

A Methodological Research on Evaluating Competitive Capability and Growth Rate of Basketball Players

Jingyi Zhang¹, Sirui Wu², Mingxian Zhou³, Xinghao Liang⁴, Keqi Wang⁵, Qiulin Wu⁶, Chenhan He⁷

¹Huili School Shanghai, Shanghai 200120, China

²WLSA Shanghai Academy, Shanghai 200243, China

³Shanghai Southwest Weiyu Middle School, Shanghai 200030, China

⁴North Cross School Shanghai, Shanghai 200040, China

⁵Shanghai Pinghe Bilingual School, Shanghai 200120, China

⁶Darlington School Rome, Georgia, USA 30161, China

⁷Minhang Crosspoint Academy at Shanghai Wenqi Middle School, Shanghai 200120, China

Abstract: This paper aims to conduct a quantitative analysis of the factors influencing the growth rate of basketball players. To achieve this goal, the paper adopts a method of entropy weight combined with ideal solution approximation comprehensive evaluation method, to analyze the growth rate of the athletes. Furthermore, to explore the relationship between various factors and the players' growth rate, a backpropagation (BP) neural network is utilized to fit the impact level of each independent variable and the dependent variable. Therefore, the method proposed in this paper can provide a more accurate, practical, and flexible evaluation method for ranking basketball players, thus supporting training, and offering reference for other similar sports competitions.

Keywords: Basketball Players; Ideal Solution Approximation Method; Entropy Weight Method; Backpropagation Neural Network.

1. Introduction.

With the rapid progress of the world basketball competition level, the game's confrontation is more and more intense, which puts forward more stringent requirements on the athletes' competitive ability. In a relatively short career, in order to achieve better game results, for team managers, team coaches, players themselves and from other perspectives, how to systematically cultivate and train the players through a systematic program to promote the growth of players, and then fully enhance the growth rate is of great significance. Therefore, it is necessary to establish a scientific and reasonable evaluation index system of basketball players' athletic ability and analyze the factors that have an impact on the growth rate, and then train them in a targeted way.

In recent years, many advanced evaluation schemes of basketball players' competitive ability have been proposed. S. Trninic studied the positional resistance of players on the court and proposed a performance scheme based on positional resistance; Lin Li constructed a hybrid evaluation model based on the analytic hierarchy process (AHP) and intelligent fuzzy comprehensive evaluation to analyze the player's ability on the court; and the most classic is the extended application of Delphi technique proposed by Claire M. Goodman in NBA in 1987. In addition, some statistical analysis and intelligent methods based on hierarchical analysis, principal component analysis, gray correlation method and their fusion methods have also been proposed in recent years.

2. Data source and data cleaning

The data for this study comes from basketball-reference players' performance in every season from 1970 to 2015, including but not limited to:

Due to the absence of some data from data sources before

1980, the data from 1970 to 1979 are only used as reference in data processing and model construction. In addition, the data source also provides detailed data of each match, which can be further analyzed in subsequent studies.

Due to the difference in the scope, scale and impact of the data on the evaluation of athletes, the data need to be preliminarily cleaned, including data forward processing, and data standardization processing.

For aggregate data sources, indicators can be divided into the following types:

- 1) Defensive metrics: front court rebounds, back court rebounds, steals, blocks
- 2) Offensive metrics: points, assists, attempts and percentage, 3-point attempts and percentage, free throw attempts and percentage
- 3) Turnovers: turnovers, personal fouls
- 4) Other data: age, minutes played

Among them, offensive metrics and defensive metrics are extremely large indicators. Turnover fouls are extremely small indicators, and age is an interval type indicator in other data. The longer the playing time is, the more important this member is, so the playing time is regarded as an extremely large indicator. According to the following formula, each indicator is positively normalized and standardized:

Very large indicators:

$$x = \frac{x - x_{\min}}{x_{\max} - x_{\min}}$$

Extremely small indicators:

$$x = \frac{x_{\min} - x}{x_{\max} - x_{\min}}$$

Intermediate indicators: $M = \max(28 - x_{\min}, x_{\max} - 30)$

$$x = \begin{cases} 1 - \frac{28 - x_{\min}}{M}, & x < 28 \\ 1 - \frac{x_{\max} - 30}{M}, & x > 30 \\ 1, & 28 \leq x \leq 30 \end{cases}$$

