ISSN 1750-9823 (print)
International Journal of Sports Science and Engineering
Vol. 07 (2013) No. 02, pp. 079-086

Applying the Markov Chain theory to Analyze the Attacking Actions between FC Barcelona and Manchester United in the European Champions League final

Tianbiao Liu [+], Andreas Hohmann

Institute of Sports Science, University of Bayreuth, 95447 Bayreuth, Germany

(*Received June 27, 2012, accepted September 9, 2012*)

Abstract. Several researchers have applied Markov chain methods in sport game analysis. This study analyzed the European Champions League final between Manchester United and FC Barcelona based on a first order Markov chain model. The results describe important passing strategies and tactical connections between certain players from each team.

Keywords. Soccer, European Champions League final, Markov chain, Stochastic process

1. Introduction

For a long time, researchers have shown interest in studying offensive actions in soccer games, and as a result, lots of methods and tools in soccer game analysis were developed. The majority of them are based on traditional statistical procedures [1], [2]. Later researchers incorporated mathematical methods into sports game research. The Markov Chain theory has already been applied to analyze Table tennis matches [3]. Concerning a soccer game, making use of the Markov Chain theory, the optimal timing of substitution and tactical decisions [4], as well as evaluation of team characteristics were determined by Hirotsu and Mike Wright [5]. In addition, during the 2006 FIFA World Cup team tactical features were researched by Pfeiffer, Hohmann and Buehrer [6].

In this case study, the Markov chain theory was applied to analyze offensive actions for both teams (Manchester United and FC Barcelona) in the 2011 European Champions League Final.

2. Methods

2.1. Data collection

The game was held on May 29th, 2011 and was videotaped so that all elements could be marked afterwards. Every detail of the game was observed and used for analysis to be able to draw comparisons between the two great teams.

2.2. Division of soccer field areas

As shown in Figure 1, the whole playing field was divided into 30 zones. In the first half, Manchester United attacked from right to left in the first half so their attacking area included zones #1-10. The midfield area consisted of zones #11-20, and the backfield area of zones #21-30. Each area covered the same amount of space, 35m out of 105m. In contrast, FC Barcelona attacked from left to right so their attacking area referred to zones #21-30, their midfield area to zones #11-20, and their backfield area to zones #1-10. In order to facilitate the recording after the teams changed ends at halftime, the numbering of the zones remained unchanged.

In addition, middle zones were defined by # 2-4, 7-9, 12-14,17-19,22-24, and 27-29. Side zones were those besides middle zones.

[+] Tianbiao Liu. Tel.: +49-921-55 3479; fax: +49-921-55 5806
E-mail address: ltbvane@yahoo.com.cn.

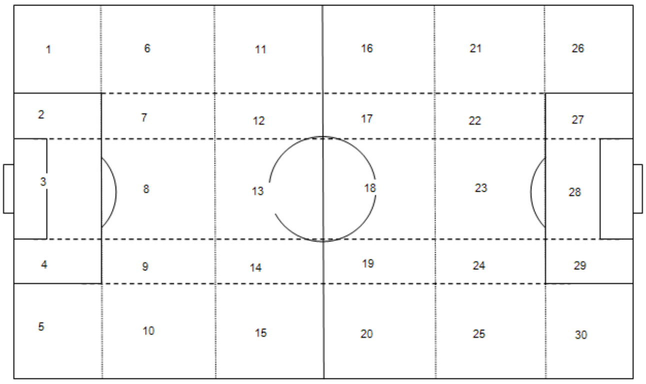

Fig. 1: Division of soccer field areas

2.3. Markov Chain

2.3.1 Markov chain theory

Mathematically, a Markov chain is a discrete random process with the Markov property. A discrete random process means a system which can be in various states, and which changes randomly in discrete steps.

A Markov chain is a sequence of random variables X1, X2, X3, ... with the Markov property, namely that, given the present state, the future and past states are independent. Formally,

$$\Pr(X_{n+1} = x | X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) = \Pr(X_{n+1} = x | X_n = x_n)$$

The possible values of Xi form a countable set S called the state space of the chain.

Markov chains are often described by a directed graph, where the edges are labeled by the probabilities of going from one state to the other states.

The changes of the states of the system are called transitions, and the probabilities associated with various state-changes are called transition probabilities.

