A Neural Network Method for Prediction of 2006 World Cup Football Game

Kou-Yuan Huang, *Senior Member, IEEE* and Wen-Lung Chang

***Abstract*—A neural network method is adopted to predict the football game's winning rate of two teams according to their previous stage's official statistical data of 2006 World Cup Football Game. The adopted prediction model is based on multi-layer perceptron (MLP) with back propagation learning rule. The input data are transformed to the relative ratios between two teams of each game. New training samples are added to the training samples at the previous stages. By way of experimental results, the determined neural network architecture for MLP is 8 inputs, 11 hidden nodes, and 1 output (8-11-1). The learning rate and momentum coefficient are sequentially determined by experiments as well. Based on the adopted MLP prediction method, the prediction accuracy can achieve 76.9% if the draw games are excluded.**

I. INTRODUCTION

NEURAL network methods have been used in the analysis of sport data. Purucker employed the supervised and unsupervised neural networks to analyze and predict the winning rate of NFL [1]. Park and Newman used network to show the ranking of American college football teams [2]. Rotshtein, Posner, and Rakityanskaya combined genetic algorithm and neural network fuzzy model to predict the soccer game of Finland [3]. In this paper, we adopt the supervised multi-layer perceptron neural network (MLP) with error back propagation learning rule (BP) to predict the winning rate for 2006 World Cup Football Game (WCFG).

According to the schedule of 2006 WCFG, shown in Fig. 1, there are 32 teams enter into this competition and overall 64 matches at 5 stages in this tournament from the beginning to the end. The competition rules in each stage are explained as follows.

1) Stage 1 is the group match, also known as round robin tournament. There is no extending time after 90 minutes regular time. In this stage, there are 32 teams in 8 groups (Group A-H), each group has 4 teams, and each team has 3 matches. There are 6 matches in each group and there are 8 groups, so totally it has 48 matches (Match 1 - 48) in stage 1. The criterion of gaining points is that winning one game gains 3 points, losing one game gains 0 point, and drawing one game gets 1 point. After stage 1, two teams that get the higher points in each group enter into next stage. Table I lists the league table for 32 teams in 8 groups after 48 matches finished at stage 1.
2) The competition rule of stage 2~5 is single elimination tournament. That means it is necessary to have penalty kick until the victory is decided for each match if two teams tied after regular time (90 minutes) and additional time (30 minutes) finished. The winning teams will get the qualification to enter the next stage and the losing teams will be eliminated from the competition. Stage 2 is the round of 16, and there are 8 matches (Match 49 – 56) for 16 teams. Stage 3 is the quarter-finals, and there are 4 matches (Match 57 - 60) for 8 teams. Stage 4 is the semi-finals, and there are 2 matches (Match 61 - 62) for 4 teams. Stage 5 is the final-game, and there are 4 teams (the same teams as in stage 4) for 2 games (one is the third place game (Match 63) and the other is the final game (Match 64)).

From the website of 2006 WCFG held in Germany [4], we get the official 64 matches' statistical records provided by FIFA [5]. From the report of each match, there are 17 statistical items, *Goals For, Goal Against, Shots, Shot On Goal, Penalty Kicks, Fouls Suffered, Yellow Cards, Red Cards, Corner Kicks, Direct Free Kicks to Goal, Indirect Free Kicks to Goal, Offside, Own Goals, Cautions, Expulsions, Ball Possession, and Foul Committed*, which represent the ability index to win the game. From these statistical data, we apply a neural network method to predict the winning rate of two teams at the next stage games by means of their previous games' statistic data. Fig. 2 shows the supervised prediction system, which is composed of two parts: training part and prediction part.

Kou-Yuan Huang is with the Department of Computer Science, National Chiao Tung University, Hsinchu, Taiwan. (corresponding author to provide e-mail: kyhuang@cs.nctu.edu.tw)

Wen-Lung Chang is with the Department of Electrical and Computer Engineering, National Chiao Tung University, Hsinchu, Taiwan. (e-mail: eltonchang2001@gmail.com)

