Research on RoboCup passing ball based on partial least square

Yuan Song
Department of Electronic Information
Maanshan Technical College
Ma'anshan, Anhui, China
songyuan@mastc.edu.cn

Ze-Kai Cheng
School of Computer Science
Anhui University of Technology
Ma'anshan, Anhui, China
chengzk@ahut.edu.cn

***Abstract*—Passing ball action runs through the RoboCup simulation games, both teams spare no effort to win the game, the paper is going to study the connection between passing ball action and the game deeply. It puts forward to adopt the idea of Data Mining, analyzes games' log files by C language program in order to collect the required passing ball data, divides passing ball into 5 types which are seen as independent variables and see score as dependent variable, then establishing mathematical model combining with partial least-square(PLS).A few relevant figures are used to analyze and verify the experimental result which comes from SIMCA-P. The VIP values of 5 independent variables to dependent variable are as follows: 0.07980, 0.98735, 1.0.888, 1.10082, 1.33668 under the circumstance that carrying 73.2% of independent variables information and 74.3% of dependent variable information. After linking theoretical result with practical scene, it is concluded that long pass plays a major role in a game for passing ball.**

Keywords—multi agent system;log files;mathematical modeling;partial least-square; regression analysis;multicollinearity

I. INTRODUCTION

RoboCup simulation soccer game is a multi-agent platform for collaboration and confrontation, it provides us with a real-time environment which is dynamic and noisy, it executes by C/S mode, and each side has 11 clients that exchange information with server by UDP/IP protocol[1].The RoboCup simulation game embodies the characteristics of human soccer and provide an important experimental platform for distributed artificial intelligence and multi agent system. Users can use various programming language under different operating system and construct a team with different technology like Mathematical Modeling, Search Reasoning, Machine Learning, Dynamic Programming etc [2].

In the simulation game, the basic coordinated action between agents is pass ball which plays a vital cohesive role both in the attack and defensive state. Based on the literature, now many scholars optimize the model of pass ball based on the thought of projecting, literature[3] use the method of geometric modeling to improve pass accuracy, literature[4] use the method of off-line learning to train pass through combining the Q-learning and neural network, literature[5] use decision-tree algorithm to find the player who has the optimal pass success rate, literature[6] use pass evaluation function to determine the optimal direction which is safe and good for attacking, literature[7] use fuzzy logic algorithm and heuristic search algorithm to plan the pass route.However, it makes the research on pass still confined to the scope of the underlying decision.The quality of cooperation strategy between the agents largely determines the strength of a team,this paper proposes to adopt the thought of data mining, as a result, by mining the pass data some internal and implied information can be achieved which will be seen as reliable theoretical guidance for a team to make effective high-level decision.

Supported by Major project of Nature Science Research of Department of Education of Anhui Province(No:KJ2014ZD05)

II. SINGLE DEPENDENT VARIABLE PARTIAL LEAST SQUARES

A. The application of Partial Least Squares

Partial Least Square is a new multivariate statistical analysis method which was first proposed[8] by S. Word and C. Albano in 1983, it attracts researchers' attention since it can solve the problem that ordinary multivariate regression analysis can not ,and its practical application involves many fields like chemical, mechanical, biological, geological, social sciences and economics etc. Zhou qiang use PLS regression analysis to find abnormal value in data mining, establishing corresponding regression analysis model combining with the real data, then the experimental conclusion got reasonable explanation[9]. Luo wei use PLS to forecast the cost of military UAV development,and making the results compared with results which get from the method of SMR,BP or RBF neural network, according to the comparative result, the conclusion shows that method of PLS get a higher accuracy[10]. Qi haoping applies PLS into research on the relational model between urban land utilization and traffic volume, by analyzing, we know that residential estate , building volume rate and overall floorage have marked influence on traffic volume, the predictive model can be tested by the real physical data, and we can find that the traffic volume model based on PLS has an excellent precision[11]. Liu gui xiong study the main influence factor of ultra-weak luminescence of the fish in breeding stage, by analyzing PLS regression model of ultra-weak luminescence of the carp we concluded that the most significant index that affect ultra-weak luminescence of carp is temperature and gonad maturation coefficient [12]. In this paper, PLS is applied to mathematical modeling and regression analysis of pass data in RoboCup in order to improve the ability to explain and

overcome the problem of multicollinearity among the explanatory variable.