2.3.2 Observation model

The main purpose of a soccer team is to score goals, specifically, score more goals than its opponent. In order to be successful, players organize attacks as soon as they take control of the ball from their opponent. Next, they keep on dribbling and passing the ball to each other until they lose control of the ball or other scenarios as selected by the researchers. In this case, the attacking/backfield third (35m in front of the goal) was selected. The kind of chain system is applied in this research as an analysis unit, and the observation unit (state) is defined as the player who controls the ball in the chain. The objectivity of the observation model is confirmed by the agreement of two independent observers by using Cohen's Kappa statistic [7].

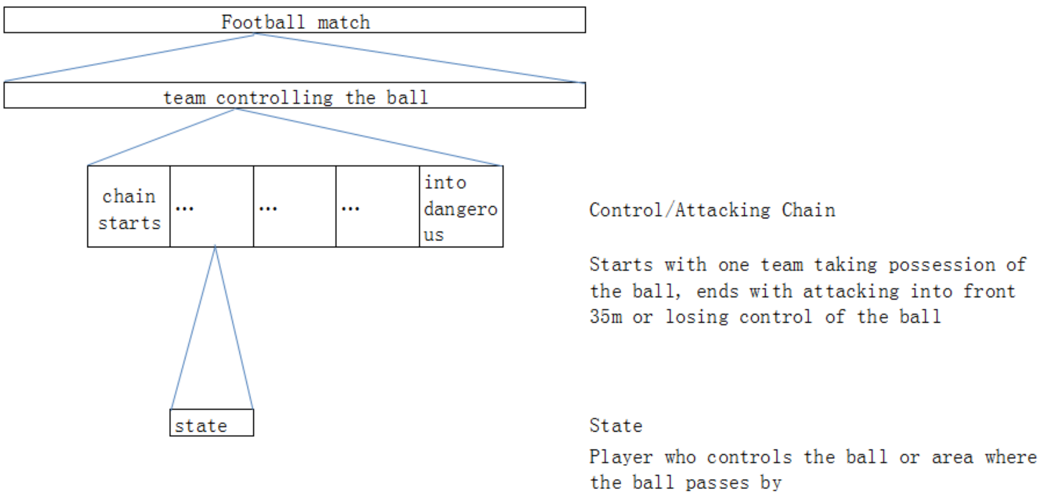

Fig. 2: Data structure of football game

2.3.3 Stochastical model

The transition probabilities between two states describe the soccer match as a process that can be understood as a first order Markov Chain. The following two properties are given: a) the probability for the next state depends solely on the current state (Markov property), it has nothing to do with any earlier state (player); and 2) the transition probability from one state to another is independent of its chronological position in the match process (Chain-property).

The transition probabilities between the states can be transformed into a two-dimensional transition matrix. Each element in the matrix $P_{ij}(n) \geq 0$ and the sum of line is equal to 1[8]. In the Markov chain, absorbing states are important, since the process ends in these states and a new process starts in these states. In this current research, the state “the front 35m” is defined as the absorbing state. One can calculate the Attacking Probability (AP) by using all the transition probabilities (see Table 1).

Table 1: Example of transition matrix by using observation model “Passes” (ManU in the first half)

	Van Der Sar 1	Patrice Evra 3	Rio Ferdinand 5	Wayne Rooney 10	Ryan Giggs 11	Park Ji-Sung 13	Javier Hernandez 14	Nemania Vidic 15	Michael Carrick 16	Fabio 20	Antonio Valencia 25	into 35m 35
Van Der Sar 1	5.00	5.00	15.00	5.00	5.00	5.00	0.00	0.00	0.00	5.00	5.00	5.00
Patrice Evra 3	0.00	19.05	0.00	9.52	4.76	28.57	4.76	4.76	14.29	0.00	0.00	4.76
Rio Ferdinand 5	21.05	10.53	10.53	0.00	5.26	5.26	0.00	15.79	15.79	0.00	5.26	5.26
Wayne Rooney 10	0.00	3.45	3.45	17.24	13.79	6.90	6.90	0.00	10.34	6.90	6.90	13.79
Ryan Giggs 11	0.00	2.94	5.88	5.88	23.53	11.76	2.94	2.94	2.94	8.82	2.94	8.82
Park Ji-Sung 13	0.00	16.67	0.00	16.67	4.17	12.50	0.00	4.17	4.17	4.17	0.00	20.83
Javier Hernandez 14	0.00	0.00	0.00	15.39	7.69	7.69	23.08	0.00	23.08	0.00	7.69	15.38
Nemania Vidic 15	13.04	4.35	26.09	4.35	13.04	4.35	0.00	13.04	8.70	4.35	0.00	0.00
Michael Carrick 16	4.17	0.00	0.00	8.33	12.50	0.00	12.50	4.17	12.50	20.83	0.00	4.17
Fabio 20	0.00	0.00	8.71	13.04	21.74	0.00	0.00	0.00	0.00	13.04	17.39	4.35
Antonio Valencia 25	0.00	0.00	13.33	6.67	6.67	0.00	0.00	13.33	6.67	20.00	6.67	6.67

