Modelling a football match as a Markov process: Estimating teams' strengths relating to the degree of the division of the pitch

01506960 Juntendo University * HIROTSU Nobuyoshi
Juntendo University INOUE Keita
Juntendo University YAMAMOTO Kenji
Juntendo University YOSHIMURA Masafumi

1. Introduction

Data analysis in sports for academic research has been expanding with the increase in available data and the help of a variety of methods. A Markov model is one of the useful tools for the analysis. In association football, each play is part of a continuous time process, and the series of plays during a game continues without any predetermined pauses. With the nature of football, modelling a game was mainly focused on the goal-scoring aspect, but some previous studies [e.g.1-3] introduce ball possession to the Markov model. We propose the Markov process models, which provide a macro view of a game and to present a method for evaluating teams' characteristics.

2. The Markov process model

We assume a Markov property in the transitions of football match, and propose a Markov process model, which seems appropriate to a football match as an approximation. We divide the pitch into up to 9 areas and define the states as follows in the case of the 9 areas:

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.

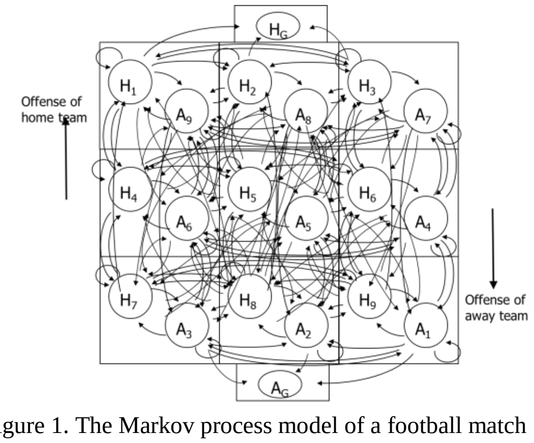

Figure 1. The Markov process model of a football match

We further define the following notations:

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

We also define the following factors as fixed effects for possession:

μ^{ij}: Intercept for the propensity of changing from state i to state j;
$\mu^{ij}{}_h$: Home advantage for the propensity of changing from state i to state j;
$\mu^{ij}{}_{off}$ (X): Offensive strength of team X for the propensity of changing from state i to state j;
$\mu^{ij}{}_{def}$ (X): Defensive strength of team X against the propensity of changing from state i to state j.

We use the possession times as offsets, and the rate of transition from state i to state j, r_{ij} can be explained by the above factors for the home team in the full model in terms of a game of X_h playing at home against X_a as follows:

$$\log(r_{ij}) = \mu^{ij} + \mu^{ij}{}_h + \mu^{ij}{}_{off}(X_h) + \mu^{ij}{}_{def}(X_a) + \mu^{ij}{}_h * \mu^{ij}{}_{off}(X_h) + \mu^{ij}{}_h * \mu^{ij}{}_{def}(X_a) + \mu^{ij}{}_{off}(X_h) * \mu^{ij}{}_{def}(X_a) + \mu^{ij}{}_h * \mu^{ij}{}_{off}(X_h) * \mu^{ij}{}_{def}(X_a)$$

$$N_{ij} \sim Poisson(r_{ij} T_i)$$

In the above expressions, offensive and defensive strengths for a team are estimated as fixed effects. As these strengths may come from a distribution, we also analyse them as random effects with associated variance components.

We look at the 2015 season of J-League Division 1, which has 18 teams. Each team plays 34 games in the season, for a total of 306. S. Hiroshima won the league championship with the highest number of goals scored, 73, and the lowest number of goals conceded, 30. A large amount of the data were provided by Data Stadium Inc. for the 306 games in the season, and we extracted several items of the data for our analysis. Table 1 shows the observed numbers of transitions between states for each game and the time spent in each state.

