

Analysis of Wimbledon men's final based on entropy weight method and gray correlation degree

Wenyu Yan¹, Haotian Wu², Yi Wei², Zhongshuai Wang^{1,*}, Junhua Lv³

¹ School of Management Engineering, Qingdao University of Technology, Qingdao, China, 266520

² School of Information and Control Engineering, Qingdao University of Technology, Qingdao, China, 266520

³ School of Architecture and Construction, Qingdao city university, Qingdao, China, 266106

* Corresponding Author Email: 18622723289@163.com

Abstract. In the men's final of the 2023 Wimbledon Open, the match took a dramatic turn as Novak Djokovic had the upper hand for a while, but then Alcaraz took control. Such striking fluctuations are usually attributed to "momentum". This paper uses an evaluation model based on the entropy weighting method to analyze an athlete's performance over a specific time period of a match. The paper used the evaluation model to derive a composite score, which was found to correlate with the athlete's wins and losses in the match. The resulting changes in the composite score are consistent with the actual wins and losses, which indicates that the results of the model are more accurate. In addition, this paper also establishes a correlation analysis model based on gray correlation, and finds that the correlation between the athlete's "motivation" and the actual win/loss situation is as high as 0.789, which indicates that the athlete's motivation level is highly correlated with the win/loss situation in the current time period.

Keywords: Entropy Weighting, Evaluation Model, Gray Correlation, Momentum.

1. Introduction

In recent years, sports analysis has become more and more popular, and more and more research has been done in tennis, and many scholars have been using some tools and techniques to study all aspects of tennis players' performance in matches [1]. By examining different physical fitness factors, we try to find the key factors for athletes to achieve victory or defeat in a match, and provide scientific basis for sports training and competition strategy [2]. From the perspective of physical exertion characteristics of women's singles matches, the athletes of both sides are relatively close to each other in various data, which proves that women's singles matches are highly competitive today; from the correlation analysis of physical exertion indexes and total scores of the matches, scores at each stage are closely related to the athletes' physical fitness [3]. The literature suggests that there is a relationship between athlete training and competition loads, which provides some reference for our study^[4]. However, in the past, coaches mainly relied on athletes' feedback and their own intuition to assess athletes' physical fitness and endurance, which might be biased.

In fact, there is a deep mathematical mechanism behind these tennis matches. We wondered what mechanism causes the incredible fluctuations in seemingly dominant players, which we usually attribute to "momentum". (Data from Wimbledon_featured_matches. csv-data set of Wimbledon 2023 Gentlemen's singles matches after second round.)

To help coaches recognize when a match will shift from one player to another through metrics such as momentum. The paper analyzed the literature [5] and wanted to build a predictive model using data from all men's matches after the first two matches of the 2023 Wimbledon Open and extend it to other sports.

2. Evaluation Model Based on Entropy Weight

2.1. Define the evaluation factors of players

Considering all factors, the paper chooses the factor x_i as an indicator to evaluate the performance of a player in a certain period of time (5-minute intervals are chosen as the time unit), and the definition of the factor is shown in Table 1:

Table 1. Definition of x_i .

The values of i	Description
1	Scoring efficiency per unit time
2	Server-to-winner ratio per unit time
3	Ace percentage per unit time
4	Untouchable winning shot ratio per unit time.
5	Missed serve and lost point ratio per unit time
6	Unforced error rate per unit time
7	The ratio of players makes it to the net per unit of time
8	Ratio of net reach to player's score per unit time in the current game
9	Winning ratio when the opponent serves
10	Average serving speed of the player

2.2. Data analysis

As an example, the data 2023-wimbledon-1301 was selected, player1 was Carlos Alcaraz, and player2 was Nicolas Jarry. Evaluation metrics for that time period were selected for processing at five-minute intervals from the start of the match.

The score gain ratio per unit time can be expressed as:

$$x_1 = \frac{\text{scores}}{\text{games}} \quad (1)$$

2.3. Construction of evaluation model

Entropy weight method is an objective assignment method, according to the variation of each indicator degree, the entropy weight of each indicator is calculated, and then the entropy weight of each indicator is weights, and then correct the weights of the indicators through the entropy weights, so as to arrive at a more objective weight of the indicators [6].

