Factor Analysis based on RoboCup Midfielder Position*

Zekai CHENG, Liang NIE, Wenwen JIN, Zhaolong LING, Feng QIN
Anhui University of Technology, Ma Anshan, China

ABSTRACT: Research on midfield position has an important meaning for the result of RoboCup Simulation 2D. This paper divides the whole region into eight parts, then count the cycle's number of the both player 6, 7, 8 midfielders in the game in the eight regional positions. Next, three common factors were extracted, which were named as midfielder factor, restricted factor and border defense factor. Finally calculate three factors' scores and get the comprehensive evaluation scores. It is found that the top few teams get good results in recent RoboCup World Cup. It confirms the reasonableness of the factor analysis method.
KEYWORD: RoboCup simulation 2D; Factor analysis; Log files; Comprehensive evaluation scores

1 INTRODUCTION

Simulation 2D group games run in the environment of a standard computer, using Client/Server mode, competition both sides between Client and Server through the UDP/IP protocol to communicate information[1-3]. To realize players best decisions in the team of Wright Eagle in University of science and technology of China use the action driven of the Markov Decision Process and reveal in sequence ,build Markov Transition Matrix[4-6].

Factor Analysis is a kind of common factors extracted from the variable group of statistical techniques, the earliest proposed by the British psychologist Charles Spearman in 1904[7]. Wei-kun LIAO, guo-liang CAI, wen-tao TU[8] using factor analysis method evaluate the level of economic development of the cities in jiangsu province in 2002; Lun RAN, jin-lin LI[9] and others using the factor analysis method on 22 listed companies to evaluate the comprehensive performance; Xue-min WANG[10], faced those listed companies in the Shanghai market in 604 financial statements of the ten main financial indicators for factor analysis in 2001, then according to the situation of each factor score of the stock, has made a comprehensive evaluation.

2 FACTOR ANALYSIS

2.1 *Factor analysis principle*

Factor analysis is a multivariate statistical analysis method that concentrate on the study of the relationship among variables, it can transform the variables with complex relationship into a fewer integrated variables.

2.2 *Factor analysis model*

Step 1 Defining the experimental sample variables. Its matrix is shown as expression (1):

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix} \quad (1)$$

Where n is the times of experimental sample, p is the number of variables to be analyzed. Each row of the matrix represents a group of experimental data, the nth row means the nth group experimental data.

Step 2 Standardizing the data, let the mean value of data be zero and the variance of data be one, then the processed data follows Gaussian distribution. For expressing simple, the standardized matrix is marked as X.

* Fund Supported: Major projects supported by natural science in Anhui province in 2014(KJ2014ZD05).

The correlation coefficient matrix is shown as expression (2):

$$A = X'X \tag{2}$$

Supposing the matrix A has eigenvalue $\lambda_1, \lambda_2, \ldots, \lambda_p$, and the corresponding orthogonal eigenvectors are shows as the matrix (3).

$$V = \begin{bmatrix} v_{11} & v_{12} & \ldots & v_{1p} \\ v_{21} & v_{22} & \ldots & v_{2p} \\ \vdots & \vdots & & \vdots \\ v_{p1} & v_{p2} & \ldots & v_{pp} \end{bmatrix} \tag{3}$$

Step 3 Find the component matrix b_{ij} by the method of PCA in this paper despite of so many other methods. The formula is shown as expression (4).

$$b_{ij} = v_{ij} . \sqrt{\lambda_j} \tag{4}$$

Where b_{ij} is the component of the i th variable on the j th common factor.

Take the common factor matrix as F, then

$$F = VX' \tag{5}$$

Where X' is the transposed matrix of X.

It is not difficult to draw a conclusion.

$$FF' = diag(\lambda_1, \lambda_2, \cdots, \lambda_p) \tag{6}$$

In the expression (6), $diag(\lambda_1, \lambda_2, \cdots, \lambda_p)$ represents this matrix which is a diagonal matrix and the diagonal elements are $\lambda_1, \lambda_2, \cdots, \lambda_p$ respectively, and $F_i = V_i X'$, i=1,2,3,...p, that is , F_i is the factor score of the i th group experimental sample. Now choose m (m<p) main factors, and extract the first main factor from the correlation matrix between the variables, and drown as F_1, and let the common factor F_1 have a biggest variance proportion in all the variables. Then, the effect of F_1 could be ignored; Next, choose the factor F_2 which is not relate to F_1 from correlation matrix. And so on, it is not until the communality of all variables could be resolved finished.

