日本オペレーションズ・リサーチ学会
2019年 秋季研究発表会

Analysing the Effect of a Change of Transition Rates Related to Possession on Probability of Winning a Soccer Game

01506960 Juntendo University * HIROTSU Nobuyoshi
Japan Sport Council KOMINE Ayako

1. Introduction

Modelling an association football match is a topic of interest for evaluating teams' characteristics, predicting the outcome of a match, or analyzing optimal tactical changes. In a previous paper [1], we extended the Markov process model [2] by considering the location of the ball on the pitch, in order to analyze teams' characteristics. In this paper, we use their Markov process model in which the pitch is divided to 9 areas, and analyze the effect of a change of transition rates related to possession on scoring a goal or winning a game.

2. The Markov process model

A football match can be seen as progressing through a set of stochastic transitions occurring due to a change of possession of the ball or scoring a goal. We assume a Markov property in these transitions and propose a Markov process model, which seems appropriate to a football match as an approximation.

We define the states as follows:

State H_G: Home team scores a goal;
State H_I: Home team is in possession of the ball and the ball is located in the "I" area (I=1,…, 9);
State A_I: Away team is in possession of the ball and the ball is located in the "I" area (I=1,…, 9);
State A_G: Away team scores a goal.

There are two states for goal scoring (states H_G and A_G) and 18 states relating to the location and team's possession of the ball. Figure 1 illustrates the Markov process model, and the relation between the states.

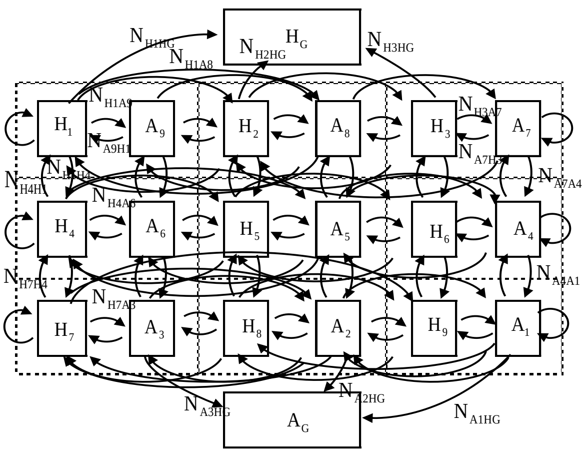

Figure 1. The Markov process model of a soccer game

T_i is the total time for which the game is in state i in a game (i=H_1,H_2,…,A_1);
N_{iHG} is the total number of goals scored by home team from state i in a game (i=H_1,H_2,…,H_6);
N_{ij} is the total number of transitions from state i to state j in a game (i,j=H_1,H_2,…,A_1);
N_{iAG} is the total number of goals scored by for away team from state i in a game (i=A_1,A_2,…,A_6).

The transition probabilities between them are defined in Table 1. In this table, a_{H1HG} is interpreted as the transition rate from state H_1 to H_G (i.e. scoring a goal from the "1" area by home team). The probability of a transition from H_1 to H_G and a transition from H_1 to H_2 in the next small time dt is expressed by $a_{H1HG}\cdot dt$ and $a_{H1H2}\cdot dt$, respectively. Other transitions are also expressed in a same manner. The probability of remaining in state H_1 is thus $1-(a_{H1HG}+a_{H1H2}+\cdots+a_{H1A1})dt$. Similarly, a_{ij} (i,j= H_1,H_2,⋯,A_1) is defined as the transition rate from state i to state j. While this is clearly a simple model to represent a very complex process, the model does reflect the most fundamental aspects of a game (goals and possession) and therefore may nonetheless be useful.

Table 1. Definition of transition probabilities from state i

Transition	Probability	Remarks
$i \to H_G$	$a_{i\,HG}\,dt$	Transition to scoring a goal for home team
$i \to j$	$a_{ij}\,dt$	Transition from state i to state j
$i \to A_G$	$a_{i\,AG}\,dt$	Transition to scoring a goal for away team

Using the transition probabilities shown in Table 1, we can obtain the probability distributions of the number of goals scored. The above method is also extended to obtain the probability of winning. Let $W_i(r|t)$ be the probability of home team winning from a position of leading by r goals with time t remaining, starting from state i (i = H_1,H_2,…,A_1). Then it can be seen that:

$$W_i(r\,|\,t+dt)=W_{H_G}(r+1\,|\,t)\cdot a_{iH_G}dt+W_{H_1}(r\,|\,t)\cdot a_{iH_1}dt+W_{H_3}(r\,|\,t)\cdot a_{iH_3}dt+\cdots$$
$$+W_{A_1}(r\,|\,t)\cdot a_{iA_1}dt+W_i(r\,|\,t)\cdot\{1-a_{iH_G}-a_{iH_1}-a_{iH_3}-\cdots-a_{iA_1})dt\} \quad (1)$$

In order to obtain the probability of winning, we need to solve this equation with the boundary conditions at the end of the game such that $W_i(r|0)$= 1 if r > 0 and 0 if r < 0. In this paper, we set $W_i(r|0)$=0.5 only if r = 0 in the case of drawing.

