

サッカーにおける Bradley-Terry モデルの適用法の研究

M2011MM016 平手良典

指導教員：松田眞一

1 はじめに

現在，FIFA 加盟国は 6 大陸ごとに連盟が存在し，全 207ヶ国が加盟している。加盟国各国代表チームの強さを推定する指標として，FIFA ランキングがある。

本研究では Bradley-Terry モデルについて実際にサッカーデータを用いて検証することを目的とする。今回は，FIFA 加盟国各国代表チームの強さを推定し，実際の FIFA ランキングと比較する。このモデルを用いると直接対戦のないチーム間についてもその勝敗の確率を予測することができる。

また，過去の研究（服部・加藤 [2] 参照）ではサッカーの試合と同じように対戦の試合がそれほど多くないテニスの試合のデータを用いて上位 20 選手と 3 つの区分に分けたそれ以下の選手についての対戦結果を元にそれぞれの上位選手の強さを推定し，実際の世界ランキングと比較した。本研究の目標として，サッカーにおいてもそのような推定法を適用し，「服部・加藤の方法」が有効かどうか，安定しているかどうかの検討およびチームの強さの推定として役に立つのかどうか適合度検定での確認が挙げられる。

2 Bradley-Terry モデル

球技をはじめとする多くのスポーツにおいては，2 人あるいは 2 つのチームが対戦して勝ち負けを争う。いくつかのチームが互いに何回か対戦するとき，その結果に基づいて各チームの「強さ」を推定するときに Bradley-Terry モデルが用いられる。（竹内・藤野 [3] 参照）

2.1 強さについて

サッカーや野球のように 2 つのチームが対戦して勝ち負けを争う中で，チームの「強さ」というものについて考える。一般にスポーツにおいては「強い」ほうが必ず勝つとは限らない。あるチームが「強い」ということは，そのチームが必ず勝つということではなく，そのチームの勝つ「確率」が大きいことを意味すると考える。各チームの「強さ」が一義的に定められるためには，それらの確率の間に，単に例えば，チーム 2 がチーム 1 より強く，3 が 2 より強く，1 が 3 より強い，いわゆる「三すくみ」の関係が成り立たないだけでなく，より強い関係が成り立たなければならない。そのための自然な条件として考えられるのは，各チームに対応して m 個の量 $\pi_1, \pi_2, \dots, \pi_m$ が存在して，すべての i, j の組み合わせに対し，確率 p_{ij} が，

$$p_{ij} = \frac{\pi_i}{\pi_i + \pi_j}$$

と表されるものと想定する。このとき， π_i はチーム i の「強さ」を表すものと解釈できる。このような構造のことを Bradley-Terry モデル（以下 BT モデル）と呼ぶ。

2.2 先行研究

BT モデルの先行研究として「服部・加藤の方法」が挙げられる。テニスプレーヤー上位 20 選手とそれ以下の選手を 21～30 位，31～50 位，51 位以下を一まとめとして考え，4 大会の勝敗のデータから選手の強さを推定し，実際の世界ランキングと比較し検証した。加えて，セット数での勝ち負け，ゲーム数での勝ち負けも分析した。（服部・加藤 [2] 参照）

その他の先行研究は BT モデルの拡張として，今井・松田 [1] の引き分けを考慮した BT モデルがある。引き分けを 0.5 勝 0.5 敗として考えず，引き分けのある対戦ゲームにおいてもそれを考慮しながら強さを求める方法を示している。

3 BT モデルのデータについて

本研究で使用するデータは国際サッカー連盟加盟国のサッカー国際 A マッチの試合のデータと FIFA ランキングデータである。国際 A マッチとは，年齢制限のない代表チーム同士の国際公式試合であり，この試合結果でのみ FIFA ランキングは計算される。

3.1 試合のデータ

2010 年 1 月 1 日から 2012 年 6 月 30 日までに実施された FIFA 加盟 207ヶ国の代表チームによる国際 A マッチの試合をデータとする。集計した項目は，対戦国，試合の重要度および試合結果である。

