

축구경기의 경기력분석

민대기¹ · 이용수² · 김용래³

¹덕성여자대학교 정보통계학과 · ²세종대학교 체육학과 · ³세종대학교 스포츠과학연구소

접수 2014년 12월 6일, 수정 2014년 12월 28일, 게재확정 2015년 1월 19일

요약

축구경기에서 승패를 결정 하는 것은 골득실이고 경기에 대한 분석은 일반적으로 득점은 공격력으로, 실점은 수비력으로 평가한다. 본 연구에서는 축구경기력에 대한 분석을 함에 있어서 승패와 득점, 실점에 미치는 요인이 무엇인가를 밝혀내고자 하였다. 경기의 승패를 결정하는 요인들을 밝혀내기 위하여 의사결정나무, 로지스틱 회귀모형 그리고 판별함수 등을 이용한다. 그 결과 공격보다는 수비와 관련된 요인이 승부에 더 결정적인 영향을 미치는 것으로 나타났다. 공격력과 수비력에 대한 분석을 실행하기 위하여 득점과 실점을 반응 변수로 사용해 본 결과, 공격력에 있어서는 논스톱패스와 공격속도가 주요한 요인이었고, 수비력에서는 수비 시 공수거리와 좌우 폭이 주요한 결정요인으로 나타났다.

주요용어: 로지스틱, 의사결정나무, 판별분석, 포아송 회귀분석.

1. 서론

우리나라의 축구는 2002년 한일월드컵을 계기로 하여 세계축구의 추세에 발맞추어 많은 발전을 해왔다. 당시의 월드컵 국가대표를 맡은 허딩크 감독은 생각하는 축구, 빠른 공수전환, 그리고 강한 압박을 통한 수비 등을 한국 대표 팀에게 교육시켰고, 그 결과 우리 대표팀은 4강에 이르는 쾌거를 이루었다. 허딩크 감독은 비디오 기술 분석을 이용하여 선수들을 지도하였고, 한국축구는 처음으로 과학적인 분석을 도입하는 계기가 되었다. 2000년도 IT (information technology) 기술은 색깔인식을 기반으로 한 로봇 축구를 가능하게 할 정도의 수준이었기에 선수들의 움직임을 추적하여 경기에 대한 자세한 기록을 자동으로 하는 것은 불가능 하였고 비디오를 이용한 분석기술이 가장 최선의 선택이었다. 그러나 비디오 경기분석은 녹화 된 동영상을 다시 시청하는 방법과 그것을 항목별로 정리하여 활용 하는 정도였기 때문에 경기력 설명에 대하여 부족함이 많았다. 무용이나 골프 등의 자세교정을 위하여 개발되었던 비디오 기술 분석에 대한 한계를 극복하고자 노력한 결과 동적인 경기종목에서 선수들의 일반적인 움직임을 기록하여 경기분석으로 활용하려는 방법에 대한 연구가 처음으로 시도 되었다. 이 방법은 오디오테이프 기록된 내용을 수기로 입력, 분석하는 시스템 이었다 (Reilly와 Thomas, 1976). 이후에도 축구에 대한 스포츠관련 산업이 발달한 유럽에서는 지속적인 움직임의 기록이 가능한 notation system을 이용한 연구가 지속되었다 (Olsen와 Larsen, 1997). Notation system은 경기에서 선수들의 움직임을 기록하는 시스템으로 초기에는 22명의 움직임을 한꺼번에 기록 하는 것이 불가능하여 볼을 소유한 선수들만의 움직임을 기록하였고, 또한 전술적으로 가장 중요한 선수들의 정확한 위치정보가 불가능하여 컴퓨터 모니터

¹ 교신저자: (132-714) 서울시 도봉구 쌍문동 419, 덕성여자대학교 정보통계학과, 교수.

E-mail: dkmin@duksung.ac.kr

² (143-747) 서울시 광진구 능동로 209, 세종대학교 체육학과, 교수.

³ (143-747) 서울시 광진구 능동로 209, 세종대학교 스포츠과학연구소, 선임연구원.