In addition, the entropy power method reduces the experts' preference for attributes, and at the same time strives to reduce the subjective arbitrariness of empowerment, so that the evaluation results avoid subjectivity and limitations, and achieve the effect that the results are based on evidence [7].

(1) Indicator normalization can be written as the following formula:

$$x'_{ij} = \frac{|X_{ij} - \min(X_{1j}, X_{2j}, \dots, X_{nj})|}{\max(X_{1j}, X_{2j}, \dots, X_{nj}) - \min(X_{1j}, X_{2j}, \dots, X_{nj})} \quad (2)$$

(2) There are n objects to be evaluated i.e. momentum y in various time intervals and m evaluation indicators i.e. x_j ($j=1, 2, 3, \dots, 10$) Calculate the weight of the i th object under the j th indicator and consider it as the probability used in the relative entropy calculation.

$$y_{ij} = \frac{x'_{ij}}{\sum_{i=1}^m x'_{ij}} \quad (i = 1, 2, \dots, n; \quad j = 1, 2, \dots, m) \quad (3)$$

(3) Where e_j is the information entropy, j refers to the j th indicator; w_j is the weight of the j th indicator. And the weight is an objective representation of the importance of the indicator so that we can evaluate the degree of player performance.

$$\begin{cases} e_j = -\frac{1}{\ln m} \sum_{i=1}^m y_{ij} \ln y_{ij} \\ w_j = \frac{1 - e_j}{m - \sum_{j=1}^m e_j} \\ s_i = \sum_{j=1}^n y_{ij} w_{ij} \end{cases} \quad (4)$$

Finally, the paper calculates the comprehensive score s_i of the i th player, and the comprehensive score is the indicator for evaluating the degree of player performance. The results of the weights of the relevant indicators are organized in Table 2.

Table 2. Results of entropy weight method.

index of correlation	Income weight
x_1	0.011644
x_2	0.031414
x_3	0.162098
x_4	0.056018
x_5	0.225218
x_6	0.061388
x_7	0.079291
x_8	0.093800
x_9	0.259166
x_{10}	0.019962

2.4. Analysis of model results

The paper apply python to the solution, which allows us to calculate the weights of the indicators x_i needed for the entropy weighting method and to calculate the composite score of each player; the higher the player's composite score, the better the performance.

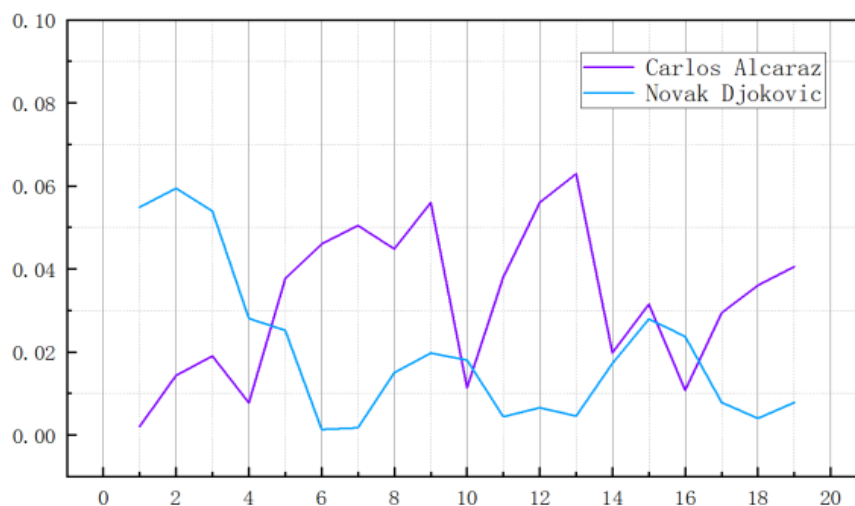


Fig 1. Combined scores of the two players.

Taking the 2023 Wimbledon men's final between Carlos Alcaraz and Nicolas Jarry as an example, the comprehensive scoring curve drawn is shown in fig.1.

