ISSN : 0974 - 7435

Volume 10 Issue 9

2014

BioTechnology

An Indian Journal

FULL PAPER

BTAIJ, 10(9), 2014 [3433-3441]

Matlab multi-dimensional model-based 2012-2013 Chinese football association super league football teams strength research

Yingcheng Zhang
Northeast Normal University, Changchun, Jilin, (CHINA)

ABSTRACT

In order to make comprehensive evaluation on Chinese football teams strength, the paper analyzes Chinese 12 football teams' performances in year 2012-2013national football first division team league matches, establishes four models from simple to complex, from rough to relative accurate, firstly successive calculate each team total score, and meanwhile make statistics of each team number of games, rank total score /number of games, obtained result can approximately be used as each team ranking. Secondly, according to game data, establish a 12×12 digital matrix A $=(a_{ij})_{1\times 12}$, use C++ programming, input obtained matrix, solve Hamilton opening path, and rank it. In the following, use three-point to calculate any i team and j team ($i \neq j$) score ratio b_{ij}, from which b_{ij}=1, and get score matrix $B=(b_{ij})_{12\times 12}$, solve score matrix maximum feature value, and solve corresponding feature vector. Compare component vector size that can solve ranking. Finally use analytic hierarchy process, take average score, number of goal difference and ratio between winning games numbers and participation games numbers as criterion layer influence factors, according to their proportional relationships, construct positive reciprocal matrix (inverse matrix), by solving maximum feature value and its feature vector, and then solve ranking.

KEYWORDS

Football team strength; Graph theory model; Score matrix; Analytic hierarchy process.

INTRODUCTION

In recent decades, football such sports event is relative popular in China, is favored by lots of ball fans, more and more large-scale football games have been organized in domestic, from which national football league match is a relative formal game organization with relative precise game requirements. Score principles being just, fair and open is particularly important.

In modern football techniques, tactics analysis and evaluation, it often adopts ball controlling percentage, pass number and other original data to make analysis and evaluation, in fact, original data and game result inconsistency possibility is larger. Wang Kai, Lv Xiao-Wei, He Jiang-Chuan [1](2010)adopted factor analysis method, established year 2009 season Chinese football association super league team matches four commonality factors influential score standardization lineal combination estimation formulas and factor total score standardization linear combination estimation formula, computed and got year 2009season Chinese football association super league team matches' 16 teams' factor scores season evaluation rankings, made analytic discussion on their differences, and provided evidence for scientific and effective pre-season training control and performance prediction; They established a set of new evaluation system. Xu Lei[2](2013)applied mathematical statistics, sum of ranks ratio comprehensive evaluation method to make quantitative analysis of 2012 season participated Chinese football association super league sixteen teams attack and defense indicators data, made variance analysis and multiple comparisons of analysis results, and applied rank correlation method to detect analysis result, implemented quantization evaluation on techniques and tactics abilities of Chinese football association super league teams, so as to pursuit objective and realistic reflecting participated teams comprehensive techniques and tactics abilities. But literatures that study from multiple perspectives are little, the paper tries to apply multiple methods to establish team strength evaluation model and ranks the teams.

EVALUATION MODEL ESTABLISHMENTS

Average score method model

According to national football league matches rule, win a game will get three points, draw gets one point, lose a game doesn't get point.

The paper's used symbols illustration: a_i——the i team total number of games; a_{i1}——the i team winning number of games; a_{i2}——the i team draw number of games; a_{i3}——the i team losing number of games; w_i——the i team total score; φ_i——the i team average score;

Successively compute every team total score and average score:

Objective function: $\varphi_i = w_i \div a_i$

Constraint condition: $\begin{cases} w_i = 3 * a_{i1} + 1 * a_{i2} \\ \sum_{j=1}^{3} a_{ij} = a_i \end{cases}$

Graph theory model

Establish a 12×12 digital matrix $\mathrm{A} = (\mathrm{a}_{ij})_{12\times12}$, when T_i defeats T_j, make marks $\mathrm{a}_{ij} = 1$;when the two draw or the two have no games, don't make any marks ;when T_i is defeated by T_j, mark $\mathrm{a}_{ij} = 0$;

