野球における捕手の配球予測に関する方法

桑原 崇† 岸 義樹‡
茨城大学大学院理工学研究科† 茨城大学工学部情報工学科‡

1．はじめに

野球における配球とは、打者を打ち取るために最適な球種やコースとなる投球の組み立てであるが、その内容は配球を決定する捕手の主観と経験から組み立てられるものであり、一般的な理論は確立されていない。しかしながら、状況別のセオリーや選手の特徴等から、捕手の組み立てる配球には傾向が存在すると考えられるが、その有効性については研究例が少ない。

そこで、本研究では配球傾向の有効性を検証するアプローチとして、次に投球される球種がストライクコースに来ると仮定したとき、速球か変化球かを予測する問題を扱う。これを速球・変化球の２クラス線形分類問題と考え、投球記録から抽出可能な数値データ要素を入力値とした決定境界を求める手法の提案と検証を行う。

2．予測内容

2.1 予測対象

予測対象は投球データ量の関係から先発投手に限定する。

2.2 予測球種

球種は「速球」と「変化球」とする。これは各投手の持ち球が異なることや、同一球種でも速度や変化量を変えて使い分けている球種を判別できないことへの対処である。

2.3 予測コース

本論文では、投球がストライクコースにくると仮定した上での球種を予測する。これは、投球コースが制球ミスによるものなのか、捕手の配球の意図に合っているかをデータから判断するのが困難であることによる。

Technique concerning forecast of catcher's combination of pitches of pitches in baseball.
Takashi Kuwabara, Yoshiki Kishi
Ibaraki University
4-12-1 Nakanarusawa, Hitachi, Ibaraki,316-8511, JAPAN

2.4 予測状況

捕手が配球決定に考慮すると考えられる状況を示す。走者とアウトカウントによる状況はバッテリーが打者に集中するものとして以下の状況に限定する。

・無死あるいは一死で走者なし
・二死におけるすべての状況

また、以下の状況は適切な数値化ができないため、予測要素から除外した[1]。

・球場の構造、立地、気象状況
・審判
・守備力
・点差とイニング

3．予測に必要な要素

本章では決定境界の入力値とする要素を述べる。要素の内容は、投球傾向と捕手が配球に考慮すると考えられる情報からなり、個人的に記録した 2007 年度日本プロ野球の公式試合データから抽出する。

3.1 バッテリーの投球傾向

過去の試合と現在の試合から、バッテリーの投球傾向を抽出する。過去の傾向は最近４試合の投球記録から表 1 に示した要素を抽出する。ただし、投手のコンディションは考慮しない。

1	対左右打者別のストライクに投じた球種の傾向
2	カウント別のストライクに投じた球種の傾向[2]
3	直前の球種およびコースと次の球種の相関
4	同打席同一打者における同球種の連続投球確率[3]
5	現行試合の直前 20 球でストライクに投げた球種の傾向

表１：抽出するバッテリーの投球傾向

3.2 打者の情報

打者の情報は、１ヶ月分の試合から打撃傾向を抽出する。本研究で用いた要素を以下に示す。ただし 50 打席以上の者に限る。

- 打率、IsoP(Isolated Power)、カウント別スイング率、球種別打率から求める当該カウントにおける打者の危険値

4．決定境界

３章で述べた要素を入力ベクトル $\boldsymbol{x}$ として線形識別関数に与える。このとき、$\boldsymbol{w}$ は重みベクトル、D はベクトルの次元である。$\boldsymbol{w}$ はバッテリー及びカウント毎に異なる値を用いている。

$$y(\boldsymbol{x}) = \sum_{i=0}^{D} x_i w_i \qquad (1)$$

4.1 入力ベクトル $\boldsymbol{x}$

入力ベクトル $\boldsymbol{x}$ は３章で述べた値の一部を前処理して扱う。その方法を例示する。

1）左打者
2）カウント 1-2
3）直前の１球が速球
4）0-0 から２球連続で速球

以上の条件を満たす場合、入力ベクトルは表２のようになる。

x_i	内容
x_0	左打者に対して速球を投げる確率
x_1	カウント 1-2 から速球を投げる確率 $- x_0$
x_2	速球を投げてカウント 1-2 になったとき、そこから速球を投げる確率 $- x_0$
x_3	３球連続で速球を投げる確率 － 0.5
x_4	直前 20 球の内、速球のストライクを投げた確率 － 0.5
x_5	1-2 での打者の危険値

表２：入力ベクトル $\boldsymbol{x}$

4.2 決定境界の学習

決定境界は、重みベクトル $\boldsymbol{w}$ に関する二乗和誤差関数 J を最小化して最適解を求める。t は正解データであり、$y(\boldsymbol{x})$ は決定境界の出力値である。この関数は式(2)で与え、重みの更新は訓練用データ１つに対し１度の更新を行う逐次学習を用いる。このとき、学習した重みは $\boldsymbol{w}$ は、各要素の相対的な重要度を表しているとも考えられる。

$$J = (t - y(\boldsymbol{x}))^2 / 2 \qquad (2)$$

$$\nabla \boldsymbol{w} = \frac{\partial J}{\partial \boldsymbol{w}} - \boldsymbol{x}(y(\boldsymbol{x}) - t) \qquad (3)$$

$$\boldsymbol{w}_{new} = \boldsymbol{w}_{old} - \eta \nabla \boldsymbol{w} \qquad (4)$$

5．シミュレーション

３組のバッテリーを取り上げ、1)すべての投球データを対象にしたときの分類精度、2)十分に $\boldsymbol{w}$ を学習させてから未知の投球データに対する予測精度を検証した。

1)での精度は３組のバッテリーとも 70%程度だったが、2)での精度は 60%ほどに落ち込んだ。しかし、2）にカウント別での精度に注目すると、高いもので 80%の精度で予測したバッテリーのカウントも複数あった。このことから、適切な要素を用いれば線形識別関数で球種を予測することが可能であると考えられる。

また、どのバッテリーにも共通して精度の悪かったのがカウント 0-0 であった。これは、バッテリーが初球を打たれることを特に嫌うため、配球の傾向が出ないようにしていると考えられる。

6．おわりに

本研究では、線形識別関数を用いて球種の分類と予測に関するシミュレーションを行った。その要素に投手および打者の情報を用いることで、バッテリーやカウントによっては２値の予測が可能であること確認した。

今後は、バッテリーが走者を考慮する状況での予測や、詳細な球種とコースの予測など、本研究で取り上げなかった要素を加えた上での手法を検討していきたい。

参考文献

[1]荒川俊一, 田沢健一郎, 石川哲也, データスタジアム. 野球の見方が 180 度変わるセイバーメトリクス. 宝島社. 2008
[2]野村克也. 野村ノート. 小学館. 2005
[3]山崎武司. 野村監督に教わったこと. 講談社. 2008