3. Estimation of Transition Rates

By solving Equations (1), the probability of winning the game can be derived using the estimators for the parameters such as a_{H1HG} and a_{H1H2}. If appropriate data are available, it is possible to deduce an estimate for these parameters for the game. If the total numbers of transitions and the time spent in each state are all known, the transition rates can be estimated thus:

$a_{i\mathrm{HG}} = \mathrm{N}_{i\mathrm{HG}} / \mathrm{T}_i$ $(i=\mathrm{H}_1,\mathrm{H}_2,\ldots,\mathrm{H}_6)$
$a_{ij} = \mathrm{N}_{ij} / \mathrm{T}_i$ $(i,j=\mathrm{H}_1,\mathrm{H}_2,\ldots,\mathrm{A}_1)$
$a_{i\mathrm{AG}} = \mathrm{N}_{i\mathrm{AG}} / \mathrm{T}_i$ $(i=\mathrm{A}_1,\mathrm{A}_2,\ldots,\mathrm{A}_6)$.

We can obtain the total numbers of transitions between states with the total time spent in each state for each game from the real data of the 2015 season of the J-League Division 1. In J-League Division 1, there are 18 teams and 306 matches played in a season. For this study, Data Stadium Inc., which helps to supply official data to the J-League, provided a large amount of data for the 306 games in the season. We extracted several data for our analysis, in which the events occurred during the game are recorded with the time. Time is measured from the beginning of the game and the location of the ball is identified as a *x* and *y* coordinate. For each game, around two thousand events are recorded, and we can use this basic information regarding to goals and possession of the ball with time and location for our analysis using Markov process models.

The total number of goals was 818, and total possession time was 16,833 minutes in the season. This total possession time corresponds to 55.2 (=16,833/306) minutes per game. This is not 90 minutes because in the measurement of possession time, we extracted consecutive possession time from the data, and did not count the following events toward the time of possession: Ball-out, Foul, Penalty, Offside, Substitution, and Goal.

4. Calculation Result

Using the transition rates estimated in the above, we calculated the probability distribution of scoring goals, the expected number of goals scored by home team, and the probability of home team winning. Table 2 represents the summary of the calculation results.

Table 2. Calculation result

				+1SD		-1SD	
				H_4H_1	H_7H_4	H_4A_9	H_7A_6
Goals	H_2	H_5	H_8	H_5	H_5	H_5	H_5
0	0.244	0.250	0.251	0.240	0.247	0.247	0.248
1	0.346	0.348	0.349	0.344	0.347	0.347	0.348
2	0.244	0.241	0.240	0.245	0.242	0.242	0.242
3	0.114	0.110	0.109	0.116	0.112	0.112	0.111
4	0.039	0.038	0.037	0.041	0.038	0.038	0.038
Exp. Num.	1.401	1.376	1.370	1.417	1.388	1.388	1.382
Prob. Win.	0.540	0.535	0.534	0.543	0.538	0.539	0.538

In Table 2, from state H_2 the probability of the home team scoring no goals is 0.244 in a match. This probability increases a little in the case from state H_5, and further increases a little from state H_8 to 0.251. That is, when the home team is in possession of the ball, the difference of the location between "2" and "8" area affects the difference of probability of scoring no goal by 0.007. In the same situation, the expected number of goals scored decreases from 1.401 to 1.370, and the probability of winning the game also decreases from 0.540 to 0.534.

As an advantage of using the Markov process model, we can calculate the probabilities in terms of scoring or winning, and evaluate the effect of the change of transition rates on them. Concretely to see the sensitivity of the transition rate, we here look at the transition from the "4" to "1" area and from the "7" to "4" area. We change the transition rates by the amount of its 1SD. The calculation result of this effect has been presented in the right side of Table 2. As shown in Table 2, if we increase the transition rate from H_4 to H_1 from 4.1 to 5.5 (=4.1+1.4) times/minutes (i.e. the increase of 1SD of the transition rate), the probability of scoring no goals decreases from 0.250 to 0.240 by 0.010, and the expected number of goals scored in a game increases from 1.376 to 1.388 by 0.012, when home team kick off in state H_5. The probability of winning the game also increase from 0.535 to 0.543. Similarly, the case of the change of transition rate from H_4 to A_9, is also calculated by changing the transition rate from 0.6 to 0.2 (=0.6-0.4). We note that A_9 is the "1" area from the aspect of the home team (corresponding to the "9" area from aspect of the away team). Decreasing this transition rate by 1SD results the increase of the expected number of goals from 1.376 to 1.388 by 0.012. We also present the effect of the change of transition rate from H_7 to H_4 and H_7 to A_6, as shown in Table 1.

Actually, the time spent in state H_4 were observed as 1426.1 minutes which corresponds to 4.66 minutes/game, and the change of the transition rate was just 1.4 or 0.4 times/minutes as 1SD. Although the effect of the changes looks small, we demonstrated how to calculate the effects by this type of approach, which would be useful for analysis of the game.

5. Further study

In this paper, we have just shown the calculation result of the change of transition rate from state H_4 and H_7 by 1SD. As this study is still in progress, we plan to present more in the Fall National Conference. Further, we plan to estimate the transition rates using log-linear models which explain such factors as home advantage, offensive and defensive strength, in terms of goals and possession, according to the location, and discuss the effect quantitatively.

References

[1] Hirotsu, H., Inoue, K., & Yoshimura, M. (2017). An analysis of characteristics of soccer teams using a Markov process model considering the location of the ball on the pitch, In MathSport International 2017 Conference Proceedings, 176-183.
[2] Hirotsu, N. & Wright, M. (2003). Determining the Best Strategy for Changing the Configuration of a Football Team. Journal of the Operational Research Society, **54**, 878-887.