According to obtained 12×12 matrix, make statistics of sum total that every line as 1 that every team defeats opponents numbers, record as a vector $\alpha = (\mathrm{a}_1, \mathrm{a}_2, \mathrm{a}_3, \mathrm{a}_4, \mathrm{a}_5, \mathrm{a}_6, \mathrm{a}_7, \mathrm{a}_8, \mathrm{a}_9, \mathrm{a}_{10}, \mathrm{a}_{11}, \mathrm{a}_{12})$;

If vectors have same elements as $a_i = a_j$, then respectively solve sum total of all teams a_j that are defeated by T_i from 1 to 12 (that is N), and use them as new vector $\beta = (a_1^{(1)}, a_2^{(1)},) a_3^{(1)}, a_4^{(1)}, a_5^{(1)}, a_6^{(1)}, a_7^{(1)}, a_8^{(1)}, a_9^{(1)}, a_{10}^{(1)}, a_{11}^{(1)}, a_{12}^{(1)})$ $a_i^{(1)}$ value, it gets new vector β;if it still has same elements, then randomly let one party to be 1, the other to be 0 according to principle of drawing lots, finally it gets 0-1 matrix;

According to obtained matrix, execute in well compiled C++ programming, and get Hamilton opening path;

Every Hamilton path is a kind of ranking result, but its dependence on matrix is too strong, which needs us to further analyze comprehensive data, and get final ranking result.

Score matrix method model

For Model one average score method, it has its irrevocable irrationality; when compute game scores, it doesn't consider opponents are strong or weak. Such as, strong team wins strong team, it gets three points, and strong team wins weak team, it similar gets three points. So adopt score ratio matrix similarly is use three-point to compute any j team and j time (i is not equal to j) score ratio b_{ij}, from which $b_{ii}=1$

According to score matrix $B = (b_{ij})_{12\times12}$ (from which b_{ij} is i team average score and j team average score ratio), solve score matrix maximum feature value, and further get corresponding feature vector. Compare component vector size that can solve ranking.

Analytic hierarchy process model

In the model, we adopt analytic hierarchy process. In the topic, we thought team ranking influences are mainly as following three factors: average score, goal difference, win/ total.
According to analytic hierarchy process, we establish following hierarchical relationship Figure 1.

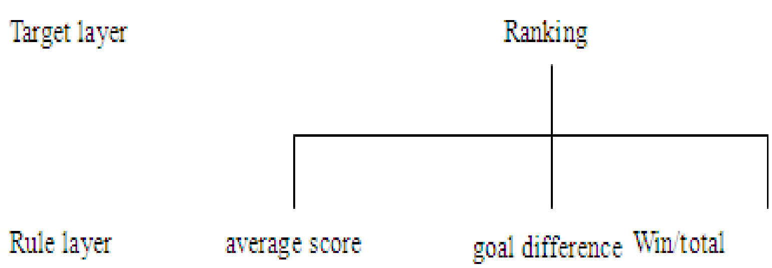

Figure 1 : Hierarchical relation

Each factor x_1, x_{2_1}, x_3, importance with respect to target y (from which $y = w_1x_1 + w_2x_2 + w_3x_2$) uses following TABLE 1 numerical values to express.

TABLE 1 : Importance

x_i / x_j	Equal	Relative strong	Strong	Very strong	Absolute strong
a_{ij}	1	3	5	7	9

If it is between above two, then take 2, 4, 6, 8[5].

By three factors impacts on ranking, it construct matrix C, from which $C = (c_{ij})_{3*3} = (x_i / x_j)_{3*3}$, for above data we can write matrix C, and then solve maximum feature value and its corresponding feature vector. Make normalization processing with feature vectors, then it can get w_1, w_2, w_3 values and we can solve ranking.

MODEL SOLUTIONS

Average score method model solution

Computed result is as TABLE 2 shows:

TABLE 2 : Each team game data

Team	Total score	Number of games	Average score
T1	34	19	1.7895
T2	21	15	1.4000
T3	27	15	1.8000
T4	9	19	0.4737
T5	8	9	0.8889
T6	6	5	1.2000
T7	39	17	2.2941
T8	22	17	1.2941
T9	23	17	1.3529
T10	24	17	1.4118
T11	5	9	0.5556
T12	10	9	1.1111

Ranking result: $T_7 - T_3 - T_1 - T_{10} - T_2 - T_9 - T_8 - T_6 - T_{12} - T_5 - T_{11} - T_4$

Model expansion: For any N team, by competition obtained data, we can rank according to average score, in case average scores are the same, it can consider rank on goal difference, total goal rate and so on, if these factors are still the same, only rank on these considerable equal level teams by drawing lots.