2.3.4 Calculating the performance relevance

Based on the ball moving matrix and player passes matrix above, it is possible to calculate the AP on the basis of a simulated transition matrix. In order to determine the performance relevance of a tactical behavior pattern, each cell in the initial matrix will be firstly modified by a certain percentage resulted from a function:

$$\delta TP = C + B \times 4 \times TP(1 - TP) \text{ (Lames, 1991).}$$

In this function, TP is the transition probability; δTP is the change of element transition probability. The constant values applied in the study are C = 1, B = 5, which were determined by Lames [9] and tested by Pfeiffer [10].

	Van Der Sar_1	Patrice Evra_3	Rio Ferdinand_5	Wayne Rooney_10	Ryan Giggs_11	Park Ji-Sung_13	Javier Hernandez_14	Nemania Vidic_15	Michael Carrick_16	Fabio_20	Antonio Valencia_25	into 35m	out of 35m
Van Der Sar_1	5.00	5.00	15.00	5.00	5.00	5.00	0.00	0.00	0.00	5.00	5.00	5.00	45.00
Patrice Evra_3	0.00	19.05	0.00	9.52	4.76	28.57	4.76	4.76	14.29	0.00	0.00	4.76	9.52
Rio Ferdinand_5	21.05	10.53	10.53	0.00	5.26	5.26	0.00	15.79	15.79	0.00	5.26	5.26	5.26
	25.38	9.94	9.94		4.97	4.97		14.92	14.92		4.97	4.97	4.97
Wayne Rooney_10	0.00	3.45	3.45	17.24	13.79	6.90	6.90	0.00	10.34	6.90	6.90	13.79	10.34
Ryan Giggs_11	0.00	2.94	5.88	5.88	23.53	11.76	2.94	2.94	2.94	8.82	2.94	8.82	20.59
......													

$AP_{into\ 35m}$(empirical matrix) = 30.31%

$AP_{into\ 35m}$(manipulated matrix) = 30.29%

$\boxed{\delta AP_{into\ 35m} = -0.02\%}$

Fig. 3: Example of simulative calculation of the Attacking relevance (δAP) of model “passes”

In order to keep the line sum in the matrix still equal to 1.00, the other values in the same row must be proportionally compensated; in this case, a compensation function is introduced and applied:

$$\delta TP_{yi} = -\left[T_{yi}/_{(1-TP)}\right] \times \delta TP_x$$

After this, another AP is calculated from the new value in the cell which helps to define the performance relevance (δAP) of a tactical behavior pattern in terms of the difference between the attacking probability (AP) as calculated by the original transition-matrix and the attacking probability as calculated by the modified transition-matrix (see Figure 3).

The higher the δAP is, the more positively the element (game state transition) affects the game, correspondingly, the lower the δAP is, the more negatively the element (game state transition) affects the game.

Similarly, all other δAPs can be calculated and comparisons can be made which are used to identify which passes (giving or receiving) are more efficient and also effective in the game.

3. Results

3.1. Objectivity of the game observation models and model validity

In this study, all match events were recorded and included into analysis. Their consistency was examined by the inter-rating consistency of two observers (inter observer agreement) quantified in Cohen's Kappa. Manchester United first half record was selected for examination of Cohen's kappa. The Cohen's kappa values (κ) of the model were found to be: $\kappa = 0.766$ for "player's number", $\kappa = 0.625$ for "Zone", indicating their usefulness [11], [12]. But these κ values also reflected the problems to identify all data while collecting them from TV.