978-1-4244-8126-2/10/$26.00 ©2010 IEEE

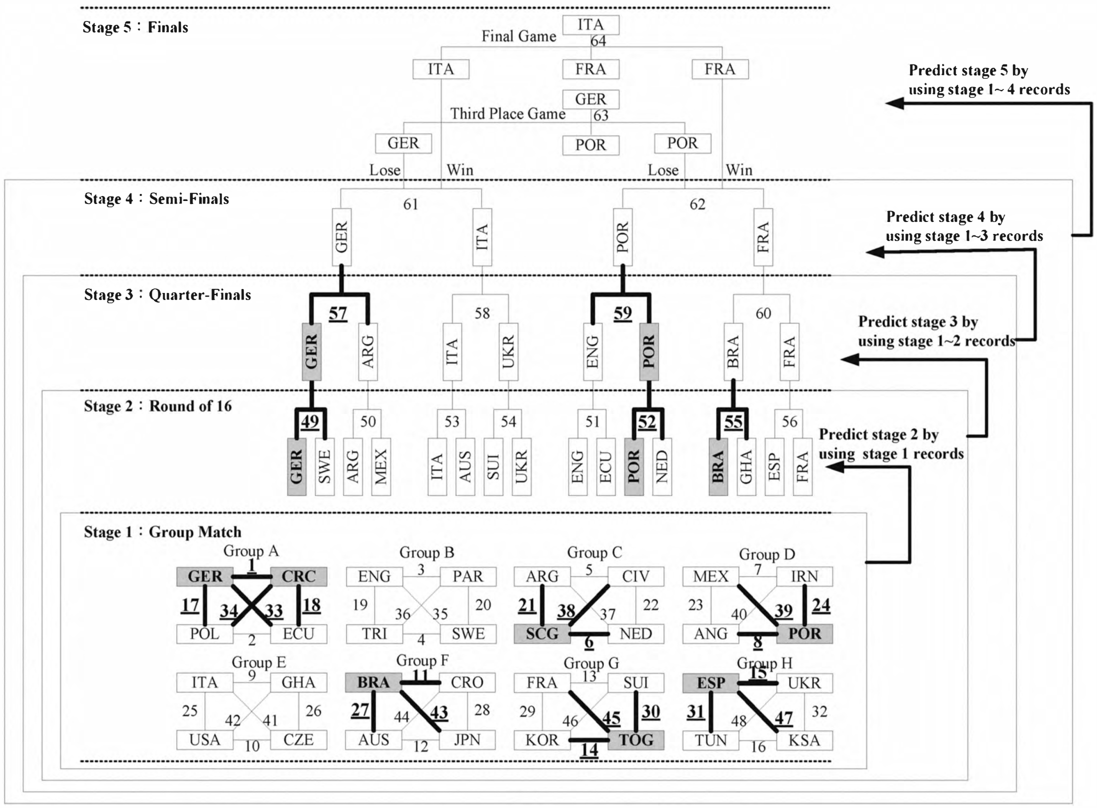

Fig. 1. Total 64 matches at 5 stages for 32 teams in the schedule of 2006 WCFG.

TABLE. I

LEAGUE TABLE AFTER FINISHED 48 MATCHES OF STAGE 1

GROUP	KEYS (Team)	Wins	Losses	Draws	Games Plays	Goals For	Goals Against	Winning rate	Points
A	Germany	3	0	0	3	8	2	1	9
	Ecuador	2	1	0	3	5	3	0.667	6
	Poland	1	2	0	3	2	4	0.333	3
	Costa Rica	0	3	0	3	3	9	0	0
B	England	2	0	1	3	5	2	0.667	7
	Sweden	1	0	2	3	3	2	0.333	5
	Paraguay	1	2	0	3	2	2	0.333	3
	Trinidad and Tobaga	0	2	1	3	0	4	0	1
C	Argentina	2	0	1	3	8	1	0.667	7
	Netherlands	2	0	1	3	3	1	0.667	7
	Cote d'Ivoire	1	2	0	3	5	6	0.333	3
	Serbia-Montenegro	0	3	0	3	2	10	0	0
D	Portugal	3	0	0	3	5	1	1	9
	Mexico	1	1	1	3	4	3	0.333	4
	Angoia	0	1	2	3	1	2	0	2
	Iran	0	2	1	3	2	6	0	1
E	Italy	2	0	1	3	5	1	0.667	7
	Ghana	2	1	0	3	4	3	0.667	6
	Czech Republic	1	2	0	3	3	4	0.333	3
	United States	0	2	1	3	2	6	0	1
F	Brazil	3	0	0	3	7	1	1	9
	Australia	1	1	1	3	5	5	0.3333	4
	Croatia	0	1	2	3	2	3	0	2
	Japan	0	2	1	3	2	7	0	1
G	Switzerland	2	0	1	3	4	0	0.667	7
	France	1	0	2	3	3	1	0.333	5
	South Korea	1	1	1	3	3	4	0.333	4
	Togo	0	3	0	3	1	6	0	0
H	Spain	3	0	0	3	8	1	1	9
	Ukraine	2	1	0	3	5	4	0.667	6
	Tunisia	0	2	1	3	3	6	0	1
	Saudi Arabia	0	2	1	3	2	7	0	1

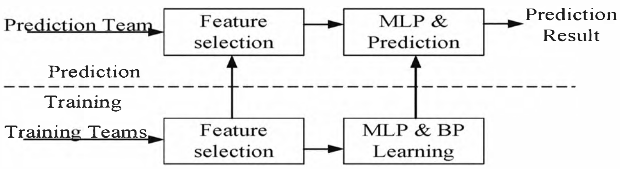

Fig. 2. Supervised prediction system.

