http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 12 (19): 5172-5176, 2013
ISSN 1812-5638 / DOI: 10.3923/itj.2013.5172.5176

Study of Application of Factors of Volleyball Game Based On Data Mining

Lei Sheng
School of Sports Science, Nantong University, Nantong 226019, China

Abstract: To volleyball game, both athletes and spectators focus on results. Its factors and data are obviously subject and obscured. Therefore this research complement the accuracy of data envelopment analysis and fuzzy of fuzzy comprehensive evaluation, to process the given data. Then study the situation and trends of volleyball groups in data mining. The result shows, to comparable teams, the key to win is whether the current winner can turn the winning advantages to winning result, when both the scores are higher than 20. Meanwhile, the team falling behind should reverse the situation and overcome the rival by functioning tactics properly and the key is to adopt advantages and avoid disadvantages.

Key words: Data envelopment analysis, volleyball game, data mining

INTRODUCTION

Nowadays, powerful sports countries pay more and more attention to improve their competitive sports levels, increasingly input human force, substance force and finance. With the computer multimedia technology and artificial smart technology developing continuously, some systems are applied to net antagonistic ball games gradually, such as computer video diagnosis system, computer tactic analysis, assisting training and competition support system which gains more and more attentions (Guo and Wang, 2012). From the native perspectives, the advanced scientific methods for high-level volleyball training and games are few, in a low level. Besides, the coaches', athletes' and managers' scientific consciousness should be enhanced. Nowadays, data mining is applied widely (Zhang *et al*., 2006). It can get and process the data in reality, to predict the results. It serves readers but not coaches for analyzing and deciding (Li and Qiao, 2012).

Nowadays, data mining is widespread. It can get and process the data in reality, so as to find important information. This research, objecting as scores, analyzes and predicts the results in such methods. Meanwhile, this research gets the factors of scores. Therefore, the performances can be improved by improving the competitive process.

JUDGMENT MODEL BASED ON FUZZY DATA ENVELOPMENT ANALYSIS

Data envelopment analysis is to utilize mathematical programming models (including linear programming, multi-objective programming, etc), to evaluate relative efficiency of multiple inputs, especially the comparative reliability of multiple outputs among decisive units. It has advantages of accuracy of objective data. However it is difficult to find accurate index factors in reality. So it is fuzzy (Zhang and Gu, 2004). This research complements the accurate data envelopment analysis and fuzzy comprehensive evaluation, resulting in a fuzzy comprehensive evaluation model for data envelopment analysis (Zhang, 2013).

If there are m evaluation units, (c+d) indexes, c quantitative indexes and d non-quantitative indexes.

Fuzzy operation for non-quantitative weights: If C = (c_1, c_2,…,c_q) is factor set and V = (v0, v1, …, v_{p-1}) is comment set, the comprehensive evaluation matrix is:

$$R_j = \begin{bmatrix} r_{j10} & r_{j11} & \cdots & r_{j1(p-1)} \\ r_{j20} & r_{j21} & \cdots & r_{j2(p-1)} \\ \cdots & \cdots & \cdots & \cdots \\ r_{jq0} & r_{jq1} & \cdots & r_{jq(p-1)} \end{bmatrix}, \; j = 1, 2, \cdots, m$$

Aj = (a_{j1}, a_{j2}, …, a_{jg}) is weight matrix. So the non-quantitative index weight of jth decisive unit after fuzzy operating is:

$$B_j = A_j R_j = (a_{j1}, a_{j2}, \cdots, a_{jq}) \begin{bmatrix} r_{j10} & r_{j11} & \cdots & r_{j1(p-1)} \\ r_{j20} & r_{j21} & \cdots & r_{j2(p-1)} \\ \cdots & \cdots & \cdots & \cdots \\ r_{jq0} & r_{jq1} & \cdots & r_{jq(p-1)} \end{bmatrix}$$
$$= (b_{j1}, b_{j2}, \cdots, b_{jp})$$

