

MLB 試合 DB を用いたベイジアンネットワークによる 試合状況に応じた投球戦略推定

Pitching strategy estimated in accordance with the game situation
by the Bayesian network using the MLB game DB

中田祥希[†]
Yoshiki Nakata

荒井秀一[†]
Shuichi Arai

1. まえがき

野球等のスポーツの試合において勝つためには戦略が重要である。従来は監督等が経験や考え方に基づいてそれぞれの戦略をたてていた。しかし、これらは主観的に策定された戦略のため、有効か否かの判断が困難である。そこで近年では、コンピュータを用いたデータ分析が盛んに行なわれている。年棒等の指標とするために打率を用いて、選手を3つのカテゴリーに分類し、その結果から評価を行った研究 [1] やアウトカウント等で攻撃状態を定義し、マルコフ連鎖を用いて1試合の期待得点を求める Bukiet らの研究 [2] 等が存在する。しかし、これらの研究は選手評価や長期的戦略は行っているが、時々刻々と変化する試合状況に適した戦略策定の研究は行っていない。野球の試合は1投球の積み重ねで構成されており、1投球の結果から優位性は変化する。よって1投球毎の戦略推定が可能な対戦モデルを構築する必要がある。しかし、1投球の戦略決定には複雑な要因があり、因果関係を探るのは困難である。そこで本稿では、MLB 試合詳細スコアデータより観測結果から要因を探ることのできる確率的グラフィカルモデルであるベイジアンネットワークを用いて学習する。これにより、対象投手、打者の対戦モデルを構築し、ベイズ確率推定により試合状況に応じた投手の戦略推定を行う手法を提案する。

2. ベイジアンネットワークによる投球戦略推定

本研究では、MLB 詳細スコアデータからベイジアンネットワークにより対戦モデルを構築し、投球戦略推定を行う。構造学習には K2 アルゴリズムを用いた。まず対象投手、打者を決定し、類似選手の選択を行う。次に対象選手、類似選手の対戦データからベイジアンネットワークにより対戦モデルを構築し、試合状況を入力しベイズ確率推定を行うことで投球結果ノードの事後確率を計算する。投手の取り得る戦術毎にベイズ確率推定を行うことで、あるコースにある球種を投じた結果、その戦略が投手にとって利得の有無を推定する。評価するスコアには赤池情報量規準 (AIC) を用いた。

2.1. ベイジアンネットワークに用いるノードの決定

ノードは MLB 詳細スコアデータから取得した情報に基づいて決定する。ノード名とノードの取り得る状態を表 1 に示す。球種は図 1 に示すように、MLB 公式サイト [3] の試合詳細スコアデータで用いられている球種を変化方向別に分類し、定義した。コースは打者が見逃した場合、概ねストライクと判定される領域を4分割したコース (図 2 中コース番号 6, 7, 10, 11) とその周りのストライクと判定されるかボールと判定されるか際どい 12 コース (図 2 中コース番号 1 ~ 5, 8, 9, 12 ~ 16) で構成した合計 16 コースとした。

3. 類似選手の抽出

ベイジアンネットワークを学習する際に各試合状況のデータが大量に必要となるが、推定対象選手の対戦データのみでは充分ではないため、投手特徴量、打者特徴量を定義し類似選手の抽出を行う。類似投手の選択は対象

表 1: 試合解析により挙げたノード名と取り得る状態

ノード名	要素
表/裏	表, 裏
イニング	1~9
得点差	勝ち, 負け, 同点
1 塁走者	いる, いない
2 塁走者	いる, いない
3 塁走者	いる, いない
投球数	1 打席での投球数
ボールカウント	0, 1, 2, 3
ストライクカウント	0, 1, 2
アウトカウント	0, 1, 2
球種	S, R, RD, D, LD, L
投球コース	1~16
スイング	打者がスイングした, していない
投球結果の評価値	+, 0, -

選手と同じ、あるいは似ている戦略を取ることが望ましい。そこで、各球種の使用率を左右の打者別に求めることで特徴量を求める。一方、類似打者の選択はスイングするコースが似ていることと長距離打者、中距離打者などの打撃の傾向が似ていることが望ましい。そこで打者がスイングするコースの割合と全打席に対してのヒット、長打、アウトの割合を求めることで特徴量を求める。