II. FEATURE SELECTION AND NORMALIZATION

A. *Feature Selection*

We get 64 matches' reports from [4] and there are 17 statistical items in each match report. Base on experience and general sense, we select 8 items which effectively represent the significant capability to win the game as the input features. These 8 features are marked as: x_1 = Goals For (GF), x_2 = Shots (S), x_3 = Shots On Goal (SOG), x_4 = Corner Kicks (CK), x_5 = Direct Free Kicks to Goal (DFKG), x_6 = Indirect Free Kicks to Goal (IDFKG), x_7 = Ball Possession (BP), and x_8 = Fouls Suffered (FS).

B. *Normalization: Relative Ratio as input feature*

We consider relative ratio as input feature. Training samples and prediction samples are normalized by relative ratio as (1).

$$\text{If } x_{iA} = x_{iB}, \text{ then } y_{iA} = y_{iB} = 0.5$$
$$\text{else } y_{iA} = \frac{x_{iA}}{x_{iA}+x_{iB}},\ y_{iB} = \frac{x_{iB}}{x_{iA}+x_{iB}},\ i = 1,\ldots,8 \quad (1)$$

The input features of $x_1 \sim x_8$ are converted into $y_1 \sim y_8$ by (1), and then the $y_1 \sim y_8$ are fed into the prediction model for neural network training or prediction. In (1), the symbol "A" indicates team A and the symbol "B" indicates team B. The symbol "*i*" is the index of 8 features. The input values of $y_1 \sim y_8$ are between 0~1 after normalization. We set "*If* $x_{iA} = x_{iB}$, *then* $y_{iA} = y_{iB} = 0.5$" that includes "*If* $x_{iA} = x_{iB}$=0, *then* $y_{iA} = y_{iB} = 0.5$". The example of normalization result of Germany (GER) versus Costa Rica (CRC) is listed in Table II.

TABLE. II
DATA NORMALIZATION OF GER VERSUS CRC

Acronym	Features	Before Normalization		After Normalization	
		Team-A (GER)	Team-B (CRC)	Team-A (GER)	Team-B (CRC)
		x_{iA}	x_{iB}	y_{iA}	y_{iB}
GF	x_1	4	2	0.6666	0.3333
S	x_2	21	4	0.84	0.16
SOG	x_3	10	2	0.8333	0.1666
CK	x_4	7	3	0.7	0.3
DFKG	x_5	1	0	1	0
IDFKG	x_6	0	0	0.5	0.5
BP	x_7	63%	37%	0.63	0.37
FS	x_8	12	11	0.5217	0.4782

III. MULTI-LAYER PERCEPTRON WITH BACK PROPAGATION LEARNING ALGORITHM

MLP model with BP learning algorithm is important since 1986 [6] - [7]. The weighting coefficient adjustment can be referred in [6] - [8]. Fig. 3 shows the 8-11-1 MLP architecture with one hidden layer used in this study to predict the winning rate of the football games. Each symbol is explained as below: **y** is the input data vector with 8 features that have been normalized, w is the connection weights between nodes of two layers, net is the value which is the sum of the product of inputs and weighting coefficients, $f(net)$ is the transfer function and the value is in 0~1, o is the output value, d is the desired output, and e is the error value. The transfer function used in hidden layer and output layer is log-sigmoid function, shown in (2). By using gradient descent method, we can get the adjustment of the weighting coefficient equations by (3) and (4). After using the momentum term in the inertia effect of the previous step adjustment, the final adjustment equations are modified as (5) and (6), where η is the learning rate, t is the index of iteration, and β is the momentum parameter.

$$f(net) = \frac{1}{1+e^{-net}} \qquad (f \text{ is in } 0\sim1.) \qquad (2)$$

$$\Delta w_{kj} = \eta(d_k - o_k)f'_k(net_k)o_j \qquad (3)$$

$$\Delta w_{ji} = \eta[\sum_{k=1}^{K}(d_k - o_k)f'_k(net_k)w_{kj}]f'_j(net_j)o_i \qquad (4)$$

$$\Delta w_{kj}(t) = \eta(d_k - o_k)f'_k(net_k)o_j + \beta\Delta w_{kj}(t-1) \qquad (5)$$

$$\Delta w_{ji}(t) = \eta[\sum_{k=1}^{K}(d_k - o_k)f'_k(net_k)w_{kj}]f'_j(net_j)o_i + \beta\Delta w_{ji}(t-1) \qquad (6)$$