As the literature[13] says, the higher the correlation degree of explanatory variable, the worse multicollinearity problem will be, and the bigger the estimated value of variance of regression coefficient will be, therefore, the accuracy of the model will be sharply reduced, even causing severe situation that the sign of regression coefficient maybe inverted, now the application value of traditional least square estimate formula $\hat{\beta}=(X^TX)^{-1}X^Ty$ decreased according. PLS adopt the idea of component extraction, it can not only ensure that reflect the information of original explanatory variables as much as possible, but ensure that has a strong explanatory ability to the dependent variable, moreover, since the components are independent each other, so the multicollinearity among explanatory variables can be effectively avoided. The single variable partial least squares is an special case of partial least squares, and the set of dependent variable is just a vector[14].

B. The Deduction of Single Dependent Variable Partial Least Squares Algorithm[15]

Step 1 : Data standardization. Record $F_0(F_0\in R^n)$ as the standard vector of dependent variable y, E_0 is the standard matrix of independent variables X.

Step 2 : Extract a component t_1 from E_0, $t_1=E_0w_1$, $w_1=\frac{E_0{}^TF_0}{\|E_0{}'F_0\|}$.

Step 3 : E_0 and F_0 is implemented regression on t_1, that is $E_0=t_1p_1{}^T+E_1$, $F_0=t_1r_1+F_1$ (p_1 and r_1 are regression coefficients).According to the theory of partial least squared regression, we can get the regression coefficient $p_1=\frac{E_0{}^Tt_1}{\|t_1\|^2}$, $r_1=\frac{F_0{}^Tt_1}{\|t_1\|^2}$, record residual matrix as $E_1=E_0-t_1p_1{}^T=(E_{11},E_{12}\cdots E_{1p})$, $F_1=F_0-t_1r_1$.

Step 4 : Check the convergence, if the regression equation of y on t_1 has already reached a satisfactory accuracy, go to the next step; if not, replace E_0 with E_1,and replace F_0 with F_1, then back to step 2, extract new components from residual matrix.

Step 5 : Component $t_1,t_2,\cdots t_m(m<Rank(X))$ have been achieved, the regression model of F_0 on t_h will be $\hat{F}_0=r_1t_1+r_2t_2+\cdots r_mt_m$ .Furthermore, E_{0j} is the linear combination of $t_1,t_2,\cdots t_m$, so regression model F_0 on t_h is $\hat{F}_0=r_1E_0w_1{}^*+r_2E_0w_2{}^*+\cdots r_mE_0w_m{}^*$,thereinto, $w_h{}^*=\prod_{j=1}^{h-1}(I-w_jp_j{}^T)w_h$, and I is unit matrix. If record $x_j{}^*=E_{0j},y^*=F_0$, so $\hat{y}^*=\alpha_1x_1{}^*+\alpha_2x_2{}^*+\cdots\alpha_qx_q{}^*$, the regression coefficient of $x_j{}^*$ is $\alpha_j=\sum_{h=1}^{m}r_hw_{hj}{}^*$, $w_{hj}{}^*$ is the first j component of $w_h{}^*$.

C. The Discrimination of The Cross Validation

In general, regression model does not need to extract all the components, we investigate that adding a new component and whether the explanatory ability of independent variable can obviously improved .

Record y_i as original data, $t_1,t_2,\cdots t_m$ are the extracted components during the regression process of partial least squares. $\hat{y}_{hi}$ is the fitted value of numbered i sample point after using all sample points and take number h component for regression modeling. Delete number i sample point, take number h component to do regression modeling, using this model to calculate the fitted value of y_i, record its result as $\hat{y}_{h(-i)}$. Record $S_{SS,h}=\sum_{i=1}^{n}(y_i-\hat{y}_{hi})^2$, $S_{PRESS,h}=\sum_{i=1}^{n}(y_i-\hat{y}_{h(-i)})^2$, $Q_h^2=1-\frac{S_{PRESS,h}}{S_{SS,h-1}}$, only if $Q_h^2\geq 0.0975$ (that is $\sqrt{S_{PRESS,h}}\leq 0.95\leq\sqrt{S_{SS,h-1}}$), we think the new introduced component t_h would have an obvious improvement for the model.

D. Variable Importance in Projection analysis

The importance of x_j when explaining y can be measured by variable importance in projection (VIP). $VIP_j=\sqrt{\frac{p}{Rd(y;t_1,t_2,\cdots t_m)}\sum_{h=1}^{m}(Rd(y;t_h)\times w_{hj}^2)}$, there into, w_{hj} is the number j component of w_h axis; $Rd(y,t_h)=r^2(y,t_h)$ represents the explanatory ability of t_h to y ; $Rd(y;t_1,t_2,\cdots t_m)$ represents the cumulative explanatory ability of $t_1,t_2,\cdots t_m$ to y .

