

2010 남아공 월드컵 축구 예측모형 분석

홍중선¹ · 정민섭² · 이재형³

¹성균관대학교 통계학과 · ^{2,3}성균관대학교 응용통계연구소

접수 2010년 9월 13일, 수정 2010년 10월 27일, 게재확정 2010년 11월 1일

요약

승부를 예측하는 많은 연구방법들이 있고, 지금도 많은 연구가 진행되고 있다. 이러한 많은 예측 방법들 중에서, 경기력에 영향을 미치는 현실적인 변수를 고려한 통계적인 모형을 사용하여 예측한다면 다른 예측방법들 보다 정교한 예측을 기대할 수 있다. 본 연구에서는 2010년 남아공 월드컵 축구 결과 예측을 위하여 Bradley-Terry 모형을 고려한다. 이 예측모형은 경기력에 영향을 미치는 확률변수들을 포함하고, 쌍별 비교 방법을 사용하였다. 모형에 포함된 각 국가의 가치모수는 Newton-Raphson 알고리즘을 이용하여 얻은 수렴한 값이다. 이 모형을 사용하여 32개국 중 16강 진출하는 국가를 예측하였고, 8강, 4강, 결승진출, 우승팀까지 예측하였다. 2010년 남아공 월드컵 축구의 최종 결과와 이 예측자료를 비교하고 향후 연구에 대해 토론한다.

주요용어: 가치모수, 경기모수, 쌍별 비교, 예측력.

1. 서론

전 세계에서 가장 인기 있는 스포츠를 꼽는다면 많은 사람들은 축구라고 말할 것이다. 세계 축구경기를 통괄하는 국제단체인 FIFA (Federation Internationale de Football Association; 국제축구연맹)는 1886년 영국에서 탄생하였다. 현재 FIFA에 등록되어있는 나라는 200여 나라가 넘고 전 세계의 등록선수는 2억명을 넘는다. 목적은 경기 추진, 각국 협회 간 우호 증진, 경기규칙의 준수 등이며, 4년마다 열리는 세계선수권대회를 주관한다. 올림픽 중간 연도를 택해 4년마다 개최하는 세계선수권대회 중 하나가 바로 월드컵이다. 단일종목으로서는 세계에서 가장 큰 스포츠 행사이자 제일 먼저 탄생한 세계선수권대회이다. 참가 자격은 프로나 아마추어를 불문하고, 선수는 소속 클럽이나 팀의 국적이 아니라 선수 개인의 국적에 따라서 출전한다. 제1회 대회는 1930년 남미 우루과이의 몬테비데오에서 13개국이 참가한 가운데 개최되었고, 1938년 제3회 프랑스로부터 이후 제2차 세계대전으로 12년간 중단되었다가 1950년에 제4회 대회가 브라질에서 다시 열렸다 (대한축구협회, 2007).

한국에서 축구가 전파된 것은 지금으로부터 100여년 전인 19세기말이다. 영국을 모태로 하는 근대 축구가 한국에 전파된 것은 1882년 (고종 19년) 인천항에 상륙한 영국 군함 플라이잉홀스의 승무원들을 통해서인 것으로 전해지며, 정식 축구의 보급은 1904년 서울의 관립 외국어 학교에서 체육 과목의 하나로 채택하면서부터이다. 한국 최초의 공개 축구경기는 1905년 6월 10일 서울훈련원 (오늘날의 동대문운동장)에서 열린 대한체육구락부와 황성기독교청년회간의 시합이라고 할 수 있다. 국제적으로 통용

¹ 교신저자: (110-745) 서울 중로구 명륜동 3-53, 성균관대학교 경제학부 통계학전공, 교수.
E-mail: cshong@skku.edu

² (110-745) 서울 중로구 명륜동 3-53, 성균관대학교 응용통계연구소 연구원, 일반대학원 통계학과 석사과정.

³ (110-745) 서울 중로구 명륜동 3-53, 성균관대학교 응용통계연구소 연구원, 일반대학원 통계학과 석사과정.

되는 규칙 하에 경기에 필요한 각종 장비를 갖추고 경기를 한 것은 1920년대부터이다 (대한축구협회, 2007).