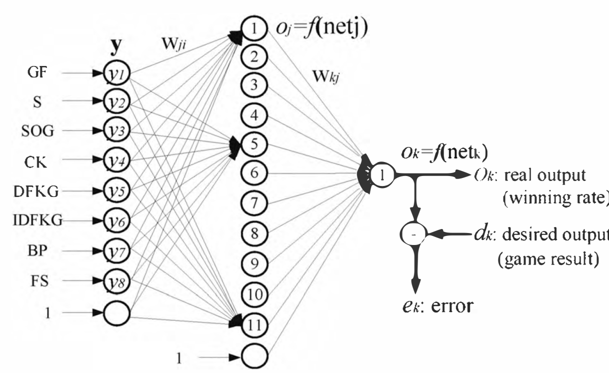

Fig. 3. MLP network architecture used in predicting the winning rate of 2006 WCFG.

IV. WINNING RATE PREDICTION

A. *Training Samples*

At stage 1, we select the teams which win or lose all the three games as our training teams. They are colored by gray in the background of Table I. We set the desired output to 1 for winning the game, and set to 0 for losing the game.

From stage 2, the winning team's record will be added into as training samples for all subsequent stages' training process if the team had won three games at stage 1.

The selected training teams (background color is gray in Fig. 1) and training matches (the bold line and bold number with under line in Fig. 1) from stage 1 to stage 4 are as follows.

1) The selected training samples for predicting the games in round of 16 (stage 2) — we select the teams whose score is either 9 (win 3 games) or 0 (lose 3 games) at the stage 1 as the representatives of winning team or losing team. Then transfer those game's records into relative ratios as the training data. Consequently we can find 21 samples from the 20 matches of 7 teams as the training data. We have 4 teams with 3 wins: GER, POR, BRA and ESP, and 3 teams with 3 losses: CRC, SCG and TOG. GER and CRC have one match and their records are selected as two training samples.
2) The selected training samples for predicting the games in quarter-finals (stage 3) — besides the 21 training data at the stage 1, we add the game's data of stage 2 from those representative teams, which are qualified to be the training samples at stage 2 and which are not only the winner at stage 2, but also have all won at the stage 1. We can find 3 teams' records (GER, POR, and BRA) as the training data at stage 2. Therefore, currently we totally get 24 (21+3=24) training data for the training to predict in stage 3.
3) The selected training samples for predicting the games in the semi-finals (stage 4) — besides the above 24 training data, we add the winners' record of stage 3 as the training data. The same reason as last stage, we can find 2 teams (GER and POR) as the training data from stage 3. Therefore, we totally get 26 (24+2=26) training data for the training to predict in stage 4.
4) The training samples for predicting the games in the finals (stage 5) — because the two teams (FRA and ITA) have not all won at stage 1, therefore, the training data for the training to predict in stage 5 are the same as above stage 4 (26 training data).

B. *Input Team Data for Predicting*

We do not have to predict the game result in stage 1, but the record in stage 1 will be extracted in order to be used for next stage prediction. Therefore, the input data used to predict the game result at stage 2~5 are described as follows.

1) The input data for predicting round of 16 (stage 2) — we respectively take the average value from the 3 games' records of each winning team at stage 1 that enter stage 2 as the input data. Therefore, we get 16 input team data for the 8 games that we want to predict at stage 2. For example, the input team data for predicting the winner of GER versus SWE are listed in Table III.
2) The input data for predicting the quarter-finals (stage 3) — the input data are got from the records of stage 1~2. We respectively take the average value from the 4 games' records (3 games from the stage 1 and 1 game from stage 2) of each team as the input data. Therefore, we get 8 input team data for the 4 games that we want to predict in quarter-finals.
3) The input data for predicting the semi-finals (stage 4) — the input data are got from stage 1~3. We respectively take the average value from the 5 games' records of each team as the input data. Therefore, we totally get 4 input team data for the two games we want to predict in semi-finals.
4) The input data for predicting the finals (stage 5) — the input data are got from stage 1~4. We respectively take the average value from the 6 game's records of each team, which have entered into the stage 5, as the input data. Therefore, we totally get 4 input team data for the last two final games we want to predict.

C. *Sequential Determination of Parameters for the Back-Propagation Learning*

The back-propagation learning algorithm used in the MLP is gradient descent with momentum. In order to determine the number of hidden neurons, momentum coefficient β, learning rate η for BP learning rule, we use the 21 training team samples got in stage 1 to train the prediction model and to keep the observation of the relationship between above mentioned parameters. The procedures of determining the parameters setting for BP are listed in Table IV.