터에 가상의 축구장을 그려놓고 마우스로 입력을 하였다 (Hughes, 1996). 경기력연구에 대한 관련문헌은 Perl (2001), Suzuki와 Nishijima (2004) 등이 있다. 그러나 이러한 시스템은 자료에 대한 정확성과 객관성이 부족하여 일반화하지 못하였다. 그러나 최근 들어 비디오 인식, 추적에 대한 기술과 IT분야의 무선기술 개발은 자료수집방법의 놀라운 발전으로 이어졌고 (Goldsberry, 2012) 이를 바탕으로 축구에서 경기력의 분석이 더욱 체계화되었다 (Taylor 등, 2008).

IT 기술의 발전은 과거에는 불가능 했던 많은 것들에 대하여 자동으로 기록하고 전달, 저장하는 것이 가능하게 하였다. 이를 바탕으로 스포츠에 있어서는 정확한 위치정보를 실시간으로 감지기를 통해서 저장, 분석하는 시스템을 가능하게 하였다. 2014년 브라질 월드컵에서 독일 팀이 우승하게 된 원동력이 되었던 SAP의 기록 분석시스템과 FIFA (Federation International Football Association)의 항목별 자세한 경기기록 (<http://www.fifa.com/worldcup/matches/index.html>)은 이런 과학적인 축구발전의 결과물이라 할 수 있겠다. 본 연구에서는 이러한 IT기술을 기반으로 개발된 비주얼스포츠사의 자료를 활용하여 과거의 득점과 실점만을 이용한 Keller (1994)와 같은 제한적인 방법에서 벗어나 실질적인 경기력을 분석하고자 한다.

2. 연구내용

2.1. 자료설명

축구는 세계에서 가장 인기가 있는 종목으로 프로팀간의 승부에 대한 경쟁은 날로 치열해 지고 있다. 또한 IT기술의 발달로 경기력향상을 지원하는 장비들이 새롭게 개발 보급되고 또 이를 바탕으로 하는 연구들이 활발하게 진행되고 있다. 축구종목은 야구나 배구와 달리 몸이 접촉하는 매치 플레이기 때문에, 농구처럼 상대방에 대한 철저한 방어가 중요하고 이를 뒷받침하는 스피드와 체력이 있어야 가능하다. 농구처럼 빨라지고 지능화 되어가는 현대축구에서 경기력의 발전을 위해서 고려해야 하는 요인들은 공격, 수비, 전술, 체력 등이 있다. 이러한 요인들 중에서 직접적인 것은 무엇일까를 의사결정나무와 같은 예측모형을 이용하여 분석해 보았다. 또한 경기력을 공격력과 수비력으로 나누어 이들 각각을 결정하는 요인은 무엇일까 하는 것을 일반화 선형모형 (generalized linear model)을 통하여 분석해 보았다. 과거의 경기분석은 경기내용에 대한 자료부족으로 단지 녹화 된 동영상을 다시 시청하는 비디오 분석과 함께 주로 감독과 코치의 주관적 판단에 의존하였다. 최근에는 디지털기기의 발달로 FIFA (Federation International Football Association)에서 경기에 대한 자료를 구체적으로 수집하여 다양한 자료를 제공하기는 하지만 기록에 대한 정이가 체계화 되어 있지 않고 제공되는 경기수도 한정되어 있어서 경기내용에 관한 분석은 아직도 활발하게 진행되지 못하고 있는 실정이다. 이번 연구에 사용된 비주얼 스포츠사의 시스템은 8대의 카메라를 경기장에 배치하여 영상 인식 추적 장치가 가능한 프로그램을 이용하여 실시간으로 90분간 움직인 22명의 선수들의 경기내용을 기록 할 수 있다. 이와 같은 방법으로 수집된 2013년 K-리그 76 경기 중 무승부 경기는 승패와 관련이 없어서 이번 분석대상에서 제외하고 51경기에 대해서만 연구를 하였다. 또한 자료에서 수집된 변수 중 득점, 실점이 변수정의에 관련이 있는 골 결정력, 페널티킥, 킬패스 (kill pass) 등의 요소는 배제하였다. 그 이유는 득점과 실점이 경기력분석에서 반응변수로 설정되었기 때문이다.