Calculation for data envelopment with quantitative weights: Imagine X = (s_{1j}, x_{2j},…, x_{nj})^T and Y_j = (y_{1j}, y_{2j},…, y_{sj})^T as input and output vectors of ith evaluation unit DMU_i(1≤i≤m), of which j = 1, 2, …, m. Each vector coordinate is positive. If demonstrate weight vectors of input and output as $v = (v_1, v_2, \ldots, v_n)^T$, $u = (u_1, u_2, \ldots, u_s)^T$, the linear programming model after Charnes-Cooper transforming is:

$$\begin{cases} \max \mu^T Y_{j0} \\ \text{s.t.} \quad \omega^T X_j - \mu^T Y_j \ge 0, j = 1, 2, \cdots, m \\ \omega^T X_{j0} = 1 \\ \omega \ge 0, \mu \ge 0 \end{cases}$$

Take data into this model and get the optimal solution B_j' which is the accurate quantitative index weight.

Although these data are more objective and persuasive, the motional cognition like "outstanding, fine, qualified and disqualified" and membership of fuzzy comprehensive evaluation don't exist. Therefore, this research applies membership function to fuzzy results.

The operation results of data envelopment can be considered as the membership degree of comment set $V = (v_0, v_1, \ldots, v_{p-1})$. Imagine $r = (r_0, r_1, \ldots, r_{p-1})$ as membership, then:

$$r_j = \begin{cases} \dfrac{x - (j-1)\frac{1}{p-1}}{\frac{1}{p-1}} & ,(j-1)\frac{1}{p-1} \le x < j\frac{1}{p-1} \\ \dfrac{(j+1)\frac{1}{p-1} - x}{\frac{1}{p-1}} & ,j\frac{1}{p-1} \le x < (j+1)\frac{1}{p-1}, r_j \in [0,1], j = 0, 1, \cdots, p-1 \\ 0 & \end{cases}$$

Take B_j' into the formula above and get membership degree as $B_j = (b_{j1}, b_{j2}, \ldots, b_{jp})$.

Comprehensive evaluation: Conduct comprehensive evaluation on such results. The comprehensive evaluation matrix is:

$$R_j = \begin{bmatrix} B_{j1} \\ B_{j2} \\ \cdots \\ B_{jk} \end{bmatrix}, j = 1, 2, \cdots, m$$

of which, k is the number of indexes (quantitative and non-quantitative). Imagine $Aj = (a_{j1}, a_{j2}, \ldots, a_{jk})$, $j = 1, 2, \ldots, m$ as weight, then B = A and:

$$R \Rightarrow B_j = (a_{j1}, a_{j2}, \cdots, a_{jk}) \begin{bmatrix} B_{j1} \\ B_{j2} \\ \cdots \\ B_{jk} \end{bmatrix} = (b_{j1}, b_{j2}, \cdots, b_{jp}), j = 1, 2, \cdots, m$$

Utilizing the principle of maximum membership degree, the final result is v_i in $(v_0, v_1, \ldots, v_{p-1})$ correspond to maximum b_{ji} in $Bj = (bj1, \ldots, b_{jp})$

DATA MINING

To volleyball game, daily training and preparation affects the results. As a group sport, the real situation of each athlete also influences the final results. A factor may have different effect on different athletes. Each athlete has distinct features. Therefore, a model is not available to predict. According to different situations and exertion of different people, we get lots of factors and complex relations exit among the factors. So the data is obscured and subjective. This research uses association rule.