Graph theory model solution

Establish matrix

$$A=\begin{bmatrix} 0 & & 0 & 1 & 1 & 1 & 0 & 0 & 1 & & & \\ & 0 & 0 & 1 & & 1 & & & 1 & 0 & & \\ 1 & 1 & 0 & 1 & 1 & 1 & 1 & 0 & 1 & 0 & & \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & & \\ 0 & & 0 & 1 & 0 & 0 & & & & & & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & & & & & & \\ 1 & & 0 & 1 & & & 0 & 1 & 1 & 1 & 1 & 1 \\ 1 & & 1 & 1 & & & 0 & 0 & 0 & & 1 & \\ 0 & 0 & 0 & 1 & & & 0 & 1 & 0 & 1 & 1 & 1 \\ & 1 & 1 & 1 & & & 0 & & 0 & 0 & 1 & 1 \\ & & & & & & 0 & 0 & 0 & 0 & 0 & 0 \\ & & & & 1 & & 0 & & 0 & 0 & 1 & 0 \end{bmatrix}$$

It gets $\alpha=(4,3,7,0,1,2,7,4,5,5,0,2)$

Step two:

$$A=\begin{bmatrix} 0 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 & 1 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & \\ 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

It gets $\beta=(8,7,22,0,0,1,20,11,11,12,0,1)$.

From above 1, 2, we still cannot decide winning or losing between T_4 and T_{11}, T_6 and T_{12}, by principle of drawing lots, we assume that T_4 is defeated by T_{11}, T_{12} is defeated by T_6, finally perfect matrix :

$$A=\begin{bmatrix} 0 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 & 1 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

By Model one, it can get: T_3 and T_7 team strength are strongest, while T_4 and T_{11} strength are relative weakest.

From program running result, it selects Hamilton opening path with T_3 and T_7 as initials, result is as TABLE 3 and TABLE 4 shows.

TABLE 3 : T_3 team Hamilton opening path

3	7	1	2	9	10	8	6	12	5	11	4
3	7	1	9	10	8	2	6	12	5	11	4
3	7	2	9	10	8	1	6	12	5	11	4
3	7	8	1	2	9	10	6	12	5	11	4
3	7	8	1	9	10	2	6	12	5	11	4
3	7	8	2	9	10	1	6	12	5	11	4
3	7	9	10	8	1	2	6	12	5	11	4
3	7	10	1	2	9	8	6	12	5	11	4
3	7	10	1	9	8	2	6	12	5	11	4
3	7	10	2	9	8	1	6	12	5	11	4
3	7	10	8	1	2	9	6	12	5	11	4

TABLE 4: T_7 team Hamilton opening path

7	1	2	9	10	8	3	6	12	5	11	4
7	1	9	10	8	3	2	6	12	5	11	4
7	2	9	10	8	3	1	6	12	5	11	4
7	8	1	2	9	10	3	6	12	5	11	4
7	8	1	9	10	3	2	6	12	5	11	4
7	8	2	9	10	3	1	6	12	5	11	4
7	8	3	1	2	9	10	6	12	5	11	4
7	8	3	1	9	10	2	6	12	5	11	4
7	8	3	2	9	10	1	6	12	5	11	4
7	8	3	9	10	1	2	6	12	5	11	4
7	9	10	8	3	1	2	6	12	5	11	4
7	10	1	2	9	8	3	6	12	5	11	4
7	10	1	9	8	3	2	6	12	5	11	4
7	10	2	9	8	3	1	6	12	5	11	4
7	10	3	1	2	9	8	6	12	5	11	4
7	10	3	1	9	8	2	6	12	5	11	4
7	10	3	2	9	8	1	6	12	5	11	4

Data analysis :

(1) From above two tables, it gets that T_6, T_{12}, T_5, T_{11}, T_4 are surely the bottom five;

(2)Rank T_1, T_2, T_8, T_9, T_{10} : Combine with vector α and β, rank them and get $T_{10}-T_9-T_8-T_2-T_1$