2.2. 변수의 정의 및 설명

분석대상 변수의 설정 방법은 각 팀의 득점과 실점을 계산하여 득점이 실점보다 많은 경우를 이긴 경기로, 적은경우는 패한 경기로 정의하였다. Table 2.1에 SPSS의 Modeler의 Feature 노트를 이용하여 이번 분석에 선택한 변수를 정리 하였다. sprint_dist는 스프린트 거리로 팀 전체선수들의 시속

25.2 이상 달린 거리를 90분간 측정하는 것이다. block은 골키퍼 외의 수비수가 슈트를 막아낸 횟수를 측정하는 수비에 관련된 변수이다. pe_shots는 페널티에어리어 지역에서의 슈팅 횟수를 말하며 승패와 관련된 변수이다. pass_freq는 경기 중에 이루어진 팀의 전체패스 횟수를 측정하는 것이다. def_depth는 수비 시 공수간격으로 최전방 공격수와 최후방 수비수와의 거리를 측정하는 것으로 수비에 관련된 변수이다. ave_wide는 경기 중 우측, 좌측 선수간의 거리를 측정하는 것이다. ave_depth는 90분간 공격수와 최후방 수비수와의 평균거리를 뜻한다. ksave_freq는 골키퍼가 막아낸 결정적인 슈트를 측정하는 것이고 sprint_freq는 스프린트 횟수를 측정하는 것이다. dpass_freq는 공격 진영에서 이루어진 논스톱패스 횟수로 공격과 관련 있는 변수이다. sog는 골대 안으로 슈팅 횟수를 측정하는 것이다. tempo는 경기 중 총 볼터치시간을 총 볼터치횟수로 나눈 것으로 볼 처리속도를 나타낸다. ha_shots는 페널티에어리어 밖의 슈팅 가능지역에서 발생한 슈팅 수를 측정하는 것이다.

Table 2.1 Definition of variables

Variable	Definition (unit)	role
sprint_dist	the whole team sprint distance traveled (meter) when the team owns the ball	predict variable
block	number of blocking the shots except goal keeper	
pe_shots	shooting times in the penalty area	
pass_freq	total number of passes	
def_depth	distance between front line attacker and rearmost defender when opponents own the ball (meter)	
ave_wide	distance between right and left player (meter)	
ave_depth	distance between front line attacker and rearmost defender during the game (meter)	
ksave_freq	number of blocking the dangerous shots by the goal keeper	
sprint_freq	sprint frequency when the team owns the ball,	
dpass_freq	the number of non-stop pass from attacking zone	
sog	shot on goal	
tempo	total ball possession time/number of total ball touch from attacking zone (second)	
ha_shots	attempts shooting in available area	
results	the winning game 1, lost game to display to 0	target variable

3. 연구결과

오늘날과 같이 정보통신기술이 발달하지 못한 과거에는 축구에 대한 계량적인 연구가 다른 종목인 야구나 농구에 비하여 한계적이었다. 그 근본적인 이유는 경기내용에 대한 자료수집의 어려움 때문이었다. 프리킥, 슈트 등과 같이 세기 (count)가 가능한 결과에 대한 기록만 가능 할 뿐 경기내용에 대한 기록이 수기로 불가능하여 경기결과에 대한 과학적 분석이 불가능했다. 2012년 FIFA에서는 종전 월드컵보다 많은 경기에 대한 자료를 제공하고 있지만 아직도 기술력을 분석하기엔 미비한 수준이며 통계에 대한 명확한 정의와 설명이 없다. 그러므로 과거 축구경기의 분석은 설명변수가 없이 승패에 관한 경기기록만을 토대로 하여 승부를 예측하였다. 최근 축구에 관한 연구로는 Hong 등 (2010)이 확률모형을 이용하여 월드컵 경기예측을 한 것과 Choi 등 (2011)의 사회네트워크를 이용한 패스에 관한 분석 등이 있다. 이에 비하여 구체적인 경기기록이 가능한 다른 종목의 경우를 보면 골프에서는 성장곡선을 이용한 경기력변화 연구 (Kim과 Min, 2014)가 있고, 야구에서는 투수에 관한 평가지표 (Lee, 2014) 등이 있다. 축구에 대한 스포츠산업이 상대적으로 발달한 미국, 영국, 기타 유럽국가에서는 일찍부터 경기내용을 기록하는 방법과 승부예측에 대한 연구가 진행되고 있다. 승부예측에 관한 연구를 보면 득점과 실점을 이용한 이변량포아송분포는 Karlis와 Ntzoufras (2003) 등이 있다.