우리나라는 1933년 조선축구협회 발족한 이후 1947년 6월에 FIFA에 가입하였고, 북한은 지난 1958년에 가입하였다. 한국 대표팀은 1986년 멕시코 월드컵부터 2010년 남아공 월드컵까지 7회 연속으로 월드컵 본선에 진출함으로써 아시아 최다 월드컵 본선 진출국 (통산 8회)이 됨은 물론 세계 수준에 근접한 한국 축구의 우수성을 지구촌의 모든 축구팬에게 과시했다. 한국 축구의 역사는 세계 축구의 최고 제전인 월드컵 개최로 화려하게 꽃피우게 되었다. 축구 사랑과 나라 사랑의 한 마음으로 온 국민이 보여준 뜨거운 열기와 함께 추진된 월드컵 유치운동은 결국 1996년 5월 31일 한국과 일본의 2002년 월드컵 공동 개최라는 성과로 이어졌다. 2002년 한일 월드컵에서 한국은 7전 3승 2무 2패로 아시아 최초로 4강에 진출하였다.

많은 사람들은 세계 축구의 최고 제전인 월드컵 경기의 결과를 예측하고자 한다. 그러나 국가대표 간의 경기인 A매치는 일년동안 많지 않은 경기가 열려 각 국가 간의 경기결과를 예측하는 것은 쉽지 않다. 여기서 A매치란 FIFA가 인정하는 국가대표팀 간의 경기이며 정식 영어 표기로는 'International A Match'라고 한다. 여기서 'A'란 한 국가의 가장 우수한 선수들로 구성된 팀이라는 뜻이다. 또한 경기가 없었던 국가 간의 경기를 예측하는 것은 더더욱 쉽지 않을 것이다. 그러나 우리는 경기가 없었던 국가 간의 경기도 그 두 상대국가와 공통으로 경기를 가진 국가를 이용하여 국가 간 경기의 결과를 예측할 수 있다.

2010년에 월드컵은 최초로 아프리카 대륙인 남아공에서 열렸다. 7회 연속으로 월드컵 본선에 진출한 우리나라 대표 팀의 선전을 기원하고 좋은 성적을 기대하는 것은 대한민국의 국민으로 당연한 관심사이다. 시합 전 대한민국 대표 팀의 성적을 예측하는 것은 통계학자뿐만 아니라 온 국민에게 매우 흥미롭다. 홍중선 등 (2010)은 2010 남아공 월드컵 대회 전 5월에 결과를 예측 발표하였고, 본 연구에서는 예측 결과와 대회가 종료된 최종 성적을 비교하여 토론하고자 한다.

일반적으로 두 팀 간의 비교를 통한 승패예측을 하는 쌍별 비교 (paired comparison) 모형으로는 로지스틱 모형 (logistic model)과 휴리스틱 모형 (heuristic model), 유전자 알고리즘 (gene algorithm)을 사용한 모형 등이 있으며 여러 가지 알고리즘을 복합적으로 사용한 모형들도 많이 사용되어지고 있다 (김혁주와 김정현, 2009; 신상근 등, 2009; 신양규, 1995; 김재희와 이경원, 2000). 2006 독일 월드컵을 예측하고 분석한 김도현 등 (2007)의 논문을 참고하여 본 연구에서는 2010 남아공 월드컵을 분석하고자 하며 무승부가 포함된 수정된 Bradley-Terry 모형을 이용하고 여러 개의 의미 있는 확률변수들을 포함시켜 분석한다. 월드컵 개최국인 남아공의 특수성을 고려하여 고도변수와 대륙 간 거리변수를 포함하여 모형을 설정한다.

본 연구의 구성은 다음과 같다. 2절에서는 Bradley-Terry 모형과 수정된 Bradley-Terry 모형에 대해 알아보고 최대가능도추정량을 이용하여 모수를 추정하는 방법을 설명한다. 3절에서는 실제 설계된 모형의 모수를 추정하고 이를 이용하여 모의실험을 통해 구한 각각의 확률로 예측한 예측결과를 설명하고 실제결과와 비교해 봄으로써 모형의 타당도를 확인한다. 마지막으로 4절에서는 결론을 유도한다.

