

最終球への配球推移に基づくキャッチャー成績分析

Analysis of Catcher Grade Based on Transition of Pitches

稲福 和史[†] 伏見 卓恭^{††} 佐藤 哲司^{†††}

[†] 筑波大学図書館情報メディア研究科 〒305-8550 茨城県つくば市春日 1-2

^{†††} 筑波大学図書館情報メディア研究系 〒305-8550 茨城県つくば市春日 1-2

^{††} 東京工科大学コンピュータサイエンス学部 〒305-0982 東京都八王子市片倉町 1404-1

E-mail: [†]{inafuku,satoh}@ce.slis.tsukuba.ac.jp, ^{††}takayasu.fushimi@gmail.com

あらまし 野球において打者を打ち取るための要素として重要な配球は、主に捕手のリードによって決定されることが知られている。本研究では、打席終了時への配球に着目し、球種、球速、座標の推移から捕手の配球傾向を解析する。日本プロ野球のデータを対象に、捕手の配球傾向を抽出、自責点より算出される防御率（通常は投手に対して適用されるが、本研究では捕手に対して算出する）と併せて分析を行ったところ、他の捕手と異なる配球傾向を示す捕手ほど自責点が小さいという傾向にあることが示された。このことから、提案手法の有効性が確認されたので報告する。

キーワード プロ野球、配球、成績分析

1 はじめに

野球において、守備側の投手は打者からアウトを取るために、球種や球速、コースを変化させながら投球を行う。打者を打ち取るための要素として、投球の一連の組み合わせである配球の役割は大きい。加えて、日本プロ野球において、配球は主に捕手によって決定されることから、捕手に着目した分析は重要だと考えられる。

捕手の数は投手の 1/5 程度と少ない。そのため、多くの捕手が複数の投手とバッテリーを組んでいるが、捕手毎に組む投手に偏りがある。投手と捕手のバッテリー関係のネットワークを図 1 に示す。赤いノードが捕手、緑のノードが投手を示し、ノードの大きさは回数に対応する。また、リンクしている選手同士はバッテリー関係にあり、リンクの太さはそのバッテリーにおける投球数である。いずれのチームにおいても回数の大きい中心的な捕手が 2 名ずつ存在していること、控えの選手が数名存在することなどが見てとれる。また、捕手によってバッテリーを組む投手が大きく異なることや、複数の捕手とバッテリーを組んでいる投手であっても、その頻度に大きな差異がある投手とない投手が存在することが分かる。これは、特定のバッテリーが好まれていることや、逆に相手に依存しない選手が存在することを示す。これらのことから、捕手の性質には何らかの差異があると考えられる。バッテリーの決定には監督の意向やシーズンを通しての試合状況、疲労状況など様々な要因が考慮されるが、そのうちの一つにこの捕手毎の性質の差異があるのではないかと考えられる。

そこで、本研究では捕手の性質を配球で表現し分析することを試みる。配球について、特定のボールカウントや走者の有無といった局面ごとに分析した研究がいくつかあるが、これらは主に「この条件下でどのような球が来るか」というアプローチ

である。一方、打者を打ち取るためには、直前の投球に比べどのように変化したか、という配球の流れ（以下、配球推移）が重要であるが、この配球推移に着目した研究は少ない。また、多くの研究では、特定の選手がどのような配球を行うか、という点に着目しており、配球傾向が類似する選手同士の比較など、選手を横断的に分析した事例も少ない。本研究では捕手ごとの配球推移傾向を特徴量とし、自責点から算出する防御率¹との関連を調査する。

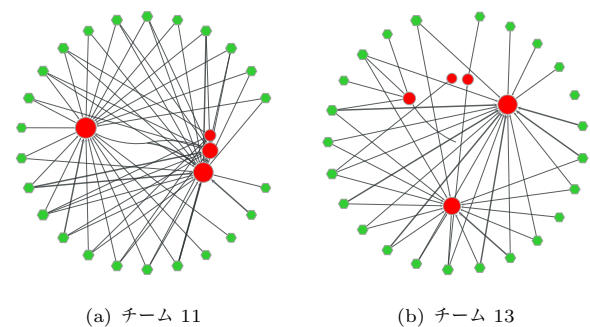


図 1 バッテリーのネットワーク

2 関連研究

配球の分析は、大きく分けて特定の局面においてどのような投球がなされるかという点に着目した分析、次にどのようなボールが投げられるかという配球予測、配球の傾向可視化などが存在する。本章ではこれらについて説明する。