3.2. Analysis of offensive actions

3.2.1 Performance relevance of both teams' tactical behaviors

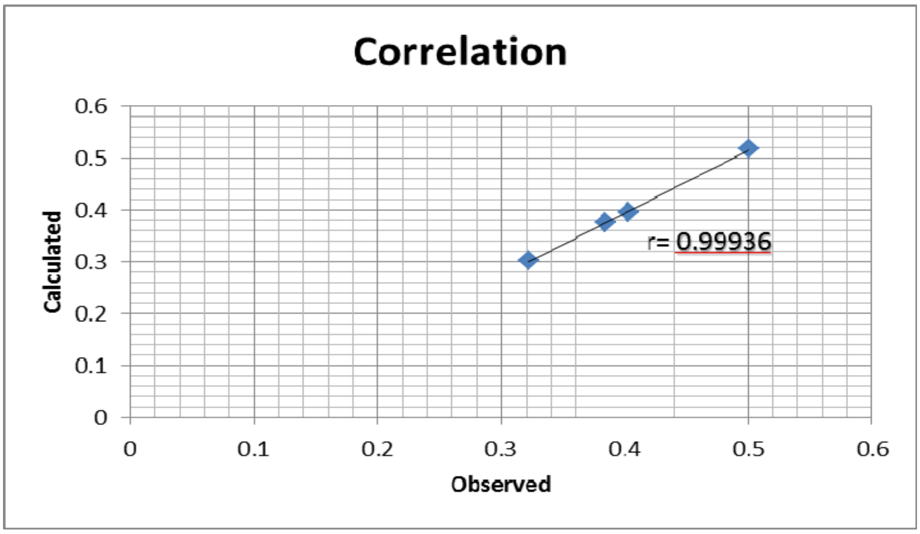

Fig. 4: Correlation between observed and calculated value

The correlation coefficient between observed and mathematically modeled attacking probability (AP) for the observation system was calculated to verify model validity. The value is more than 0.999 (see Figure 4), which indicates the validity of the Markov Chain.

3.2.2 Performance relevance of Manchester United

Figure 5 shows the Performance relevance of passes for Manchester United in the European Champions League Final's first half. The diagram (Transition matrix see Table 1) shows that the passing combinations "Patrice Evra – Park Ji-Sung"; "Michael Carrick - Javier Hernandez" and "Patrice Evra - Javier Hernandez", as well as "Giggs – Park Ji-Sung" were an important contribution and therefore very effective for the team's attack, especially in the front 35m, which is considered the opponent's dangerous area. On the other hand, the passing combinations "Patrice Evra – Michael Carrick"; "Javier Hernandez – Michael Carrick" and "Patrice Evra - Vidic" had a negative impact on the attack in the front 35m, hence reducing the mistake rate among them would contribute much more to the team's attack. Furthermore, both statistical and mathematical simulation results showed that although Carrick did not pass very often to Hernandez, his passes to him contributed more than his other passes to create an advantageous situation.

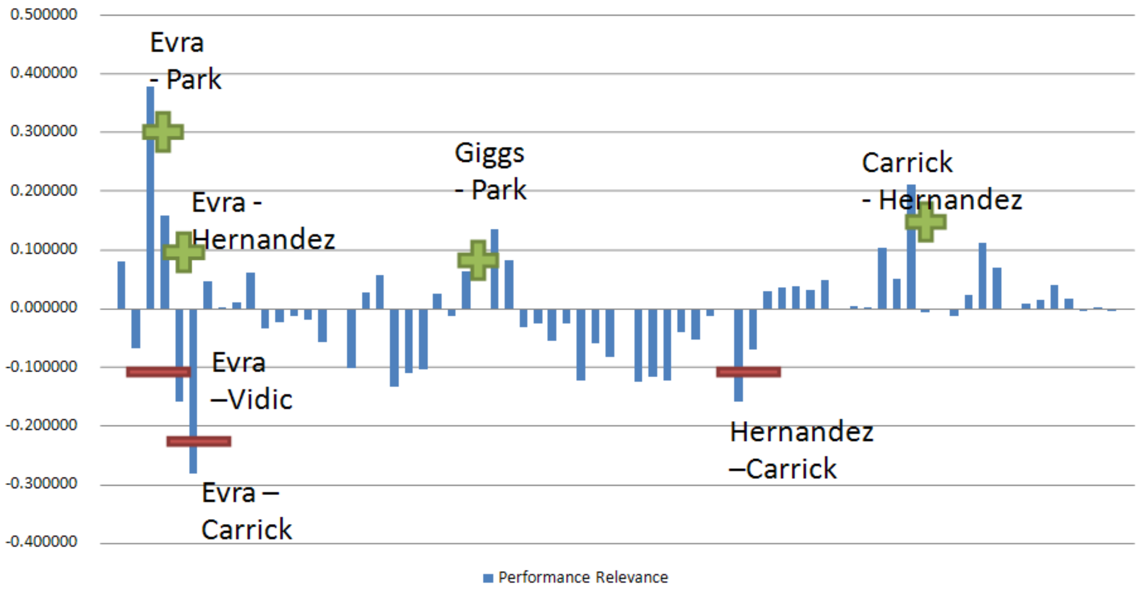

Fig. 5: Performance relevance of passes for Manchester United in the first half

In addition, the table also states that dribbling in the first half was not very important in the team's attacking, however, passes were more effective and efficient.