3.1. 중요변수의 평균검정

승패 (results)를 반응변수로 설정한 다음 주요변수를 선택하여 경기력차이를 분석하고자 평균검정을 하였다. 승리한 팀과 패한 팀의 경기력 차이를 difference로 표시하였다. Table 3.1에서 보여준 변수나 열 순서는 반응변수와 상관도에 따라 정리한 것이다. 총 패스횟수 (pass_freq), 수비 시 공격수와 수비수의 거리 (dep_depth) 그리고 경기 중 좌측 우측선수 간의 거리 (ave_wide) 등도 패한 팀이 높게 나타났다.

Table 3.1 Average test between wins and losses

Variable	t value	degree of freedom	difference	standarderror	p-value
sprint_dist	2.891	49	0.156	0.054	0.006**
block	2.405	49	1.141	0.474	0.02**
pe_shots	2.325	49	1.92	0.826	0.024**
pass_freq	-2.305	49	-58.294	25.295	0.025**
def_depth	-2.119	49	-2.28	1.076	0.039**
ave_wide	-2.016	49	-1.782	0.884	0.049**
ave_depth	1.78	49	2.891	1.624	0.081*
ksave_freq	1.771	49	0.675	0.381	0.083*
sprint_freq	2.25	49	8.378	3.724	0.029**

p-value < 0.05 ** p-value < 0.1 *

3.2. 예측모형을 이용한 승패요인분석

Table 3.2에 각 예측 모형의 결과를 정리 하였다. 3가지 방법의 정확도는 매우 근사하게 나타났으나 의사결정나무 (decision tree)는 13/25 (52%)로 민감도가 다른 결과에 비하여 낮게 나타났다. 민감도와 특이도는 이진분류법에 대한 통계적 방법이다. 민감도를 일반적으로 양의 비율이라 하며 실제 양성 이 양성으로 분류되는 비율을 말한다. Table 3.2에서는 실제 승리한 경기를 승리한 것으로 분류된 비율을 뜻한다. 특이도는 음의 비율이라 할 수 있으며 실제 패한 경기가 패한 것으로 분류되는 비율이다. Table 3.2에서는 패한 경기를 패한 것으로 분류한 비율을 말한다.

Table 3.2 Prediction results evaluation

Methods	prediction	W	L	
Decision Tree	W	13 (25)	13	37/51 (72.5%)
	L	1	24 (26)	
Logistic	W	18 (24)	7	37/50 (74.0%)
	L	6	19 (26)	
Discriminant	W	19 (25)	7	38/51 (74.5%)
	L	6	19 (26)	

3.2.1. 의사결정나무

SPSS Modeler의 C5.0 의사결정나무방법을 이용하여 자료를 분석 하였다. 과적합을 방지하기 위하여 가지치기 강도를 75%로 하고 최소가지 개수를 5로 설정하였다. 분석결과 전체 51게임 중 37게임을 정확하게 분류하여 정확도는 72.55%이다. 정확도는 다른 예측방법과 비슷하나 Table 3.2에서 보는 바와 같이 민감도가 13/25=52%로 이진 경기에 대한 오차가 크다는 것을 알 수 있었다. Figure 3.1의 나무도표 결과에 의하면 좌우 폭이 35.5미터보다 크면 분류된 6경기가 모두 패하고 35.5미터이하이고 블록이 3보다 크면 14경기 중 13경기 이진 경기로 분리가 되었다. 블록이 3이하이면 나머지 31경기 중 18경기가 패한 경기로 분류 되었다. 블록은 수비력을 측정하는 변수이고 단지 이것에 의하여 승리한 경기를 분류한 것이므로 민감도에 대한 오차가 크게 나타난 것으로 해석 할 수 있다.