2. 예측모형

본 연구에서는 두 팀 간의 비교를 통한 승패를 예측하기 위하여 쌍별 비교 (paired comparison) 모형 중 하나인 Bradley-Terry 모형 (Bradley와 Terry, 1952)을 기본으로 수정된 Bradley-Terry 모형 (Davison, 1970)을 고려한다. Bradley-Terry 모형은 A팀이 다른 B팀을 이길 확률을 다음과 같이 나타낸다.

$$\begin{aligned}
 P_{AB} &= P(\text{A 팀이 B 팀을 이긴다}) \\
 &= \frac{\pi_A}{\pi_A + \pi_B} \\
 &= \frac{e^{\gamma_A}}{e^{\gamma_A} + e^{\gamma_B}},
 \end{aligned}
 \tag{2.1}$$

여기서 모수 π_A 는 A팀의 상대적인 능력을 나타내는 가치모수 (worth parameter)로 정의하며, $\gamma_A = \ln\pi_A$ 이며 무승부를 고려하지 않으므로 $P_{AB} + P_{BA} = 1$ 이다.

Davison (1970)은 여러가지 방법 중 극한분포와 로지스틱분포를 사용하여 다음과 같은 방법으로 이 모형을 유도하였다. 우선 A팀의 점수를 나타내는 확률변수 S_A 는 위치모수 $\ln\pi_A$ 를 가진 극한분포 (extreme distribution)를 가정하고 누적분포함수는 다음과 같다.

$$F_{S_A}(s) = \exp\{-\exp(s - \ln\pi_A)\}, -\infty < s < \infty, \pi_A > 0. \tag{2.2}$$

두 팀의 점수 차 $S_A - S_B$ 의 분포는 평균이 $\ln\pi_A - \ln\pi_B$ 인 다음과 같은 로지스틱분포를 따른다.

$$F_{S_A - S_B}(s) = \frac{1}{1 + \exp\{-(s - (\ln\pi_A - \ln\pi_B))\}}, -\infty < s < \infty, \pi_A > 0. \tag{2.3}$$

그러므로 A팀이 다른 B팀을 이길 확률은 다음과 같으며 이것은 식 (2.1)과 동일하다.

$$\begin{aligned}
 P(S_A > S_B) &= P(S_A - S_B > 0) \\
 &= 1 - \frac{1}{1 + \exp(\ln\pi_A - \ln\pi_B)} \\
 &= \frac{\pi_A}{\pi_A + \pi_B}.
 \end{aligned}
 \tag{2.4}$$

또한 Davison (1970)에 의해 제안된 무승부를 고려한 수정된 Bradley-Terry 모형 (modified Bradley-Terry model)은 A팀이 다른 팀 B팀을 이길 확률, 질 확률, 비길 확률을 다음과 같이 나타낸다.

$$\begin{aligned}
 P_{AB} &= P(\text{A 팀이 B 팀을 이긴다}) = \frac{e^{\gamma_A}}{e^{\gamma_A} + e^{\gamma_B} + e^{\lambda + (\gamma_A + \gamma_B)/2}}, \\
 P_{BA} &= P(\text{B 팀이 A 팀을 이긴다}) = \frac{e^{\gamma_B}}{e^{\gamma_A} + e^{\gamma_B} + e^{\lambda + (\gamma_A + \gamma_B)/2}}, \\
 P_{AB,0} &= P(\text{A 팀과 B 팀은 동점}) = \frac{e^{\lambda + (\gamma_A + \gamma_B)/2}}{e^{\gamma_A} + e^{\gamma_B} + e^{\lambda + (\gamma_A + \gamma_B)/2}},
 \end{aligned}
 \tag{2.5}$$

여기서 모수 λ 는 무승부와 관련된 모수이며, 이 값이 클 경우 두 팀 간의 경기에서 무승부가 나올 확률이 높다 (Singh와 Gupta, 1978). 본 연구에서 분석할 자료는 FIFA에 등록되어 있는 208개 나라의 20년 (1990.1.1 ~ 2009.12.31)의 자료로 14,000여 개의 경기기록으로 구성되어 있으며, 고려하는 확률 변수로는 경기장소 (h_i), 대륙 간 거리 (X_1), 최근 5회의 걸친 월드컵 참여 수 (X_2), 최고성적 (X_3)과 고도 (X_4)로 구성되어 있다. 개최국인 남아공은 특히 한국에서부터의 거리가 멀리 떨어져 있는 나라이며, 남아공의 수도인 요하네스버그에 있는 축구장은 매우 높은 고도에 위치하고 있기 때문에 변수 X_1 과 X_4 를 고려하였다. 그리고 FIFA 순위 (FIFA ranking)는 분석결과에 많은 영향을 주기 때문에 고려하