¹：防御率は通常、投手に対して算出される評価指標であるが、本研究では捕手に対して適用する

2.1 特定の局面における投球に関する研究・報告

配球は、アウトカウントや走者の有無など様々な試合状況によって大きく異なる。そのため、実際の試合の配球について分析した研究では、特定の状況を想定したものが多い。

菊地らはカウント 0-0 からの投球（すなわち打席の初球）において、どのような投球がストライクを取りやすいのかに着目し、コースや球種別にカウント 0-0 からの投球結果を集計・分析している [2]。また、山本らは高校野球において、攻撃が予想される場面（送りバントやスクイズなど）における配球に着目し、攻撃側の有効な攻め方を提案している [5]。小宮山らは日本のプロ野球におけるデータ活用の状況について報告しており、カウント別など局面毎にどのような投球がなされるかに着目した活用事例が多いとされている [6]。その他にも、同様の現状報告があるが、いずれも局面毎の投球に着目したものである [7] [1]。

本研究は、一連の配球である配球推移に着目する点で異なる。

2.2 配球予測に関する研究

次に投げられるボールの球種やコース、球速を予測する配球予測に関する研究について説明する。

桑原らは、3 組のバッテリーを対象とし、次に投球される球種がストレートか変化球かを予測する二値分類問題を扱い、バッテリーやボールカウントによっては分類が可能であることを示している [3]。判別する球種をストレートと変化球としているのは、投手によって持ち球が異なることへ対処するためである。また、田淵らは、投手 1 名を対象とし、決定木を用いて配球傾向の分析を行い、打者を打ち取れる確率の高い配球パターンについて明らかにしている [8]。

本研究は、試合経験の少ない選手を除くすべての捕手を対象にしている点、投手毎に得意な変化球を算出し、球種をストレート、得意な変化球、その他の変化球の 3 つに分けている点などで異なる。

2.3 配球可視化に関する研究

配球の分析結果は、主に選手や監督へと提供されることが多い。このことから、直感的に分析結果を伝達するための可視化研究が行われている。光廣らは捕手ごとの配球傾向について、コースと球種（ストレートか変化球かの二値）の推移関係を可視化する手法を提案している [4]。

本研究は、捕手同士の配球傾向を比較するなど選手間の横断的な分析を行うという点で異なる。

3 配球傾向の特徴量構築手法

守備側である捕手は攻撃側である打者に打たれないために、球種、球速、コースなどを変化させながら配球を行う。そしてその配球は、打者をアウトにする投球に向かって組み立てられると考えられる。そこで本研究では、打席が終了するまでの数球の配球推移に着目し、捕手の配球傾向を分析する手法を提案する。具体的には、球種、球速、座標についてそれぞれの推移パターンを数えあげ、捕手の配球傾向とみなす。例えば、変化球の球種を V 種類とした時、最終球を含めた N 球の配球推移

パターンは V^N 通りになる。この V^N 通りの出現回数を数え上げ正規化を行うことで、捕手の配球傾向とみなすことができる。これらの配球傾向と自責点から算出される防御率を照らし合わせ、捕手の配球傾向と成績について考察する。

3.1 球種推移ベクトルの構築

今回利用するデータの球種は 8 種類存在する²が、投手毎に投げられる球種は異なる。そこで、各投手の持ち玉をストレート、得意な変化球、その他の変化球の 3 つに置き換える。球種毎に得意度 m を算出し、最も大きい m を示す変化球を得意な変化球とする。球種 b の得意度 m_b は、球種 b の投球数 c_b と球種 b でアウトにした（三振あるいは凡打にした）数を o_b を用いて

$$m_b = \frac{o_b}{c_b}$$

と定義する。

この時、考える球種推移パターンは 3^N 通りである。各捕手について、打席毎に球種推移パターンの出現回数を数え上げ、正規化 (L_1 norm) を行い、球種推移パターンの出現確率ベクトルを取得する。これにより、捕手ごとに頻出する配球傾向を特徴量化できる。また、本研究では、3 種類以上の球種を持たない投手のデータについては分析対象外とした。