From this, we get the scoring matrix: Z

Table 1. Data source fields

Abbreviations	Full name	Abbreviations	Full name
Season	Season	2PA	
Age	Player's age on February 1 of the season	2P%	
Tm	team_id	eFG%	Effective Field Goal Percentage This statistic adjusts for the fact that a 3-point field goal is worth one more point than a 2-point field goal.
Lg	league	FT	Free Throws Per Game
Pos	pos	FTA	Free Throw Attempts Per Game
G	Games	FT%	
GS	Games Started	ORB	Offensive Rebounds Per Game
MP	Minutes Played Per Game	DRB	Defensive Rebounds Per Game
FG	Field Goals Per Game	TRB	Total Rebounds Per Game
FGA	Field Goal Attempts Per Game	AST	Assists Per Game
FG%		STL	Steals Per Game
3P	3-Point	BLK	Blocks Per Game
3PA		TOV	Turnovers Per Game
3P%		PF	Personal Fouls Per Game
2P	2-Point	PTS	Points Per Game

Table 2. Examples of some data sources

Rk	Player	Pos	Age	Tm	G	GS	MP	FG	FGA	FG%
0	Shareef Abdur-Rahim	PF	29.0	SAC	72.0	30.0	27.2	4.6	8.8	0.525
1	Alex Acker	SG	23.0	DET	5.0	0.0	7.0	0.8	3.2	0.250
2	Malik Allen	PF	27.0	CHI	54.0	20.0	13.0	2.2	4.6	0.490
3	Ray Allen	SG	30.0	SEA	78.0	78.0	38.7	8.7	19.2	0.454
4	Tony Allen	PG	24.0	BOS	51.0	9.0	19.2	2.5	5.4	0.471

Table 3. Scoring Matrix Z: Sample of raw data after normalization and normalization

	Player	Pos	Age	Tm	G	GS	MP	FG	FG%
0	Shareef Abdur-Rahim	PF	0.9	SAC	0.877	0.366	0.631	0.377	0.525
1	Alex Acker	SG	0.5	DET	0.049	0.000	0.162	0.066	0.250
2	Malik Allen	PF	0.9	CHI	0.654	0.244	0.302	0.180	0.490
3	Ray Allen	SG	1.0	SEA	0.951	0.951	0.898	0.713	0.454
4	Tony Allen	PG	0.6	BOS	0.617	0.110	0.445	0.205	0.471

3. An Integrated Evaluation Method Combining TOPSIS with Entropy Weight Method

The TOPSIS method (Technique for Order Preference by Similarity to Ideal Solution), which can be translated as the approximate ideal solution sorting method, is often referred to as the good and bad solution distance method across the country.

After all the solutions are presented, an ideal best solution and worst solution can be constructed based on the data of these solutions. The idea of TOPSIS is to evaluate the comprehensive distance of any solution in the solution system from the ideal best solution and the worst solution through

$$z^+ = [z_1^+, z_2^+, \dots, z_m^+] = [\max\{z_{11}, z_{21}, \dots, z_{n1}\}, \max\{z_{12}, z_{22}, \dots, z_{n2}\}, \dots, \max\{z_{1m}, z_{2m}, \dots, z_{nm}\}]$$

Similarly, we calculate the ideal worst solution vector by taking the smallest number in each column:

$$z^- = [z_1^-, z_2^-, \dots, z_m^-] = [\min\{z_{11}, z_{21}, \dots, z_{n1}\}, \min\{z_{12}, z_{22}, \dots, z_{n2}\}, \dots, \min\{z_{1m}, z_{2m}, \dots, z_{nm}\}]$$

Here, z^+ is z_{\max} and z^- is z_{\min} .

