

A Method for Evaluating the Ability of the Players

Zonghan Li^{1, a}

¹ North China Electric Power University, Beijing, 102206, China

^a310153630@qq.com

KeyWords: ability of the players; principal component analysis; weight; prediction

Abstract: this paper uses principal component analysis method to analyze the available data and personal ability information of players, establish evaluation model of player ability based on statistics data at the end of season, it adopts entropy method of objective weight method to confirm weight of each index, so that it can reduce effect of human subjectivity and select evaluation with obvious difference and rich information according to weight value, so it indicates that this model and algorithm is feasible through test result, the evaluation made on ability of the players is effective.

Introduction

The outcome of sports race not only depends on strength of both parties, it also has bigger uncertainty, this kind of randomness make it difficult^[1, 23] for prediction on outcome and score. In recent years, many scholars hope to strengthen comprehension of people on complicated object by relying on research on score prediction. Of which, probabilistic statistics is universally applied in the outcome prediction: firstly, we should establish one model used for prediction and it calculates predicted outcome probability of race according to strength of team. Secondly, we should make training according to historical outcome data of team and update rule to calculate new strength value of team according to actual race outcome and model strength. But the existing method is predicted on grade of team, but it has no detailed refinement^[4] on grade of players. In fact, because race between teams is revival among players, ability of the players usually will have critical effect in the offensive and defensive process, it will determine outcome of race, but factors such as player exertion, playing time, injury condition, player transfer etc will affect strength of team, so these factors should not be neglected. We should establish reasonable casual relationship on offense and defense condition, ability value of players, make modeling from ability value grade of players, it uses many indexes such as score, hit rate, backboard, steal and error etc to make evaluation on ability of players.

This paper completely uses 2 payers as research object to analyze the main problems of race and rectify model of this paper based on principal component analysis method, it also establishes one mechanism model between victor data of basic data and ability data of player, which is net score in basket ball race and score difference of victor and revival, finally it gets the ability ranking of player by optimizing calculation, it also make prediction on score condition of each offense and defense process during race.

Problem analysis and model assumption

If we want to establish one mathematical model that can evaluate the ability of trainer, index system of comprehensive evaluation method all has it own opinion, while index choice has certain effect(sometimes it confirms index system, while it has to abandon because it is difficult to collect index), for example, it is difficult to quantify and measure data such as the main aspect of trainer, the command ability in race involves player transfer, layout tactics and adjustment on time out etc, while time is different corresponds to different index system, it will have obvious effect on evaluation conclusion.

The main idea of this paper is to use net score as one measurement on mutual comparison of player ability, and then establish network model to carry our network analysis and determine ability value of team, and then analyze mutual effect and balance relations between player ability and team ability.

Suppose unit score, offense and defense ability of one player has little change in one race season, on the other hand, in the statistics, data at the end of race season, the ability of this player appears as one sample point in the overall normality, so under this assumption, the ability of this player is regarded as constant at any time of race season, ability at different time can be regarded as the sample of overall normality from spastics data at the end of race season(overall distribution of each ability is different), so this sample belongs to this player.

Data collection and preparation

We choose data collection and integration of race club in Super League Bundesliga from 2010 to 2014 and set race season from 2010 to 2014 as S_1, S_2, \dots, S_5 , there are 20m teams participate in German Bundesliga, there are totally 28 teams at least participate in German Bundesliga during 4 race seasons from 2010 to 2014, we make its number as X_1, X_2, \dots, X_{28} by the initial character sequence.

According to planet of race data, each score needs offense time of average 1 to 2 minutes. We extract score time of player at site commentary, if there is score in the 76th minutes, and then we record the 75th and 76th minute as one effective time section. We divide this time score into $1\Delta_1$. If there is score includes key words, and then we regard this time section as effective time section, on the contrary we regard it as ineffective time section.

During race, change of player at site will undoubtedly produce effect of different degree on race even race result. We usually regard P_i^k is better than P_j^k of the same team, if we change P_j^k into P_i^k , the situation at site is much more superior to their D_k . On the other hand, although the total playing time of every player is the important factor of producing effect on race condition, but it is undoubtedly that race site has designated lineup at ant moment. That is to say although D_k 的 P_i^k of D_k and 和 P_j^m of D_m are both on site when racing, but there is also the possibility of changing player, 2 players have not met each other during race. So it is one scientific method to divide race schedule by changing player and then make evaluation on player ability, we can get division $2(\Delta_2)$ according to change condition of players.