3.2 球速推移ベクトルの構築

球速推移は、直前の投球との球速差が重要である。ここで、球速差を基準に、球速差が +10km/h の時はラベル A、+20km/h の時にはラベル B というように、しきい値を決定しラベルを付与することを考える。しかし、投手ごとに投げられる速度には違いがあるので、投手毎にラベル付与のしきい値を決定する必要がある。そこで、各投手について球種別の平均球速を算出し、最大の平均球速と最小の平均球速との差 δ を該当投手が投球可能なおおよその球速範囲とみなす。図 2 に δ のヒストグラムを示す。横軸が δ 、縦軸が投手数を示す。15km/h から 30km/h に選手が集中していることが分かる。

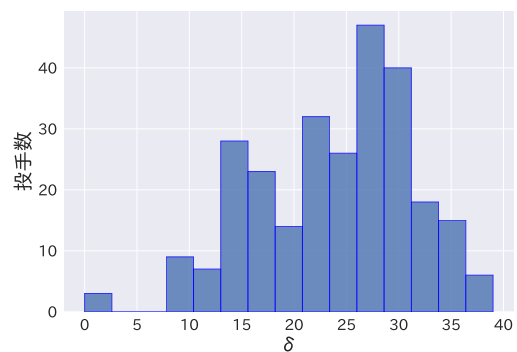


図 2 投球可能な球速範囲 δ の分布

2: 厳密には、ストレート、カット、シュート、フォーク、スライダー、カーブ、シンカー、チェンジアップ、特殊球の計 9 種類存在する。また、球種の計測が出来なかった「無し」のデータも存在する。本研究では、「無し」と特殊球を除いた 8 種類を分析対象とする。

球速差は遅い球から速い球への推移のときに正の値、速い球から遅い球への推移のときに負の値を取るため、球速差の取りうるおおよその範囲は $-\delta$ から $+\delta$ の間である。これをラベル数で除したものをラベル付与のしきい値とする。本研究では表 1 のように 5 種のラベルを付与し、3.1 と同様に球速推移パターンの出現確率ベクトルを構築する。投球毎の差であるので、 5^{N-1} 通りの球速推移パターンが存在し、球速推移ベクトルは 5^{N-1} 次元となる。

表 1 球速差ラベル

球速差	ラベル
$(\delta \times 1.5)\text{km/h}$ 以上	+大
$(\delta \times 0.5)\text{km/h}$ 以上 $(\delta \times 1.5)\text{km/h}$ 未満	+中
$(\delta \times -0.5)\text{km/h}$ 以上 $(\delta \times 0.5)\text{km/h}$ 未満	小
$(\delta \times -1.5)\text{km/h}$ 以上 $(\delta \times -0.5)\text{km/h}$ 未満	-中
$(\delta \times -1.5)\text{km/h}$ 未満	-大

3.3 座標推移ベクトルの構築

座標推移についても、球速推移同様、直前の投球との差に着目する。投球座標のデータは、 x (横) が 0 から 240, y (縦) が 0 から 250 の値を取る二次元平面の座標として記録されている。座標については、球種や球速と異なり、投手ごとに投げられない箇所というものほとんど存在しないので、投手全体に対して一定のしきい値に基づくラベル付与を行う。直前の投球の移動量とその方向を用いて図 3 のように 9 種のラベルを付与する。移動量ラベルの基準となる黄色の円の半径 r はストライクゾーンの中心から移動できる最大量を移動量のラベル数で除したものとす。ここではラベル数を 3 と設定し、

$$r = \frac{\sqrt{120^2 + 125^2}}{3} \simeq 57.759$$

とする。これにより、 9^{N-1} 次元の座標推移ベクトルを構築できる。また、打者から見て内角側に移動するか外角側に移動するかが重要であると考えられるので、左打者の場合にはラベルを左右反転させる。