Table 4. The best and worst solutions obtained by the TOPSIS method

GS_Dmax	GS_Dmin	MP_Dmax	MP_Dmin	ORB_Dmax	ORB_Dmin	FT%_Dmax	FT%_Dmin	DRB_Dmax	DRB_Dmin
0.634	0.366	0.369	0.631	0.595	0.405	0.216	0.784	0.646	0.354
1.000	0.000	0.838	0.162	0.946	0.054	0.287	0.713	0.919	0.081
0.756	0.244	0.698	0.302	0.784	0.216	0.395	0.605	0.818	0.182
0.049	0.951	0.102	0.898	0.757	0.243	0.097	0.903	0.667	0.333
0.890	0.110	0.555	0.445	0.838	0.162	0.254	0.746	0.848	0.152
0.232	0.768	0.104	0.896	0.838	0.162	0.308	0.692	0.646	0.354
0.976	0.024	0.587	0.413	0.486	0.514	0.524	0.476	0.707	0.293
0.915	0.085	0.636	0.364	0.838	0.162	0.195	0.805	0.869	0.131
0.866	0.134	0.436	0.564	0.865	0.135	0.162	0.838	0.717	0.283
0.988	0.012	0.691	0.309	0.892	0.108	0.278	0.722	0.869	0.131

Next, we calculate the distance of each player to the best and worst solutions:

$$d_i^+ = \sqrt{\sum_{j=1}^m (z_j^+ - z_{ij})^2}$$

$$d_i^- = \sqrt{\sum_{j=1}^m (z_j^- - z_{ij})^2}$$

$$S_i = \frac{d_i^-}{d_i^+ + d_i^-}$$

Where d_i^+ is the distance of the i player to the best solution, d_i^- is the distance of the i player from the worst solution, and S_i is the score of the i player.

The TOPSIS method views each factor as having the same weight, so the entropy weight method is introduced to assign weights to each factor. In information Theory, entropy is a measure of uncertainty. The larger the amount of information

certain calculations. If a solution is closer to the ideal best solution and farther from the worst solution, we have reason to believe that this solution is better. The ideal best solution is that each indicator value of this ideal solution takes the best value of the evaluation indicator in the system, and the worst solution is that each indicator value of this ideal worst solution takes the worst value of the evaluation indicator in the system.

The score matrix Z , which has been processed by normalization and standardization, contains only maximum type data. We can extract the ideal best solution and the worst solution from it.

Therefore, we take out the largest number in each indicator, that is, each column, to form the ideal best solution vector, namely:

is, the smaller the uncertainty and the entropy will be; The smaller the amount of information is, the greater the uncertainty and the larger the entropy will be. According to the characteristics of entropy, we can judge the randomness and disorder degree of an event by calculating the entropy value. We can also use the entropy value to judge the dispersion degree of an indicator. The greater the dispersion degree of the indicator is, the greater the impact of the indicator on the comprehensive evaluation will be.

According to the definition of information entropy, for a certain indicator, we can use its entropy value to judge its dispersion degree. The smaller the entropy value is, the greater the dispersion degree of the indicator and the impact (i.e., the weight) of the indicator on the comprehensive evaluation will be. If the values of an indicator are all equal, then this indicator does not play a role in the comprehensive evaluation.

$$f = \ln\left(\frac{1}{P}\right) = -\ln P$$

f is the amount of information, and P is the probability of an indicator occurring?

Information entropy is the expectation of information. Information entropy can be understood as the size of uncertainty. The greater the uncertainty is, the greater the information entropy will be.

After normalizing and standardizing the data source, calculate the proportion of the j indicator in the i solution:

$$y_{ij} = \frac{x_{ij}}{\sum_{i=1}^n x_{ij}}, i = 1, \dots, n, j = 1, \dots, m$$

Indicators	Weights
GS	2.75%
MP	5.31%
ORB	5.00%
FT%	4.17%
DRB	5.44%
FG	5.16%
2P	5.07%
PTS	5.05%
3P	2.84%
2P%	6.47%
FT	4.35%
AST	4.10%
FG%	6.54%
Age	3.96%
BLK	3.33%
3P%	5.62%
PF	5.59%
eFG%	6.42%
G	3.44%
STL	5.22%
TOV	4.16%