Mode design and evaluation based on weight principal component analysis method

Weight principal component analysis model

We can confirm the players of host team and guest team start and stop time as well as the total time of every player according to starting lineup and transfer information. We record the total time of P_j^k for D_k in S_l home revival with $D_m (m \neq k)$ is $Y_j^k \{m, k : l\}$, the total playing time in S_l is as follows:

$$\sum [Y_j^k \{k, m : l\} + Y_j^k \{m, k : l\}] \quad (1)$$

Score, offense and defense action are respectively recorded as α, β, γ without dimension.

Under Δ_1 , it regards the key words of technical actions appear in the effective time section as effective action, it is also divided into effective score action, effective offense action and effective defense action according to score, offense and defense. Although race score is completely determined by effective score action, but the successful score of one player needs great cooperation of the team players, so the contribution made by key words of other technical action should also be taken into consideration and given corresponding weight. In addition, it regards key word of technical action in the ineffective time section as ineffective action; it is divided into ineffective offense action and ineffective defense action according to offense and defense. Record the score action of players as α ,

effective offense action and effective defense action are respectively recorded as β , γ , ineffective offense action and ineffective defense actions are respectively recorded as β , γ .

Every ability average value and variance of all the players in single race season is indicted as follows:

$$P_j^k(\alpha; l) \sim N(\mu_{\alpha;l}, \sigma_{\alpha;l}^2) \quad (2)$$

$$P_j^k(\beta; l) \sim N(\mu_{\beta;l}, \sigma_{\beta;l}^2) \quad (3)$$

$$P_j^k(\gamma; l) \sim N(\mu_{\gamma;l}, \sigma_{\gamma;l}^2) \quad (4)$$

Of which, $\mu_{\alpha;l}$, $\mu_{\beta;l}$, $\mu_{\gamma;l}$ are respectively the mathematical expectation of unit score, offense and defense ability of all the Premier League players in SI from the statistics data at the end of S_l , $\sigma_{\alpha;l}^2$, $\sigma_{\beta;l}^2$, $\sigma_{\gamma;l}^2$ are corresponding variance. Suppose all the clubs have N_l players have played on the site, it can get maximum likelihood estimation from the maximum likelihood estimation method.

$$\mu_{\alpha;l} = \frac{1}{N_l} \sum_{i=1}^{N_l} P_i(\alpha; l) \quad (5)$$

$$\sigma_{\alpha;l}^2 = \frac{1}{N_l} \sum_{i=1}^{N_l} [P_i(\alpha; l) - \frac{1}{N_l} \sum_{i=1}^{N_l} [P_i(\alpha; l)]]^2 \quad (6)$$

The expression formula of $P_j^k(\alpha; l)$, $P_j^k(\beta; l)$ and $P_j^k(\gamma; l)$ contain unknown parameter ω^α , ω^β , ω^γ , so it can not get $\mu_{\alpha;l}$ and $\sigma_{\alpha;l}^2$ etc. For this, we start from perspective of respective score in every race and get the following result. Suppose the constant parameter under comprehensive effect of host advantage and transfer is θ , dimension is score, then when $S_{l1} \leq l \leq 4$, and home field competes with D_m .

Weight principal component analysis method

The weight PCA algorithm proposed by this paper is to make improvement on defect in classical PCA algorithm by adding weight value and data pretreatment on the basis of classical PCA algorithm, the algorithm flow chart after improvement is indicated by diagram 1.

The basic algorithm and step of principal component comprehensive analysis method is indicated as follows:

(1) Collect the N necessary sample data of random vector $X = (x_1, x_2, \dots, x_p)$ of P dimension, so that it can list the necessary data matrix of standard sample阵 $X = (x_{ij})_{n \times p}$.

(2) It makes pretreatment on data of collected sample data matrix, that is to make positive dimension and standard procession of data, the standardization formula of data is as follows:

$$d_{ij} = \frac{f_{ij} - \min f_{ij}}{\max f_{ij} - \min f_{ij}} \quad \text{else} \quad d_{ij} = \frac{-f_{ij} + \max f_{ij}}{\max f_{ij} - \min f_{ij}} \quad (13)$$

So that it can get matrix after standardization.