Due to the large amounts of time-varying data, the early data is invalid. So this research uses main factors to build prediction model as follows:

$$v = b_1 v_1 + b_2 v_2, \ b_1 + b_2 = 1$$

$v_1 = E(I)$ is the mathematical anticipation of team I. V_2 is predicted result. b_1 and b_2 is regulated according to predicted results as follows:

Set V as predicted result and the real result is T for team I. When $|V-T|>\sigma$, set b_1 and b_2 according to equation:

$$\begin{cases} b_1 + b_2 = 1 \\ b_1 v_1 + b_2 v_2 = T \end{cases}$$

Utilizing rank of augmented matrix:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & v_1 - v_2 & T - v_1 \end{pmatrix}$$

when v_1-$v_2 \neq 0$, solve the equation and get b_1 and b_2; when v_1-v_2 = 0 and T-v_1 = 0, there are infinity solutions, b_1 and b_2 is not regulated; when v -v_1 $\overline{\overline{2}}$ 0 and T-v ≠0, no solution exists, then set:

$$\varepsilon = \frac{2}{5}(T - v_1)$$

Deal with data in neural network model. When L is in the middle level, v_2>T, v_2, is available and the error of v_1 is big. Then increase b_1 and decrease b_2. Set v_1' = T+∊L to replace v_1 and get the equation system:

$$\begin{cases} b_1 + b_2 = 1 \\ b_1 v_1' + b_2 v_2 = T \end{cases}$$

and solve b_1 and b_2.

When L is in the low level, v_2<T, v_1 is available and the error of v_2 is big. Then increase b_2 and decrease b_1. Set v_1' = T+∊L to replace v_2 and get the equation system:

$$\begin{cases} b_1 + b_2 = 1 \\ b_1 v_1 + b_2 v_2' = T \end{cases}$$

and solve b_1 and b_2.

If one of b_1 and b_2 is kept in the low level, the prediction is not correct. b_1 Also exhibits the state of athlete (peak or low).

All in all, the gap between the predicted performance of a team in the next game and recent performance is not big, indicating E(I) is stable. So consider E(I) as a factor is scientific.

APPLICATION OF DATA MINING BASED ON HOMOGENEOUS MARKOV MODEL

Homogeneous markov model: Markov process is a special description of developing process. Obviously, the development of a matter is time-varying. To some matter, it must observe its past, current to predict. However, to some matter, the current state can predict the future. For example, to chess, the next step only relates to current position, not need to know its previous position. Such situation is called ineffectiveness. The developing process of matter with ineffectiveness is called Markov process.

Imagine there are n state, P_{ij} is transition probability from state i to state j in a step. Arrange such data and construct a matrix called transition probability matrix:

$$P = \begin{bmatrix} P_{11} & P_{12} & \cdots & P_{1n} \\ P_{21} & P_{22} & \cdots & P_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ P_{n1} & P_{n2} & \cdots & P_{nn} \end{bmatrix}$$

On such basis, the multi-step transition probability can be deduced. If the system is in state i at time t_0, after n steps, it is in state j at time t_n, its quantitative index is called n-step transition probability, written as $P(x_n = j|x_0 = i| = P_{ij}(n))$. n-step transition probability matrix:

$$P(n) = \begin{bmatrix} P_{11}(n) & P_{12}(n) & \cdots & P_{1n}(n) \\ P_{21}(n) & P_{22}(n) & \cdots & P_{2n}(n) \\ \cdots & \cdots & \cdots & \cdots \\ P_{n1}(n) & P_{n2}(n) & \cdots & P_{nn}(n) \end{bmatrix}$$

Markov chain is an important random process, whose state space can be finite or infinite. After certain time, system transfers from a state to another state which is only relates to current state.

{X(n), n = 0, 1, 2,…} is a random process and E = {0, 1, 2,…} is state space. If the condition probability is, to any integer time set $0 \le n_1 < n_2 < \ldots < n_k$ and any state $i_1, i_2, \ldots, i_k \in E$:

$$\begin{aligned} &P\{X(n_k) = i_k \mid X(n_1) = i_1, X(n_2) = i_2, \cdots, X(n_{k-1}) = i_{k-1}\} \\ &= P\{X(n_k) = i_k \mid X(n_{k-1}) = i_{k-1}\} \end{aligned}$$