Table 1. Observed numbers of transitions between states and observed length of possession time in each state for each game

No.	Home	Away	Goals $N_{H_p^1 H_G}$	$N_{H_p^2 H_G}$	$N_{H_p^3 H_G}$	$N_{A_p^1 A_G}$	$N_{A_p^2 A_G}$	$N_{A_p^3 A_G}$	Possession $N_{H_p^1 H_p^2}$	$N_{H_p^1 A_p^8}$	$N_{H_p^3 H_p^2}$	$N_{H_p^5 H_p^2}$	$N_{H_p^5 A_p^5}$	…	$N_{A_p^1 A_p^2}$	$N_{A_p^2 H_p^8}$	$N_{A_p^3 A_p^2}$	$N_{A_p^5 A_p^2}$	$N_{A_p^5 H_p^5}$
1	V. Sendai	M. Yamagata	0	2	0	0	0	0	13	2	10	3	2		12	11	22	6	6
2	M. Yamagata	V. Sendai	0	1	0	0	1	0	14	2	6	6	5		14	16	10	5	7
3	S.Hiroshima	V. Sendai	0	2	0	0	0	0	14	3	17	3	2		7	15	11	8	3
4	V. Sendai	S. Hiroshima	0	3	0	0	3	1	22	2	19	8	2		7	8	9	5	4
5	S. Hiroshima	M. Yamagata	0	5	0	0	1	0	13	4	14	9	5		17	17	18	9	4
6	M. Yamagata	S. Hiroshima	0	1	0	0	2	1	20	4	11	10	2		9	13	14	3	5
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
305	A. Niigata	V. Kofu	0	0	0	0	2	0	16	5	11	12	3		10	14	6	6	2
306	Shimizu S. P.	V. Kofu	0	0	0	1	1	0	27	8	22	12	3		8	12	3	2	5
	Total		14	404	12	6	364	18											

No.	Time (min.) $T_{H_p^1}$	$T_{H_p^2}$	$T_{H_p^3}$	$T_{H_p^4}$	$T_{H_p^5}$	$T_{H_p^6}$	$T_{H_p^7}$	$T_{H_p^8}$	$T_{H_p^9}$	$T_{A_p^1}$	$T_{A_p^2}$	$T_{A_p^3}$	$T_{A_p^4}$	$T_{A_p^5}$	$T_{A_p^6}$	$T_{A_p^7}$	$T_{A_p^8}$	$T_{A_p^9}$
1	1.4	1.3	2.1	3.2	2.9	4.5	0.8	3.5	1.6	2.1	3.6	2.5	4.7	3.1	3.7	2.5	3.0	2.0
2	3.0	2.1	1.8	4.3	2.1	2.7	1.5	2.7	1.7	2.0	2.5	1.8	4.0	2.1	3.1	1.3	3.7	2.0
3	2.9	1.9	2.8	3.4	3.4	6.0	2.6	4.8	2.0	1.9	1.5	3.7	4.5	6.7	6.4	1.1	3.5	1.2
4	4.2	3.5	3.5	6.6	6.9	4.8	1.5	2.6	1.0	1.8	1.4	1.7	4.3	3.3	2.8	3.3	4.0	2.1
5	2.3	2.2	1.4	4.3	3.3	3.6	3.2	4.7	3.4	2.9	2.8	2.8	5.4	4.8	3.7	1.4	2.8	0.8
6	2.5	2.0	2.6	4.1	5.9	6.0	1.7	2.8	1.2	1.6	1.6	2.3	4.0	4.1	4.8	3.3	6.7	2.0
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
305	3.3	4.1	1.7	6.4	6.4	4.9	2.0	2.3	1.4	2.1	1.6	2.7	2.8	3.6	4.6	2.0	4.7	1.9
306	3.5	5.3	2.6	7.3	5.3	4.8	1.7	3.1	1.1	1.4	0.9	1.4	3.0	1.9	2.7	2.4	4.2	2.1
Total	820.0	783.4	746.4	1,422.5	1,247.4	1,343.6	568.3	1,064.0	564.9	697.9	724.7	719.5	1,366.1	1,193.2	1,304.6	573.0	1,109.8	562.3

3. Result and discussion

We selected the suitable model for expressing the number of transition using fixed effects log-linear models. The expected frequency of goals and possession changes are calculated by using the maximum likelihood estimators of these factors in the model that achieves the minimum AIC. We tested the Poisson assumption by comparing it to the observed frequency with p-values of the χ2-goodness of fit test. Figure 2 shows the case of transition from H_p^1 to H_p^2 as an example. Model $\log(r_{H_1^p H_2^p}) = \mu^{H_1^p H_2^p} + \mu^{H_1^p H_2^p}{}_h + \mu^{H_1^p H_2^p}{}_{off}(X_h)$ minimises AIC for transition from '1' area to '2' area with keeping possession.