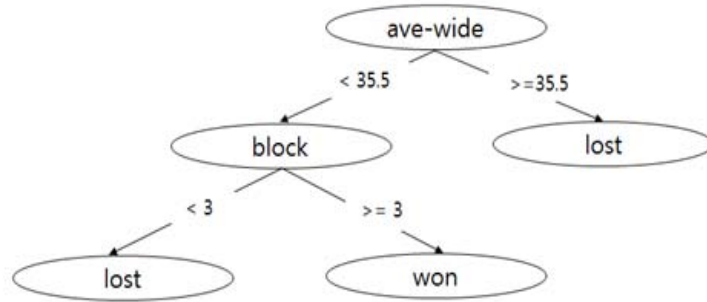


Figure 3.1 Prediction results by decision tree

Table 3.3 Prediction rules by decision tree

ave_wide	block	count	percentage	result
> 35		6/6	100%	lost
≤ 35	> 3	13/14	93%	won
	≤ 3	18/31	58%	lost

Table 3.3은 나무도표 Figure 3.1을 정리한 것이다. 분리규칙에 따르면 좌우 폭이 35이상이 되면 100% 패한 경기이다. 이 그 만큼 밀집하게 방어를 못하면 패할 가능성이 많다는 것을 의미한다. 블록 변수 역시 수비에 관련된 요인으로 슈팅을 3번 이상 방어하면 41경기 중 13번이 승리 한 것으로 분류 되었다. 반면에 수비수가 3번 이하로 블록 한 경우는 전체 31경기의 18번이 패한 것으로 분류가 되어 60%의 정확도를 나타냈다. 이는 전체 50% 승률보다 10% 정도 높은 것이다. 결론적으로 수비력이 승패를 결정하는 열쇠가 되는 것을 알 수 있다.

3.2.2. 로지스틱 모형을 이용한 분석

승리한 경기를 1로 설정한 로지스틱모형의 정확도는 37/50=74%이다. 민감도는 18/24=75%이고 특이도는 19/26=73%이다. 승과 패에 대한 각각의 통계 값은 정확도와 비슷하고 특정 값의 오차가 크지 않아 균형적으로 예측함을 알 수 있다. 결과를 정리한 아래의 Table 3.4에 의하면 블록의 단위 증가는 경기에 승리할 오즈가 80% 증가하고, 수비 시 공수간격이 1 단위 증가하면 패할 오즈가 37% 증가함을 알 수 있다. 의사결정나무에서는 좌우 폭의 길이가 넓으면 패한 것으로 분류가 된 것에 비하여 아래 Table 3.4에서는 종적인 폭을 나타내는 수비 시 공수간격, 즉 공격수의 수비가담이 적극적이지 못하면 패할 경우가 많은 것으로 분석되었다. Table 3.4에 의한 모형식은 (3.1)이다.

$$\text{Log} \left(\frac{\pi}{1 - \pi} \right) = 7.98 - .32\text{dep_depth} + .59\text{block} \tag{3.1}$$

Table 3.4 Prediction model of game using logistic regression

parameter	estimate	standarderror	wald	p-value	exp(estimate)
intercept	7.98	3.79	4.44	0.035	
def_depth	-0.32	0.13	5.93	0.015	0.73
block	0.59	0.24	6.20	0.013	1.81

3.2.3. 판별분석을 이용한 승패분석

Table 3.2의 승패요인분석에 사용한 설명변수들이 연속 형이고 의사결정나무는 질 단점에 따라 결과가 달라지는 단점이 있어 판별분석으로 그 결과를 비교하였다. 정확도의 결과는 $38/51=74.5\%$ 이다. 정확도, 민감도 그리고 특이도는 로지스틱회귀모형과 거의 같은 결과가 나왔다. Table 3.5의 표준화계수로 판별함수를 표시하면 식 (3.2)이다.

$$\text{판별함수} = .42\text{pass_freq} + .483\text{def_depth} + .343\text{ave_wide} - .579\text{block} \quad (3.2)$$

모형의 변수구성은 의사결정나무와 로지스틱 회귀에 유의했던 변수들에 총 패스횟수만 추가 한 것과 같다. 판별식의 계수를 살펴보면 패스횟수, 수비 시 공수간격, 선수간의 좌우 폭이 커지면 패할 가능성이 커지고 블록의 값이 커지면 승리 하는 것으로 나타났다. 특히 총 패스횟수는 일반적인 통념과 달리 많을수록 경기에 패하는 것으로 나타났다. Table 3.6은 판별함수의 유의성을 나타낸다.