It makes the same and normalization, it makes applicability test on original data and judge correlation of index data, calculate relevant parameter matrix R.

$$R = (\rho_{ij})_{p \times p} = \begin{bmatrix} \rho_{11} & \rho_{12} & \cdots & \rho_{1p} \\ \rho_{21} & \rho_{22} & \cdots & \rho_{2p} \\ \rho_{p1} & \rho_{p2} & \cdots & \rho_{pp} \end{bmatrix} \quad (7)$$

ρ_{ij} is the relevant parameter of original variable x_i and x_j , $\rho_{ij} = \rho_{ji}$, $i, j = 1, 2, 3, \dots, p$, calculation formula is as follows:

$$\rho_{ij} = \frac{1}{n} * \frac{\sum_{k=1}^n (x_{ij} - \bar{x}_i)(x_{ki} - \bar{x}_j)}{s_i s_j} = \frac{\sum_{k=1}^n (x_{ij} - \bar{x}_i)(x_{kj} - \bar{x}_j)}{\sqrt{\sum_{k=1}^n (x_{ki} - \bar{x}_i)^2 \sum_{k=1}^n (x_{kj} - \bar{x}_j)^2}} \quad (8)$$

Calculate the characteristic value λ_i of relevant parameter matrix R and unit characteristic vector \vec{a}_i corresponds to it, and then confirm the principal component.

It can get p characteristic roots λ_i ($i=1,2,3,\dots,P$) by characteristic equation, it is arranged as $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$ according to its size, from mathematics it can demonstrate that characteristic root λ_i is the standard deviation of principal component. We usually think that we should extract the principal component its characteristic value larger than 1, if characteristic value is less than 1, which indicates that explanation strength of principal component is less than that of directly citing the original variable. Every characteristic root corresponds to one characteristic vector $\vec{a}_i = (a_{i1}, a_{i2}, \dots, a_{ip})$. The expression formula of principle component is as follows:

$$N_i = a_{i1}z_1 + a_{i2}z_2 + \dots + a_{ip}z_p, i = 1,2,3, \dots, p \quad (9)$$

Of which, it uses relations between factor load parameter μ_{ij} and characteristic value λ_i to obtain characteristic vector a_i , so that it gets each principal component.

$$\vec{a}_i = \frac{u_{ij}}{\sqrt{\lambda_i}} (i = 1,2,3, \dots, m; j = 1,2,3 \dots, p) \quad (10)$$

Calculate contribution ratio of standard deviation and accumulated contribution ratio; confirm the number of principal component.

Contribution ratio of standard deviation can be used to reflect information of each principal component. It can be indicated as follows:

$$F_1 = \frac{\lambda_i}{\sum_{i=1}^p \lambda_i} \quad (11)$$

Accumulated contribution ratio can be used to reflect comprehensive ability of m principal component information. It can be expressed as follows:

$$F_2 = \sum_{i=1}^m \frac{\lambda_i}{\sum_{i=1}^p \lambda_i} \quad (12)$$

It is thought that when accumulated contribution ratio is bigger than 85%, then it can be thought it can adopt new principal component to replace original variable, meanwhile it can also include most information of original variable. Here we take m components conform to condition.

