

# 成績データにみる投手と対戦チームの相性の分析

2017SS092 山口宗泰

指導教員：小市俊悟

## 1 はじめに

日本プロ野球は国内でも人気の高いスポーツの一つである。各チームは毎年 143 試合行い、交流戦を除くと、125 試合が各リーグ内で行われる。同じチームと 25 試合行う中で、チームとして得意不得意の投手が出てくる。本研究では、機械学習の手法を用いて、このようなチームと投手の対戦成績における相性を明らかにすることを旨とする。具体的には、クラスタリング手法を用いて、投手成績データから投手を似たタイプが同じグループになるように、いくつかのグループにわけると同時に、その後、各チームと各投手グループごとに相性を検証し、各チームについて、成績を上げるための最適なローテーションを探る。

## 2 データについて

本研究で利用するデータは、2019 年度のセ・リーグとパ・リーグの各 6 球団について、投手の各チームに対する成績と総合的な個人成績のデータである [1]。

各投手について、対戦チームごとに登板数、勝ち数、負け数、セーブ数といった項目が集計されている。

また、これとは別に各投手について、総合的な個人成績として 24 項目ほどが集計されているが、そのうち、分析に利用した項目は、(対戦) 打者数、投球回、被安打数、被本塁打、与四球数、奪三振数である。

## 3 分析方法

### 3.1 データの規格化

2 節に挙げた投手の総合的な個人成績の項目のうち、打者数と投球回を除く項目は、打者数もしくは投球回が多ければ、自然に大きくなると考えられる。そこで、表 1 に示すように、打者数もしくは投球回で割ることで規格化した数値を利用することにした。同時に、投球回が 0 であるよ

表 1 規格化した投手の総合的な個人成績

項目
被安打数/打者数
被本塁打数/打者数
与四球数/打者数
奪三振数/打者数
被安打数/投球回
被本塁打数/投球回
与四球数/投球回
奪三振数/投球回

うな投手は除いている。

### 3.2 階層的クラスタリング

投手をグループに分けるために、階層的クラスタリングを用いることにした。階層的クラスタリングとは、あらかじめデータ間に距離を定めておき、基本的な考えとしては、最も距離が近い組み合わせから順にまとめていくことで、徐々に、クラスタと呼ばれるデータのまとまりを大きくしていく方法である。クラスタがまとめられていく構造は、デンドログラムと呼ばれる木構造で表現できる。ただし、このような階層クラスタリングを実行するには、データ間の距離はもちろんのこと、クラスタリングの途中では、クラスタ間の距離も必要となる。このようなデータ間の距離およびクラスタ間の距離には、いずれも様々な定義・計算法が存在し、どのような距離を採用するかは、クラスタリングを行う上で、熟考しなければならない。本研究では、ユークリッド距離を用いた最小分散法と呼ばれる計算法を採用して、階層的クラスタリングを行うこととした。

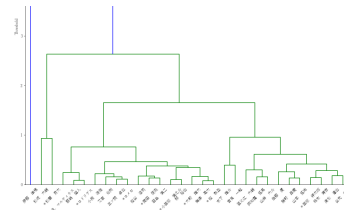


図 1 中日投手のデンドログラム

図 1 は、中日の投手について、規格化した投手の総合的な個人成績から得られたデンドログラムである。

## 4 全リーグ投手のクラスタリング

セ・リーグの全投手をまとめて対象にしたクラスタリングを行う。これにより、各投手が、チーム内ではなく、リーグ全体で示す特性に基づいて分類されると考えられる。また、そうすることで、所属チームの成績となるべく連動しないような投手の評価が可能ではないかと考えた。

登板数が少ない投手などは対象から削除することにし、残る投手を 6 個のクラスタに分けた。

## 5 チームと投手との相性

次に、投手の各クラスタに属す投手が、各チームに対してどのような対戦成績であるかを集計した。具体的には、各チームと各クラスタの組み合わせについて、勝ち数、負け数、セーブ数、勝ち数/登板数、負け数/登板数、セーブ数/登板数を集計した。勝ち負けは、投手から見た成績である。これにより、各チームについて、どのクラスタの投手が苦手であり、得意であるかが明らかになることを期待

する。

得られた結果の一部として、中日と巨人の対戦成績を示す。表2は、中日に対して、表3は巨人に対して各クラスターの投手がどのような成績をあげているかを集計したものである。この表から、中日に対して、クラスター2に属す投手とクラスター3に属す投手は、ともに勝ち越しているが、巨人に対して、クラスター2は勝ち越しているものの、クラスター3は負け越していることがわかる。

表2や同様の表から投手の各クラスターと各チームの相性を算出することにした。本研究では、先発投手のローテーションに主に興味があるため、勝ち数と負け数を利用して算出する。要旨では、算出の方法として、次を紹介する。