Kecman recommended $\beta = 0.5 \sim 0.7$ and $0 < \eta < 1$ in the BP learning [9]. Therefore, we set $\beta=0.5$, $\eta=0.1$, and test three different Mean-Square-Error settings (MSE = 0.1, 0.05, and 0.01) for the number of hidden neurons from 2 to 40. The testing result for three MSE settings is shown in Fig. 4. From Fig. 4, we can find out that the training process always can converge no matter how the number of hidden neurons is set from 2 to 40 or how the MSE value is set to be 0.1, 0.05, or 0.001. However, there is an inclination to show the less MSE setting, the less training time for convergence and it is obviously found out that the rapidly decreased training epochs become slow down when the neurons increase from 2 to 11 under three different MSE settings. Therefore, we decide to select 11 neurons as the number of hidden neurons in the MLP model for winning rate prediction.

Based on above experimental result, we set hidden neurons=11, MSE=0.01 and set the learning rate $\eta = 0.1$, then sequentially to test three different momentum coefficient β (β= 0.5, 0.6, and 0.7). Each β setting is tested for 40 tests. The testing result is shown in Fig. 5. From Fig. 5, we can find out the convergent epochs are more stable in the 40 tests when β is set to 0.6. Therefore, we decide to set momentum coefficient β to be 0.6 in the MLP model for winning rate prediction.

From the above experimental results, we set hidden neurons=11, MSE=0.01, β=0.6, and then sequentially to test five different learning rate η (η=0.1, 0.3, 0.5, 0.7, and 0.9). Each η setting is tested for 40 tests. The testing result is shown in Fig. 6. From Fig. 6, we find out that when η is set as 0.9 will get a less convergent time than other η settings. Therefore, we decide to set the learning rate η to be 0.9 in the MLP model for winning rate prediction.

TABLE. III
INPUT DATA USED TO PREDICT THE WINNING RATE OF GER VERSUS SWE AT STAGE 2

Statistic Item	Team-A (GER)				Team-B (SWE)			
	Rec. 1	**Rec. 2**	**Rec. 3**	**Average**	**Rec. 1**	**Rec. 2**	**Rec. 3**	**Average**
GF	0.6664	1	1	0.8889	0.5	1	0.5	0.6667
S	0.84	0.7619	0.6818	0.7929	0.75	0.7692	0.4286	0.6493
SOG	0.8333	0.7619	0.8182	0.7612	0.75	0.5152	0.3913	0.5522
CK	0.7	0.7143	0.45	0.6214	0.4737	0.6667	0.6667	0.6023
DFKG	1	0.5	0	0.5	1	0.5	0.5	0.6667
IDFKG	0.5	0.5	0	0.3333	0.5	0.5	0.5	0.5
BP	0.63	0.58	0.43	0.5467	0.6	0.57	0.37	0.54
FS	0.5217	0.5526	0.538	0.5374	0	0.4412	0.4194	0.2868

TABLE. IV
PROCEDURES OF DETERMINING THE PARAMETERS FOR BP LEARNING RULE

Determined Parameters	Fixed Conditions	Variable Conditions	Observation items
Number of hidden neurons	(1) β=0.5 (2) η=0.1	(1) MSE = 0.1, 0.05, 0.01 (2) Hidden neurons = 2~40	Neurons versus epochs
Momentum coefficient β	(1) hidden neurons=11 (2) MSE=0.01 (3) η=0.1	(1) β=0.5, 0.6, 0.7 (2) testing times = 40	β versus epochs
Learning rate η	(1) hidden neurons=11 (2) MSE=0.01 (3) β=0.6	(1) η=0.1, 0.3, 0.5, 0.7, 0.9 (2) testing times = 40	η versus epochs