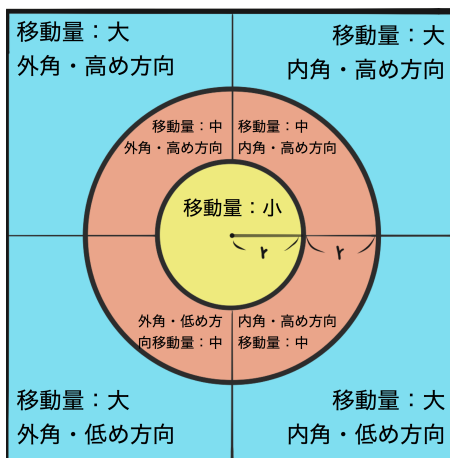


図 3 座標推移ラベル (右打者の場合)

3.4 主成分分析による次元削減と可視化

球種、球速、座標推移ベクトルは多次元ベクトルであるため、そのままでは可視化できない。そこで、主成分分析による次元削減を行い、二次元平面上に表現する。この際、捕手の成績指標として防御率を算出し、重ねて可視化する。防御率とは、9 イニングを投げた時の平均自責点であり、通常は投手の評価指標として算出される。本研究では、この防御率を捕手について算出する。自責点を l , 出場したイニング数を i で表わすと防御率 g は次のように表され、低い値を示すほど、失点が少ないことを意味する。

$$g = \frac{l * 9 * 3}{i * 3}$$

これにより、捕手の配球傾向と成績の関係について分析できる。

4 分 析

4.1 データセット

本研究では、楽天株式会社から提供された「2016 年プロ野球パ・リーグ全試合 全球データ」を利用する。全球データとは投手がどのような投球を行い、打者がどのように対応したか、その打撃内容や守備内容などを記録したデータである。一球毎に球種や球速、コースに加え、打撃結果、選手のマスク ID、走者の有無、球場や当日の天候など様々な情報が記録されている。本データセットには捕手 48 名が登録されている。図 4 に捕手の捕球数と防御率を示す。横軸が捕球数、縦軸が防御率、一つの点が一の選手に対応する。この内、ある程度の試合経験がある選手を抽出するため、捕球数が下位 10%の選手を分析対象外とした。また、極端に防御率の高い選手 (防御率 8.0 以上) についても除外した。結果、青い点で示された 39 名の捕手を分析対象とした。

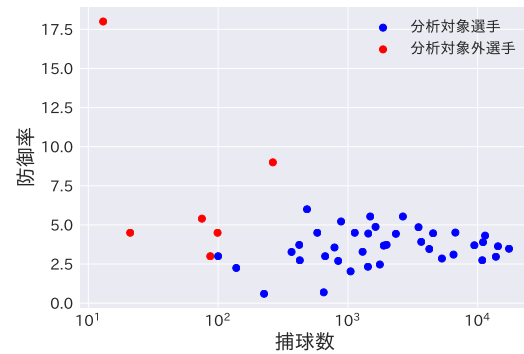


図 4 捕球数と防御率

また、本研究では最終球を含めた N 球の配球推移について取り扱う。 N の決定にあたり、図 5 に打席毎の投球数分布を示す。横軸は投球数、縦軸は打席数を示す。例えば、4 球で終了 (被安打あるいは奪アウト) した打席が約 7,000 打席存在することが分かる。本稿では、 N 球に満たない打席は分析対象外とする。そのため、 N の値を大きく設定しすぎると、分析対象の打席が減少し捕手の特徴を十分に表せなくなるので、4 球よりも

大きく設定することは好ましくない。加えて、推移ベクトルの次元数はラベル数の N 乗あるいは $N - 1$ 乗となるため、 N の値が大きいほど特徴量の粒度自体は細くなるが、1 クラスあたりの打席数は減少するため捕手同士の比較がしにくくなる。また、応用面から考えると、どのような配球に偏っているかなどを選手やチームに説明しうる次元数である必要がある。以上の理由から、本稿では $N = 3$ と設定し分析を行った。また、上述したように、3 球に満たない打席は分析対象外とした。