지 않았다. 이러한 확률변수를 고려하여 식 (2.5)의 분자를 (2.6)식과 같이 설정하여 A와 B팀의 번째 경기에 대한 가치모수를 추정한다. 즉 i 번째 경기에서 A팀이 B팀을 이길 확률은 A팀과 관련된 가치모수 Δ_{AB}^i 와 B팀과 관련된 가치모수 Δ_{BA}^i 그리고 A, B팀 모두와 관련이 있어서 두 팀이 비기는 가치모수 $\Delta_{AB,() }^i$ 로 다음과 같이 설정한다.

$$\begin{aligned}\Delta_{AB}^i &= \exp\left\{\gamma_{A_i} + \frac{\delta}{2}h_i + \sum_{j=1}^4 \alpha_j x_{jA_i}\right\}, \\ \Delta_{BA}^i &= \exp\left\{\gamma_{B_i} - \frac{\delta}{2}h_i + \sum_{j=1}^4 \alpha_j x_{jB_i}\right\}, \\ \Delta_{AB,() }^i &= \exp\left\{\frac{1}{2}(\gamma_{A_i}\gamma_{B_i} + \sum_{j=1}^4 \alpha_j(x_{jA_i} + x_{jB_i})) + \lambda\right\},\end{aligned}\tag{2.6}$$

여기서 경기장소 변수 (h_i)가 홈이면 1, 원정이면 -1 , 제3국이면 0이며, δ 는 홈 이점, λ 는 무승부, α_j 는 j 번째 변수에 관련된 모수이다. δ 가 클수록 홈경기 이점이 경기결과에 많은 영향을 미친다는 의미이다. 따라서 A와 B팀의 i 번째 경기에서 A팀이 B팀을 이길 확률은 다음과 같이 정의한다.

$$\begin{aligned}P_{AB}^i &= P(i\text{번째 경기에서 A팀이 B팀을 이긴다}) \\ &= \frac{\Delta_{AB}^i}{\Delta_{AB}^i + \Delta_{BA}^i + \Delta_{AB,() }^i}.\end{aligned}\tag{2.7}$$

모수 추정을 위한 최대가능도 추정량 (maximum likelihood estimates)은 일반적으로 잘 알려진 방법인 Newton-Raphson 알고리즘에 의해 구해질 수 있으며 모수추정을 위한 가능도함수는 다음과 같다.

$$L = \sum_{i=1}^N \left(\frac{\Delta_{AB}^i}{\Delta_i}\right)^{W_i} \left(\frac{\Delta_{BA}^i}{\Delta_i}\right)^{L_i} \left(\frac{\Delta_{AB,() }^i}{\Delta_i}\right)^{T_i},\tag{2.8}$$

여기서 $\Delta_i = \Delta_{AB}^i + \Delta_{BA}^i + \Delta_{AB,() }^i$ 이다. 또한 여기에서 A와 B팀 사이에 총 N 번의 경기가 이루어졌다고 할 때, i 번째 경기에서의 A팀이 승리하면 $W_i = 1, L_i = 0, T_i = 0$ 이고, i 번째 경기에서 B팀이 승리하면 $W_i = 0, L_i = 1, T_i = 0$ 이며, 그리고 두 팀이 비기면 $W_i = 0, L_i = 0, T_i = 1$ 이다.

3. 예측과 분석결과

2절에서 설명한 모형을 기반으로 분석한 예측결과를 제시한다. 고려해야할 확률변수들을 선정하여 모형에 적합시킨 뒤, 가치모수와 경기모수들의 추정량을 구하고 이를 기반으로 모의실험을 통하여 각 국가의 16강 진출할 확률과 8강, 4강, 준결승 그리고 우승확률을 구한다.

3.1. 가치모수와 경기모수의 추정

가치모수 추정을 위한 최대가능도 추정량은 앞서 언급한 방법인 Newton-Raphson 알고리즘에 의해 구해질 수 있으며 이를 통하여 208개 국가에 대한 가치모수 ($\gamma_1, \gamma_2, \dots, \gamma_{208}$)와 경기력에 영향을 미치는 경기모수 ($\delta, \lambda, \alpha_1, \alpha_2, \alpha_3, \alpha_4$)를 추정하였다. 208개의 국가에 대한 가치모수 중 2010 남아공 월드컵에 참여하는 32개국의 가치모수와 경기력에 영향을 미치는 홈 이점 (Home), 무승부 (Tie), 대륙 간 거리 (Distance), 최근 월드컵 참여 수 (Recent 5), 월드컵 최고 성적 (Best Rank), 고도 (Altitude)의 경기

모수 ($\delta, \lambda, \alpha_1, \alpha_2, \alpha_3, \alpha_4$)의 추정값은 표 3.1에서 나타내었다. 국가에 대한 가치모수를 추정할 때 초기 값을 복한을 0으로 고정시키고 추정하였다.