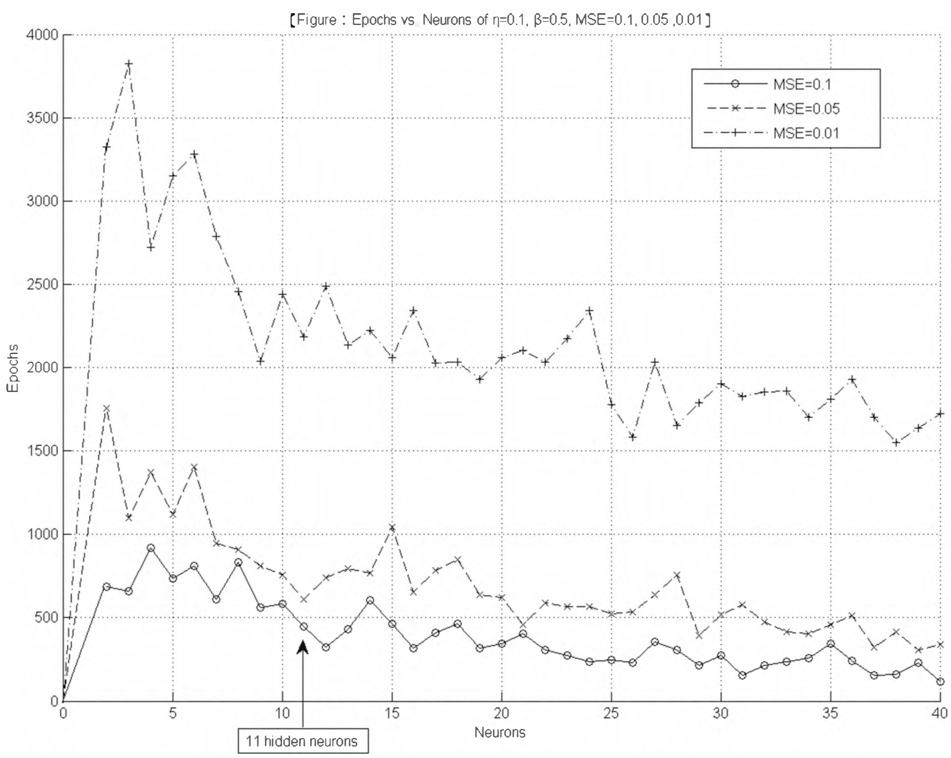

Fig. 4. Set β= 0.5, η = 0.1. Test the relationship between epochs vs. 2~40 neurons for three different MSE settings.

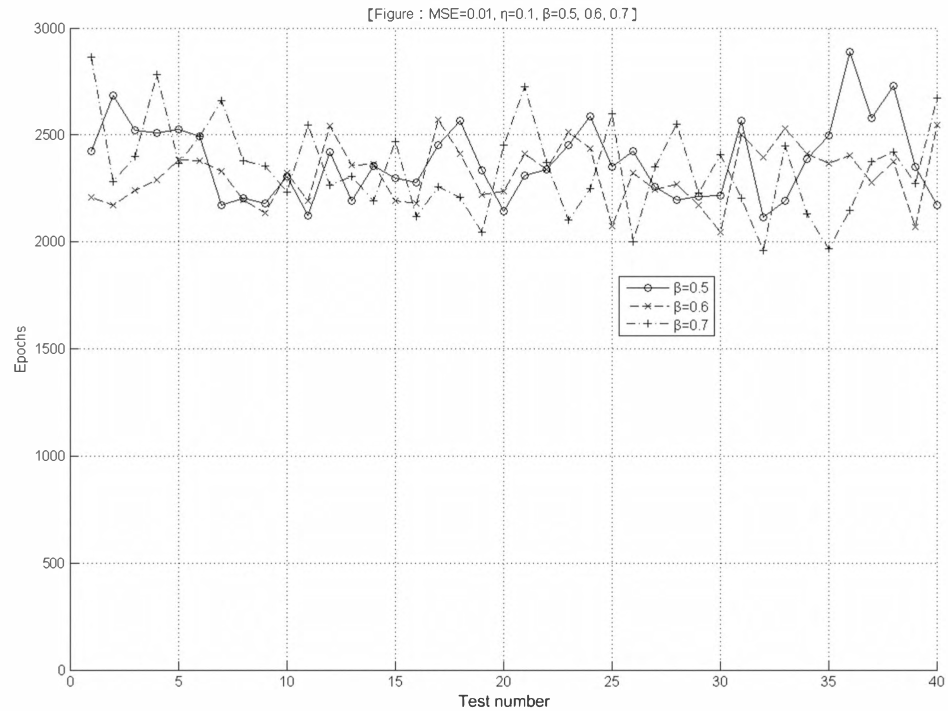

Fig. 5. Set MSE=0.01, hidden neurons=11, $\eta = 0.1$, then test three different β settings. Each β is tested for 40 tests.

