

# サッカーリーグの順位推定

Yasuhiro

June 25 2020

## 1 動機付け

2020 年初め、新型コロナの影響で、スポーツを含めて様々なイベントは中止しなければならない状態で、私がいつも見てるイングランドのサッカーリーグ「Premier League」も 38 節 380 試合中、29 節 287 試合 (3 試合は時間調整) だけが行われて中止になりました。新型コロナの影響がますます大きくなるにつれて、Premier League が今シーズン再開できなくなる心配もありますが、リーグの順位結果は来年のチャンピオンズリーグの資格 (前四位) につながります。

もし再開できない場合、それに対して二つの計画が提案されました：

1. 今シーズンの試合をすべてキャンセル、チャンピオンズリーグの資格は去年と同じにする。
2. 今の順位表のままで、チャンピオンズリーグの資格を決める

多くの人が二番目の提案に賛成するが、この方法では以下の問題があります：

- 3 試合が時間調整の原因で行わなかったので、今の順位表の 20 チームの内、6 チームが 27 試合で、14 チームが 28 試合を行った状況である。順位表の決定方法は勝ち点は勝利 3 点、引き分け 1 点、敗北 0 点で、勝ち点が多い順に順位を決めている。このまま計算すると、6 チームに不公平。
- 日程によって、残りの試合相手の強さは違う。

このレポートは以上の問題を解決するために、数理的な方法で中止したリーグの順位を決めることが目的です。具体的に、20 チームがお互いにホームとアウェイで試合をする、380 試合が行われた時と同じように、287 の試合結果からチームの順位を決めます。

## 2 データセット

試合結果は様々なウェブサイトから見られますが、まとめられたファイルがないので、今回は BBC スポーツから試合結果をクローラーで取得しました。クローラーの詳細は [https://github.com/yasuhiro0802/Soccer-team-ability-estimator/blob/master/get\\_scores.py](https://github.com/yasuhiro0802/Soccer-team-ability-estimator/blob/master/get_scores.py) で参照してください。

データは表 1 のように、home, home\_score, away, away\_score 四つのコラムで、home と away はチーム名で、home\_score と away\_score は試合の得点です。

## 3 問題の定式化

### 3.1 仮説

サッカー試合を分析してチームの実力を計算するのはとても複雑な問題なので、簡単のため以下の仮説を置きます。

- チームの実力は攻撃力を示す  $\alpha$  と防御力を示す  $\beta$  の足し合わせとして考える。

Dataset				
id	home	home_score	away	away_score
1	Southampton	1	Manchester United	1
2	Chelsea	2	Sheffield United	2
3	Crystal Palace	1	Aston Villa	0
...	...	...	...	...
285	Burnley	1	Tottenham Hotspur	1
286	Everton	1	Manchester United	1
287	Tottenham Hotspur	2	Wolverhampton Wanderers	3

表 1 データセット

- 各チームの  $\alpha$  と  $\beta$  は全シーズン固定とする。つまり、メンバーのローテーション、状態、監督の戦術、ホームかアウェイかなどは考慮しない。
- 試合で得点する確率は得点側の攻撃力  $\alpha$  と相手側の防御力  $\beta$  だけが関係する。
- 毎試合の時間は 90 分で固定とする。
- 各試合は独立で、試合中両方の得点も独立とする。

### 3.2 定式化

リーグに参加したチームはチーム 1, チーム 2, ..., チーム N と呼びます。チーム  $i$  の攻撃力は  $\alpha_i$ 、防御力は  $\beta_i$  と表します。

チーム  $i$  とチーム  $j$  の試合で、チーム  $i$  が得点する確率  $p$  は、仮説の通り、チーム  $i$  の攻撃力  $\alpha_i$  およびチーム  $j$  の防御力  $\beta_j$  だけで決まります。チーム  $i$  の攻撃力が高いほど、チーム  $j$  の防御力が低いほど、チーム  $i$  が得点する確率が高くなります。これより様々な関係式が考えられます、例えば  $p \propto (\alpha_i - \beta_j)$ 、 $p \propto \frac{\alpha_i}{\beta_j}$  です。