이 추정값을 식 (2.7)에 대입하면 A팀이 B팀에 이길 확률을 구할 수 있고, 같은 방법으로 비길 확률 질 확률을 구할 수 있다. 32개국의 가치모수 추정량과 이 추정값을 토대로 3.2절에서 설명할 모의실험을 하여 상위 경기로의 진출확률을 추정한다.

3.2. 모의실험 과정

본 연구에서 구한 각 국가의 16강 진출확률과 8강, 4강, 준결승, 우승확률은 다음의 모의실험 과정을 통하여 구한다.

1. Newton-Raphson 알고리즘 방법을 이용하여 구한 최대가능도 추정량이 포함된 모형으로부터 각 국가 간의 승, 무, 패의 확률을 구한다.
2. 위에서 구한 국가 간의 승, 무, 패 확률을 토대로 다항확률 난수발생 (Multinomial random number generate)을 통하여 각 조의 16강 예선전의 모든 경기결과를 얻는다.
3. FIFA의 승점 산출 방식에 따라 모든 예선전 경기에 승점을 부여하고 그 승점을 계산하여 각조의 순위를 정한다.
4. 단계 2, 3을 10,000번 반복한다.
5. 10,000번의 실험 중 2위안에 들어간 빈도수로 16강 확률 (2위안에 들어간 횟수/10,000)을 구한다.
6. 조건부확률을 이용하여 각 10,000번의 경기결과에서 만날 수 있는 상대를 대진표에 의해 구하고 각 국가 간의 승, 무, 패 확률을 이용하여 8강, 4강, 준결승, 우승확률을 각각 구한다.

표 3.1 가치모수와 경기모수 추정값

Algeria	Argentina	Australia	Brazil	Cameroon
-0.51931	-19.6599	-15.3644	-19.4445	-49.5019
Chile	Cote d' Ivoire	Denmark	England	France
13.21924	-15.8044	-1.57449	-33.1428	12.429
Germany	Ghana	Greece	Honduras	Italy
-20.1907	-16.2609	-15.6718	0.671135	-19.8711
Japan	Korea DPR	Korea Republic	Mexico	Netherlands
-33.1631	0	-18.471	-34.2578	-18.5752
New Zealand	Nigeria	Paraguay	Portugal	Serbia
-2.43371	-18.5752	-32.2764	12.66977	-14.429
Slovakia	Slovenia	South Africa	Spain	Switzerland
-0.49751	-16.5966	-48.2431	-16.9525	-16.6341
Uruguay	USA			
-16.6341	-64.6242			

δ	λ	α_1	α_2	α_3	α_4
0.59919	-0.09603	0.132033	15.94421	-14.4199	-0.20003

3.3. 예측결과

모의실험결과 먼저 한국이 속한 B조뿐만 아니라 A부터 H조까지의 16강 진출 확률은 아래와 같이 추정한다. 이때 각 조의 16강 진출 확률의 합은 200%인데 그 이유는 각 조당 두 나라가 16강에 진출하기 때문이다. 모든 조 각각의 16강 진출확률은 표 3.2와 같으며 우리가 일반적으로 알고 있는 축구 강국들이 높은 확률을 가지고 있는 것을 알 수 있다.

하위경기에서 상위경기로의 진출확률은 각 국가가 대진표에 의한 경기에서 상대국에 대한 승리확률의 합으로 구할 수 있다. 16강부터는 32강의 승점제와 다르게 승패만 가리므로 수정된 Bradley-Terry 모형이 아닌 Bradley-Terry 모형을 사용하였으며, 여기서 상대국과의 승리확률은 상대국과 경기할 확률과 경기에서 승리할 확률의 곱으로 구한다. 4강, 결승, 우승확률도 같은 방법으로 추출하였으며 월드컵 참가하는 어느 국가라도 우승할 확률이 있고 32개국 중 하나의 우승국이 나오므로 32개국의 우승확률의 총합은 100%이고 브라질이 15%로 가장 높은 확률을 나타낸다. 32개국의 결승확률은 두 국가가 남게 되어 200% 4강은 400% 8강은 800%이다. 각국의 8강 4강 결승진출 우승확률은 표 3.2부터 표 3.6과 같다.