対戦国については，上位 30 チームを中心にデータを取り，それ以下のチームについては区分ごとに一まとめりとして考える。試合の重要度については，FIFA 公式のランキング算出方法の割合を用いて，勝ち点と勝ち数で考える際に，どちらも FIFA ワールドカップ本大会での試合を基準とし「勝ち」を「1 勝」または「勝ち点 1」とする。試合結果については，対戦国それぞれに対して「勝ち」「負け」「引き分け」の 3 通りとする。なお 90 分間で勝敗が着いた場合を「勝ち」「負け」とし，同点または延長戦，PK で勝敗が着いた場合を「引き分け」とする。「引き分け」については勝ち点で考える場合は「0.33」とし，勝ち数で考える場合は「0.5 勝」とする。

3.2 FIFA ランキングデータ

FIFA ランキングデータとは国際 A マッチの成績をポイント化した FIFA 加盟国のランキングであり，FIFA によって原則毎月発表される。ランキングの算出（よしびた [6] 参照）については，勝ち点，試合の重要度，対戦国間の強さ，大陸連盟間の強さの 4 つの値を元に計算され，過去 4 年間の対戦成績が対象となる。なお実際に FIFA ランキングが算出されるとき「勝ち点」は全ての国際 A マッチにおいて，勝ち 3 点，引き分け 1 点，負け 0 点とし，PK 戦での勝敗については勝ち 2 点，負け 1 点となっ

ていて、勝ち点制である。今回は勝ち点と試合の重要度を中心に考える。今回使用する FIFA ランキングデータについては、2012年7月4日に更新されたランキングを比較の対象として使用した。

4 BT モデルを用いた強さの推定

実際に BT モデルを用いてチームの強さを推定する。BT モデルの特徴を利用し、直接対戦のないチーム同士でも強さを比較していく。なお、BT モデルの推定は鬼頭・高田 [5] が作成した R 関数を利用した。

4.1 推定方法

まず勝ち数と勝ち点を試合の重要度に絡めて、上位 30 チームに、31~50 位、51~70 位、71~100 位、100 位以下の区切りで以下の 4 通りについて考える。

1. 勝ち数 (重要度なし)
2. 勝ち数 (重要度あり)
3. 勝ち点 (重要度なし)
4. 勝ち点 (重要度あり)

今回、ワールドカップの試合を基準とし、一まとめにした 4 つチームについては、上位 30 チームとの対戦結果のみを考慮し、一まとめにしたチーム同士の対戦結果は考慮しない。試合の重要度を絡めた場合の割合ポイントは表 1 の通りであり、この値を用いて各チームの推定値を計算した。

パターン	勝ち数		勝ち点	
	勝ち	引き分け	勝ち	引き分け
a	1	0.5	1	0.33
b	0.75	0.375	0.75	0.25
c	0.63	0.315	0.63	0.315
d	0.25	0.125	0.25	0.125

表 1 4 通りの割合ポイント

この表にある「引き分け」は、90 分間で同点の場合と、延長戦、PK で勝敗が着いた場合も「引き分け」とする。表にある a~d は、試合の重要度であり、

- a FIFA ワールドカップ本大会
- b 大陸選手権本大会
- c FIFA ワールドカップ、大陸選手権の予選
- d 親善試合

である。

4.2 勝ち数での推定

結果が良かった勝ち数の方を示す。勝ち数で重要度を考えない場合と考える場合の 2 通りでの結果を表 2 に示す。

表 2 勝ち数 (引き分けは 0.5 勝とする)

チーム	重要度なし	重要度あり
スペイン	190.0	374.0
ドイツ	113.4	179.5
ウルグアイ	76.9	68.8
イングランド	106.0	68.4
ポルトガル	52.4	59.2
イタリア	40.8	41.1
アルゼンチン	96.0	80.7
オランダ	106.8	117.9
クロアチア	51.6	36.9
デンマーク	40.1	31.3
ブラジル	91.0	79.9
ギリシャ	28.2	23.0
ロシア	34.3	27.5
フランス	48.0	26.4
チリ	58.0	61.6
コートジボワール	46.1	46.6
スウェーデン	64.0	36.7
チェコ	27.3	22.3
メキシコ	48.2	30.9
日本	27.4	23.3
スイス	31.7	26.6
コロンビア	27.6	21.1
オーストラリア	31.3	26.5
ノルウェー	36.5	25.5
パラグアイ	28.8	26.4
アイルランド	32.3	22.8
エクアドル	30.6	22.2
韓国	34.9	25.1
トルコ	34.4	16.0
ボスニア	13.9	13.4
31~50 位	22.8	18.7
51~70 位	16.8	10.2
71~100 位	8.9	7.3
101 位以下	3.0	2.1
順位相関係数	0.8272	0.8974