That is process {X(n), n = 0, 1, 2,…} only relates to current state and independent of previous states. Process {X(n), n = 0, 1, 2,…} is called time-discrete Markov chain. Probability Pij(m, k) = P{m+k} = j|X(m), I, j∈E called k-step transition probability of Markov chain. If k-step transition probability only relates to k, independent of time start m, process {X(n)} is called time-discrete homogeneous Markov chain. k-step transition probability matrix is:

$$P(m,k) = \begin{bmatrix} P_{00}(m,k) & P_{01}(m,k & \cdots & P_{0n}(m,k) & \cdots \\ P_{10}(m,k) & P_{11}(m,k) & \cdots & P_{1n}(m,k) & \cdots \\ \vdots & \vdots & & \vdots & \\ P_{j0}(m,k) & P_{j1}(m,k) & \cdots & P_{jn}(m,k) & \\ \vdots & \vdots & & \vdots & \end{bmatrix}$$

Process {X(t), t≥0} is time-continuous random process and E = {0, 1, 2…} is state space. If the condition probability is, to any n $0 < t_1 < t_2 < \ldots < t_n < t_n+1$ and $i_1, i_2, \ldots, i_n, i_{n+1} \in E$:

$$\begin{aligned} &P\{X(t_{n+1}) = i_{n+1} \mid X(t_k) = i_k, k = 1, 2, \cdots, n\} \\ &= P\{X(t_{n+1}) = i_{n+1} \mid X(t_n) = i_n\} \end{aligned}$$

That is process {X(t)}, t≥0 is called time-continuous Markov chain.

Probability $P_{ij}(s, t)$ = P{X(s+t) = j|X(s) = i}, i, j∈E is called transition probability function. If transition probability function only relates to time interval t, independent of time start s, process {X(t), t≥0} is called time-continuous homogeneous Markov chain.

Data mining algorithm based on homogenous Markov model: Consider mining object as a state consisting of many states and the transition conforms to half Markov

Table 1: Statistical Data of Bayi team vs. Beijing team

			Spike dead		Buckle		Blocked dead		Blocked back		Blocked up	
Attacking type	Time	(%)	Time	(%)	Time	(%)	Time	(%)	Time	(%)	Time	(%)
Fast Break	18	69.2	9	46.8	1	6.6	1	6.6	2	12.4	5	26.8
Storm	2	5.3	-	-	2	100	-	-	-	-	-	-
Three-dimensional attack	6	28.5	2	41.8	-	-	1	14.5	1	14.5	2	28.8

process. Get state transition probability matrix by summarizing and calculate the reliability. Then set the increment and calculate the reliability difference. Then analyze the sensitivity of reliability to transition rate. The method to calculate reliability is: set Q_{ij} as state transition probability from state i to state j, C_{1j} is the reliability from initial state to state j. Solve C_{1n} (reliability of transition from initial state to successful state):

$$\begin{bmatrix} C_{1n} \\ C_{2n} \\ \vdots \\ C_{nn} \end{bmatrix} = \begin{bmatrix} Q_{11} & Q_{12} & \cdots & Q_{1n} \\ Q_{21} & Q_{22} & \cdots & Q_{2n} \\ \vdots & \vdots & & \vdots \\ Q_{n1} & Q_{n2} & \cdots & Q_{nn} \end{bmatrix} \begin{bmatrix} C_{1n} \\ C_{2n} \\ \vdots \\ C_{nn} \end{bmatrix}$$

The method to calculate difference is: after the reliability calculation, add a tiny increment to each item in transition probability matrix. Then recalculate the reliability. The difference between twice results is the reliability difference. The bigger difference indicates the greater effects of the item in the transition probability on reliability.

Taking an example as two games of league, this research analyzes and studies the results in above algorithm. In the first two games, in the first round, Beijing gives priority to fast break, with three-dimensional attack complementary and minimizes the strong attack, of which the number of fast break is 69.2% of the total number of attacks and 28.5% is the number of three-dimensional attacks. It indicates Beijing tries to overcome the blocking of bayi by rapidly attacking, to get initiatives. The specific data is shown as Table 1.

