

ビッグデータを用いた野球選手の評価

Evaluation of baseball players with big data

大澤 清
筑波大学

1. 研究目的

野球選手を評価する指標として、打撃成績としては打率・本塁打等、投手成績としては防御率・奪三振等、走塁成績としては盗塁・盗塁死等、守備成績としては守備率・失策等が長年にわたり用いられてきた。近年セイバーメトリクスと称する新しい指標がいくつか提案され、従来の指標よりも正確に野球選手の能力を測る試みが行われている。長らく低迷していた日米の一部球団が新しい指標を用いることで成績が改善され、ポストシーズン進出の常連となった例も存在する。

一方、近年球場内に軍事用のドップラーレーダーとステレオカメラを設置して詳細なデータを記録・分析するシステムをすべての MLB の球団と一部の NPB の球団が導入するなど、野球界にもビッグデータの活用を模索する動きが始まりつつある。

本研究ではインターネット上に公開されている大量のデータを用いた野球選手の新たな評価手法の提案を目的とする。

2. 研究成果の内容

野球の最適な打順の計算方法について、その概略は以下のようになる。マルコフ連鎖モデルを野球の試合に適用し、選手毎にその打撃成績、走塁成績、守備成績から定義される状態遷移行列を用いて行列積の演算を行い、試合に勝つ確率を計算する。より具体的には、行列積の演算から 9 イニング終了時点でチームが挙げている得点の確率分布が求められ、これを対戦している 2 チームについて求め、これらから一方がもう一方に対して勝つ (= 得点が上回る) 確率が計算される。与えられた選手の集合から構成し得るすべての打順について勝つ確率を計算し、その値が最大になる打順が最適な打順となる。最適な打順の計算においてこれまで使用されていた各種成績は以下の通りである。

打撃成績：打球処理野手を含む打撃結果（レフトへの単打、センターへの二塁打、セカンドへのゴロ、四球、三振など）

走塁成績：上記打撃結果に伴う進塁結果（一塁走者としてレフトへの単打で二塁へ進塁、三塁へ進塁、三塁でアウトなど）

守備成績：上記打撃結果に伴い発生する失策（打球に対する失策）

投手成績：使用しない

本研究ではインターネット上に公開されている大量のデータから投手成績として被打撃成績を取得し、打者の打撃成績と案分することで投手の能力を考慮に入れた最適な打順を求める。

マルコフ連鎖モデルに基づいた 70 状態間の状態遷移確率を表す状態遷移行列を上記各成績から定義し、 n (≥ 9) 人の選手の集合から抽出した 9 人で構成可能な打順に沿って選手毎に定義された状態遷移行列の行列積演算を行い、得点の確率分布を対戦する二チームについて計算してそれぞれが勝つ確率を計算する。

日本のプロ野球 (NPB) の広島東洋カープと対戦相手の読売ジャイアンツのそれぞれ 9 選手の 4 シーズン (2015 年~2018 年) 各種成績を使用し、広島の 9 選手から構成可能な 362,880 ($=9!$) 通りの打順から最適な打順を求めた。最適な打順と対戦相手の打順 (固定)、投手の被打撃成績 (抜粋) を以下に示す。

Carp's optimal batting order in 362,880 (=9!) orders

Pos	Player	AVG	HR	OBP
2	Aizawa	.305	13	.401
8	Maru	.306	39	.468
9	Suzuki	.320	30	.438
5	Nishikawa	.309	6	.364
3	Matsuyama	.302	12	.368
6	Tanaka	.262	10	.362
7	Noma	.286	5	.343
4	Kikuchi	.233	13	.301
1	Osera	.175	0	.217

Opponent Giants' order (fixed)

Pos	Player	AVG	HR	OBP
6	Sakamoto	.345	18	.424
4	Yoshikawa	.253	4	.304
5	McGehee	.285	21	.336
3	Okamoto	.309	33	.394
9	Kamei	.254	13	.315
7	Guerrero	.244	15	.325
8	Chono	.290	13	.359
2	Kobayashi	.219	2	.300
1	Sugano	.130	1	.145

Pitching results

Player	AVG	HR	OBP
Sugano allowed	.221	14	.260
Osera allowed	.215	22	.263

3. 学際共同利用が果たした役割と意義

本研究は萌芽的な研究であり、このような研究に対して計算資源を提供してもらえらることには大変意義があると考えている。またスポーツ科学分野の研究者に対して並列計算機を用いたアプローチを用いた研究手法を紹介することにも繋がり、同分野の発展にも寄与するものと考えらる。

4. 今後の展望

本年度に研究を再開したこともあり、年度中に対外発表を行うことができなかった。そのため来年度は研究成果を発表する機会を設けることにする。

また、研究内容については 9 人を超える選手の集合から最適な打順を求めるように問

題を拡張し、より大規模な計算によるメリットが享受できるような研究へと進展させる予定である。

使用計算機	使用計算機 に○	配分リソース*	
		当初配分	追加配分
Cygnus			
Oakforest-PACS	○	50,000	
※配分リソースについてはノード時間積をご記入ください。			