

# 野球チームの最適打順決定手法の高速化

## — 合併球団, 新球団の戦力評価 —

申請中 東京工業大学 \*大澤 清 OSAWA Kiyoshi  
非会員 東京工業大学 合田 憲人 AIDA Kento

### 1 はじめに

球団の合併やオーナーの相次ぐ辞任など, 2004年は日本プロ野球界にとって激動の年となったが, その結果 2005年のパシフィックリーグには合併球団オリックスバファローズと, 新球団東北楽天ゴールデンイーグルス加わることになった. 球団の合併は46年振り, 新規参入は50年振りの出来事であり, その戦力分析は興味深いテーマと考えられる.

球団の戦力分析手法としてマルコフ連鎖を用いた手法 [1] が知られている. 本手法では, 対象の球団から打者9人を選出し, 構成し得るすべての打順について D'Esopo and Lefkowitz モデル [1] を適用してその期待得点を算出する. しかし, 解を得るためには  $9! = 362,880$  通りの打順について期待得点を算出する必要があり, その中から最大値をとる打順を最適打順とするため, その計算量が大きいという問題がある. 例えば文献 [1] では, 期待得点を算出する打順の数を減らして得た最適打順と真の最適打順を比較するといった手法が試されている.

本稿では文献 [1] の手法の計算時間を短縮することを目的とし, 本手法に対して以下の高速化を適用した結果について報告する.

- 打順の循環に着目した計算結果の再利用
- 最適打順決定計算の並列化

比較的小規模な PC クラスタにおいてこれらの手法を適用して最適打順を求めた結果, 平均で5分半程度の計算時間で最適打順とその期待得点を得ることが確認された.

### 2 最適打順決定手法

D'Esopo and Lefkowitz モデルでは各打者について安打 (単打, 二塁打, 三塁打, 本塁打) を放つ確率, 四球を選ぶ確率, 凡打を放つ確率によって求められる状態遷移行列を定義し, 得点, アウトカウントおよび走者の初期状態を表す行列にそれらを打順に従って掛け合わせることで期待得点を算出する.

各行があるイニングにおける得点を表し, 各列が走者とアウトカウントの組合せを表す行列  $U$  を定義し, イニングにおいて  $n$  人の打撃が終了した時点の  $U$  を  $U_n$  と定義する.  $U_0$  はイニングの初期状態とする. 打者  $b$  の打撃結果により  $r$  点を挙げる場合の状態遷移行列を  $P_b^{(r)}$  とすると,  $U_n$  の第  $i$  行  $U_n|_i$  は

$$U_n|_i = \sum_{r=0}^4 U_{n-1}|_{i-r} P_b^{(r)}$$

で表される. 打順に従って行列  $P_b^{(r)}$  の乗算を繰り返して, 3アウトを意味する  $U_n$  の最右端の列の要素の和が1に近い基準値を超えた時点でイニングの終了と判定する.

あるイニングの先頭打者が  $i$  番打者で, その次のイニングの先頭打者が  $j$  番打者である確率を  $t_{ij}$ , その場合の期待得点を  $e_{ij}$  とすると,  $n$  回の攻撃が  $m$  番打者で始まる確率  $p_{m,n}$  ( $m = 1, 2, \dots, 9, n = 2, 3, \dots, 10$ ) は

$$p_{m,n} = \sum_{k=1}^9 p_{k,n-1} t_{km}$$

で表される ( $n=10$  は期待得点を算出する際に必要となる).  $a_{m,n}$  を  $n$  回の攻撃が  $m$  番打者で始まる場合

の初回からの合計得点とすると、これは

$$a_{m,n} = \frac{1}{p_{m,n}} \sum_{k=1}^9 p_{k,n-1} t_{km} (a_{k,n-1} + e_{km})$$