Final ranking: $T_7-T_3-T_{10}-T_9-T_8-T_2-T_1-T_6-T_{12}-T_5-T_{11}-T_4$

Score matrix model solution

Score matrix:

$$B = \begin{bmatrix} 1.0000 & 1.2782 & 0.9942 & 3.7777 & 2.0132 & 1.4913 & 0.7800 & 1.3828 & 1.3227 & 1.2675 & 3.2208 & 1.6106 \\ 0.7823 & 1.0000 & 0.7778 & 2.9555 & 1.5750 & 1.1667 & 0.6103 & 1.0818 & 1.0348 & 0.9916 & 2.5198 & 1.2600 \\ 1.0059 & 1.2857 & 1.0000 & 3.7999 & 2.0250 & 1.5000 & 0.7846 & 1.3909 & 1.3305 & 1.2750 & 3.2397 & 1.6200 \\ 0.2647 & 0.3384 & 0.2632 & 1.0000 & 0.5329 & 0.3948 & 0.2065 & 0.3660 & 0.3501 & 0.3355 & 0.8526 & 0.4263 \\ 0.4967 & 0.6349 & 0.4938 & 1.8765 & 1.0000 & 0.7408 & 0.3875 & 0.6869 & 0.6569 & 0.6296 & 1.5999 & 0.8000 \\ 0.6706 & 0.8571 & 0.6667 & 2.5332 & 1.3500 & 1.0000 & 0.5231 & 0.9273 & 0.8869 & 0.8500 & 2.1598 & 1.0800 \\ 1.2820 & 1.6386 & 1.2745 & 4.8429 & 2.5808 & 1.9117 & 1.0000 & 1.7727 & 1.6954 & 1.6249 & 4.1290 & 2.0647 \\ 0.7232 & 0.9244 & 0.7189 & 2.7319 & 1.4558 & 1.0784 & 0.5641 & 1.0000 & 0.9564 & 0.9166 & 2.3292 & 1.1647 \\ 0.7561 & 0.9665 & 0.7517 & 2.8564 & 1.5222 & 1.1276 & 0.5898 & 1.0456 & 1.0000 & 0.9584 & 2.4354 & 1.2178 \\ 0.7889 & 1.0084 & 0.7843 & 2.9804 & 1.5883 & 1.1765 & 0.6154 & 1.0910 & 1.0434 & 1.0000 & 2.5410 & 1.2706 \\ 0.3105 & 0.3969 & 0.3087 & 1.1729 & 0.6250 & 0.4630 & 0.2422 & 0.4293 & 0.4106 & 0.3935 & 1.0000 & 0.5000 \\ 0.6209 & 0.7936 & 0.6173 & 2.3456 & 1.2500 & 0.9259 & 0.4843 & 0.8586 & 0.8212 & 0.7870 & 1.9998 & 1.0000 \end{bmatrix}$$

Use matlab software, it can solve B maximum feature value and its corresponding feature vector, it can get matrix B maximum feature value as 12.0000, its corresponding feature vector is:

$[0.3718\ 0.2909\ 0.3740\ 0.0984\ 0.1847\ 0.2493\ 0.4767\ 0.2689\ 0.2812\ 0.2933\ 0.1154\ 0.2309]^T$ So we get each team ranking result as:

$T_7 - T_3 - T_1 - T_{10} - T_2 - T_9 - T_8 - T_6 - T_{12} - T_5 - T_{11} - T_4$

Analytic hierarchy process model solution

We can get each team average score, goal difference, win/total, as TABLE 5 shows.

TABLE 5 : Each team game data processing result

Team	Average score	Goal difference	Win/Total
T1	1.7895	8	10/19
T2	1.4000	2	1/3
T3	1.8000	8	8/15
T4	0.4737	-20	1/19
T5	0.8889	-5	2/9
T6	1.2000	-4	2/5
T7	2.2941	25	13/17
T8	1.2941	2	6/17
T9	1.3529	-6	7/17
T10	1.4118	-2	6/17
T11	0.5556	-7	1/9
T12	1.1111	-3	2/9

We can write matrix $C = \begin{bmatrix} 1 & 3 & 2 \\ 1/3 & 1 & 1/2 \\ 1/2 & 2 & 1 \end{bmatrix}$

In Matlab software, it can solve C maximum feature value as $\lambda_{\max} = 3.0092$, feature value $\lambda_{\max}$

corresponding feature vector as $\begin{bmatrix} 0.8468 \\ 0.2565 \\ 0.4660 \end{bmatrix}$, normalize it and get vector $\begin{bmatrix} 0.5396 \\ 0.1634 \\ 0.2970 \end{bmatrix}$,