推定値を求めて推定値の大きさの順位と実際の FIFA ランキングのスピアマンの順位相関係数をそれぞれ求めた。順位相関係数は、重要度なしの場合 0.8272 となり、ありの場合 0.8974 となり、試合の重要度がある程度ランキングに関係していることが分かる。上位のチームではイングランドのように重要度ありだと大幅に下がっているチームもあった。これは、重要な試合では勝っていないことを示す。守備的な特徴を持つイタリアやギリシャなどは、上位にしては引き分けが多いのが、推定値が低い要因と考えられる。下位のチームでは、重要度なしは少し推定値が高めに出てしまったチームが目立ったが、重要度ありではある程度ランキングに近い結果となった。これは、日本や韓国など、親善試合や予選で勝ち数を増やしているチームが影響しているせいだろう。試合の重要度を絡

めると、上位に不利、下位に有利があまりなくなり、力の差が少し大きくなることでバランスがとれて全体としてある程度ランキングに近い結果となった。

5 服部・加藤の方法の安定性

実際に「服部・加藤の方法」の安定性について、前章の勝ち数の推定の順位の区切りをさらに細かくすることに加え、ワールドカップ前の試合結果からもある程度推定できるか検証する。

5.1 推定方法

次に相関が高かった「勝ち数(重要度あり)」について、さらに以下の3通りについて考えて検証する。

1. 31位~40位, 41位~50位に分ける
2. 上記に加え, 51位~60位, 61位~70位に分ける
3. 1.についてWCを含めない場合

5.2 勝ち数(重要度あり)改良(WCを含む)

次に相関が高かった「勝ち数(重要度あり)」について31~50位を, 31~40位, 41~50位に分けて考える場合と, それに加えて51~70位を51~60位, 61~70位に分ける場合について同じように推定値を出した。順位相関はそれぞれ, 31~50位を2分割した方が0.912、さらに51~70位を2分割した方が0.884となり, 30~70位を細かく区切った場合元より低い相関となった。この2つの結果から分割に対する影響として, 下位チームにいくにつれ上位チームとの対戦が少なくなるので, 下位チームは細かく区切りすぎず, 適度に区切る必要があると考えられ, 最終的な区分として, 上位30チーム, 31~40位, 41~50位, 51~70位, 71~100位, 101位以下が良いと分かった。よって, 区分に対する「服部・加藤の方法」の安定性は試合数に応じて上位チームにいくほど細かな区切りでよくなるといえる。

5.3 勝ち数(重要度あり)改良(WCを含まない)

勝ち数(重要度あり)の31~50位を2分割の区切りでワールドカップの試合を含めずに推定値を出した。7章に示す, 適合度検定で実際のワールドカップの試合を予測するときに, ワールドカップの試合結果を含めた場合と含めない場合でどれだけ結果に差がでるか検証するために用いる。順位相関は, 0.919となった。ワールドカップの試合を含めないことで, 全体的に推定値の差が小さくなった。