Model: $$\log(r_{H_1^p H_2^p}) = \mu^{H_1^p H_2^p} + \mu^{H_1^p H_2^p}{}_h + \mu^{H_1^p H_2^p}{}_{off}(X_h)$$

Residual Degrees of Freedom	593
Number of Parameters	19
$\log L^*$	-1649.3
Deviance -2logL*	3298.6
Residual Deviance	610.8
AIC (Minimum)	3336.6

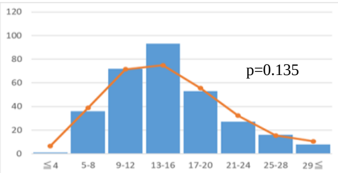

Figure 2. Observed and expected number of transition from H_p^1 to H_p^2 with statistical information for calculating AIC

Table 3 presents some of the results of the maximum likelihood estimation of the parameter values in terms of fixed effect model, with regard to the intercept and home advantage and offensive and defensive strengths of the 18 teams. In defensive strength, more negative values correspond to greater strength, thus conceding fewer goals or allowing the opposing team to move the ball between areas (whilst maintaining possession) less frequently. In Table 3, the parameter values are converted to set the sum of these values to zero (i.e. the parameter values of an average team are converted to zero) owing to easy interpretation of these values. Now, we consider S. Hiroshima's goal-scoring process through transitions, that is, moving the ball from the '5' area to the '2' area whilst maintaining possession. S. Hiroshima seems not particularly good at such transitions; the offensive strength for such transitions is -0.373 (95%C.I [-0.504, -0.240]). However, we can observe that S. Hiroshima is good at transitions from '1' to '2', with a value of 0.046 (95%C.I [-0.045, 0.137]) and from '3' to '2', with a value of 0.134 (95%C.I [0.047, 0.221]). In other words, S. Hiroshima is not particularly good at delivering the ball from the centre of the pitch to the area in front of the goal, but he is good at delivering the ball from the right side to the area in front of the goal.

Although we here did not present the result of random effect model, we will show it in the conference.

Table 3. Results of the maximum likelihood estimations of parameter values in terms of the intercept, home advantage, and offensive and defensive strengths of J-league teams

	Transition from H_p^5 to H_p^2		from H_p^1 to H_p^2	from H_p^3 to H_p^2	
Intercept	0.482		1.704	1.710	
Home advantage	-0.003		0.031	0.054	
Team	Offense	Defense	Offense	Offense	Defense
S. Hiroshima	-0.373	-0.036	0.046	0.134	-0.047
Urawa R. D.	-0.050	0.120	0.142	0.138	0.185
G. Osaka	-0.252	0.103	-0.169	-0.100	0.015
F. C. Tokyo	-0.098	-0.123	-0.055	-0.144	0.021
Kashima A.	0.041	-0.182	-0.061	-0.007	-0.083
Kawasaki F.	0.258	0.031	-0.044	0.125	0.024
Yokohama F.	-0.192	-0.078	-0.085	-0.049	-0.101
Shonan B.	0.203	0.078	-0.127	-0.057	-0.075
Nagoya G. E.	0.069	0.029	0.122	0.098	0.069
Kashiwa R.	-0.069	-0.075	-0.088	-0.118	0.082
S. Tosu	0.103	-0.042	0.068	0.076	0.013
V. Kobe	0.197	0.124	-0.013	-0.133	-0.015
V. Kofu	-0.218	-0.025	0.028	-0.208	0.020
V. Sendai	-0.001	-0.014	-0.078	-0.023	0.064
A. Niigata	0.054	0.109	0.049	0.002	-0.094
Matsumoto Y.	0.226	0.171	0.221	0.143	-0.076
Shimizu S. P.	0.132	-0.202	0.102	0.129	0.031
M. Yamagata	-0.029	0.012	-0.058	-0.004	-0.031

4. Further study

We did not show the details of the analyses for other teams or other transitions in other games, but we will present details about the characteristics of other teams in a further work.

References

[1] Hirotsu, N. & Wright, M. (2003) An evaluation of characteristics of teams in association football by using a Markov process model. Stat., 52, 591–602.
[2] Heiner, M., Fellingham, G. W. & Camille, T. (2014) Skill importance in women's soccer. J. Quant. Anal. Sports, 10, issue 2, 16.
[3] Woodfield, D.N. & Fellingham, G.W. (2018) Quantifying the probability of a shot in women's collegiate soccer through absorbing Markov chains. J. Quant. Anal. Sports, 14, issue 3.