Table 3.5 Standardized canonical discriminant function coefficients

pass_freq	def_depth	ave_wide	block
.423	.483	.343	-.579

Table 3.6 Summary of canonical discriminant functions

Function	Wilks' Lambda	Chi-square	degree of freedom	p-value
1	.741	13.768	4	0.008

3.3. 득점을 이용한 공격력분석

3.3.1. 득점의 분포

Table 3.5에서 보듯이 득점의 분포는 0, 1 그리고 2에 심하게 편중되어 있다. 축구경기에서는 득점이 자주 발생하지 않으므로 포아송 분포를 따르는 것으로 예측할 수 있다. 분포에 대한 검정은 Kolmogorov-Smirnov 방법을 이용하였다. 검정통계량으로 최대차이를 이용한 p값은 .103으로 유의수준 0.05에 비하여 크므로 득점은 포아송분포를 따른다고 할 수 있다. Table 3.6에 의하면 평균이 1.18인 포아송 분포를 따른다.

Table 3.7 Frequency of gain

	0	1	2	3	4	5
frequency	24	26	18	5	2	1
percent	31.58	34.21	23.68	6.58	2.63	1.32

Table 3.8 Summary statistics of gain

	mean	variance	maximum	minimum
gain	1.18	1.25	5	0

3.3.2. 평균득점비율 포아송 회귀모형

포아송 회귀모형분석은 SAS의 Genmod를 이용하였다. 예측모형에서 사용한 변수와 공격력에 관련된 주요변수로 최적의 결과를 얻었다. Table 3.7의 결과에 의하면 dpass.freq (논스톱 패스)와 tempo가 증가 할수록 득점의 기회는 감소된다. 그 이유는 볼 처리 스피드가 느려져 수비수나 골키퍼가 준비할 시간이 많기 때문이다. 경기스피드 외에 유효 슛과 골이 가능한 지역에서의 슛이 증가 할수록 득점의 기회

가 많아진다. 유효 슈트 한 번에 득점이 20% 증가하는 반면 일반적인 슈트는 9% 증가하는 것으로 나타나 정확도의 중요성을 알 수 있었다. 득점 (gain)의 빈도에 관한 포아송 회귀모형식은 (3.3)이다. 그러나 이 모형식의 설명변수는 유의하나 Table 3.10에서 보듯이 적합도에 부족함이 있어 득점을 전력 질주한 거리 (sprint_dist)에 대한 평균비율로 설정하여 모형식 (3.4)을 설정하였다. Table 3.9에서 보듯이 적합도가 향상되어 최종 모형식으로 결정하였다.

$$\text{Log}(\text{gain}) = -0.0125\text{dpass_freq} + .181\text{sog} - 1.04\text{tempo} + \text{ha_shots} \quad (3.3)$$

$$\text{Log}(\text{gain}) = -0.0125\text{dpass_freq} + .181\text{sog} - 1.04\text{tempo} + \text{ha_shots} + \text{Log}(\text{sprint_dist}) \quad (3.4)$$

Table 3.9 Prediction model of gain in game using Poisson regression

parameter	estimate	standarderror	wald	p-value	exp(estimate)
dpass_freq	-.0125	0.006	4.98	0.026	0.988
sog	.1806	0.063	8.24	0.004	1.197
tempo	-1.04	0.387	7.16	0.007	0.353
ha_shots	.085	0.039	4.76	0.029	1.089

3.4. 모형 선택 및 최종모형 적합도

1차적으로 득점 (gain)을 반응변수로 하여 분석을 실시한 결과, 설정한 모형이 적합도에 부족함이 나타나는데 이는 Table 3.10의 적합도 통계량으로 알 수 있었다. Table 3.9에서 유의한 설명변수 논스톱 패스 횟수 (dpass_freq)와 tempo가 속도와 관련이 있어 모형을 수정하였다. 골을 전력 질주한 거리에 대한 평균비율로 설정하여 모형분석을 다시 하였다. Table 3.11의 결과 4개의 적합도 통계량이 모두 1에 가까워서 수정된 모형이 자료를 잘 설명하고 있음을 알 수 있었다.