표 3.2 각 조의 16강 진출 확률

A조	France 71%	Mexico 61%	Uruguay 52%	South Africa 16%
B조	Argentina 85%	Nigeria 46%	Greece 41%	Korea 28%
C조	England 82%	USA 60%	Slovenia 32%	Algeria 26%
D조	Germany 76%	Serbia 66%	Australia 38%	Ghana 20%
E조	Netherlands 80%	Denmark 60%	Cameroon 31%	Japan 29%
F조	Italy 95%	Paraguay 80%	Slovakia 24%	New Zealand 2%
G조	Brazil 88%	Portugal 75%	Cote d'Ivoire 24%	Korea DPR 12%
H조	Switzerland 58%	Chile 56%	Spain 52%	Honduras 34%

표 3.3 8강 진출 확률 (상위 10개국)

Brazil 67%	Argentina 53%	Italy 51%	England 50%
Netherlands 49%	Germany 45%	Spain 43%	France 41%
Paraguay 40%	Serbia 37%	남은 국가들의 확률합 324%	

표 3.4 4강 진출 확률 (상위 6개국)

Brazil 47%	Spain 33%	Netherlands 31%	Argentina 29%
England 27%	Germany 22%	남은 국가들의 확률합 211%	

표 3.5 결승 진출 확률 (상위 3개국)

Brazil	Spain	England	남은 국가 확률합
30%	23%	17%	130%

표 3.6 우승 확률 (상위 2개국)

Brazil	England	남은 국가 확률합
15%	11%	74%

3.4. 실제결과와 비교

대회결과가 확정된 후 16강, 8강, 4강, 결승에서의 우승예측과 실제경기 결과를 비교해보면 표 3.7과 같다. 각 칸의 값은 진출할 예측확률이고 실제로 진출한 경우에는 대응하는 칸에 # 표시를 하였다.

표 3.7 상위경기 진출 확률과 실제와 비교

Group	Country	Top 16	Q-F	S-F	Final	Victory
A	France	71	41.7	15	6.4	3.2
	Mexico	61 [#]	33	10.7	5.7	2.7
	Uruguay	52 [#]	23.8 [#]	9.5 [#]	3.8	1.8
	South Africa	16	2.7	0.5	0	0
B	Argentina	84.6 [#]	53.1 [#]	29.8	13.8	6.9
	Mexico	46.4	21.5	7.7	3.7	1.3
	Uruguay	41	14.3	7.1	3.6	0.9
	South Africa	28 [#]	9.9	3.5	1.3	0.4
C	England	82 [#]	50.5	27.1	17.9	11.2
	U. S. A	60 [#]	31.2	8.9	4.9	3.1
	Slovenia	32	10.3	2.9	0.5	0
	Algeria	26	7.9	1.5	0.2	0
D	Germany	76 [#]	45.2 [#]	22.7 [#]	11.9	7.1
	Serbia	66	37.2	9.4	4.2	2.5
	Australia	38	12.4	4.2	2.1	1.2
	Ghana	20 [#]	5.3	1.1	0	0
E	Netherlands	80.2 [#]	49.9 [#]	31.5 [#]	16.9 [#]	10.8
	Denmark	60.2	32	11.2	3.8	1.8
	Cameroon	30.6	10.1	4.1	1.6	0.7
	Japan	29 [#]	8.3	3.2	1.2	0.5
F	Italy	95.8	51.3	21.5	9.4	5.8
	Paraguay	80 [#]	40	8.8	3.8	1.2
	Slovakia	24 [#]	8.4	2.1	0.3	0
	New Zealand	0.2	0	0	0	0
G	Brazil	88 [#]	67.4 [#]	47.7	30.5	15.9
	Portugal	75 [#]	34.6	31.3	6.2	5.8
	Cote d' Ivoire	24.8	5.5	2.7	0.8	0.3
	Korea DPR	12.2	3.5	0.5	0	0
H	Switzerland	58	20.2	19.6	8.4	3.2
	Chile	55.8 [#]	19.6	18.2	6.6	2.1
	Spain	61 [#]	42.7 [#]	32.9 [#]	28.9 [#]	9.5 [#]
	Honduras	34.2	6.5	3.1	1.6	0.1

표3.7의 결과를 살펴보면, 모형으로 예측한 결과와 실제 결과와의 차이가 크지 않은 것을 파악할 수 있다. 16강 진출은 16개 국가 중 11개 국가의 예측이 적중하여 정확도는 68.8%이며, 결과 중에서 이번

이라고 평가할 수 있는 것은 최근 전력이 급상승한 나라들이 대부분이다.