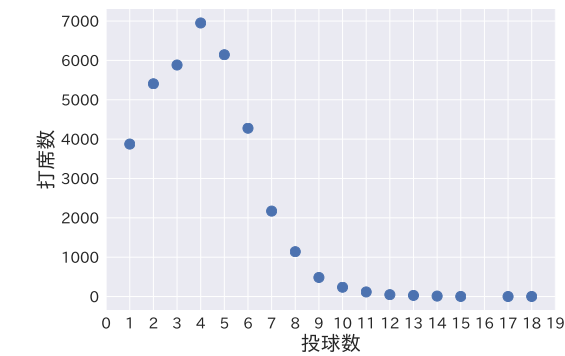


図 5 打席毎の投球数分布

4.2 上位主成分の可視化

4.1 の捕手 39 名のデータから、球種、球速、座標の推移ベクトルを構築し、それぞれに対して主成分分析を行った。上位主成分の累積寄与率を表 2 に示す。一般に、累積寄与率が 0.8 を超えるとデータを十分に説明できている状態だとされているが、紙面の都合上、第三主成分までの累積寄与率を記載する。なお、各推移ベクトルの累積寄与率が 0.8 に達したのは、球種推移ベクトルで表 2 に示すように第三主成分、球速推移ベクトルで第五主成分、座標推移ベクトルで第九主成分である。

表 2 累積寄与率

	球種	球速	座標
第一主成分 (PC1)	0.586	0.400	0.277
第二主成分 (PC2)	0.707	0.550	0.441
第三主成分 (PC3)	0.812	0.663	0.537

球種、球速、座標推移ベクトルの第一主成分と第二主成分の散布図を図 6 に示す。横軸は第一主成分、縦軸は第二主成分、プロット点は捕手に対応し、色は防御率を示す。また、近くに位置する捕手ほど互いに配球傾向が類似しており、防御率が低い (青い) ほど自責点の少ない好成績の捕手である。球種、球速、座標のいずれにおいても、防御率成績の低い捕手 (赤いプロット点) ほど近くに位置し、類似した配球傾向を持つことが観察できる。一方で、好成績な捕手 (青いプロット点) ほど、他の捕手とは離れた場所に位置する、すなわち異なる配球傾向を示すことが分かる。特に球種については、配球傾向の類似性と防御率成績の関連がより強く表れていることが分かる。

4.3 配球傾向と防御率の関連

4.2 において防御率の高い (低成績な) 捕手ほど他の捕手と類似した配球傾向を示し、逆に防御率の低い (好成績な) 捕手ほど他の選手とは異なる配球傾向を示すことが示唆された。そこで、分析対象の捕手を、防御率が中央値以上の 20 名 (上位選手) と中央値未満の 19 名 (下位選手) の 2 グループに分け、それぞれの上位主成分における標準偏差を比較した (表 3)。球種、球速、座標のすべての上位主成分において、上位選手の標準偏差が大きい。これは、上位選手は下位選手と比べて、互いにより異なる配球傾向を持っており、逆に下位選手同士はより類似した配球傾向にあることを意味する。

表 3 上位選手と下位選手の標準偏差比較

		上位選手	下位選手	上位-下位
球種	第一主成分 (PC1)	0.102	0.066	0.036
	第二主成分 (PC2)	0.053	0.022	0.031
	第三主成分 (PC3)	0.043	0.023	0.019
球速	第一主成分 (PC1)	0.053	0.032	0.021
	第二主成分 (PC2)	0.033	0.021	0.012
	第三主成分 (PC3)	0.031	0.011	0.020
座標	第一主成分 (PC1)	0.062	0.008	0.054
	第二主成分 (PC2)	0.045	0.015	0.030
	第三主成分 (PC3)	0.035	0.014	0.021

ここで、球種、球速、座標推移ベクトルについて、それぞれの重心と各捕手の特徴量間のユークリッド距離を求め、相関係数を算出した。重心からのユークリッド距離が大きいほど、他の選手と異なる配球傾向を持つ。重心からの距離と防御率の散布図を図 7 に示す。横軸は重心からの距離、縦軸は防御率、一つの点は一人の捕手に対応する。球種、球速、座標のいずれにおいてもユークリッド距離が大きいほど、防御率が低くなる傾向にあることが分かる。また、球種推移の相関係数は-0.550、球速推移の相関係数は-0.370、座標推移の相関係数は-0.258 であり、いずれにおいても弱い相関あるいはやや相関ありという結果が得られた。これらの上位選手と下位選手の標準偏差の比較及び重心からの距離と防御率の相関関係から、好成績の捕手ほど他の選手と異なる配球傾向を持つと考えられる。