We can see average score proportion is larger, so when we rank teams, we firstly consider average score, when average score is about the same, we then calculate winning number of games and the number of games ratio. Therefore, we get each team ranking as :

$T_7 - T_3 - T_1 - T_{10} - T_2 - T_9 - T_8 - T_6 - T_{12} - T_5 - T_{11} - T_4$

MODEL TEST

Adopt computer simulation method to do model test. Specific method is as following: set it has n pieces of teams to attend the games, adopt random function to generate n pieces of number in the interval [0, 1] respectively record them as M_i, it shows the n teams overall strength level, rank the n numbers from big to small then can get n each team ranking. According to generated n numbers, it can generate a group of game data, for any T_i and Tj, firstly use random function to generate their number of games b_{ij}(value as one among 0, 1, 2, 3), it should also note number of games selection should ensure graph connectivity that for any T_i, it should play one game with other teams at least. Then, generate game data, it might as well set T_i is stronger than Tj, we get a game result probability experience formula by consulting information[4]:

$$P\{T_i win\} = 0.3 + 0.7\sqrt{M_i - M_j}$$

$$P\{T_j win\} = 0.3 - 0.3\sqrt{M_i - M_j}$$

$$P\{\text{dogfall}\} = 1 - P\{T_i win\} - P\{T_j win\} = 0.4 - 0.4\sqrt{M_i - M_j}$$

Record above three formulas probability respectively as P_1, P_2, P_3.

According to above probability algorithm, it can respective divide interval [0, 1] into three segments according to above probability size to use as computer random simulation game result. Finally we simulate every game score, set T_i and Tj the q game score is $a:b$, then

$T_i$wins that is when random number X drops into [0, P_1]

$b = \text{rand()}\%3$, $a = b + 1 + \text{rand()}\%(\text{int})(1 + 2(M_i - M_j))$

$T_j$wins, that is when random number X drops into $[P_1, P_1 + P_2]$ $a = \text{rand()}\%3$, $b = a + 1 + \text{rand()}\%2$

Draw, that is when random number X drops into $[P_1 + P_2, 1]$ $a = b = \text{rand()}\%5$

After model finishing, it can get any group data, carry on simple screening on data then can select some rough data to test, analyze and evaluate established model.

Record random generated ranking order is Q_i ($i = 1,2,...,n$), model generated ranking order is q_i ($i = 1,2,...,n$). We adopted test formula is

$$E = \frac{1}{n}\sum_{n}(Q_i - q_i)$$

Obviously E gets small, it shows model is more reasonable, in order to eliminate random factors impacts on model testing, we simulate enough more data to test, and E takes average level.

When $n = 12$, due to data amount is very big, we only take five groups of data to carry on simple testing, test result is as following TABLE 6.

TABLE 6 : Test result

Model	Model one	Model two	Model three	Model four
E average value	4.50	5.43	3.13	4.04

From TABLE 6, it is clear that model three E average value is smaller. To model usage condition, we need to further consider variance and so on.

CONCLUSION

The paper establishes four models from simple to complex, from rough to relative accurate; their respective advantages are as following:

(1) Computation is simple, operation is convenient;

(2) From operation result, it can distinguish every team rough strength range in short time, and distinguish them between different levels;

(3) It can relative comprehensive and comprehensive compare each sub team strength level;

(4) Consider multiple factors impacts on result;

REFERENCES

[1] Wang Kai, Lu Xiao-wei, He Jiang-chuan; Multivariate Statistical Analysis on Tactical and Technical Structural Difference of 16 Teams in 2009 China Football Super League[J]. China Sport Science and Technology, **46(5)**, **(2010)**.

[2] Li Qiang, Zheng Chang-Jiang; The Factor Analysis and Hierarchical Cluster Analysis of Techniques Strategies of the last 32 teams in European Championship Cup Football In 2010[J]. Journal of Physical Education Institute of Shanxi Teachers University, **27(2)**, 89-95 **(2012)**.

[3] Zhong Jian-ming, Jia Hong, Yang Xiao-hongl; Application Study on the Factors Affecting Teams' Score from Multivariate Statistics in the 2010 FIFA World Cup[J]. Journal of Guangzhou Physical Education Institute, **32(3)**, 69-74 **(2012)**.