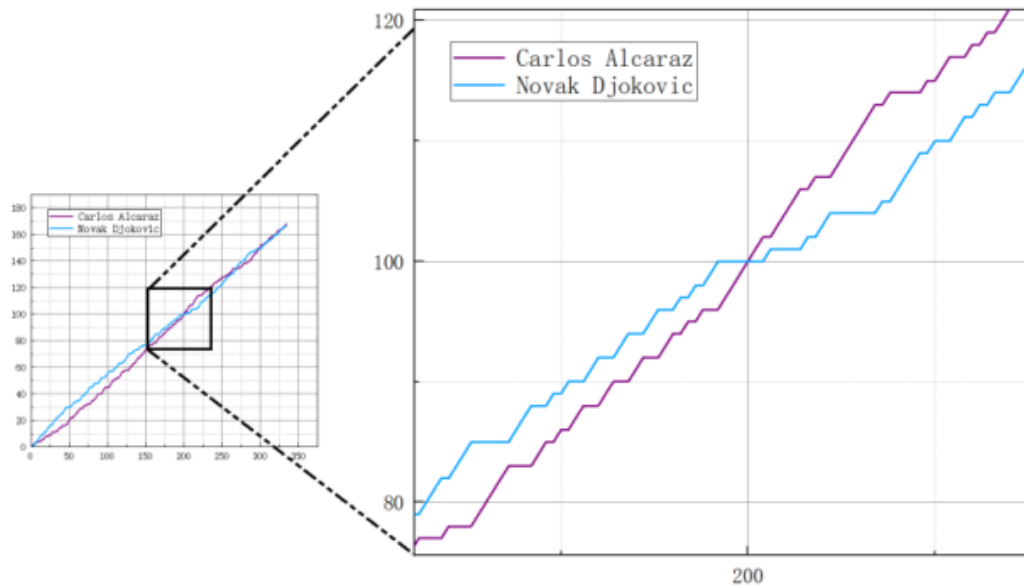


Fig 2. Scoring of two players.

Analyzing the combined score graph, it can be seen that Carlos Alcaraz's combined score is higher than Novak Djokovic's in the time period between 5-15 (1:00:04-3:45:09), which shows that the player is gradually catching up with the score.

From the figure of the scoring situation of the match, it can be seen that before the 200th round (point), the increase rate of Carlos Alcaraz's score is higher than that of Novak Djokovic, which coincides with the fact that the player's combined score is higher than that of his opponent's at this time, and in the 200th round (point), which is located in the turning point of the score catching up, Carlos Alcaraz succeeded in catching up with the score. Score, at which point the time interval was 2:54:24.

From the comparison analysis of the two graphs, it can be seen that the performance of the players coincides with the overall score, and the evaluation model is more reliable.

Fig.2 gives a clear picture of the performance of the two players in a given time period. The model is more reliable as it agrees with the performance of Carlos Alcaraz and Novak Djokovic given in the title.

In addition, the model can further reflect the magnitude of the player performance advantage and game trends.

We do this by considering z_i , the percentage of combined player scores. This is used as a metric to assess performance advantage and match trend, defined as:

$$Z_i = \frac{s_i}{\sum_{j=1}^m s_j} (i = 1, 2; m = 2) \quad (5)$$

The plotted percentage of combined scores is shown in fig.3. Analyzing fig.3, it can be concluded that when a player is more dominant, the percentage of the corresponding area is greater than 50%. The larger the percentage accounted for, the greater the advantage of that player.

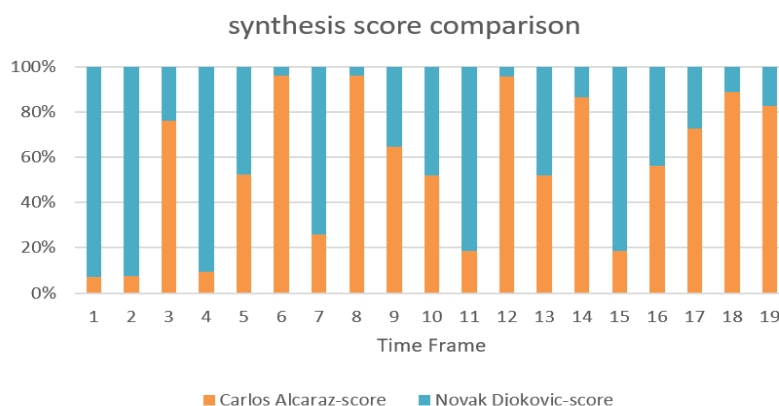


Fig 3. Combined score ratio.