3.1. 投球結果の評価値定義

投球結果が利得のある結果かを評価するために投球前の試合状況と投球後の試合状況の各評価値の和を算出し、その差を投球の評価値として定義した。そのために、得点に関わる試合状況となるボール、ストライク、アウトカウント、各塁の走者の有無に対し、野球の得点と対応した以下のような重みを定義することにした。1 失点を -1 とし各塁は 1 塁 $-\frac{1}{4}$, 2 塁 $-\frac{2}{4}$, 3 塁 $-\frac{3}{4}$ とし、ボールは 4 つで 1 塁に進むため 1 塁に $\frac{1}{4}$ 倍した $-\frac{1}{16}$ とした。ストライクはボール 1 つと同じ比率にするため $\frac{3}{4}$ を乗じて符号を正にした $\frac{1}{12}$ とし、アウトはストライク 3 つで 1 アウトのため 3 倍した $\frac{1}{4}$ と定義した。

4. 投球戦略推定実験

本章では、走者の有無による戦略の変化と構築した戦略モデルの推定が適切であることを確認するため、実験を行った。投球戦略を推定したい試合状況と同一のデータを類似対戦選手のものだけ MLB 試合データベースから抽出し、ベイジアンネットワークを学習する。学習したネットワークに対し、球種 - コースを指定し、その戦略の利得を推定する。戦略推定は以下の 2 種類について行った。

実験 (1) 同一選手で異なる試合状況。

実験 (2) 実際の投球結果が投手にとって利得のあった

[†] 東京都市大学 知識工学部 Faculty of Knowledge Engineering, Tokyo City University

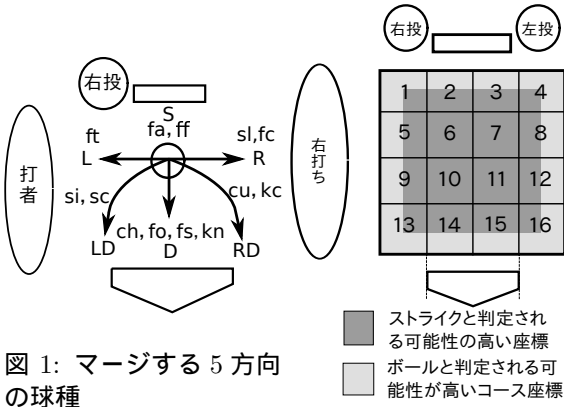


図1: マージする5方向の球種

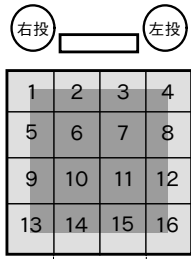


図2: コース定義

時, なかった時を推定結果と比較し妥当性の確認.

最初に実験(1)では, 投手: Felix Hernandez と打者: Josh Hamilton の対戦を走者無しと走者2, 3塁の試合状況で行った. 走者なしでは3回裏, 0ストライク, 0ボール, 2アウト, 負けている状況で, 走者2, 3塁では1回裏, 同点以外は走者無しと同じ状況で行った. 実験結果より, Felix Hernandez と Josh Hamilton の対戦データを抽出し学習したベイジアンネットワークを図3に示す. 表2から走者の居ない状況での投球はコース6, 7等のカウントを取る投球戦略が多く上位に推定していることがわかり, 表3から走者の居る状況での投球はカウントを取る投球戦略が推定されているだけでなく, 高めを見せる戦略や低めに投げて打ち取るための戦略も推定していることがわかる. 以上より走者の居る状況と居ない状況で異なる戦略を推定できたことを確認した.