Finally, the model is shown as expression (7).

$$\begin{cases} X_1 = b_{11}F_1 + b_{12}F_2 + \cdots + b_{1m}F_m + \varepsilon_1 \\ X_2 = b_{21}F_1 + b_{22}F_2 + \cdots + b_{2m}F_m + \varepsilon_2 \\ \vdots \\ X_p = b_{p1}F_1 + b_{p2}F_2 + \cdots + b_{pm}F_m + \varepsilon_p \end{cases} \tag{7}$$

The residual $\varepsilon_1, \varepsilon_2 \ldots, \varepsilon_p$ are special factors. It should be ignored.

2.3 *Calculating the scores of factors*

Because the common factors could reflect the correlation of original variables, and make these original variables expressed as a linear combination of the common factor. If these factors are expressed as a linear combination of original variables, just as the expression (8) shows.

$$\begin{cases} F_1 = \beta_{11}X_1 + \beta_{12}X_2 + \cdots + \beta_{1p}X_p \\ F_2 = \beta_{21}X_1 + \beta_{22}X_2 + \cdots + \beta_{2p}X_p \\ \vdots \\ F_m = \beta_{m1}X_1 + \beta_{m2}X_2 + \cdots + \beta_{mp}X_p \end{cases} \tag{8}$$

2.4 *Comprehensive evaluation scores*

Take each factor variance contribution rate as weighting, and use factor rotation method, combining the variance contribution rate after factor rotation and these factor extracted constitute the linear equations. Finally establish a comprehensive evaluation index function, as shown in type (9):

$$F_{\text{total}} = \frac{\alpha_1 F + \alpha_2 F_2 + \cdots + \alpha_m F_m}{\alpha_1 + \alpha_2 + \cdots + \alpha_m} \tag{9}$$

Where α_i is the contributing rate of variance of i th common factor before rotating or after rotating.

3 EXPERIMENT

3.1 *Data modeling*

3.1.1 *Obtaining log files*

In RoboCup simulation 2D game, the server generates log files, RCG and RCL files. The RCG files records the state of ground in each period, including the ball's coordinate, ball's speed and the coordinate and stamina of all agents, and so on. The RCL files records some command information, etc. The experiment data in this paper is obtained from analyzing log files.

3.1.2 *Choosing variables*

The field is divided into eight regions, just as figure1 shows.

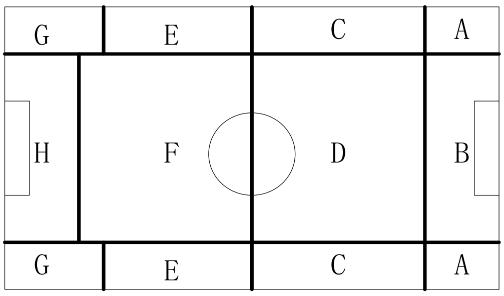

Figure 1. The division of field

Now analyzing the data of player 6, player 7, player 8 in one game randomly. As is shown in Table 1.

Table1. Data of midfield player(unit: cycles)

	Team1 Player number			Team2 Player number		
	6	7	8	6	7	8
X_1	0	127	64	0	36	0
X_2	0	163	123	0	6	42
X_3	197	377	1400	3	1028	1006
X_4	2497	3111	2077	2648	1712	1474
X_5	406	117	1079	0	533	529
X_6	2263	2108	1131	2848	2147	2382
X_7	328	0	118	0	372	112
X_8	308	0	6	499	175	453

3.2 *Data analysis*

Taking 60 games' log files as experiment data, the results of experiment KMO and Bartlett Test are shown in table 2.

The result of table 2 shows that the KMO statistics equals 0.721 which is greater than 0.7.

Table 2. KMO and Bartlett Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy		.721
Bartlett's test of Sphericity	Approx. Chi-Square	763.309
df		28
Sig.		.000

The contributing rate of variance and cumulative contribution of variance in each common factor is shown as figure 2. From the figure, the eigenvalue of three common factors are 3.226,1.724,1.087, respectively, the cumulative contribution of three common factors is 75.453%.

	Component		
	1	2	3
X1	.234	.813	.009
X2	.070	.860	-.048
X3	.865	.100	-.007
X4	.736	.213	-.524
X5	.087	.032	.816
X6	-.969	-.079	-.098
X7	-.228	-.420	.631
X8	.025	-.707	.447

Figure 2. Factor loading matrix after rotation

3.3 *Factor rotation*

From figure 2, the correlation between F_1 and X_3, X_4, X_6 is higher, and named midfielder factor. The correlation between F_2 and X_2, X_8 is higher, and named restricted area factor. The correlation between F_3 and X_5, X_7 is higher, and named border defense factor. These three factors could reflect clearly the significance of common factor which is found through the method of factor analysis, and table 3 shows the results.