3. Correlation Analysis Model Based on Grey Relational Analysis Method

3.1. Correlation analysis modeling

In order to further evaluate the player's performance, we chose another game and calculated the player's score for each time period of the game according to the evaluation model. The greater the score, the greater the “momentum” and the greater the probability of winning.

By calculating the percentage of games won by each athlete in each time period. The degree of correlation between momentum scores and game results is assessed through statistical analysis, and determining the spearman correlation coefficient between momentum scores and game results can be transformed into calculating the spearman correlation coefficient between the ratio scores between p1-score/p2-score and the probability of winning the game, which in turn determines whether there is a correlation between momentum and winning and the degree of that correlation.

Gray correlation analysis (GRA) is a quantitative comparative analysis method used to determine the magnitude of the correlation between the target variable and the input variables to determine the model's input variables [8]. It is a method for estimating and evaluating in advance the combined effects of different factors on the indicators of a system [9]. The gray correlation optimal reference sequence and the worst rate and peak shape are better, so the above method is chosen [10].

3.2. Analysis of Model Results

First, the p1-score/p2-score and the probability of winning the game are used as the comparison and reference sequences, respectively, and are dimensionless, and then the gray correlation coefficient is calculated with the help of python according to equation (6), where ρ is taken as 0.5, the results of the gray correlation analysis are plotted as shown in fig.4:

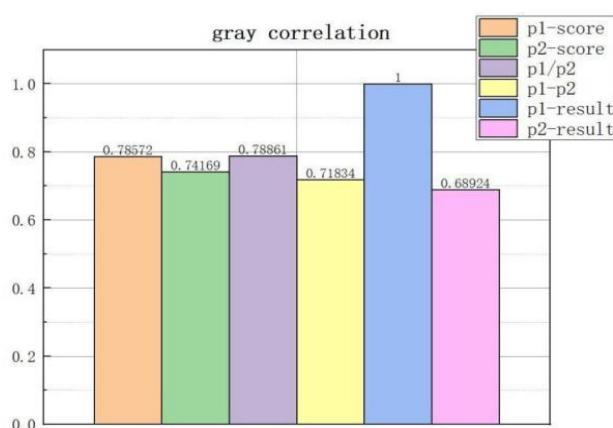


Fig 4. Gray correlation coefficient.

The closer the gray correlation coefficient is to 1, the stronger the correlation is, as seen in the above figure, the correlation coefficients of the gray correlation coefficients between p1-p2 and p1-result and p1-score/p2-score are 0.78861 and 0.71834, respectively, which proves that there is a strong correlation between the two and p1-result. Therefore, the statement that “swings in play and runs of success by one player are random” is ill-considered.

$$\xi(X_i) = (\Delta_{min} + \rho\Delta_{max}) / (\Delta O_{i(k)} + \rho\Delta_{max}) \quad (6)$$

4. Result

By establishing a correlation analysis model based on gray correlation, this paper found that the correlation between the athlete's "motivation" and the actual win/loss situation is close to 1, indicating that the athlete's motivation level in the current time period is highly correlated with the win/loss situation. And the following three recommendations were made.

When serving, you can try to stay away from the line as far as possible, the farther away from the net, the greater the range of control, when the opponent player counterattacks, you can leave yourself room to hit back. When receiving the ball, you can get closer to the net, so you can move in and out freely and hit the ball relatively well.

Try to hit topspin during your game, this is because topspin increases the arc of the ball and therefore the range. This increases the distance you can run around the court and the topspin is more challenging and difficult for your opponent to deal with.