4. 결론

2010 남아공 월드컵에서도 우리는 원정 16강 진출이라는 쾌거를 이루었다. 이는 2002년의 4강신화가 단지 홈 이점에 의한 결과가 아닌 우리나라 축구의 실력수준이 높아졌음을 보여주는 예라고 볼 수 있다. 이에 대한 의문에서 출발한 본 연구는 고려해 볼 수 있는 모든 변수들을 고려해서 최대한 객관적인 추정값들을 계산하였고, 이에 대한 16강, 8강, 4강, 결승 진출의 확률을 계산해 보았다. FIFA 순위를 배제한 이유는 FIFA 순위는 FIFA 자체에서 순위를 정하는 기준이 있으므로 고려할 변수로 모형에 추가하면, 추정값이 편의 (Bias)되기 때문이다. 본 연구에서에서 고려한 모형으로 모수를 추정하여 우승까지의 모의실험을 해본 결과, 브라질이나 영국이 결승에서 우승할 것이라고 예측하였으며, 대체적으로 우리가 알고 있는 강팀들이 예측결과에서도 높은 순위에 위치하고, 이는 국가모수 값이 크게 계산되었기 때문이다. 위 경기에서 상위 경기로 갈수록 줄어드는 확률의 크기가 같지 않음은 대진표에 따라 각국들이 경기할 수 있는 상대국가가 다르기 때문에 나타나는 현상으로, 이는 월드컵경기에서 각 나라의 경기력도 중요하지만 조 편성이나 대진 운 또한 중요 변수가 될 수 있음을 의미한다. 또한 최근 경기력이 급성장한 칠레, 멕시코, 우루과이, 일본 같은 국가들도 상위경기에 많이 진출하여 예상외의 선전을 보여준 것을 표 3.7로 확인할 수 있다.

세계 각국의 여러 기관들이 매회 월드컵마다 승부를 예측한다. 글로벌 투자회사들 또한 월드컵을 앞두고 나름대로의 전망을 내놓는다. 글로벌 금융기관들이 경제와 무관해 보이는 월드컵 성적 예측에 골몰하는 이유는 세계인의 관심이 쏠린 월드컵 성적에서 좋은 예측력을 보이는 일은 곧 자신들의 예측 능력을 과시할 수 있는 기회가 되기 때문이다. 2006년 독일 월드컵 성적을 놓고 보면, UBS가 골드만삭스에 비해 높은 확률로 예측을 했다.

4년 뒤인 2010년 남아공 월드컵, 그 중에서도 한국의 성적에 대해서 골드만삭스는 “6강, 8강 진출 가능하다”, UBS는 “16강 진출 확률 29%, 8강까지 가는 놀라운 일이 벌어질 수도 있다”고 전망했다. 이번엔 처음 월드컵 성적 전망 보고서를 낸 JP모간은 “한국은 B조에 속한 국가 중 우승확률이 가장 낮다”고 예상했다. 한국이 16강 진출함에 따라 한국 성적과 관련해서는 골드만삭스와 UBS는 예측을 잘 했으며, JP모간은 잘하지 못했다. 글로벌 투자사의 전망이 엉터리냐 아니냐는 4강 관문에서 또 한 번 가려지게 되었다. 세 금융기관이 4강 진출 팀으로 꼽은 국가로 골드만삭스는 브라질, 스페인, 독일, 잉글랜드로 예상하였고, JP모간은 잉글랜드, 스페인, 네덜란드, 슬로베니아로 예측하였으며, UBS는 브라질, 독일, 이탈리아, 네덜란드로 예측하였다. 또한 JP모간은 브라질이 강력한 우승 후보지만 8강에서 네덜란드에 덜미가 잡힐 것으로 전망했다. 그러나 결과는 골드만삭스는 스페인과 독일만을 맞춰 두 나라를 맞췄고 JP모간 또한 스페인과 네덜란드 두 나라를 맞췄다. 하지만 UBS는 네덜란드만 맞추는데 그쳐 상대적으로 낮은 적중률을 보였으며 본 연구결과에서는 스페인과 네덜란드를 맞춰 골드만삭스와 JP모간과 비슷한 정확도를 보였다.