4.4 被安打時及び奪アウト時の配球傾向と防御率の関連

本研究では、配球推移を打者を打ち取るために重要な要素だとみなし、捕手の特徴量を構築、分析を行ってきた。ここで、打席の結果に着目し、打者を打ち取ったケース (奪アウト時) と打者に安打を許したケース (被安打時) に分けて同様の分析を行った。

図 8 に各ベクトルについて、奪アウト時と被安打時それぞれの、重心からのユークリッド距離と防御率の相関係数を示す。横軸が各種ベクトル、縦軸が相関係数を示す。この相関係数は、ユークリッド距離がどれだけ防御率と関連が強い値であるか、すなわち指標としての有効性評価だと解釈することができる。座標推移ベクトルの相関係数に着目すると、奪アウト時と被安打時で 2 倍近くの差があることが分かる。これは、座標推移に

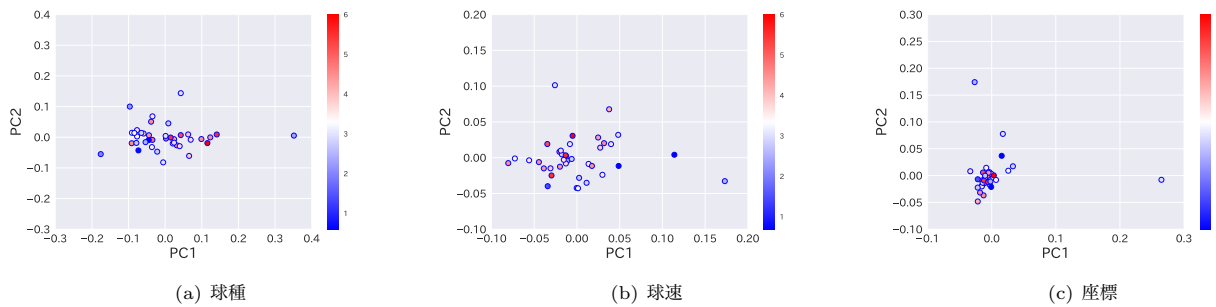


図 6 各種推移ベクトルの上位主成分及び防御率の可視化結果

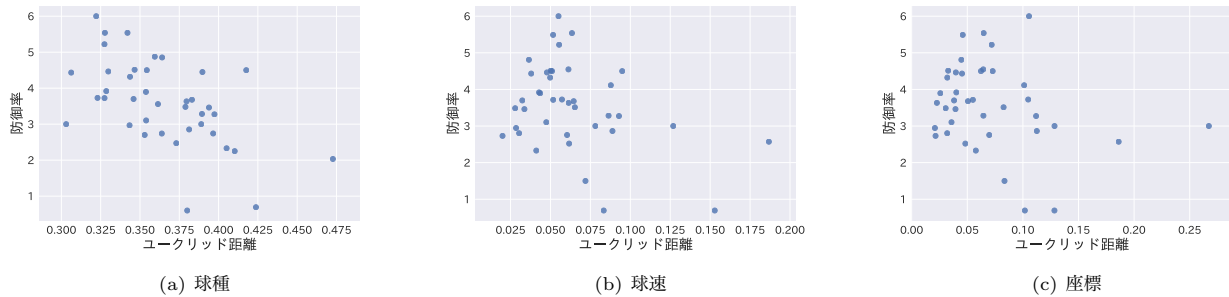


図 7 重心からの距離と防御率

着目した特徴量が、被安打時の分析において特に有効であることを示している。逆に、球種推移ベクトルにおいては、奪アウト時の相関係数が高いことが分かる。すなわち、アウトを取るための分析であれば球種推移に着目することが有効であるといえる。以上のことから、球種、球速、座標それぞれについて、より有効に機能するケースが存在することが明らかとなった。