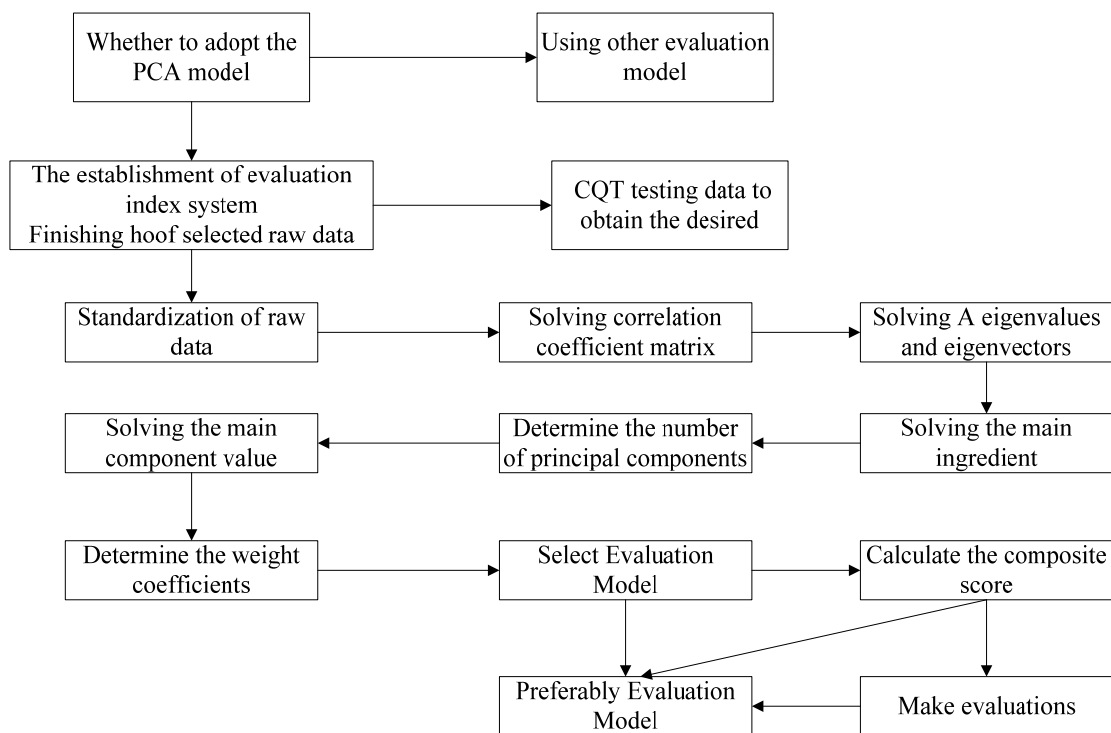


Diagram 1 PCA flow chart

Comprehensive evaluation on player ability based on principal component analysis method

This test is carried out under the Matlab environment; it adopts 3 different kinds of weight calculation method, pre-treatment method of 2 kinds of data: different combination of mean removal and equalization make classification and comparative description on test sample. Before making principal component analysis, we should make applicability test. Because principal component analysis method uses dimensionality reduction idea to obtain principal component, simplify data structure so as to solve correlation problem of data, so the variable of original data has stronger correlation, so it has possibility of simplification in data structure. The zero assumption of this test is to use relevant matrix as unit matrix, if it refuses zero assumption, then it regards relevant matrix is non-unit matrix, there exists correlation among original variable and data. KMO tests that this test stresses comparing simple relevant parameter and correlated size of partial correlation coefficient. If the quadratic sum of simple dependency number far bigger than quadratic sum of partial correlation coefficient, the less of partial correlation coefficient among variables, KMO is closer to 1 and original variable data is much more suitable to make principal component analysis. The value of KMO is between 0 and 1, Kaiser gives the measurement standard of KMO.

Table 1 Measurement standard of KMO value

KMO value	[1,0.9]	(0.9,0.8]	(0.8,0.7]	(0.7,0.6]	(0.6, 0.5]	(0.5,0]
Applicability degree	Very good	Good	common	proper	Basically proper	inconformity

It chooses 11 samples and applies SPSS19.0 to make principal component analysis method on the index data after dimension normalization, it can get analysis results such as relevant parameter matrix, characteristics value, variance contribution and accumulated contribution ratio of natural sub-system etc.

The applicability of principal component analysis method is indicated by the following table 2, the value of KMO is 0.748, bigger than 0.5, it shows that this group of index can make component analysis according to table 5.10. The associated probability of spherical degree test by Bartlett is 0.000, which is less than the obvious level 0.05, it refuses zero assumption of Bartlett spherical degree test, so it regards that it is suitable to make principal component analysis.

Table 2 Test results of KMO and Bartlett

Take enough sampling of Kaiser-Meyer-Olkin		0.748
Spherical degree test of Bartlett	Similar chi-square	87.767
	df	36.00
	Sig.	0.00

Calculate correlated parameter matrix R and primarily judge correlation degree of every index.