Table 3.10 Criteria of goodness of fit of model of the frequency of gain

Criterion	DF	Value	Value/DF
Deviance	72	70.06	0.973
Scaled Deviance	72	70.0642	0.9731
Pearson Chi-Square	72	57.8318	0.8032
Scaled Pearson X^2	72	57.8318	0.8032

Table 3.11 Criteria of goodness of fit of model of the average rate of gain

Criterion	DF	Value	Value/DF
Deviance	72	79.2908	1.1013
Scaled Deviance	72	79.2908	1.1013
Pearson Chi-Square	72	70.9836	0.9859
Scaled Pearson X^2	72	70.9836	0.9859

3.5. 실점을 이용한 수비력분석

3.5.1. 실점의 빈도 및 분포

Table 3.12에서 알 수 있듯이 실점도 득점과 같은 분포형태를 따르는 것을 짐작할 수 있다. 실점에 관한 분포검정을 Kolmogorov-Smirnov 방법으로 실시하였다. 그 결과 검정통계량 .015에 대한 p값을 0.129로 포아송 분포를 따른다고 할 수 있다. 실점은 Table 3.13에 의하면 평균 1.16인 포아송 분포를 따른다.

Table 3.12 Frequency of allowed

	0	1	2	3	4	5
frequency	25	26	17	5	2	1
percent	32.89	34.21	22.37	6.58	2.63	1.32

Table 3.13 Summary statistics of allowed

	mean	variance	maximum	minimum
allowed	1.16	1.25	5	0

3.5.2. 실점의 포아송 회귀모형

실점 (allowed)에 관한 모형은 아래의 Table 3.14에서 보는 바와 같이 속도에 관련 된 변수가 없이 수비 시 공수간격 (def_depth)과 선수들의 좌우 폭 (ave_wide)을 표시하는 변수들로 구성되었다. Table 3.14에 의한 모형식은 (3.5)이다. 모형식에 의하면 수비 시 공수간격이 넓으면 실점에 9% 기여하고 좌우 폭이 넓으면 6.5% 기여한다. 실점모형의 설명변수들은 예측모형에서 사용한 변수와 일치하였다. 수비에 중요한 변수인 블록을 모형에 추가해 보았으나 실점모형에서는 유의하지 못했다. Table 3.12의 결과에 의하면 좌우상하로 압박하는 밀착수비의 강도가 실점을 결정하는 것으로 나타났다.

$$\text{Log}(\text{allowed}) = -4.9 + .085\text{def_depth} + .063\text{ave_wide} \quad (3.5)$$

Table 3.14 Prediction model of allowed in game using Poisson regression

parameter	estimate	standarderror	wald	p-value	exp(estimate)
intercept	-4.90	1.26	15.07	0.001	
def_depth	0.085	0.027	9.91	0.001	1.089
ave_wide	0.063	0.03	3.91	0.048	1.065

3.5.3. 모형 선택 및 최종모형 적합도

실점을 발생빈도로 생각하여 반응변수로 설정하고 포아송 회귀모형으로 분석하였다. 모형의 적합도는 아래 Table 3.15에서 보는 바와 같이 Deviance는 1과 가깝지만 Pearson 카이제곱통계량이 1 값과 약간 차이가 있음을 알 수 있다. 실점에 관한 회귀모형에는 속도에 관련 된 설명변수는 없지만 득점과 같이 평균비율을 추정하는 모형으로 다시 분석하였다. Table 3.16에서 알 수 있듯이 네 개의 적합통계량이 모두 1을 초과하여 과적합을 나타낸다. 그러므로 실점을 빈도로 간주한 처음의 모형을 최종결과로 설정하였다.

Table 3.15 Criteria of goodness of fit of model of the frequency of allowed

Criterion	DF	Value	Value/DF
Deviance	73	81.4111	1.1152
Scaled Deviance	73	81.4111	1.1152
Pearson Chi-Square	73	66.8545	0.9158
Scaled Pearson X^2	73	66.8545	0.9158

Table 3.16 Criteria of goodness of fit of model of the average rate of allowed

Criterion	DF	Value	Value/DF
Deviance	73	95.6589	1.3104
Scaled Deviance	73	95.6589	1.3104
Pearson Chi-Square	73	94.8742	1.2996
Scaled Pearson X^2	73	94.8742	1.2996