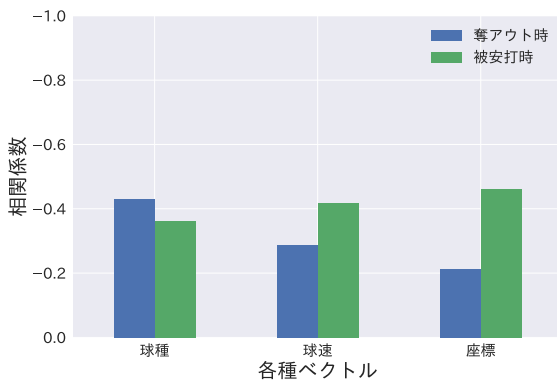


図 8 奪アウト時及び被安打時の相関係数

5 おわりに

本研究では、打席が終了するまでの打球の推移に着目し、捕手毎の配球傾向を球種、球速、座標について抽出、防御率との関連を分析した。まず、4章の結果から、配球傾向が他の選手と異なる捕手ほど、防御率が低く好成績であることを明らかにした。逆に防御率の高い低成績な捕手については配球傾向が互いに類似しているという結果が得られた。配球には当然一定の

セオリーが存在するが、これらの結果から、セオリーを踏まえた上で独自の配球傾向を持つ捕手が、失点を少なく抑えられているのではないかと考えられる。これらの好成績な捕手と低成績な捕手の違いを分析することで、配球の改善に繋げられる可能性がある。また、4.4で、奪アウト時及び被安打時のケース別に分析を行い、各特徴量にはより有効に機能するケースがあることを明らかにした。如何にアウトを取るかという観点での分析であれば球種推移に、如何に打たれないかという観点での分析であれば球速推移や座標推移に着目することが有効だと考えられる。

今後の課題として、まず、最終的に試合の勝利に貢献することを目指した具体的な配球傾向の抽出及び分析が挙げられる。監督や選手への説明を視野に入れた分析手法を考案する必要がある。また、手法についてはラベル付与に伴う閾値決定の改善が挙げられる。単純には閾値を変化させながら、相関係数や標準偏差にどのような変化が現れるかを観察することなどが考えられる。今回は成績指標として防御率を用いたが、被安打数や被長打率などその他の成績指標との関連を調査することが考えられる。最後に、実際の試合における配球は、様々な試合状況を鑑みて総合的な判断のもと決定されるものである。そのため、アウトカウントや走者の有無、また投手の疲労度や累積投球数などの諸条件を考慮した配球傾向の抽出を行う必要がある。

謝 辞

本研究の実装・評価に際し、楽天株式会社が提供する「2016年プロ野球パ・リーグ全試合 全球データ」を利用しました。本研究は JSPS 科研費 JP16H02904 の助成を受けたものです。ここに記して謝意を示します。

文 献

- [1] 猿渡康文, 安藤順三, 大山達雄. プロ野球現場の戦略: スコアラーによるデータの収集と整理と活用. オペレーションズ・リサーチ: 経営の科学, 44(3):119–124, 1999.
- [2] 菊地啓太, 中島宣行, 綿田博人. 大学野球における配球について: カウント 0-0 における投球の分析. 慶應義塾大学体育研究所, 1, 2010.
- [3] 桑原崇, 岸義樹, et al. 野球における捕手の配球予測に関する方法. 全国大会講演論文集, (人工知能と認知科学):213–214, 2009.
- [4] 光廣正基, 谷岡健資, 宿久洋. 視覚化法を用いた野球配球の特徴把握: 多元データを用いたアプローチ (session 4a (スポーツ統計)). 日本計算機統計学会シンポジウム論文集 26, pages 31–32. 日本計算機統計学会, 2012.
- [5] 山本誠二, 濱田幸二, 池田一徳, 杜小剛. 高校野球のゲーム分析: 攻撃が予測される場面での投手の配球に関する研究. 日本体育学会大会号 第 42 回 (1991), page 727. 一般社団法人日本体育学会, 1991.
- [6] 小宮山悟, 勝亦陽一, 福永哲夫. プロ野球球団におけるゲーム分析データの活用事例. スポーツパフォーマンス研究, 7:346–355, 2015.
- [7] 仁志敏久, 鳥越規央. 日本プロ野球におけるデータ活用の現状 (session 4a (スポーツ統計)). 日本計算機統計学会シンポジウム論文集 26, pages 33–34. 日本計算機統計学会, 2012.
- [8] 田淵裕崇, 西嶋尚彦. 決定木を用いた野球の配球分析 (08. 測定評価, 一般研究発表). 日本体育学会大会号, 55:432, 2004.