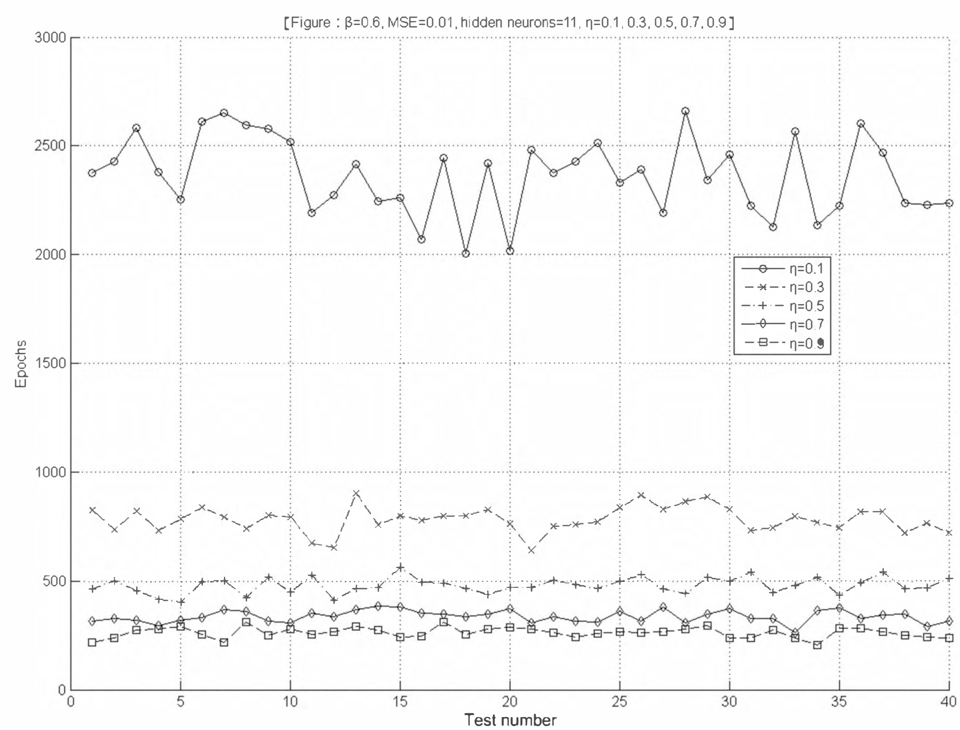

Fig. 6. Set MSE=0.01, hidden neurons=11, β= 0.6, then test five different η settings. Each η is tested for 40 tests.

V. MATCH PREDICTION RESULTS AND DISCUSSIONS

The final prediction model used in this study is 8-11-1 MLP with BP learning rule. The parameter settings for BP learning rule are η=0.9, β=0.6, and MSE=0.01. The prediction way is explained as follows. We input the average data of each team into the well trained MLP, and then we compare the output value to determine the relationship between victory and defeat. The team with bigger output value, which means the more ability to win the game, is the winner. The winning rate prediction results from stage 2 to stage 5 are listed in Table V. The symbol "W" means the team is winner whose real output value is bigger. On the other hand, the symbol "L" means the team is loser whose real output value is smaller. The symbol "Y" means the prediction result is correct. On the contrary, the symbol "N" means the prediction result is wrong. The symbol "NA" means the prediction result is ignored because of both teams tied in the end. The odds are also offered for betting reference and its calculation formula is defined as (7).

$$\text{Odds } (Team\_A \text{ vs. } Team\_B) = \frac{O_B}{O_A} \qquad (7)$$

where O_A and O_B are the real output of MLP for team A and team B.

As we know, the prediction for soccer games is not so easy because the players use feet to control the ball. Too many factors and situations are sometimes changeable, and thus the game results are usually unpredictable. Getting the goal score is not easy in soccer game, so there are many tied games in the records. Most of the time neural network can only predict the winner and loser. In fact, it is not easy to get 0.5 at the output of MLP from 8 input features, and predict the tied game. So we exclude the tied games in the calculation of prediction accuracy. From Table V, we can see the percentage of prediction accuracy at stage 2 is 85.7% (6/7). The accuracy is even higher than most of the predictions of so-called soccer game forecasters. The percentages of prediction accuracy at the following stages are 66.7% (2/3), 50% (1/2), and 100% (1/1). If we exclude three tied games (Match 54, 57, 64) and calculate the average prediction accuracy of other 13 games from stage 2 to stage 5, the percentage of the prediction accuracy is 76.9% (10/13).

TABLE. V
THE WIN RATE PREDICTION RESULTS FROM STAGE 2 TO STAGE 5
TABLE. V
THE WIN RATE PREDICTION RESULTS FROM STAGE 2 TO STAGE 5

Stage	Match	Team	Real Output	Prediction Result	Game Result	Prediction Correct	Prediction Accuracy	Odds
Stage 2	49	GER	0.9131	W	W	Y	85.7% (6/7)	0.7818
		SWE	0.7139	L	L			1.279
	50	ARG	0.813	W	W	Y		0.7043
		MEX	0.5726	L	L			1.4199
	51	ENG	0.8993	W	W	Y		0.5967
		ECU	0.5366	L	L			1.6758
	52	POR	0.7023	L	W	N		1.0623
		NED	0.7461	W	L			0.9413
	53	ITA	0.7758	W	W	Y		0.9348
		AUS	0.6227	L	L			1.0698
	54	SUI	0.9256	W	D	NA		0.7825
		UKR	0.6887	L	D			1.2779
	55	BRA	0.9213	W	W	Y		1.0475
		GHA	0.584	L	L			0.9546
	56	GER	0.9272	W	W	Y		0.9727
		SWE	0.7835	L	L			1.0281
Stage 3	57	GER	0.9556	W	D	NA	66.7% (2/3)	0.9263
		ARG	0.8933	L	D			1.0795
	58	ITA	0.9419	W	W	Y		0.7188
		UKR	0.7371	L	L			1.3912
	59	ENG	0.8785	L	L	Y		1.0381
		POR	0.9203	W	W			0.9633
	60	BRA	0.9342	W	L	N		0.9187
		FRA	0.9088	L	W			1.0885
Stage 4	61	GER	0.98	L	L	Y	50% (1/2)	1.0071
		ITA	0.9869	W	W			0.993
	62	POR	0.9719	W	L	N		0.9944
		FRA	0.9665	L	W			1.0056
Stage 5	63	GER	0.8837	W	W	Y	100% (1/1)	0.9495
		POR	0.839	L	L			1.0532
	64	ITA	0.9672	W	D	NA		0.9574
		FRA	0.926	L	D			1.0445