6 強さの推定の検証

6.1 適合度検定について

適合度検定の一般論について示す。(白旗[4](9章)参照)
 k 個の排反な分類項目 A_1, \dots, A_k があり, 一つの対象を調べたとき, それが A_i に属する確率を p_i とする。帰無仮説として

$$H_0 : p_i = p_i(\theta), i = 1, \dots, k \quad (1)$$

次に, 対立仮説として

$$H_1 : H_0 \text{ ではない} \quad (2)$$

とする。 H_0 を仮定したときの θ の推定量を $\hat{\theta} = \hat{\theta}(X_1, \dots, X_k)$ とし, n は十分大きいとする。 X_i は2項分布に従うのでその平均は np_i となる。そして, 統計量

$$\chi^2 = \sum_{i=1}^k \frac{(X_i - n\hat{p}_i)^2}{n\hat{p}_i} \quad (3)$$

を考えると, これが H_0 からの離れ具合を表している。 χ^2 は漸近的に自由度 $k - r - 1$ のカイ2乗分布に従う。ただし r はパラメータ θ の次元, すなわち p_i を決めるパラメータ数である。これにより調べたいデータのあてはまり具合を検定できる。

6.2 検証に用いるデータ

第4章で求めたチームの推定値を用いて, グループリーグから決勝トーナメントまで試合で各々勝つ確率を求めて, その値を区分けし, 区分ごとでの勝敗を集計して適合度検定を行う。ワールドカップの試合結果を含めなくても, その試合の勝敗を予測できるかどうかを検証する。対戦する2チームで, 推定値が高い方の勝つ確率を x とすると, 他方が勝つ確率を y とすると, 以下の5つに区分けして考える。

$$\begin{cases} 0.85 < x \leq 0.95 & (0.05 \leq y < 0.15) \\ 0.75 < x \leq 0.85 & (0.15 \leq y < 0.25) \\ 0.65 < x \leq 0.75 & (0.25 \leq y < 0.35) \\ 0.55 < x \leq 0.65 & (0.35 \leq y < 0.45) \\ 0.5 < x \leq 0.55 & (0.45 \leq y < 0.5) \end{cases}$$

7 適合度検定の具体的な方法について

上記の区分ごとでの勝敗を集計し, 以下のような計算法で p 値を出す。具体的な方法を示すため表3を用いて説明する。

	9区	8区	7区	6区	5区
x 勝ち数	x_1	x_2	x_3	x_4	x_5
y 勝ち数	y_1	y_2	y_3	y_4	y_5
試合数	z_1	z_2	z_3	z_4	z_5
x 期待度数	a_1	a_2	a_3	a_4	a_5
y 期待度数	b_1	b_2	b_3	b_4	b_5
χ_x^2 値	A1	A2	A3	A4	A5
χ_y^2 値	B1	B2	B3	B4	B5
$\chi_x^2 + \chi_y^2$	C1	C2	C3	C4	C5

表3 検定法の説明

ここでは、勝つ確率が高い方のチームの勝ち数、すなわち観測度を x とし、区分ごとに $x_1 \sim x_5$ とする。同じように勝つ確率が低い方のチームの勝ち数を y とし、区分ごとに $y_1 \sim y_5$ とする。ただし、 $x=y=0.5$ のデータがあった場合は 0.5 勝 0.5 敗とした。

x の期待度数は区分ごとの試合数 $z_1 \sim z_5$ にそれぞれ、0.9, 0.8, 0.7, 0.6, 0.525 を掛けた値である。 y の期待度数も同じように試合数 $z_1 \sim z_5$ にそれぞれ、0.1, 0.2, 0.3, 0.4, 0.475 を掛けた値である。

χ_x^2 値, $A_i = \frac{(x_i - a_i)^2}{a_i}$ をそれぞれの区分で計算し、 χ_y^2 値も同じように計算する。

そして区分ごとでの x と y の χ^2 値を合計した値 C_i が統計量となる。全体統計量は、 C_i の和である。その統計量と自由度 5 からエクセルを用いて p 値を計算した。

7.1 ワールドカップを含めた場合の結果

まずワールドカップを含めた場合の検定結果である。チームの推定値は、勝ち数（重要度あり）の 30～50 位を 2 分割した方の結果を用いて、勝つ確率を求め、区分けして勝敗を集計した。 p 値は、0.919 となった。

区分ごとの統計量を見ると、9 割区分、8 割区分、6 割区分については離れ具合が小さかった。7 割区分では、ドイツやスペインに対しイングランドやアルゼンチンが引き分けられなかったなど、最も離れ具合が大きかったが、高いとは言えない。5 割区分では、推定値が同じチーム同士の対戦が重なってしまい、解釈が難しくなった。ワールドカップを含めた場合であるので全体的に、実際の勝敗に近い結果となった。