次に実験(2)では, 推定結果と実際の投球を比較して

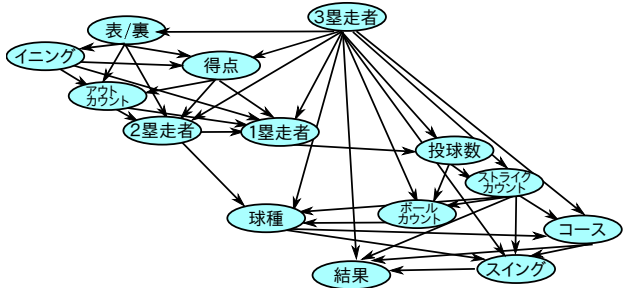


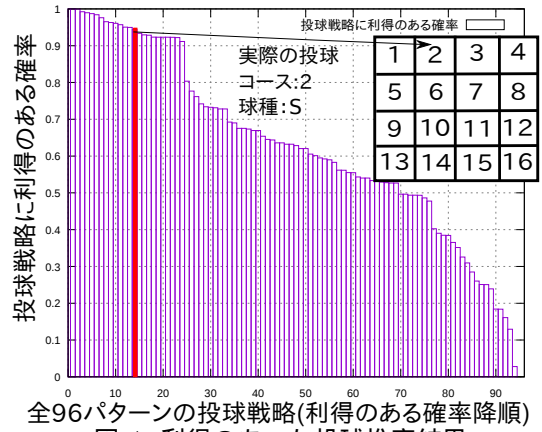
図3: 投手: Felix Hernandez vs 打者: Josh Hamilton のベイジアンネットワーク図

表2: 走者なしの推定結果 表3: 走者2, 3塁の推定結果

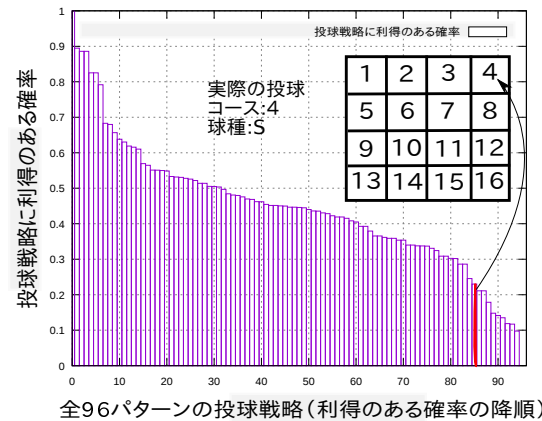
利得率 (%)	球種-コース	利得率 (%)	球種-コース
94.99	RD - 7	88.79	R - 7
90.48	R - 7	83.38	S - 3
89.12	RD - 6	79.55	RD - 10
89.01	D - 7	79.23	R - 11
88.98	LD - 7	79.12	D - 8

利得のあった投球と利得のなかった投球の推定結果の妥当性を評価した. 投手: Clayton Kershaw と打者: Jose Lopez の対戦で推定を行った. まず, 実際に利得のあった投球の状況は2回表, 走者なし, 2ストライク, 1ボール, 2アウト, 同点の状況である. 実際の投球結果はコー

ス2, 球種Sでフライアウトに打ち取った. 実際に利得のなかった投球の状況は2回裏, 走者1塁, 2ストライク, 1ボール, 2アウト, 勝っている状況である. 実際の投球結果はコース4, 球種Sでボールの判定である. 実験結果より, 実際に投手に利得のあった投球が推定結果の上位に推定できていることを図4から確認できる. また, 実際に利得のなかった投球も推定結果の下位に推定できていることを図5から確認できる. 以上より, 実際の投球と比較し推定結果の妥当性を確認することができた.



全96パターンの投球戦略(利得のある確率降順) 図4: 利得のあった投球推定結果



全96パターンの投球戦略(利得のある確率の降順) 図5: 利得のなかった投球推定結果

5. 結論

本研究では, 野球の試合が1投球の積み重ねであることに着目し, 1投球の戦略推定が可能なベイジアンネットワークにより試合状況に応じた投球戦略推定を行うことを目的とした. 結果より, ベイジアンネットワークにより走者の有無やカウント数等の様々な試合状況に応じた投球戦略推定を行えることと実際の投球との妥当性を確認した. 今後, 先発や中継ぎ, 抑え等の様々なタイプの投手において, 投球戦略推定を繰り返し行うことで, データからの投球戦略推定についての知識発見を行う.

参考文献

- [1] Hong, B.A. "Fuzzy clustering of batting averages," Fuzzy Information Processing Society, 2006.NAFIPS 2006. Annual meeting of the North American. IEEE, 2006.
- [2] Buiket, B. and Harold, E., A Markov Chain Approach to Baseball, Operational Research, Vol.45, No1(1977), pp.14-23.
- [3] MLB公式サイト <http://gd2.mlb.com/components/game/mlb>