There are three prediction errors (N) and three results excluded (NA) in our prediction. We discuss as follows.

Match 52 (POR vs. NED) – It is a 9 to 9 players' game, because two players of both teams got red cards. From the FIFA match report, we can see NED is more excellent than POR (the same as MLP real output), but POR finally wins the game by getting a direct free kick to goal because NED committed a foul. (N)

Match 54 (SUI vs. UKR) – From the FIFA match report, we can tell SUI is more excellent than UKR (the same as real output), but the final result ties the game. (NA)

Match 57 (GER vs. ARG) – From the game, we actually find ARG is stronger than GER. It is a 10 to 11 players' game because one player of ARG got a red card. If without this red card, the winning opportunity of ARG would increase, but the final result is a draw match. (NA)

Match 60 (BRA vs. FRA) –The score is BRA 0 and FRA 1. Most of people bet BRA will win FRA. The statistical records of previous stages and the prediction result can show that, but the game result is opposite. BRA is losing. That reverses our image. Ball is round. Sometimes there is a lucky. In soccer game, the player must keep the chance of every shot to get the goal, otherwise the game result may reverse. (N)

Match 62 (POR vs. FRA) – In the game, POR gets more ball possession time and more shots on goal than FRA, but POR can not keep the chance to win the game. In the final, FRA wins the game by getting a penalty kick due to POR's foul within the area of eighteen yards. (N)

Match 64 (ITA vs. FRA) – The game ends in a tie. (NA)

VI. CONCLUSIONS

In this study, we adopt multi-layer perceptron with back propagation learning rule to predict the winning rate of 2006 WCFG. We select 8 significant statistical records from 17 official records of 2006 WCFG. The 8 records of each team are transformed into relative ratio values with another team. Then the average ratio values are fed into 8-11-1 MLP for predicting the winner and loser. The training data and the input data for predicting winning rate of the games at next stages are acquired from their statistical records at their previous stages' matches. New training samples are added to the training samples at the previous stages. In the BP learning, by means of sequential testing, the best number of hidden neurons is 11, the learning rate is 0.9, and the momentum coefficient is 0.6.

It is interesting that two teams, ITA and FRA, in the final match, in fact, do not contribute at all in the training samples. Because we only select the training teams with 3 wins and 3 losses in the stage 1, they are not selected in the training samples. It is a tied game.

Because getting the goal score is not easy in soccer game, there are many tied games in the records. Most of the time neural network can only predict the winner and loser. In fact, it is not easy to get 0.5 at the output of MLP from 8 input features, and predict the tied game. If the tied games are excluded, the prediction accuracy can achieve 76.9%.

REFERENCES

[1] M. C. Purucker, "Neural networks quarterbacking - how different training methods perform in calling the games," *IEEE Potentials*, August/September 1996, pp.9-15.

[2] J. Park and M. E. J. Newman, "A network-based ranking system for US college football," *Journal of Statistical Mechanics: Theory and Experiment*, IOP Publishing, Issue 10, 2005, pp.1~14.

[3] A. P. Rotshtein, M. Posner, and A. B. Rakityanskaya, "Football predictions based on a fuzzy model with genetic and neural tuning," *Cybernetics and Systems Analysis*, Vol.41, No.4, 2005, pp.619-630.

[4] 2006 FIFA World Cup Germany – match schedule, matches and results, and statistics reports, http://fifaworldcup.yahoo.com.

[5] Official website of FIFA, http://www.fifa.com.

[6] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by backpropagating errors," *Nature*, Vol.233, 1986, pp.533-536.

[7] David E. Rumelhart and James L. McClelland, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, Vol.1, MIT Press: Cambridge, MA, 1986.

[8] Kou-Yuan Huang, *Neural Networks and Pattern Recognition*, Weikeg Publishing Co., Taipei, Taiwan, March 2003, 406 pages.

[9] V. Kecman, *Learning and Soft Computing: Support Vector Machines, Neural Networks, and Fuzzy Logic Models*, MIT Press: Cambridge, MA, 2001.