Table 3. Common factor named

Common factor	Explained variables	Named
F_1	X_3, X_4, X_6	Midfielder factor
F_2	X_1, X_2, X_8	Restricted area factor
F_3	X_5, X_7	Border defense factor

3.4 *The score of factors*

Using regression method to find factor score function and common factor and linear relationship between the original variables is shown as type(10):

$$\begin{aligned} F_1 &= 0.008X_1 - 0.078X_2 + 0.393X_3 \\ &\quad + 0.291X_4 + 0.08\mathrm{X}_5 - 0.449X_6 \\ &\quad - 0.021X_7 + 0.124X_8 \\ F_2 &= 0.445X_1 + 0.479X_2 - 0.052X_3 \\ &\quad - 0.09X_4 + 0.196\mathrm{X}_5 + 0.053X_6 \\ &\quad - 0.07X_7 - 0.311X_8 \\ F_3 &= 0.207X_1 + 0.165X_2 + 0.063X_3 \\ &\quad - 0.311X_4 + 0.633\mathrm{X}_5 - 0.143X_6 \\ &\quad + 0.371X_7 + 0.178X_8 \end{aligned} \tag{10}$$

3.5 *Comprehensive evaluation score*

By taking factor's variance contribution after rotation as weighting, then establish comprehensive evaluation index and extract the common factor F_1, F_2, F_3.The linear relation is shown as type(11):

$$\begin{aligned} F &= 0.2943F_1 + 0.26746F_2 \\ &\quad + 0.19363F_3 \end{aligned} \tag{11}$$

By this formula, calculate the 60 sets of original data's comprehensive evaluation scores. Now the scores of the top five largest are as shown in figure3.

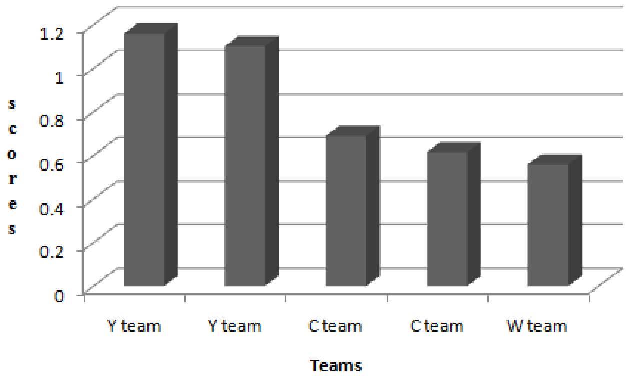

Figure 3. Comprehensive Evaluation Score

3.6 *Results*

Concluded from the combinaiton of theory and practise, the comprehensive evaluation scores of the three factors including midfielder factor, restricted area factor, border defensive factor that are extracted from the midfielder position factor analysis in Robocup Simulation 2D, can reflect the strength of the team to a certain extent.

4 CONCLUSION

In this paper, the method of factor analysis is applied to Robocup Simulation for the first time. Using the thinking of data analysis and data mining to analyse the midfield position with factor analysis, we can extract three common factors including midfield, restricted area and border defense, which can reflect the conclusion of each simulation team strength to a certain extent.

REFERENCES

[1] Peter, Stone. Veloso M. Multiagent Systems: A survey from a Machine Learning Perspective. Automous Robots, 2000, 8(3):345-383.

[2] Guo Ye-jun, Xiong rong. RoboCup robot soccer simulation game development and design. Computer Engineering and Applications，2003, 29:146-149.

[3] Yang Yu-pu, Li Xiao-meng, Xu Xiao-zing. A survey of technology of multi-agent cooperation. Information and Control, 2001, 30(4):337-342.

[4] Shi ke. Research on agent decision problems Markov Decision Process Theory. Master's thesis in University of Science and Technology of China 2010.

[5] Feng Wu, Shlomo Zilberstein, Xiaoping Chen. Online Planning for Multi-Agent Systems with Bounded Communication, Artificial Intelligence, 2011, 175(2): 487-511.

[6] Pan ling-han, Chu wei, Cheng xian-yi. RoboCup Soccer strategies based on role. Computer Engineering and Applications, 2004, 26: 66-69.

[7] Su Li-yun, Chen CaiXia, Gao Hong-xia. SPSS19 application of statistical analysis and case-based tutorial. Beijing Hope Electronic Press, 2012.

[8] Liao Wei-kun, Cai Guo-liang, Tu Wen-tao. Urban Economic Development Evaluation Based On Factor Analysis. Statistics and Decision, 2005.

[9] Ran lun, Li Jin-lin. Factor Analysis in the SME sector companies listed on the consolidated results of the evaluation of the application. Mathematical Statistics and Management, 2005, 24 (1): 75-79.

[10] Wang Xue-min. Factor analysis on the evaluation of the stock. Mathematical Statistics and Management, 2004, 23 (3):6-10.