Table 2: Transition matrix of Manchester United passes in the second half

	Van Der Sar 1	Patrice Evra 3	Rio Ferdinand 5	Wayne Rooney 10	Ryan Giggs 11	Park Ji-Sung 13	Javier Hernandez 14	Nemania Vidic 15	Michael Carrick 16	Nani 17	Scholes 18	Fabio 20	Antonio Valencia 25	into 35m 35
Van Der Sar 1	10	0	10	25	0	0	5	10	5	5	0	5	5	0
Patrice Evra 3	0	23.33333	0	6.66667	16.66667	3.33333	0	6.66667	0	0	13.33333	0	0	13.33333
Rio Ferdinand 5	6.89655	0	17.24138	10.34483	0	10.34483	0	10.34483	6.89655	3.44828	13.7931	0	13.7931	0
Wayne Rooney 10	0	2.43902	0	26.82927	7.31707	9.7561	4.87805	2.43902	4.87805	4.87805	2.43902	0	9.7561	17.07317
Ryan Giggs 11	0	18.19182	0	22.72727	18.18182	4.54545	4.54545	4.54545	0	0	0	0	4.54545	9.09091
Park Ji-Sung 13	0	0	0	10.52632	10.52632	15.78947	5.26316	15.78947	15.78947	10.52632	0	0	5.26316	0
Javier Hernandez	0	0	0	10	20	0	0	0	10	0	10	0	0	50
Nemania Vidic 15	16.67667	8.33333	20.83333	4.16667	0	4.16667	0	16.66667	4.16667	0	8.33333	0	4.16667	0
Michael Carrick 16	5.55556	5.55556	5.55556	11.11111	16.66667	0	16.66667	0	11.11111	5.55556	0	0	5.55556	5.55556
Nani 17	0	4.34783	0	4.34783	0	4.34783	4.34783	0	4.34783	39.13043	4.34783	0	4.34783	13.04348
Scholes 18	6.25	6.25	31.25	6.25	0	12.5	0	18.75	0	0	0	0	6.25	6.25
Fabio 20	0	0	50	0	0	0	0	0	50	0	0	0	0	0
Antonio Valencia 25	3.7037	0	11.11111	7.40741	0	0	0	3.7037	0	14.81481	11.11111	0	29.62963	0
into 35m 35	0	0	0	0	0	0	0	0	0	0	0	0	0	100
out of 35m 36	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table 2 shows the transition matrix of Manchester United passes in the second half, and figure 6 suggests the performance relevance of passes for Manchester United in the second half. The diagram shows that the passes between "Van Der Sar - Rooney", "Van der Sar - Hernandez", "Rooney - Hernandez" and "Carrick - Hernandez" ranked at the top, therefore affecting the team's attack into the front 35m areas very positively in the second half. However, the passes between "Rooney - Valencia", "Evra – Vidic", "Rooney – Park Ji-sung" and "Hernandez - Giggs" negatively affected the team, which suggests reducing this mistake rate would help setting up the team's offense in the opponent's dangerous areas.