7.2 ワールドカップを含めない場合の結果

次にワールドカップを含めない場合の検定結果である。 p 値は、0.407 となった。

区分ごとの統計量を見ると、9 割区分では、推定値が上がったイングランドが 3 試合入ったが、試合結果は 1 勝 2 分けとなって、それが最も高く出た離れ具合に影響したと考えられる。8 割区分、7 割区分、6 割区分では、離れ具合は小さかった。5 割区分では、上位のブラジルなどと 16 位のコートジボワールあたりまでの推定値の差が縮まったことで強豪チームが多く入っていて、勝ち数が 5 分とはならなかった。

	9区	8区	7区	6区	5区
x 勝ち数	9.5	9.5	9.5	7.5	7.5
y 勝ち数	3.5	3.5	4.5	4.5	4.5
試合数	13	13	14	12	12
x 期待度数	11.7	10.4	9.8	7.2	6.3
y 期待度数	1.3	2.6	4.2	4.8	5.7
χ_x^2 値	0.413	0.077	0.009	0.013	0.228
χ_y^2 値	3.723	0.311	0.021	0.018	0.253
$\chi_x^2 + \chi_y^2$	4.136	0.389	0.03	0.031	0.481

表 4 WC 含めない場合

8 まとめ

BT モデルでは、まず勝ち数、勝ち点に試合の重要度を絡めた 4 通りのうち「勝ち数・重要度あり」のパターンが最も相関が高かった。この 4 通りに共通することとして、上位のチームでは FIFA ランク 1 位スペイン、2 位ドイツは圧倒的に強い結果になった。これらの上位チームは、ワールドカップやユーロでの勝ち数も多く、引き分けと負けも少ない。上位だが引き分けが多いイタリアは本大会での成績も良くなかったので推定値が低めになった。逆に引き分けが少なく、本大会で好成績を残した 8 位のオランダが 3 番目に推定値が高い結果となった。20 位の日本は、4 通りとも推定値が低めとなった。

次に「服部・加藤の方法」で、30 位以下のチームの区切りを細かくした結果、31～50 位を 2 分割にしたら相関がさらに高くなった。分割に対する影響として、下位チームにいくにつれ上位チームとの対戦が少なくなるので、下位チームは細かく区切りすぎず、適度に区切る必要があることが分かった。最終的に良かったパターンから、試合数に応じて上位チームにいくほど「服部・加藤の方法」が安定しているという結果になった。強さの推定を検証する適合度検定では、ワールドカップの試合結果を含めない場合でも、ある程度実際の本大会を予測できたといえる。ただイングランドのように大会前は好成績でも本大会で結果を残せないチームなど予測が難しいチームもあった。

9 おわりに

全体的に BT モデルをサッカーのデータで検証するという目的を果たしたと思う。試合の重要度やチームの区切りなど、確かめてみたかったことを全部できたとは言えないが、引き分けのある対戦ゲームに役に立てるそれなりの結果が得られたと思う。まだ引き分けの多いチームの扱いや、30 位以下のチームもどこまで個別に区切れるかなど改善の余地があると感じた。

参考文献

- [1] 今井寛・松田眞一:「引き分けを考慮した BT モデルの性能評価」南山大学紀要アカデミア数理情報編, 第 3 巻 (2003), pp.35-45.
- [2] 服部 匡志・加藤明:「プロテニスプレーヤーの強さの統計的研究」, 南山大学数理情報学部数理科学科卒業論文要旨集, 1997.
- [3] 竹内啓・藤野和建:応用統計数学シリーズ「スポーツの数理科学 もっと楽しむための数字の読み方」, 共立出版, 1988.
- [4] 白旗慎吾:「統計解析入門」, 共立出版, 1992.
- [5] 鬼頭薫・高田涼子: S-plus における強さの推定, 南山大学経営学部情報管理学科卒業論文, 1999.
- [6] よしびた:「Jcalcio FIFA ランキングユーロメモ」, <http://www.fifaworldranking.com/wrhow.html>, 2001.