4. 논의 및 결론

본 연구에서는 위치정보를 활용하여 수집된 기록을 토대로 경기력을 승패와 공격, 수비력으로 나누어 각각에 미치는 요인을 분석하였다. 승패에 관한 분석은 3가지 예측모형을 통하여 실시하였다. 3가지 분석방법의 결과 모형의 설명변수는 수비력에 관한 변수로 구성되어 공격보다 수비가 승패에 대한 중요한 요소임을 알 수 있었다. 공격력과 수비력을 득점과 실점모형으로 분석한 결과 공격의 중요한 요소는 슈트를 제외하고 논스톱패스와 공격의 빠르기였다. 두 변수는 공격스피드와 깊은 관련이 있는 것으로 현대 축구에 있어서 빠른 공격은 골을 얻기 위한 필수적인 요소로 나타났다. 1990년 월드컵의 골 분석 통계자료에 따르면 전체 득점의 88%는 4회 미만의 패스에서 얻어졌다 (Hughes, 2004). 그만큼 수비수와 골키퍼가 준비되지 않은 상황에서 진행된 공격의 효과가 크다고 할 수 있다. 반면에 수비에서는 수비시 공수간격과 좌우 폭이 중요한 요소로 나타났다. 횡적, 종적으로 거리를 줄여 상대방을 압박하는 공간방어 능력이 공격을 좌절시켜 수비에서 가장 중요한 실점을 줄일 수 있는 요인이라 하겠다 (Harris와 Reilly, 1997). 이는 현재 정상급 팀들이 추구하는 전술이라 볼 수 있겠다.

References

- Choi, S., Kang, C., Choi, H. and Kang, B. (2012). Social network analysis for a soccer game. *Journal of the Korean Data & Information Society*, **22**, 1053-1063.
- Hong, C. S., Jung, M. S. and Lee, H. J. (2010). Prediction model analysis of 2010 South Africa world cup. *Journal of the Korean Data & Information Society*, **21**, 1137-1146.
- Hughes, M. (2004). Notational analysis-a mathematical perspective. *International Journal of Performance Analysis in Sports*, **4**, 97-139.
- Karlis, D. and Ntzoufras, L. (2003). Analysis of sports data by using bivariate Poisson models. *The Statistician*, **52**, 381-393.
- Keller, B. J. (1994). A characterization of the Poisson distribution and the probability of winning a game. *The American Statistician*, **48**, 294-298.
- Kim, N. J. and Min, D. K. (2014). The study of changes in performance in KLPGA using growth curve analysis. *Journal of the Korean Data & Information Society*, **25**, 847-845.
- Olsen, E. and Larsen, O. (1997). Use of match analysis by coaches. *Science and Football*, III, 209-220
- Perl, J. (2001). Artificial neural networks in sports: New concepts and approaches. *International Journal of Performance Analysis in Sports*, **1**, 106-121.
- Reilly, T. and Thomas, V. (1976). Motion analysis of work-rate in different position roles in professional football match-play. *Journal of Human Movement Studies*, **2**, 87-97.
- Suzuki, K. and Nishijima, T. (2004). Validity of a soccer defending skill scale using game performances. *International Journal of Sport and Health Science*, **2**, 34-49.
- Taylor, J. B., Mellalieu, S. D., James, N. and Shearer, D. A. (2008). The influence of match location, quality of opposition and match status on technical performance in professional association football. *Journal of Sports Science*, **26**, 885-895.

Performances analysis of football matches

Dae Kee Min¹ · Young-Soo Lee² · Yong-Rae Kim³

¹Department of Information & Statistics, Duksung Women's University

²Department of Physical Education, Sejong University

³Sports Science Institute, Sejong University

Received 6 December 2014, revised 28 December 2014, accepted 19 January 2015

Abstract

The team's performances were analyzed by evaluating the scores gained by their offense and the scores allowed by their defense. To evaluate the team's attacking and defending abilities, we also considered the factors that contributed the team's gained points or the opposing team's gained points? In order to analyze the outcome of the games, three prediction models were used such as decision trees, logistic regression, and discriminant analysis. As a result, the factors associated with the defense showed a decisive influence in determining the game results. We analyzed the offense and defense by using the response variable. This showed that the major factors predicting the offense were non-stop pass and attack speed and the major factor predicting the defense were the distance between right and left players and the distance between front line attackers and rearmost defenders during the game.

Keywords: Decision tree, discriminant analysis, logistic regression, Poisson regression.

¹ Corresponding author: Professor, Department of Information and Statistics, Duksung Women's University, Seoul 132-714, Korea. E-mail: dkmin@duksung.ac.kr

² Professor, Department of Physical Education, Sejong University, Seoul 147-747, Korea.

³ Senior researcher, Sports Science Institute, Sejong University, Seoul 147-747, Korea.