$$(m = 1, 2, \dots, 9, n = 2, 3, \dots, 10)$$

で表され、期待得点は  $\sum_{k=1}^9 p_{k,10} a_{k,10}$  で求められる。

本稿では、本最適打順決定方法について次に示す高速化を適用した。9人の打者が循環して打席に立つことを考慮すると、打者を  $b_n$  ( $n$ は打者インデックス) で表した場合に例えば打順  $b_1, b_2, \dots, b_9$  について求めた  $t_{ij}$  と  $e_{ij}$  は打順  $b_2, b_3, \dots, b_9, b_1$  の期待得点を求める際に再利用可能である。具体的には  $p_{1,1} = 1, p_{m,1} = 0$  ( $m = 2, 3, \dots, 9$ ) として求めていた  $p_{m,n}$  ( $n \geq 2$ ) を  $p_{1,1} = 0, p_{2,1} = 1$  に変更して求めることで  $t_{ij}$  と  $e_{ij}$  の再計算が不要になる。 $p_{m,n}$  と  $a_{m,n}$  の計算コストは  $t_{ij}$  と  $e_{ij}$  のそれに比べて小さく、 $t_{ij}$  と  $e_{ij}$  を共通して使用できる打順は9通りであるため、9倍程度の速度向上が期待される。また、個々の打順についての計算は独立に実行できるため、複数の計算機に割り当てることにより計算時間の短縮が期待できる。

### 3 実験結果

PC4台で構成したPCクラスタにD'Esopo and Lefkowitzモデルを実装し、マスタワーカ型の並列プログラミング環境を提供するNinf[2]を用いて並列化を行った。計算のカーネル部分となる行列ベクトル積には性能の自動チューニングが施された数値計算ライブラリのATLAS[3]を用いた。すべてのPCについてCPUはPentium 4 2.40GHz、メモリは512MB、OSにはRed Hat Linux 7.1(Kernel 2.4.7-10)を用い、コンパイラはgcc 2.96を使用し最適化オプションとして-O3を指定した。以上の環境で2005年に予想されるパ・リーグ6球団の最適打順とその期待得点を算出した。

表1に計算結果の再利用と並列化による効果を示す。実験の結果、計算結果の再利用により9.00倍の速度向上が図られた。また、Ninfを用いて並列化を行った結果、ほぼPC台数に比例した性能を得た。

新入団の外国人選手と2004年度に実績を残した選手より9人を選出し、各選手の状態遷移行列を定義して期待得点を算出した結果、攻撃力に関してオリックスはリーグ3位、楽天はリーグ最下位という

表 1: 計算結果の再利用と並列化による効果

実験環境 (6球団平均)	計算時間 (sec)	速度 向上比	台数 効果
再利用無し/PC1台	11,678	1.00	-
再利用有り/PC1台	1,298	9.00	1.00
再利用有り/PC4台	328	35.62	3.96

表 2: 各球団の最適打順による期待得点

球団	得点	球団	得点
日本ハム	6.39	西武	5.56
ソフトバンク	5.90	ロッテ	5.11
オリックス	5.82	楽天	4.87

結果が得られた(表2)。

さらに2005年度にオリックス、楽天それぞれに所属する全投手と、既存4球団の打者との最近3年間の対戦成績から状態遷移行列を定義し、各投手陣の既存4球団に対する期待失点を算出した。その結果2球団に対して楽天がオリックスよりも期待失点を低く抑える結果が得られた。その詳細については発表の際に述べる。

### 4 まとめ

本稿ではD'Esopo and Lefkowitzモデルを用いた最適打順決定手法の高速化手法について述べ、本手法によりパ・リーグ6球団の攻撃力、投手力の評価を行った。本高速化手法では最適打順計算において計算結果を再利用し、さらにNinfを用いて並列化を行うことにより約5分半で最適打順、期待得点が得られることを示した。その結果、新球団の攻撃力に関してはリーグ最下位であるが、投手力に関しては合併球団と比較して2球団相手に好成績を得た。

### 参考文献

- [1] Bukiet, B., Harold, E. and Palacios, J.: A Markov Chain Approach to Baseball, *Operations Research*, Vol. 45, No. 1, pp. 14-23 (1997).
- [2] Ninf: A Global Computing Infrastructure, <http://ninf.apgrid.org/>.
- [3] Whaley, R. C.: Automatically Tuned Linear Algebra Software(ATLAS), <http://math-atlas.sourceforge.net/>.