The Table 1 shows, the ratio of effective blocking is high, of which the ratio of blocked back and blocked up of rapid break reaches 39.2% and the ratio of blocked back and blocked up of three-dimensional attack reaches 43.3%. Meanwhile, it also indicates the Bay's net advantages are obvious, with clear understanding of first-round attacks of Beijing and it conducts effective blockings. To Beijing's rapid break, mainly move fast and stretches in horizon. The result is the scores of fast break and three-dimensional attacks are high, of which the three-dimensional attack mainly is at the back of the central and right area. Therefore, to the Beijing's attack features, Bayi should enhance the blocking arrangement for front fast ball and horizontal stretching.

Table 2: Statistical data of block

Team	Round	Game	Serve way	Score	Cause
Bayi vs. Beijing	37	1	Jump serve	0:1	Blocked dead
Bayi vs. Beijing	37	1	Jump serve	0:2	Re-attack mistake
Bayi vs. Beijing	37	2	Jump serve	15:9	Re-attack mistake
Bayi vs. Beijing	37	2	Jump serve	15:10	Blocked dead
Bayi vs. Beijing	37	2	Jump serve	16:12	Re-attack mistake
Bayi vs. Beijing	37	2	Jump serve	16:13	Re-attack mistake

In the first round of first game, Beijing brings Bay's three blocks. In the beginning of first round, at the scores 15:9 and 16:12, Beijing results in Bayi continuous missing 2 scores, of which Beijing athlete uses smashing jump and jump floating serve. The specific data is shown as Table 2:

The result shows, the block in the first game is caused by blocked die and re-attack mistake. And the twice blocks are caused by re-attack mistakes.

Generally, consider the time from the beginning to first pause as start stage. The data indicates, during the seven games in first two rounds, both two teams have a 8-segment leading with 1, 2, 3 points lead, leading 4 points in 3 games and the leading team wins. Therefore, winning in the start stage can forms strong physiological advantages which is of great significance. However in a game, it has 8-segment behind with 4 points. Finally it loses.

8-16 segments are the middle stage and 16 segments is just the second pause. The result in the middle stage is vital. From the data, in the 16 segment, a team has 1, 2, 3, 4 and even more than 4 points leading, in a game, 0 game, 0 game, 1 game and 3 games respectively and the leading team wins with 100% probability. In modern volleyball game, when the scores are more than 20 points, it is the key to winning. It should capture the chances and decrease the mistakes.

CONCLUSION

Through data analysis, in the key segment, the number of games that a team leading 1, 2, 3 and 4 points is 1, 2, 2, 0 respectively, but the leading team wins. Therefore, to comparable teams, the key to win is whether the current winner can turn the winning advantages to winning result, when both the scores are higher than 20. Meanwhile, the team falling behind should reverse the situation and overcome the rival by functioning tactics properly and the key is to adopt advantages and avoid disadvantages.

December 9, 2013 The amount of volleyball exercise is varying which suites people with different ages, sexes, physical qualities and training degree. To volleyball game, people pay more attention to results. Considering subjective and fuzzy factors and data, deal with data in data envelopment analysis and fuzzy comprehensive evaluation. Then analyze a volleyball game in data mining algorithm based on homogenous Markov model and get fine effects.

REFERENCES

Guo, Q.G. and X.Q. Wang, 2012. Fuzzy comprehensive evaluation model based on DEA and its application. Control Decision, 4: 575-578.

Li, G. and X.S. Qiao, 2012. Situation of application of computer technology in China volleyball area. Bull. Sports Sci. Technol., 3: 123-124.

Zhang, B., 2013. Dynamics mathematical model and prediction of long jump athletes in Olympics. Int. J. Applied Math. Stat., 44: 422-430.

Zhang, T. and X.L. Gu, 2004. Research on application of homogenous Markov chain in evaluation on sports education qualities. Sports Sci. Res., 4: 74-76.

Zhang, Y.J., H.Q. Zhao and J.W. Wu, 2006. Studies of application of data mining in volley tactics analysis. Comput. Appl., 12: 3027-3029.