세계적인 글로벌 투자회사들은 FIFA 랭킹, 과거 월드컵 성적, 개최국 여부, 국가별 경제력 등을 변수로 설정한 뒤 변수마다 가중치를 부여해 나름의 결과를 도출해 내는 복잡한 계량적 분석방법을 사용했으나, 본 연구에서는 통계적인 Bradley-Terry 모형으로 분석하였다. 서로 다른 방법으로 예측하였지만 세계적인 글로벌 투자회사들의 예측결과와 본 연구에서 낸 결과와 크게 다르지 않음을 파악할 수 있다.

본 연구는 축구 전문가나 팬들이 쉽게 접할 수 있는 A매치 결과와 여러 가지 현실적인 확률변수들을 통하여 확률을 제시한 점에서 의미있는 연구이다. 이밖에 경기결과에 영향을 줄 수 있는 다른 확률변수들을 고려해 볼 수 있을 것이며, 축구에 국한되지 않고 승패가 있는 다른 스포츠분야에도 적용하는 연구는 향후 연구과제로 남겨두기로 한다.

참고문헌

- 김도현, 이상인, 김용대 (2007). Bradley-Terry 모형을 이용한 2006 독일월드컵예측. <응용통계연구>, **20**, 205-218.
- 김재희, 이경원 (2000). 위치모수 변화 모형에서 순위함수와 평균함수를 이용한 비모수적 변화점 추정. <한국데이터정보과학회지>, **11**, 279-29.
- 김혁주, 김정현 (2009). 선형추세를 갖는 모집단에 대한 변형계통표집의 일반화와 회귀추정법. <한국데이터정보과학회지>, **20**, 1103-1118.
- 대한축구협회 (2007). www.KFA.or.kr.
- 신상근, 조용주, 조영석 (2009). 유럽리그에서 득점과 실점을 이용한 승점 추정에 관한 연구. <한국데이터정보과학회지>, **20**, 837-844
- 신양규 (1995). 확률적 명제 논리 프로그래밍. <한국데이터정보과학회지>, **6**, 13-21.
- 홍종선, 정민섭, 이재형 (2010). 2010 남아프리카 공화국 월드컵 예측모형. <스포츠통계연구회논문발표회>, 2월 5일, 원광대학교.
- Bradley, R. A. and Terry, M. E. (1952). Rank analysis of incomplete block designs: The method of paired comparisons. *Biometrika*, **39**, 324-345.
- Davison, R. R. (1970). On extending the Bradley-Terry model to accommodate ties in paired comparison experiments. *Journal of the American Statistical Association*, **65**, 317-328.
- Singh, J. and Gupta, R. S. (1978). A paired comparison model allowing for ties. *Scandinavian Journal of Statistics*, **5**, 65-68.

Prediction model analysis of 2010 South Africa World Cup

Chong Sun Hong¹ · Min Sub Jung² · Jae Hyoung Lee³

¹Department of Statistics, SungKyunKwan University

^{2,3}Research Institute of Applied Statistics, SungKyunKwan University

Received 13 September 2010, revised 27 October 2010, accepted 1 November 2010

Abstract

There are a lot of methods to predict the result of a game and many forecasting researches have been studied. Among many methods, if a statistical model including some realistic random variables is used to forecast, more accurate prediction could be expected than any others. In this work, Bradley-Terry model is considered to predict results of 2010 South Africa World Cup games via paired comparison method. This prediction model includes some random variables which affect the results of games. The worth parameters for each country in this model are convergence values obtained by using Newton-Raphson algorithm. With this model, we can forecast top 16 among 32 countries and up to who will win the victory. Final results of 2010 South Africa World Cup games are compared with this prediction and discuss further works.

Keywords: Game parameter, paired comparison, prediction power, worth parameter.

¹ Corresponding author: Professor, Department of Statistics, Sungkyunkwan University, 3-53, Myungryun-Dong, Jongro-Gu, Seoul 110-745, Korea. E-mail: cshong@skku.ac.kr

² Master's course, Department of Statistics, Researcher, Research Institute of Applied Statistics, Sungkyunkwan University, 3-53, Myungryun-Dong, Jongro-Gu, Seoul 110-745, Korea.

³ Master's course, Department of Statistics, Researcher, Research Institute of Applied Statistics, Sungkyunkwan University, 3-53, Myungryun-Dong, Jongro-Gu, Seoul 110-745, Korea.