表2 各クラスターの投手の中日に対する成績

クラスター番号	勝ち数	負け数	セーブ数
1	4	5	3
2	20	17	15
3	34	29	3
4	3	3	5
5	0	1	0
6	0	0	0

表3 各クラスターの投手の巨人に対する成績

クラスター番号	勝ち数	負け数	セーブ数
1	1	6	6
2	27	20	16
3	23	32	1
4	5	5	9
5	0	0	
6	0	0	

**勝ち数の勝ち数と負け数の和に対する比** 各チームについて、各クラスターの勝ち数/(勝ち数 + 負け数)を、そのクラスターの当該チームへの相性と考える。勝ち数/(勝ち数 + 負け数)は、投手から見た勝率のようなものを表すと考えられるので、その値が大きいほど、そのクラスターは当該チームに対して、相性が良く、勝ちやすいと考える。セ・リーグの6チームについて、その相性をまとめたものが表4である。

表4 各チームに対する各クラスターの相性 (勝ち数/(勝ち数 + 負け数))

クラスター番号	中日	巨人	DeNA	広島	阪神	ヤクルト
1	0.44	0.14	0.36	0.31	0.44	0.42
2	0.54	0.57	0.5	0.58	0.47	0.58
3	0.54	0.42	0.54	0.44	0.5	0.62
4	0.5	0.5	0.8	0	0.5	0.6
5	0	0	0	0.5	0	0.5
6	0	0	0	0	0	0

2019年に首位であった巨人に対して、クラスター2以外の投手は、かなり相性が悪かったことが読み取れる。阪神に対して勝率が5割を超えるクラスターはないこととも合わせると、クラスター1の投手では、巨人やDeNAを抑えることができなかったことが、巨人やDeNAが阪神の成績を上回った理由かもしれない。

## 6 最適ローテーション

投手(クラスター)とチームの相性分析に基づき、最適な先発投手のローテーションを考える。単純化のために、先発

表5 先発投手の最適ローテーションと相性の総和の最大値

チーム	最大値	1	2	3	4	5
中日	55.3	大野	ロメロ	柳	山井	山本
巨人	64.0	山口	菅野	高橋	メルセデス	今村
DeNA	62.8	上茶谷	井納	大貫	今永	濱口
広島	63.5	九里	大瀬良	床田	アドゥワ	ジョンソン
阪神	62.8	西	青柳	ガルシア	メッセ	高橋
ヤクルト	59.8	小川	高橋	石川	ブキャナン	山田

投手は各チーム5名とし、各チームについて、登板数に対して投球回数が4倍以上である投手の中から登板数が多い者5名を選択した。

各チームにとって、対戦日程は、基本的に与えられるものであると考えられるので、与えられた対戦日程に対して、どのような先発投手のローテーションが最適であるかを考える。単純のため、開幕時に決めたローテーション通りに順次、登板していく形で1年間戦うものとする。したがって、ローテーションを決めると、各投手の対戦相手が決まる。対戦相手が決まれば、先発投手がどのクラスターに属すかにより、対戦相手との相性も決まる。なるべく相性がよい対戦相手とあたるようなローテーションが求められるべきローテーションとなる。相性として、勝率とも考えられる勝ち数/(勝ち数 + 負け数)を採用し、各ローテーションについて、1年間の相性の総和を求める。ローテーションは5名なので、すべての順番を列挙することにし、相性の総和が最も大きかったものを最適ローテーションとする。

乱数を用いて、仮想的な対戦日程を作成し、それに対して、上述の意味で最適なローテーションを求めた。表5に、最大値とともに各チームの最適ローテーションを示す。

先発投手のクラスターの構成が同じであっても、自チームとの対戦がないため、得られる相性の総和は異なる。また、その値には、チーム成績が少なからず加味されるので、巨人の値が大きいのは、それで説明できるが、DeNAや阪神より広島の方が大きい値となるにも関わらず、成績では、DeNAや阪神より下位の4位になったことは興味深い。ただし、2019年のその3チームの成績は、ほとんど同じでもあった。

## 7 おわりに

各チームの最適ローテーションは、非常に単純な設定の下に求めたが、基本的になるべく相性のよいチームと対戦するようにローテーションが組まれる。実際に使用するには、より現実的な設定の下に求める必要があると思うが、より複雑なデータも必要となるので、本研究では、簡単なものにとどまったが、最適ローテーションを求める過程で行った分析において、各チームの様々な特徴を明らかにすることができ、それも本研究の成果になると考える。

## 参考文献

- [1] ベースボール・マガジン社：『2020 ベースボール・レコード・ブック』。東京，2019。