III. EXPERIMENT RESULT AND ANALYSIS

A. Data Modeling

*1) RoboCup Log File:*Server generated log file which recorded simulation games with the real data of ground and players. Developers always use log file to replay games for checking errors of code which control the agent, then modify the code again. There are two kinds of log files, they are RCG and RCL, each record is stored as string in fixed rules, RCG file records the ground state of each period, including ball coordinates and speed, players' coordinates and stamina, etc. RCL mainly records some command information, like Kick, Turn, Say, etc. The paper will make full use of the rich data in RCL log file.

*2) Parser Log File:*In order to get the analytical data of pass, use C language to define relevant data structure based on

the storage rule of log file for parsing strings. The pseudocode of parsing is as follow:

```
'STEP1: Get team information of each side through filename ;
    Import (rclfilename);
    Call anafilename(rclfilename,team1name,team2name, goal);    //Get goal and name of each team
'STEP2: Get the information of kick action from RCL file
    Parse the current cycle and team name tn;
    if Call strcmp(team1name,tn)=1 Then    //1 and 2 represent two teams respectively
        flag=1;
        Else flag=2;
    EndIf
    if *p='k' Then //Just study pass, so only need the information of kick command
        If flag=1 Then //Deal with team 1
            If flag=reflag Then //The current flag is the same with the flag of executing the last kick command, then regard the action as a pass.
                Begin
                    pcycle=cycle-recycle; //Current cycle minus the last cycle is the pass cycle
                    balltype ← getballtype(pcycle); //Judge the type of pass based on pass cycle
                    Count[flag][balltype]++; //Counting
                    reflag ← flag;
                    recycle ← cycle;  //Setting current cycle as previous cycle in order to parsing the next record
                End
            EndIf
        Else //Team 2 do the same dealing with team 1
        EndIf
    EndIf
```

3) The Selection of Variable: As we know, there are 5 types of pass in RoboCup, they are Holdball, Dribble, Directpass, Leadpass, Throughpass. Before parsing log files, we set the specific cycle interval according to the duration of pass experimentally, as it shown in Table1.Then counting all types of pass of each team.

TABLE I. THE CYCLE INTERVAL OF THE CLASSIFICATION OF PASS TYPE (UNIT: SECOND)

	Holdball	Dribble	Directpass	Leadpass	Throughpass
cycle interval	[0,1]	(1,3)	[3,10)	[10,20)	$[20,+\infty)$

For example,parse a whole game arbitrarily, the result is stated in Table 2.

TABLE II. THE ANALYTICAL RESULTS OF A GAME(UNIT:TIMES)

	Score	Holdball	Dribble	Directpass	Leadpass	Throughpass
Team 1	3	1	352	40	18	5
Team 2	0	3	174	50	29	3

To study the relationship between pass and the game, do subtraction of two teams' data, now a data record which has 6 properties can be obtained. Using y represent the score subtraction and see it as dependent variable, regarding five types of pass as explanatory variable and x_1, x_2, x_3, x_4, x_5 represent five types of pass severally.

B. Data Analysis

This paper will use the parse results of 60-times simulation competition as the modeling data. Firstly, do the correlational analysis about the observation data, Table 3 shows the correlative coefficient matrix of dependent variable and explanatory variable. Obviously, there is serious multicollinearity problem among the explanatory variables, such as $r(x_2,x_3)=0.8543$, $r(x_2,x_4)=0.7965$, $r(x_3,x_4)=0.8335$.

TABLE III. THE CORRELATION COEFFICIENT BETWEEN DEPENDENT VARIABLE AND EXPLANATORY VARIABLE

	y	x_1	x_2	x_3	x_4	x_5
y	1	-0.0238	0.6806	0.6662	0.7192	0.8116
x_1		1	0.0021	-0.1197	-0.0144	-0.0384
x_2			1	0.8543	0.7965	0.5000
x_3				1	0.8335	0.5923
x_4					1	0.6149
x_5						1

C. PLS component extraction

Automatic fitting by SIMCA-P, choose 2 PLS components according to the cross validity index.The standardized regression coefficient histogram of explanatory variable shown in the Fig.1,So the standardized PLS regression model is $\hat{y}^* = -0.006x_1^* + 0.172x_2^* + 0.022x_3^* + 0.203x_4^* + 0.637x_5^*$