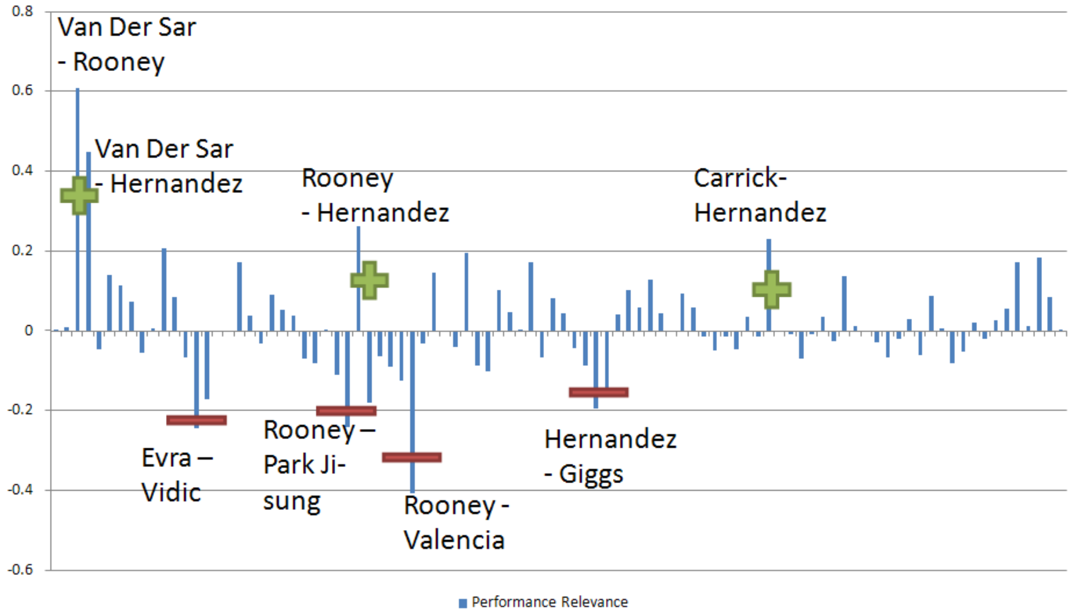

Fig. 6: Performance Relevance of passes for Manchester United in the second half

3.2.3 Performance relevance of FC Barcelona

Table 3: Transition matrix of FC Barcelona passes in the first half

	Valdes 1	Daniel Alves 2	Gerard Pique 3	XAVI Hernandez 6	David Villa 7	Andres Iniesta 8	L Messi 10	Javier Mascherano 14	Sergio Busquets 16	Pedro Rodriguez 17	Eric Abidal 22	Into 35m 35
Valdes 1	7.14286	7.14286	14.28571	0.00000	7.14286	7.14286	0.00000	28.57143	14.28571	0.00000	0.00000	0.00000
Daniel Alves 2	0.00000	16.07143	3.57143	17.85714	5.35714	10.71429	21.42857	3.57143	3.57143	0.00000	0.00000	5.35714
Gerard Pique 3	6.07061	3.03030	15.15152	24.24242	3.03030	6.06061	6.06061	12.12121	15.15152	0.00000	9.09091	0.00000
XAVI Hernandez 6	1.01010	9.09091	0.00000	19.19192	7.07071	20.20202	9.09091	11.11111	6.06061	7.07071	5.05051	3.03030
David Villa 7	0.00000	16.00000	0.00000	12.00000	20.00000	8.00000	12.00000	0.00000	0.00000	4.00000	0.00000	28.00000
Andres Iniesta 8	0.00000	8.97436	2.56410	19.23077	2.56410	12.82051	23.07692	1.28205	8.97436	5.12821	5.12821	6.41026
L Messi 10	0.00000	8.04598	0.00000	14.94253	4.59770	13.79310	28.73563	2.29885	6.89655	4.59770	1.14943	8.04598
Javier Mascherano 14	9.52381	11.90476	9.52381	21.42857	0.00000	2.38095	11.90476	21.42857	4.76190	0.00000	4.76190	0.00000
Sergio Busquets 16	0.00000	6.52174	8.69565	17.39130	0.00000	17.39130	10.86957	4.34783	13.04348	2.17391	13.04348	2.17391
Pedro Rodriguez 17	0.00000	0.00000	0.00000	7.14286	0.00000	25.00000	10.71429	0.00000	3.57143	17.85714	7.14286	17.85714
Eric Abidal 22	0.00000	0.00000	5.41541	10.81081	0.00000	21.62162	8.10811	2.70270	10.81081	13.51351	13.51351	0.00000