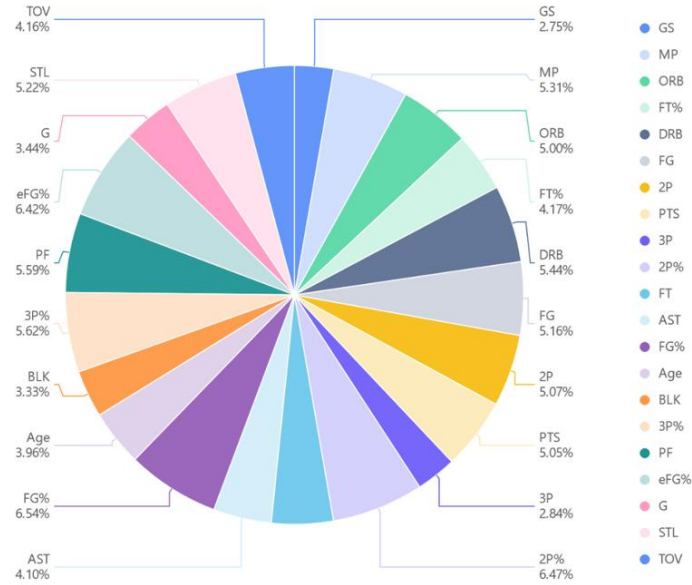


Figure 1. Weights for each indicator

Table 5. Evaluation of athletes' competitive ability (Part)

	Player	Pos	Dmax	Dmin	score
0	Shareef Abdur-Rahim	PF	0.563966	0.436034	0.436034
1	Alex Acker	SG	0.774534	0.225466	0.225466
2	Malik Allen	PF	0.614888	0.385112	0.385112
3	Ray Allen	SG	0.432573	0.567427	0.567427
4	Tony Allen	PG	0.647440	0.352560	0.352560
5	Rafer Alston	PG	0.555950	0.444050	0.444050
6	Chris Andersen	C	0.637694	0.362306	0.362306
7	Alan Anderson	SF	0.672687	0.327313	0.327313
8	Derek Anderson	SG	0.643668	0.356332	0.356332
9	Shandon Anderson	SF	0.691031	0.308969	0.308969

Based on this, we obtain the weights of each indicator in the TOPSIS method, and the above formula can be rewritten as:

$$d_i^+ = \sqrt{\sum_{j=1}^m w_j (z_j^+ - z_{ij})^2}$$

$$d_i^- = \sqrt{\sum_{j=1}^m w_j (z_j^- - z_{ij})^2}$$

$$S_i = \frac{d_i^-}{d_i^+ + d_i^-}$$

Then the competitive ability of each athlete can be evaluated.

4. Analysis of Athletic Ability Enhancement Factors Based on BP Neural Networks

For complex multivariate problems, the logic of the influence of each independent variable on the dependent variable is difficult to be accurately modeled by simple mathematical methods, but many mathematical problems in engineering do not need to obtain an accurate solution, but only need to obtain an approximate solution or even a qualitative law. Therefore, BP neural network with its powerful nonlinear fitting ability can provide us with help to understand the influence of different training modes, team composition, and other factors on the improvement of a basketball player's competitive level.

We selected several types of training factors as dependent variables for basketball players' ability improvement, namely (1) position on the court (Pos), (2) players' BMI level, (3) whether the All-Star was selected or not, and (4) age. A four-layer neural network was constructed as the network structure required for training, and the connections between layers were all fully connected, and the number of nodes in the four-layer structure was 4-20-16-1, respectively. The network structure is shown schematically in the figure.

The forward propagation of each neuron is calculated as:

$$y = f\left(\sum_{i=1}^n w_i x_i + b\right)$$

x is the computed value of the previous node, w is the weight of the edge, b is the bias value, n is the number of nodes in the previous layer, and f is the activation function. We can take a sigmoid function:

$$\text{sig}(x) = 1/(1 + \exp(-x))$$

The Loss function uses an absolute value function:

$$\text{Loss} = |y_i - \hat{y}_i|$$

During the backpropagation process, the parameter worse algorithm uses the mini-batch gradient descent algorithm, which is executed for a total of 200 epochs, and the batch size is taken as 20.

It can be seen that as the training process advances, the accuracy rate gradually rises, while the change in the loss function slows down, and the change slows down at 20 epochs, so we choose to stop training at 20 epochs.

The trained neural network can evaluate the level of player development under each parameter, taking BMI and court position as an example, as in Fig. 4 and Fig. 5:

From the calculation results, we can learn that the PF position (power forward), and the C position (center) grow faster, which is also consistent with our general perception. Meanwhile, in general, players with higher BMI develop faster, which may reflect the physical quality of individual players.