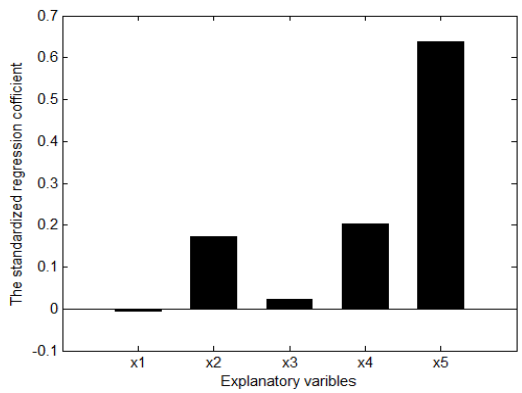

Fig. 1. The standardized regression coefficient histogram

.

D. The Precision Analysis

Table4 shows that the extracted two components have already included 73.2% information of explanatory variable, and it can explain 74.3% information of dependent variable y .

TABLE IV. THE EXPLANATORY PRECISION OF COMPONENTS TO DEPENDENT VARIABLE AND EXPLANATORY VARIABLE

	component1	component2
The value of Q^2	0.638	0.209
The explanatory ability to X	62.3%	10.9%
The cumulative explanatory ability to X	62.3%	73.2%
The explanatory ability to y	65.0%	9.3%
The cumulative explanatory ability to y	65.0%	74.3%

E. The Analysis of important index

1) The analysis of the relation between explanatory variable and dependent variable.

t_1 and u_1 are the first main component of X and y ,so the relation between t_1 and u_1 can represent the relation between explanatory variable and dependent variable in PLS analysis[15].As it shown in Fig.2, there is obvious linear relationship between t_1 and u_1 , that is to say, there is also obvious linear relationship between explanatory variable and dependent variable. It comes to conclusion that the established model is reasonable.

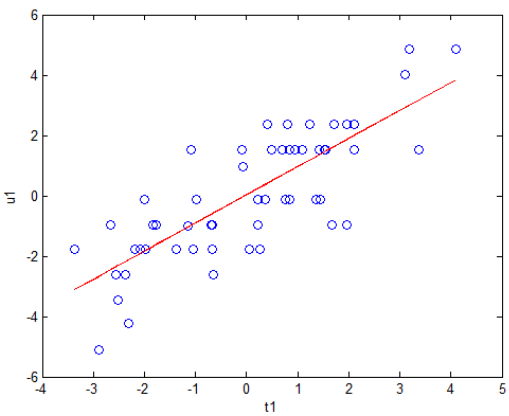

Fig. 2. t_1 - u_1 scheme

2) Importance analysis of explanatory variable to dependent variable.

The VIP value histogram is shown in Fig.3, its corresponding numeric value is shown in Table 5.

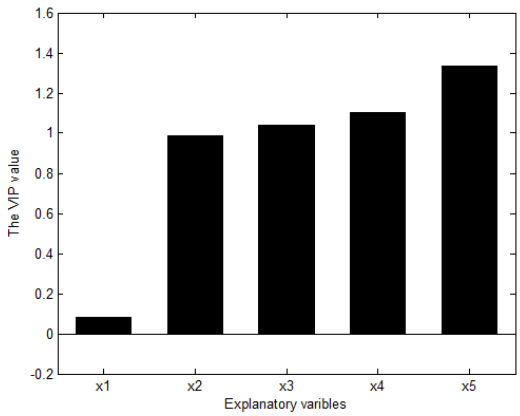

Fig. 3. Variable Importance in Projection histogram of explanatory variable

TABLE V. THE VIP VALUE OF EXPLANATORY VARIABLE

	x_1	x_2	x_3	x_4	x_5
VIP value	0.07980	0.98735	1.03888	1.10082	1.33668

From Fig.1, we assume that only x_1 showing a negative correlation with y , the others are positive correlation.In standardized regression coefficient,the maximum coefficient is x_5 and the minimum is x_1 , the descending order is x_5, x_4, x_3, x_2, x_1 . From Fig.3,we consider that x_5 have the biggest impact to y ,while x_1 have the smallest ,and the descending order x_5, x_4, x_3, x_2, x_1 .

Combining with the actual game scenes, $x_5(pcycle \geq 20)$ equals the long pass, and $x_1(0 \leq pcycle \leq 1)$ equals Holdball in the game. Throughpass could mostly stand for the long pass, the more times of long pass,the stronger the offensive situation will be, and the more likely to goal. However, the more times of HoldBall, the easier the ball will be intercepted by opponent, and that may loose offensive opportunities or cause unsuccessfully defend, even loose score. So, the theoretical result accord with the actual situation.

