

テニスの四大大会における選手の特性に関する統計的分析

2020SS065 汐見紀哉

指導教員：松田真一

1 はじめに

テニスの四大大会のすべてを勝つことがとても難しいとされている理由は、大会ごとにコートが異なり勝つ選手が変わるからである。他のスポーツではあまり考えられないことであり、テニスの面白いところでもある。

そのため、どんな選手がどのコートに強いのかを数値で表すことで、選手におけるコートの向き不向きが分かると勝敗に繋がり、どのコートに対してどんな特徴を持たせたいといけなくのかを知るために解析を行う。

2 データについて

2018年から2022年における毎年度末の選手のランキング [1] をもとに5年分集めた順位の調和平均を取る。ただし101位以下は101とした。その値を昇順にして加藤・服部 [2] のように、1位から10位、11位から50位、51位以下のように順位を分けた。1位から10位の選手は個別にし、11位から50位、51位以下の選手をそれぞれ1人の選手とみなす。1位から10位におけるランキングの結果を表1に示す。

表1 ランキングと調和平均の値

1位	1.35	Novak Djokovic
2位	1.88	Rafael Nadal
3位	4.33	Daniil Medvedev
4位	4.71	Carlos Alcaraz
5位	5.25	Alexander Zverev
6位	5.32	Roger Federer
7位	5.56	Stefanos Tsitsipas
8位	6.37	Dominic Thiem
9位	9.55	Casper Ruud
10位	9.84	Andrey Rublev

選手の強さの推定値を求めるために、直近6回分の全豪、全仏、全英、全米における獲得ゲーム数を用いる。なお、勝ち数でなく獲得ゲーム数を採用した理由はトーナメント制のため勝敗データが少なく強さの値が変動しやすいせいである。

また、重回帰分析で使用するデータは、説明変数としてサーブエースは8変数、ダブルフォルトは8変数、サーブは16変数、リターンは14変数、ポイントは9変数、パフォーマンスは17変数、タイブレークは5変数、セットは4変数、試合は7変数、ランキングは8変数、試合時間は4変数の合計で全96変数、目的変数は強さの推定値である。説明変数は標準化して用いる。

3 分析方法

各コートでどんな選手が強いのかを数値で示したい。そのために選手の強さの値を求めるのにBradly-Terryモデル

(以下BTモデル)を使用する。また、選手のプレイスタイル、コートの違いを見つける際には、BTモデルで求めた強さの推定値に対しウォード法を用いたクラスター分析を行う。(web[5]参照)

総獲得ゲーム数を収集する際に引き分けはないものとして、どちらかの選手が途中棄権の場合は棄権する前のゲーム数を集計する。なお、試合が始まる前に棄権をした場合には両選手ともに獲得ゲーム数は0とする。

3.1 Bradly-Terry モデル

BTモデルとは m チームが1対1で対戦する際にチーム i がチーム j に勝つ確率を $p_{ij} = \frac{\pi_i}{\pi_i + \pi_j}$ ($i \neq j; i, j = 1, 2, \dots, m$)と定義して強さ π_i を求めるモデルである。(竹内・藤野 [4] 参照) 使用するプログラムは鬼頭・高田 [3] が作成したもので、長さが n の総獲得ゲーム数ベクトルと、選手の各組み合わせごとの対戦ゲーム数 $n \times n$ の行列を引数として用いる。なお、強さの推定値の基準を50とする。

3.2 VIF2 関数

VIFは分散拡大要因といい、逆行列を求めるため変数の数よりもデータの数が小さいとエラーが出力される。今回は逆行列の代わりに一般化逆行列を用いて改良したVIF2関数を使用する。VIF2関数では最小値が1より小さい時に1になるよう補正をしている。そのため、変数の数が少ない時にVIFとVIF2は同値となる。残った変数全ての場合VIF2の値が10以下になる変数の組み合わせのときに多重共線性の問題はなくなったと判断する。

4 強さの推定値と群分け

総獲得ゲーム数ベクトルと選手の各組み合わせごとの対戦ゲーム数の行列の2つの引数を用いて表2の大会ごとの強さの推定値を求めることができた。

表2 強さの推定値

	全豪	全仏	全英	全米
1 NovakDjokovic	68.97	68.65	69.95	62.38
2 RafaelNadal	63.00	77.20	66.63	55.71
3 DaniilMedvedev	56.22	44.95	41.31	58.76
4 CarlosAlcaraz	53.86	55.29	67.56	46.54
5 AlexanderZverev	53.58	45.72	50.02	52.19
6 RogerFederer	54.72	49.19	64.08	55.90
7 StefanosTsitsipas	51.48	51.22	41.50	51.01
8 DominicThiem	38.30	44.28	27.38	56.15
9 CasperRuud	34.09	52.14	40.51	41.04
10 AndreyRublev	48.75	41.35	48.11	48.16
11位から50位	42.17	38.33	45.29	39.74
51位以下	34.88	31.69	37.65	32.43