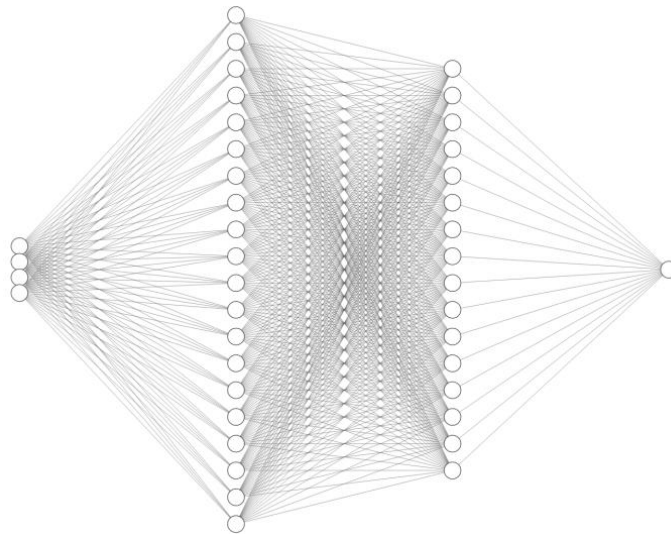


Figure 2. Schematic diagram of BP neural network structure

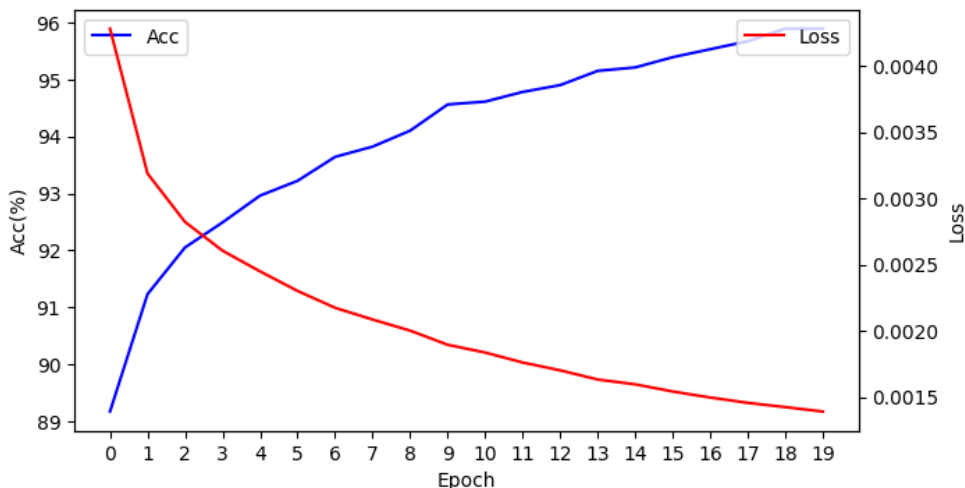


Figure 3. Changes in loss function (Loss) and accuracy (Acc) during the training process