サッカーにおける得点はポアソン分布に従うことが知られています。ポアソン分布を表す確率関数は以下の通りです。

$$P(k|\lambda) = \frac{\lambda^k e^{-\lambda}}{k!} \quad (3.1)$$

単位時間あたり平均  $\lambda$  回起こるようなランダムなイベントが、単位時間に  $k$  回発生する確率が  $P(k)$  です。サッカーの場合には、90 分 (単位時間) 内平均  $\lambda$  回の得点が、90 分内に  $k$  回得点する確率が  $P(k)$  です。簡単のため、チーム  $i$  がチーム  $j$  から得る得点の平均を

$$\lambda = \alpha_i - \beta_j \quad (3.2)$$

とします。

1 試合のホームチームの得点を  $x$ 、アウェイチームの得点を  $y$  とします。ホームチームはチーム  $i$ 、アウェイチームはチーム  $j$  とします。チーム  $i$  が 90 分内でアウェイチームのチーム  $j$  から  $x$  点取、チーム  $j$  が 90 分内

でホームチームのチーム  $i$  から  $y$  点取った場合を考えます。試合中両方の得点が独立だという仮説にしたがって、 $t$  番目の試合結果の確率は

$$P((x, y)^{(t)}) = P(x^{(t)} | \alpha_{j^{(t)}}, \beta_{i^{(t)}}) \cdot P(y^{(t)} | \alpha_{i^{(t)}}, \beta_{j^{(t)}}) \quad (3.3)$$

を表せます。 $x^{(t)}, y^{(t)}, i^{(t)}, j^{(t)}$  はそれぞれホームチームの得点数、アウェイチームの得点数、ホームチームの番号、アウェイチームの番号です。

### 3.3 解析手法

今回の目標は  $T$  個の試合結果から、 $N$  チーム  $\times 2$  のパラメータを推定することです。用いる手法は最尤推定です。尤度関数は

$$L(\Theta) = \prod_{t=1}^T P((x, y)^{(t)}) \quad (3.4)$$

で式 (3.3) を代入すると

$$L(\Theta) = \prod_{t=1}^T \frac{(\alpha_{j^{(t)}} - \beta_{i^{(t)}})^{x^{(t)}} (\alpha_{i^{(t)}} - \beta_{j^{(t)}})^{y^{(t)}} e^{(-\alpha_{i^{(t)}} - \alpha_{j^{(t)}} + \beta_{i^{(t)}} + \beta_{j^{(t)}})}}{x^{(t)}! y^{(t)}!} \quad (3.5)$$

となります。尤度の負の対数は

$$E(\Theta) = -\ln L(\Theta) = \sum_{t=1}^T [(-\alpha_{i^{(t)}} - \alpha_{j^{(t)}} + \beta_{i^{(t)}} + \beta_{j^{(t)}}) + x^{(t)} \ln(\alpha_{j^{(t)}} - \beta_{i^{(t)}}) + y^{(t)} \ln(\alpha_{i^{(t)}} - \beta_{j^{(t)}}) - \ln(x^{(t)}! y^{(t)}!)] \quad (3.6)$$

となります。この関数を最小化するため、勾配法を使います。 $\tau$  は訓練回数を表します。

$$\alpha^{(\tau+1)} = \alpha^{(\tau)} - \eta \frac{\partial E}{\partial \alpha} \quad (3.7)$$

$$\beta^{(\tau+1)} = \beta^{(\tau)} - \eta \frac{\partial E}{\partial \beta} \quad (3.8)$$

$m$  番目の  $\alpha$  との勾配は

$$\frac{\partial E}{\partial \alpha_m} = \sum_{t=1}^T (I(i^{(t)} = m) \cdot (\frac{y^{(t)}}{\alpha_{i^{(t)}} - \beta_{j^{(t)}}} - 1) + I(j^{(t)} = m) \cdot (\frac{x^{(t)}}{\alpha_{j^{(t)}} - \beta_{i^{(t)}}} - 1)) \quad (3.9)$$

となり。同じく、 $n$  番目の  $\beta$  との勾配は

$$\frac{\partial E}{\partial \beta_n} = \sum_{t=1}^T (I(j^{(t)} = n) \cdot (1 - \frac{y^{(t)}}{\alpha_{i^{(t)}} - \beta_{j^{(t)}}}) + I(i^{(t)} = n) \cdot (1 - \frac{x^{(t)}}{\alpha_{j^{(t)}} - \beta_{i^{(t)}}})) \quad (3.10)$$