Table 3 Correlated parameter matrix of evaluation index among players

Index	x ₁	x ₂	x ₃	x ₄	x ₅
x ₁	1.000	0.103	0.016	0.072	0.205
x ₂	-0.103	1.000	0.890	0.869	0.833
x ₃	-0.016	0.890	1.000	0.984	0.642
x ₄	-0.072	0.869	0.984	1.000	0.636
x ₅	0.205	0.833	0.642	0.636	1.000

It calculates variance contribution ratio, accumulated contribution ratio of characteristic value of correlated parameter matrix R to determine the number of principal component. It calculates the characteristic vector corresponds to the selected principal component, writes out expression formula of principal component. And then the calculation model of index is as follows: $I = w_1 \times N_1 + w_2 \times N_2 + w_3 \times N_3$ /, it applies entropy method to give weight to every principal component, w_1 , w_2 , w_3 are respectively weight of the first principal component N_1 , the second component N_2 and the third principal component N_3 . The result is indicated by the following table 4:

Table 4 Calculation result of every principal component

Player name	N_1	N_2	N_3	Comprehensive index and skill
QY1	-3.575	-0.974	0.986	-2.631
QY2	3.837	-1.057	-0.165	2.565
QY3	-1.440	0.303	-1.494	-1.150
QY4	2.067	-0.431	-0.403	1.371
QY5	2.968	-0.491	-0.254	2.026

While the sampling point is the unit score, offense and defense ability of player P_j^k expressed in every race. It can confirm the single ability ranking and unit comprehensive ability ranking of player, as for P_j^k , the sampling point and average value can determine standard deviation (that is the stability indication of player ability, the less of value, the stable of ability), some results are indicated by table 5.

Table 5 Ability evaluation and ranking of player under principal component analysis method

Table 5 Calculation result of each principal component

Player name	Score ability and ranking	Offensive ability and ranking	Defense ability and ranking	Comprehensive skill
QY1	113/1	544 /1	0/439	1683/1
QY2	407/2	245 /2	2.1/130	654/2
QY3	360.1/3	206.6/3	1.328/206	568/3
QY4	342.7/4	177.5/4	1.18/234	521/4
QY5	336/6	160.8/8	1.81/148	499 /5

The football world is full of every kind of ranking list and honor title, these ranking and titles, even the social status of player etc all reflect personal ability of player. From the result we can see that the selection standard of rewards is closer to model, the evaluation made on ability of player is feasible. In addition, the methods in this paper has higher transferability, based on this nature, it can further

strengthen and refine evaluation parameter of this model, so that it can get precise evaluation and prediction result.

Summary

In order to make systematic and quantized evaluation on exercise ability of football player and predict ability trend of designated player according to the known data, it applies statistics method to make research. This paper is based on ability analysis of football player and puts forward the weight PCA algorithm, but it only uses weight PCA algorithm through simulation test to make classification on sample, and it can not achieve the ideal classification effect. The evaluation result obtained by principal component analysis method is basically consistent with it. It establishes comprehensive evaluation model by constructing football evaluation and principal component analysis method and obtains it by test result, principal component analysis method can well make evaluation on ability of player, which provides instruction meanings for arrangement of athlete personnel.

References

- [1] Song Xiaoli, Yu Jing, Sun Meichuan etc. Application of Fuzzy Comprehensive Evaluation Method in Risk Evaluation. Information Safety, Vol 12, 2006 ,p71-73.
- [2] Fu Ronglin. Discussion on 7-Component Comprehensive Evaluation Model. Practice in System Theory, Vol 11, 2001, p 68-74.
- [3] Hu Yonghong, He Sihui. Comprehensive Evaluation Method. Beijing: Science Press, 2000.
- [4] Qin Fenkang. Comprehensive Evaluation and Application. Beijing: Publishing House of Electronics Industry, 2003.
- [5] X. F. He and P. Niyogi. Locality preserving projections. Proceedings of the Conference on Advances in Neural Information Processing Systems. Vancouver, Canada, 2003, p 152-160
- [6] Bai Xiaoman. Human Facial Relization of Improved Weight Principal Component Analysis Method. Beijing: Beijing Jiaotong University, 2006
- [7] Isidoro Guzmán, Stephen Morrow. Measuring Efficiency and Productivity in Professional Football Teams: Evidence from the English Premier League[J]. Central European Journal of Operations Research, Vol 15(4), 2007, p 309-328.
- [8]. Adomavicius g., tuzhilin a. Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. Ieee transactions on knowledge and data engineering, Vol 17, 2005, p 734-749.