2023 as an example, this study draws a characteristic radar map as shown in Figure 9, which can intuitively see the advantages and disadvantages of each player.



Figure 9. The top four players' comprehensive scoring radar

Before the game, the coach needs to guide the players to understand their technical advantages and disadvantages and clarify their shortcomings, to carry out targeted training and strengthening. Players' deficiencies may be reflected in strength, speed, accuracy, or endurance. Coaches need to develop a training plan suitable for players to improve their weak links. At the same time, given the advantages and disadvantages of the opponent, the coach needs to study and analyze with the players to develop a skilled preparation strategy. These strategies should include how to deal with the adversary's advantage attack, how to crack the opponent's tactical layout, etc. Through the formulation and

implementation of these strategies, players can better cope with the challenges of their opponents in the game and improve their chances of winning.

5. Conclusions

To determine the player's performance and level in the competition, this study used the entropy Weight-TOPSIS, the weight of each index, and the relative comprehensive score of the players obtained. The highest and lowest weight indexes are `break_pt_won` and `double_fault`, which are 31.515% and 0.184% respectively. The results show that the comprehensive score fluctuates regularly with time. To further explore the influence of "momentum" in competition, we construct a momentum model based on positive and negative indicators. Spearman correlation analysis results show that there is a significant correlation between momentum and scorer, and the correlation coefficient is as high as 0.884. At last, we used the Particle Swarm Optimization- -back propagation (PSO-BP) neural network to predict the results of the competition, and the accuracy of the model is 0.893, and the prediction accuracy is 81.8%. In addition, we performed correlation analysis on the factors and obtained that the total score rate and the return of serve success rate were most correlated with the results, and the correlation coefficients were 0.721 and 0.543, respectively.

This study reveals a research idea and framework applied to the field of sports science and sports psychology, which proves the feasibility that in the era of big data, we can quantitatively analyze the momentum to predict the results of the competition according to the performance of athletes in the competition.

References

- [1] L. Sheng, Application Status and Development Prospect of Big Data in Chinese Sports Field, 2019 International Conference on Communications, Information System and Computer Engineering (CISCE), Haikou, China, 2019, pp. 583-585.
- [2] 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.
- [3] Koronas V. EVALUATION OF AGONISTIC TENNIS IN GREECE: TENNIS PLAYERS' BELIEFS AND POSITION[J]. Human Sport Medicine. 2020, 20(3): 119-128.
- [4] Scamardella F, Casillo V,& Cusano P. Engagement and tennis: The applicability of occupational psychology to the world of sport[J]. Journal of Human Sport and Exercise. 2020, 15(2proc): S173-S176.
- [5] Mahammed R F, Kadhum S R. THE ASSERTIVE BEHAVIOR OF DISABLED TENNIS AND TABLE TENNIS PLAYERS.Revista iberoamericana de psicología del ejercicio y el deporte. 2023, 18(3): 248-250.
- [6] Zhuo C, Xie Q, Lin Y, et al. Multi-Dimensional Sport Physique Evaluation and Ranking for Healthcare Based on TOPSIS[J]. Journal of organizational and end user computing. 2021, 33(6): 1-9.
- [7] Huang W, Lu M, Zeng Y, et al. Technical and tactical diagnosis model of table tennis matches based on BP neural network[J]. BMC sports science, medicine & rehabilitation. 2021, 13(1): 54.
- [8] Zhao Z, Zhang C, Guo D. Analysis of Risk-Based Airport Passenger Classification with PSO-BP Neural Network[C]. Technical Committee on Control Theory, Chinese Association of Automation, 2020.
- [9] Yan L, Xin S. Design and Image Research of Tennis Line Examination Based on Machine Vision Analysis[J]. Computational Intelligence and Neuroscience. 2021, 2021: 1-11.
- [10] Jaworski J, Lech G, żak M, et al. Relationships between selected indices of postural stability and sports performance in elite badminton players: Pilot study[J]. Frontiers in Psychology. 2023, 14.