となります。ただし、 $I$  は指示関数です。

## 4 実例

第 2 章のデータに対して第 3 章で取り上げたモデルを適用し、その訓練結果を図 1 で表します。

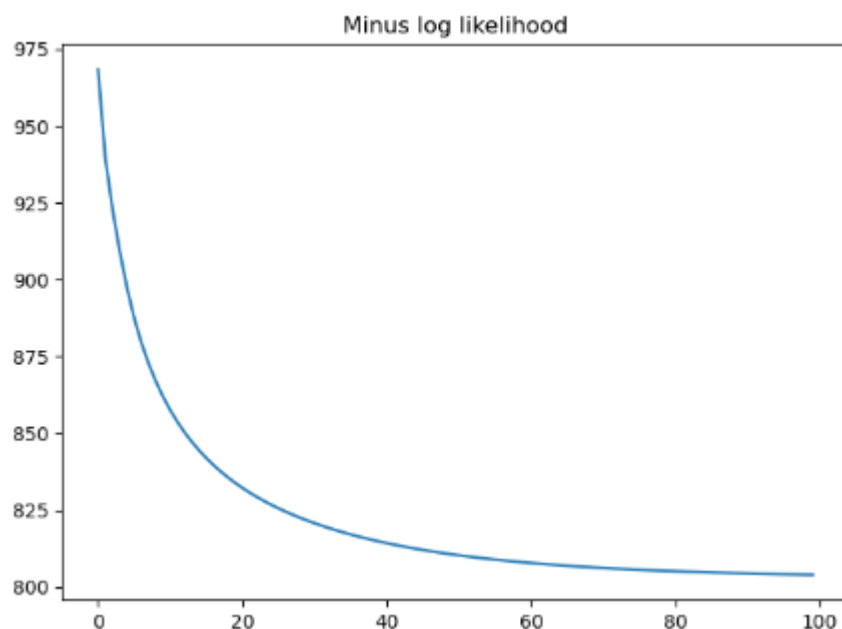


図 1 訓練結果。縦軸はロス、横軸は訓練回数です。

推定した各チームの  $\alpha$  と  $\beta$  とその足し合わせを図 2 で表します。

## 5 考察と展望

### 5.1 考察

2020 年 3 月に中止したプレミアリーグの順位表を図 3 に示します。

図 3 の順位は図 2 の推定結果と比べて少し異なります。その原因は第一章で記した順位表からそのまま実力を推定することによる問題点だと考えられます。

	team	attack	defense	sum
1	Liverpool	2.43	0.78	3.21
2	Manchester City	2.43	0.48	2.91
3	Leicester City	2.07	0.76	2.83
4	Chelsea	2.01	0.31	2.32
5	Manchester United	1.83	0.47	2.30
6	Sheffield United	1.38	0.75	2.13
7	Tottenham Hotspur	1.84	0.27	2.11
8	Wolverhampton Wanderers	1.65	0.44	2.09
9	Arsenal	1.65	0.35	2.00
10	Burnley	1.43	0.35	1.78
11	Crystal Palace	1.19	0.54	1.73
12	Everton	1.61	0.06	1.67
13	Brighton & Hove Albion	1.37	0.28	1.65
14	West Ham United	1.59	-0.03	1.56
15	Southampton	1.53	-0.03	1.50
16	Watford	1.30	0.19	1.49
17	Newcastle United	1.22	0.27	1.49
18	AFC Bournemouth	1.38	0.08	1.46
19	Aston Villa	1.53	-0.22	1.32
20	Norwich City	1.32	-0.06	1.26

図2 各チームの $\alpha$ を「attack」と表記し、そして $\beta$ を「defense」と表記します。「sum」はその足し合わせで、順番は「sum」の降順です。

## 5.2 展望

本レポートでは、簡単のため第二章で複数の現実と離れた仮説を置いた上で順位表を得ました。現実に合わせて、より複雑なモデルを考えなければなりません。例えば試合中の両チームの得点の関係、メンバーのローテーション、同じチームのホームとアウェイでの戦力差などを考える必要があります。

## 5.3 追記

2020年6月でプレミアリーグが無事に無観客で再開されました。しかし、このレポートは無意味だったわけではありません、同じ手法で他の問題への応用することもできます。例えばシーズン最後の順位を推測することや1試合の結果の予測することなどが考えられます。

ozumin と 楯くん から貴重な意見をもらい、心から感謝申し上げます。

	team	point
1	Liverpool	82
2	Manchester City	57
3	Leicester City	53
4	Chelsea	48
5	Manchester United	45
6	Sheffield United	43
7	Wolverhampton Wanderers	43
8	Tottenham Hotspur	41
9	Arsenal	40
10	Crystal Palace	39
11	Burnley	39
12	Everton	37
13	Newcastle United	35
14	Southampton	34
15	Brighton & Hove Albion	29
16	Watford	27
17	West Ham United	27
18	AFC Bournemouth	27
19	Aston Villa	25
20	Norwich City	21

図3 中止の時点の順位表