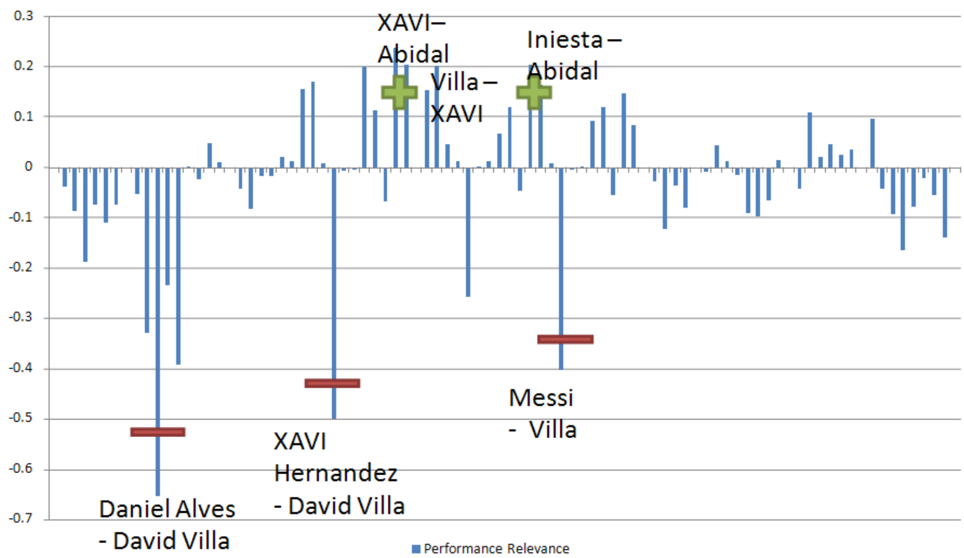

Fig. 7: Performance Relevance of passes for FC Barcelona during the first half

Table 3 shows the transition matrix of FC Barcelona passes in the first half, and figure 7 indicates the performance relevance of passes for FC Barcelona in the first half. From the chart one can see that the rising success rate of the pass combinations “XAVI Hernandez – Eric Abidal”, “David Villa – XAVI Hernandez” and “Andres Iniesta – Eric Abidal” did contribute to the team’s attack of the front 35m area. In addition, reducing the mistake rate of passes between "Daniel Alves - David Villa", "XAVI Hernandez - David Villa", "Messi - David Villa" could further enhance the performance relevance of attacking the 35m area in the first half. Since the majority of these players are playing on the right side, it can be concluded that by reducing the mistake rate on that side, FC Barcelona would make its own attack more effective.

Table 4: Transition matrix of FC Barcelona passes in the second half

	Valdes 1	Daniel Alves 2	Gerard Pique 3	Puyol 5	XAVI Hernandez 6	David Villa 7	Andres Iniesta 8	L Messi 10	Javier Mascherano 14	Keita 15	Sergio Busquets 16	Pedro Rodriguez 17	Eric Abidal 22	Into 35m 35
Valdes 1	0	21.4286	7.14286	0	0	7.14286	0	0	7.14286	7.14286	7.14286	14.2857	0	0
Daniel Alves 2	0	17.3177	0	1.92308	17.3077	11.5385	0	15.3846	7.69231	0	13.4615	0	0	5.76923
Gerard Pique 3	7.40741	3.7037	29.6296	0	25.9259	0	7.40741	0	0	0	3.7037	7.40741	11.1111	0
Puyol 5	0	20	0	0	40	0	0	0	0	0	0	20	0	0
XAVI Hernandez 6	1.09696	14.1304	5.43478	0	16.3044	6.52174	15.2174	15.2174	1.08696	0	6.52174	4.34783	2.17391	5.43478
David Villa 7	0	11.5385	0	0	11.5385	15.3846	3.84615	15.3846	11.5385	0	3.84615	0	0	7.69231
Andres Iniesta 8	0	2.8269	2.8169	1.40845	19.7183	0	22.5352	19.7183	4.22535	0	7.04225	7.04225	4.22535	5.6338
L Messi 10	0	6.09756	2.43902	0	21.9512	0	14.6342	31.7073	1.21951	1.21951	4.87805	1.21951	0	9.7561
Javier Mascherano 14	4.16667	8.33333	4.16667	0	4.16667	20.8333	0	8.33333	16.6667	0	8.33333	0	4.16667	4.16667
Keita 15	0	0	0	0	0	0	25	0	0	0	25	0	0	0
Sergio Busquets 16	0	6.25	6.25	0	8.33333	4.16667	22.9167	14.5833	2.08333	4.16667	18.75	4.16667	6.25	0
Pedro Rodriguez 17	0	0	0	6.89655	3.44828	0	20.6897	6.89655	0	0	6.89655	20.6897	20.6897	6.89655
Eric Abidal 22	6.89655	0	6.89655	0	20.6897	0	17.2414	0	0	0	10.3448	13.7931	13.7931	0