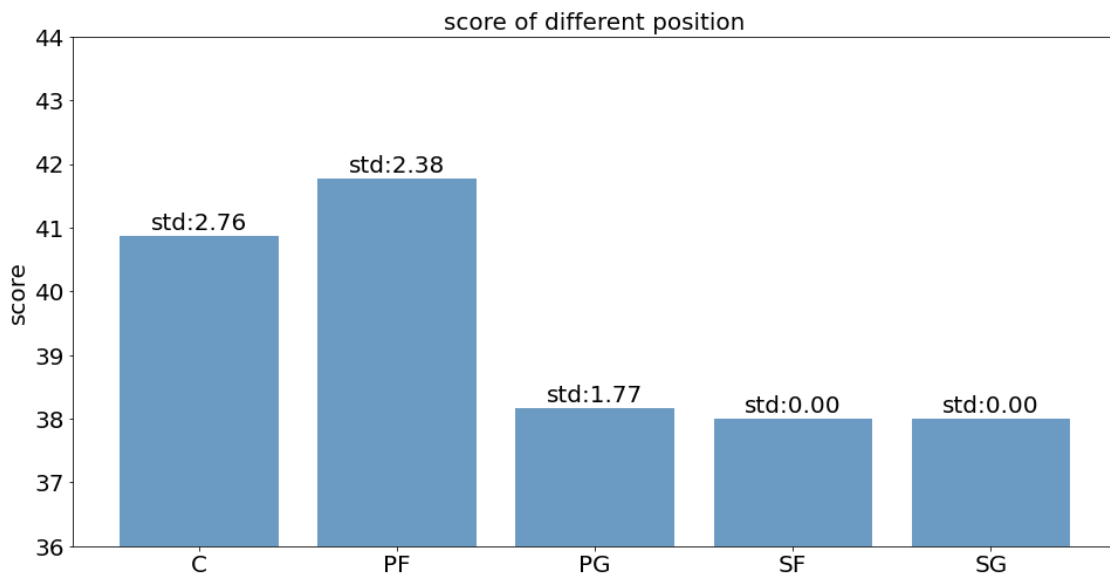


Figure 4. Growth scores for players in different positions on the field

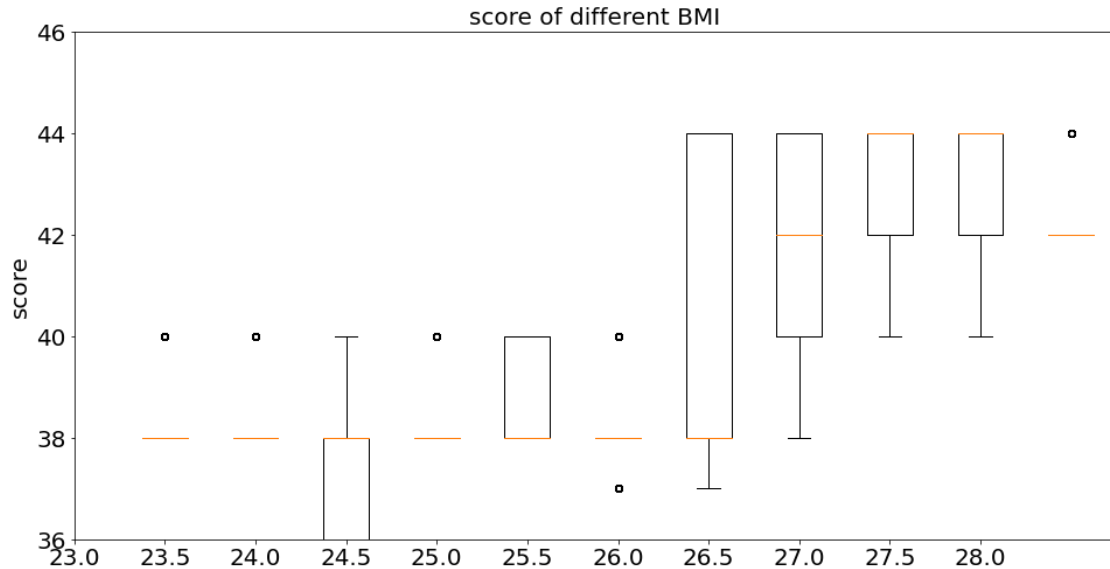


Figure 5. Growth scores of players with different BMIs

References

- [1] Fang XuXin, Zhuge Bin. Comprehensive ability analysis of badminton players based on entropy method[J]. Excellent Papers Online of Chinese Science and Technology, 2023, 16(02): 233-240.
- [2] Trninić S, Dizdar D. System of the performance evaluation criteria weighted per positions in the basketball game[J]. Collegium antropologicum, 2000, 24(1): 217-234.
- [3] Jin YingYan, Jia JunCheng, Hong MinJie, et al. Data analysis and visualization for NBA players[J]. Computer Applications and Software, 2021, 38(08): 84-91+174.
- [4] Sarlis V, Tjortjis C. Sports analytics—Evaluation of basketball players and team performance[J]. Information Systems, 2020, 93: 101562.
- [5] Fu JiaWei. A comparative study of the offensive and defensive abilities of excellent international players and American players in the 2018-2019 NBA season[D]. Sichuan Normal University, 2020.
- [6] Terner Z, Franks A. Modeling player and team performance in basketball[J]. Annual Review of Statistics and Its Application, 2021, 8: 1-23.
- [7] Trninić S, Dizdar D, Dežman B. Empirical verification of the weighted system of criteria for the elite basketball players quality evaluation[J]. Collegium Antropologicum, 2000, 24(2): 443-465.