Combining theory with real situation, it is concluded that long pass is key to the outcome of the game to pass.

IV. CONCLUSION

The paper firstly introduces partial least square into the study of RoboCup 2D, mining out that long pass plays a key role to the outcome of the game according to the thought of data mining, then analyzing the rationality of result is combining with some related figures. Our next work is regard the conclusion as the high-level decisions, making appropriate adjustments in the code of pass action evaluating, that is increasing the weight of executing Throughpass. Through observing the simulation game before this study, we subjectively deduce that Leadpass is the key which decides the outcome of game, but the theoretical result is Throughpass, the following factors may cause the different result: 1.The amount of sample data is not much enough, it may influence the experimental result. 2.The cycle interval of pass is defined by

our experience, if we do the data discretization much more better, that we will get much more accurate results. Though there are a few insufficiency, the emphasis of this paper is put forward a new idea for the study of RoboCup2D, that is data mining.

The partial least squares adopts the idea of decomposing data information, regroup information according to the variation degree of the overall data, thus the meaningless variable could be eliminated. Reduction dimension for the high dimension data by this method, there will be special application value in the today's situation that processing Big Data difficultly.

Acknowledgment

Firstly, thank you for the support from Major project of Nature Science Research of Department of Education of Anhui Province(KJ2014ZD05). And to give thanks for the friends' helpful comments as well as the support from family. The authors thankfully acknowledge the computer resources and technical expertise which is provided by anonymous researchers on network.

References

[1] Peter stone.Layered Learning in Multi-agent systems[D].Pittsburgh: Carnegie Mellon University，1998.

[2] USTC simulation 2D team WrightEagle.Robert soccer : Design and Implementation[EB/OL]. http://wenku.baidu.com/view/ae0c20ee6294dd88d0d26be0.html,2013.7.4/2013.9.7.

[3] GUO Bo,CHENG Jia-xing.Strategy of Pass in RoboCup Simulation[J],2006,16 (2):129-131.

[4] Yong Zhou,Liu Feng.Research of RoboCup Pass Strategy Based on Improved Q-Learning[J].Computer technology and development，2008，18（4）:63-67.

[5] ZHANG Jia-wang,HAN Guang-sheng,ZHANG Wei.Application of C5.0 Algorithm in Passing Ball Training of RoboCup[J].Computer Simulation,2006，23（4）:132-135.

[6] Xiaobing Zhang,Chunyan Liu,Li Chen.RoboCup Passing Strategy Based on the Passing Evaluation Function[J].J.of Anhui University of Technology,2011，28(2),171-174.

[7] Li jiyao,Chen Wei,Zhang Qi.Intelligent Algorithms in the Simulation of Robotic Soccers[J].Control Theory and Applications,2007，26（9）:7-9.

[8] S.Wold,H.Martens,H.Wold.The multivariate calibration problem in chemistry solved by the PLS method[C]//Matrix Pencils.Springer-Verlag Berlin Heidelberg, 1983:286- 293.

[9] Zhou Qiang,Ouyang Yiming,Hu Xuegang,Wanghao.Using Partial Least-Squares Regression To Found Abnormal Data in Data-Mining.Microelectronics,2005,22(1):25-31

[10] Luo Wei,Liu Lu.Prediction o fm ilitary UAV deve lopment cost based on partial leas-t square regressionmethod[J].Journal of Beijing University of Aeronautics and Astronautics,2010,36(6):667-670.

[11] Qi Haoping,Zhang Li,Wang Wei,Lu Jian.Research on Model of Relationship between Urban Land use and Traffic Volume Based on Paflial Least Squares Regression[J].Journal of Highway and Transportation Research and Development, 2011,28(3):138-143. partial leas-t square regressionmethod[J].Journal of Beijing University of Aeronautics and Astronautics,2010,36(6):667-670.

[12] Liu Guixiong,Lin Xuhong.Partial Least-Square Regressive Analysis and Modeling of Ultraweak Photon Emission of Fish[J].Journal of South China University of Technology,2006，34（11）：29-32.

[13] Xiaoqun He.Statistical analysis[M].Bei jing : Renmin University of China press.2012.

[14] Huiwen Wang.Partial least squares regression method and its application[M].Bei jing: National defence Industry Press,1999.

[15] Huiwen Wang.Partial Least-Squares Regression—Linear and Nonlinear Methods[M]. Beijing:National Defense Industry Press,2006.