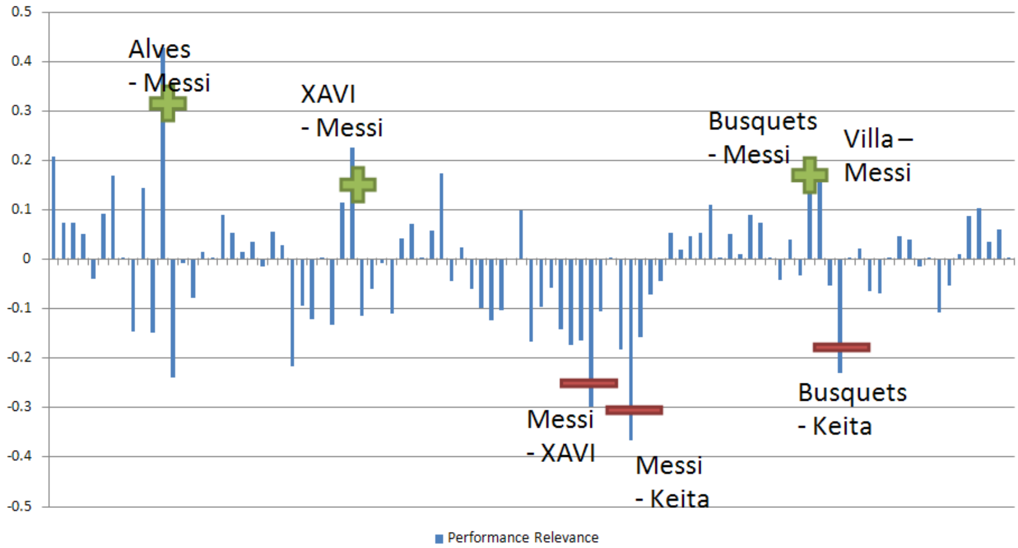

Fig. 8: Performance Relevance of passes for FC Barcelona during the second half

Table 4 shows the transition matrix of FC Barcelona passes in the second half, and figure 8 indicates the performance relevance of passes for FC Barcelona in the second half. The diagram indicates that an improving success rate of the passes between "Daniel Alves - Messi", “XAVI Hernandez - Messi”, “Sergio Busquets - Messi” and “David Villa – Messi” would produce better results in the attacking of the front 35 m. However, the pass combinations "Messi - Keita", "Messi - XAVI Hernandez", “Sergio Busquets - Keita” negatively affected the team’s attack of the front 35m area. To reduce the mistake rate of their passes would make it more efficient in the opponent's dangerous areas. The data suggested that Messi became the playmaker for FC Barcelona in the second half. Furthermore, "Messi's passes to Keita" and "Sergio Busquets' passes to Keita" played a real negative role in the second half, which was probably because of Keita's late substitution in the game.

4. Conclusion

In this study, the Markov model was built to calculate performance relevance. The data clearly indicated that the star players of each team Messi (FC Barcelona) and Rooney (Manchester United) looked to play a more important role in the second half, which was supported by the performance relevance results that were relatively high in relation to them.

5. Reference

[1] Kuhn, Werner, and Schmidt, Werner. Analyse und Beobachtung in Training und Wettkampf: Beitraege und Analysen zum Fussballsport IV. Sankt Augustin: Academia Verlag, 1991.

[2] J. Bangsbo, T. Reilly, and C. Hughes. *Science and football*. Spon Press, 1997.

[3] H.Zhang, and A. Hohmann. Performance Diagnosis through Mathematical Simulation in Table Tennis Game. *Journal of Shanghai University of Sport*. 2004, **28**(2): 68-72.

[4] N. Hirotsu, and M. Wright. An evaluation of characteristics of teams in association football by using a Markov process model. *The Statistician*. 2003, **52**: Part 4, pp. 591-602.

[5] N. Hirotsu, and M. Wright, Using a Markov process model of an association football match to determine the optimal timing of substitution and tactical decisions. *Journal of the Operational Research Society*. 2002, **53**: 88-96.

[6] M. Pfeiffer, A. Hohmann. and M. Buehrer. Computersimulation zur Bestimmung der Leistungsrelevanz taktischer Verhaltensweisen bei der FIFA WM 2006.5. Dvs-sportspiel-Symposium 2006, Universitaet Flensburg.

[7] C. Robson. *Real world research.* Oxford: Blackwell Publishers, 2002.

[8] Pfeiffer, M., Zhang, H and Hohmann, Andreas. A Markov Chain Model of Elite Table Tennis Competition. *International Journal of Sport Science and Coaching*. 2008, **5**(2): 205-222.

[9] M. Lames. Leistungsdiagnostik durch Computersimulation. Frankfurt/Main: Harri Deutsch. 1991, pp.7-257.

[10] M. Pfeiffer. *Leistungsdiagnostik im Nachwuchstraining der Sportspiele*. Koeln: Sportbuch Strauss, 2005.

[11] W. Greve. and D. Wentura. Wissenschaftliche Beobachtung: Eine Einführung, PVU/Beltz, Weinheim 1997.

[12] J. R. Landis, and G. G. Koch. The measurement of observer agreement for categorical data. *In: Biometrics.* 1977, **33**: 159-174.