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.

Published by World Academic Press, World Academic Union
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 \( X_1, X_2, X_3, \ldots \) with the Markov property, namely that, given the present state, the future and past states are independent. Formally,

\[
P(X_{n+1} \mid X_1, X_2, \ldots, X_n) = P(X_{n+1} \mid X_n)
\]

The possible values of \( X_i \) 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].
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).

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:

\[
\delta TP = C + B \times 4 \times TP (1 - TP)
\] (Lames, 1991).

In this function, TP is the transition probability; \( \delta TP \) 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].

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

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Van Der Sar 1</td>
<td>5.00</td>
<td>5.00</td>
<td>15.00</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>5.00</td>
<td>5.00</td>
</tr>
<tr>
<td>Patrice Evra 3</td>
<td>0.00</td>
<td>19.05</td>
<td>0.00</td>
<td>9.52</td>
<td>4.76</td>
<td>28.57</td>
<td>4.76</td>
<td>4.76</td>
<td>0.00</td>
<td>14.29</td>
<td>0.00</td>
</tr>
<tr>
<td>Rio Ferdinand 5</td>
<td>21.05</td>
<td>10.53</td>
<td>10.53</td>
<td>0.00</td>
<td>5.26</td>
<td>5.26</td>
<td>0.00</td>
<td>15.79</td>
<td>15.79</td>
<td>0.00</td>
<td>5.26</td>
</tr>
<tr>
<td>Wayne Rooney 10</td>
<td>0.00</td>
<td>3.45</td>
<td>3.45</td>
<td>17.24</td>
<td>13.79</td>
<td>6.90</td>
<td>6.90</td>
<td>0.00</td>
<td>10.34</td>
<td>6.90</td>
<td>6.90</td>
</tr>
<tr>
<td>Ryan Giggs 11</td>
<td>0.00</td>
<td>2.94</td>
<td>5.88</td>
<td>23.53</td>
<td>11.76</td>
<td>2.94</td>
<td>2.94</td>
<td>8.82</td>
<td>8.82</td>
<td>20.83</td>
<td>6.67</td>
</tr>
</tbody>
</table>

\[
\delta AP_{into 35m} = \delta AP_{into 35m} = -0.02%\]

Fig. 3: Example of simulative calculation of the Attacking relevance (\( \delta AP \)) 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:

\[ \delta TP_x = -\frac{TP_x}{1 - TP_L} \times \delta TP_L \]

After this, another AP is calculated from the new value in the cell which helps to define the performance relevance (\(\delta AP\)) 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 \(\delta AP\) is, the more positively the element (game state transition) affects the game, correspondingly, the lower the \(\delta AP\) is, the more negatively the element (game state transition) affects the game.

Similarly, all other \(\delta APs\) 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 (\(\kappa\)) 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 \(\kappa\) 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

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.

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

<table>
<thead>
<tr>
<th></th>
<th>Van Der Sar</th>
<th>Evra</th>
<th>Rio Ferdinand</th>
<th>Park Ji-sung</th>
<th>Javier Hernandez</th>
<th>Michael Carrick</th>
<th>Nani</th>
<th>Scholes</th>
<th>Fabio</th>
<th>Antonio Valencia</th>
<th>into 35m</th>
</tr>
</thead>
<tbody>
<tr>
<td>Van Der Sar</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Evra</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Rio Ferdinand</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Park Ji-sung</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Javier Hernandez</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Michael Carrick</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Nani</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Scholes</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Fabio</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Antonio Valencia</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>into 35m</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

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.
3.2.3 Performance relevance of FC Barcelona

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

![Fig. 6: Performance Relevance of passes for Manchester United in the second half](image1)

![Fig. 7: Performance Relevance of passes for FC Barcelona during the first half](image2)

SSci email for contribution: editor@SSCI.org.uk
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 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