Focus on serving, in the first serve you can choose according to the opponent's state, while the second serve you should choose your own most familiar way to serve to improve their own fault tolerance rate.

The plotted percentage of combined scores is shown in fig.4. It can be concluded that when a player is more dominant, the percentage of the corresponding area is greater than 50%. The larger the percentage accounted for, the greater the advantage of that player.

5. Conclusions

In this paper, the evaluation model based on entropy weighting method is used to find the weight of each index and calculate the athlete's comprehensive score in a specific time period to evaluate the athlete's sports performance. By analyzing the changes in the comprehensive score of the athlete in the game, it can be found to be in line with the actual victory and defeat of the athlete.

This paper establishes a gray correlation analysis model, and finds that the gray correlation between the athlete's "momentum" and the actual win/loss situation is as high as 0.789, which proves that the win/loss situation is highly correlated with the current time period's "momentum" level. The grey correlation between "momentum" and the actual winning and losing situation is as high as 0.789, which proves that the winning and losing situation is highly related to the level of "momentum" in the current time period.

In this paper, entropy weight method is used because it is an objective weighting method, which does not need to determine the subjective weight of each criterion in advance. After determining the evaluation factors of the players, it can utilize the concept of information entropy to calculate the weights automatically, thus reducing the burden of subjective judgment. In addition, the entropy weighting method can more comprehensively consider the amount of information contribution of each evaluation factor to the degree of player performance, which helps to improve the comprehensiveness and accuracy of the model.

When the sample is small, the accuracy of the Spearman coefficient method decreases sharply and is easily affected by extreme values, while the gray correlation analysis method has relatively low data requirements and is highly adaptable. The dataset used to build the model in this paper is

"Wimbledon_featured_match", which has a moderate sample size and is a more robust choice for gray correlation analysis.

References

- [1] Bai ZT. Analysis of the indicators related to athletes' physical exertion and match winning and losing in men's singles tennis matches [J]. Sports Science and Technology Literature Bulletin, 2022, 30(06):61-65.
- [2] Gao Jianjie. Correlation analysis of athletes' physical exertion and game winning and losing [J]. New Sports, 2024(08):89-91.
- [3] SHEN Jingyao, ZHU Yuanweng, MEN Fangyu. Characteristics of athletes' physical exertion in women's singles tennis Grand Slam matches and the correlation analysis with match wins and losses [J]. Hubei Sports Science and Technology, 2023, 42(02):184-188.
- [4] LIU Chang, HOU Huisheng, SHEN Dahai, et al. An empirical study on the training and competition loads of high-level tennis players [J]. Journal of Tianjin Sports Institute, 2024, 39(02):241-248.
- [5] DONG Yu-Cheng, MENG Yun-Chen. Analysis of the current status and hotspots of tennis training research from 2000 to 2023 based on Cite space and VOSviewer visualization[C]//China Bandy Association, Macau Physical Fitness Association, and Guangdong Physical Fitness Association. Proceedings of the 11th Chinese Conference on Physical Training Science (Next). [Publisher unknown], 2024:9.
- [6] Zhao Dongcun. Research on the application of entropy weight method in enterprise job evaluation [J]. China Market, 2024(11):93-96.
- [7] Lin Xiaoyun. Evaluation of river and lake hydrological connectivity based on entropy weight TOPSIS combination model [J]. Water Resources Technical Supervision, 2024(03):98-101.
- [8] QIU Guoxing, CAI Mingchong, ZHANG Yi, et al. Prediction of silicon content in blast furnace molten iron based on gray correlation analysis and machine learning [J/OL]. Materials Herald: 1-11[2024-04-24].
- [9] Deng Changcheng, Cheng Xueli, Li Hui, et al. Analysis of bearing temperature influencing factors based on gray correlation[J]. Bearing, 2024(03):52-57.
- [10] YANG Xiaodong, Deng Chengcheng, YANG Wenhui, et al. A study on UPLC feature mapping and quality of different commodity specification grades of Baihe and Baihe based on principal component analysis and gray correlation degree method [J]. Chinese Medicine Herald, 2024, 30(02):37-41.