BTモデルにおける各選手の各大会の強さの推定値からウォード法を用いたクラスター分析を行う。11位から50

位の選手の名称を b, そして 51 位以下の選手を c とする。また, 図 1 の縦軸の高さが 3 付近を基準に群を分けることで 4 群に分けることができる。

4.1 考察

4 群を左から順に第 1 群, 第 2 群, ... として各群の意味するものを以下のように考えた。

第 1 群: 大会の違いにかかわらず他の選手と比べると全体的に強さが劣る。特に全豪, 全米のハードコートと全仏の大会を苦手としている。

第 2 群: 全仏に強いが全英に弱い。

第 3 群: 全ての大会を得意とする。

第 4 群: 強さの推定値は 3 群よりは劣るが, 全豪と全英と全米に強くハードコートと芝を得意としている。

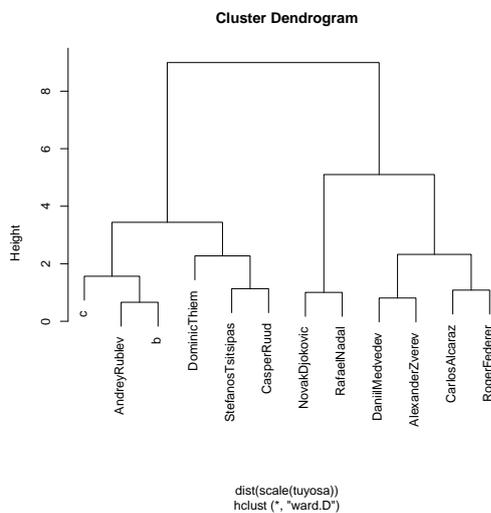


図 1 クラスタ分析

5 重回帰分析の結果

紙面の都合上, 全豪と全仏の重回帰分析の結果と考察を取り扱う。VIF2 関数を使用して VIF2 の値が 10 以上になる変数を削除して残った変数の数が 12 個以下になるまで繰り返す。12 個以下になったら次は 1 つずつ VIF2 の値が 10 以上のものを削る。変数全ての VIF2 の値が 10 以下になる変数の組み合わせのときにその大会における変数削除を終了する。そして残った変数でステップ関数を用いて解析を行い, 大会を説明するのに必要な変数から考察をする。

5.1 全豪

表 3 全豪における重回帰分析の結果

	係数	標準偏差	t 値	P 値
(Intercept)	0.000	0.131	0.000	1.000
Pts Lost per Svc Game	-0.405	0.203	-1.990	0.082
ReturnRating	0.416	0.153	2.715	0.026
SetPlayed	0.310	0.194	1.598	0.149

表 3 より係数が大きな変数に注目するとリターンでポイントを獲得する選手が総合的強さに結びついていた。また, 係数が負の変数は自分のサービスゲームでポイントを取らなくなるたびにゲームをブレイクされてしまう機会が増えるため, 1 ゲームを取られやすく試合を通じて不利な状況になりやすい。

5.2 全仏

表 4 全仏

	係数	標準偏差	t 値	P 値
(Intercept)	-0.000	0.153	0.000	1.000
Matches Won	0.864	0.160	5.416	0.000

表 4 より 96 個から残った変数は 1 個となり試合を勝ち進むことができる選手が総合的強さに直結する。

6 まとめ

選手の強さを定量的に示す過程で対戦競技は 1 対 1 で行い個々の対戦成績を考慮をし, 対戦選手によってゲームを取る難しさは変化することを BT モデルから表現することができた。重回帰分析の結果からハードコートの全豪と全米はリターンとサービスポイントからサーブが重要な項目と分かる。また, 全仏は大会の勝利数だけの変数として残った。勝つためには粘りが必要だから全英以上に体力が必要になる。そして, 全英はゲームをブレイクする技術が必要になる。全ての大会に勝つには体力と技術の 2 つを兼ね備える必要がある。いくつコートを制することができるのかが総合的強さに直結する。

7 おわりに

今回の解析を通じて選手が大会ごとに必要な特徴が理解できた。大会ごとに必要な技術が異なるため, 選手のプレイスタイルとコートとの関係からその選手が得意とする大会を見つけることで勝敗の予測などに活用できるよう今後も様々な試合を観戦していきたい。

参考文献

- [1] ATPRankings: PepperstoneATPRankings(Singles), <https://www.atptour.com/en/rankings/singles>, (2024/1 閲覧).
- [2] 加藤明・服部匡志: 『プロテニスプレーヤーの強さの統計的研究』, 南山大学経営学部情報管理学科卒業論文, 1997.
- [3] 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, 1998.
- [4] 竹内啓・藤野和建: 『スポーツの数理科学』, 共立出版, 1988.
- [5] 統計 WEB: 階層型クラスタ分析 2, <https://bellcurve.jp/statistics/course/27193.html> (2024/1 閲覧)