

最適ラインナップ構成に向けての打順の要求機能分析法

Required Function Analysis of Each Lineup Position toward Optimal Batting Order

角井勇哉 荒井幸代

Yuya Kakui Sachiyō Arai

千葉大学大学院工学研究科

Graduate School of engineering, Chiba University

It is important for organizations to realize Placing Right-Person in the Right-Position because of complication of tasks and shortage of human resources. Therefore, we suggest unifying data mining methods for Placing Right-Person in the Right-Position. We focus on baseball and propose the algorithm for composing optimal batting order. Experimental results indicated the usefulness of proposed algorithm based on analysis of required function of each lineup position and player's ability. In this paper, we introduce our method where required function of each lineup position is extracted explicitly for evaluative criterion.

1. 本研究の背景と目的

組織にとって適材適所の実現は喫緊の課題である。組織はタスク実行に必要な機能を担う部署群と、その機能を遂行する個体群からなるシステムである。適材適所は部署と個体の組み合わせ最適化であり、組織のパフォーマンスを最大化する問題である。組織に要請されるタスクが複雑化し、各部署に要求される機能が多種多様化する一方、人的資源の確保が困難な昨今の背景から、適材適所を定式化し、解を得る手続きを形式化する意義は大きい。

本研究では適材適所問題を、図1のように各部署に要求される機能の抽出と定量化(以後、要求機能分析)、各機能に対する個体の能力の定量的評価(以後、能力分析)、そして、両分析結果に基づく個体の部署への配置(以後、整合分析)の3つの部分問題に分割し、組織パフォーマンスの向上を目指す。本稿ではこのうち、要求機能分析に焦点を当てる。適材適所問題の関連研究には、部署に要求される機能を所与として、個体の部署への割り当て法や評価法を対象とした[菅原 08][橋浦 07]がある。要求機能分析については、アンケート調査などで得られた評価に基づいた経験的手法[水島 91][岡部 02]が主流であり、主観に大きく依存する。一方、タスクの実行結果や成績などの客観的データに基づいた解析は筆者が知る限り形式化されていない。形式化を困難にする理由として、各部署は他の部署との連携が必要で、部署間の相互作用を考慮しなければならない問題であることがあげられる。しかし、社会生活における多くの作業において自動化が進む状況で、相互作用を前提とした部署ごとの要求機能を陽に抽出することへの要請は大きい。

筆者らは野球を取り上げ、得点力を最大化するラインナップ構成問題を対象としてきた。本稿では、各打順で要求される機能を定量化する手法を紹介する。打順の要求機能分析においても前述した部署間の相互作用が大きく、打者の組み合わせ順序が得点力を左右するため、打順間の関係を考慮した分析が必要である。ラインナップ構成問題に関しては選手の能力評価法の提案[J. アルバート 04][Cover 77]やラインナップの得点力評価法の提案[Bukiet 97][廣津 04]があるが、各打順に要求される機能を定量化した研究はみあたらない。

連絡先: 角井勇哉, 千葉大学大学院工学研究科, 千葉市稲毛区弥生町 1-33, 043-251-1111(代表), y-kakui@graduate.chiba-u.jp

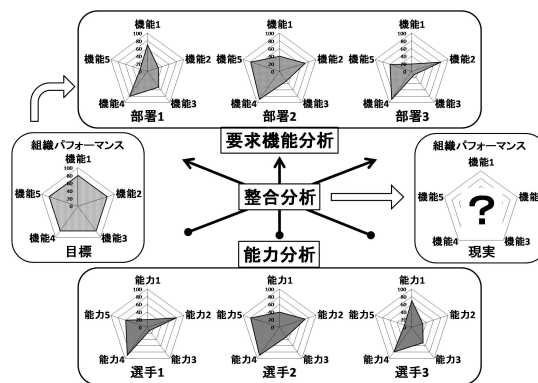


図1: 適材適所問題の構造

2. 問題設定

2.1 ラインナップ構成問題

本研究では、組織をラインナップ L 、タスクを得点、組織パフォーマンスを得点力、各部署を打順 k ($k = 1, 2, \dots, 9$), 各部署の総称を全打順, 各個体を選手 j , 各個体の総称を全選手, 打順 k に配置された選手を k 番打者とよぶ。

2.2 得点値算出モデル

提案手法の基盤理論として得点値算出モデルを用いる。本モデルは野球の試合状況を 25 個の状態集合 S へと離散化し^{*1}, マルコフ連鎖として期待得点値を算出するモデルである。代表的なモデルとして、特定の選手が単独でプレーした場合の期待得点値 (Expected Run distributions of Individual, RI 値) の算出モデルである OERA[Cover 77], SI[Bukiet 97], ラインナップでの期待得点値 (Expected Run distributions of Line-up, RL 値) の算出モデルである OERAL[廣津 04] がある。

得点値算出モデルにおける状態遷移は、選手の打撃確率 H と進塁規則によって記述される。ここでは、打撃確率 H を選手の打撃成績から非得点圏 (a) / 得点圏 (b) での { 凡打, 四死球, 単打, 二塁打, 三塁打, 本塁打 } の各確率として与え、また進塁規則は選手の盗塁数 T に応じて 2 種類を用意した。例えば、盗塁数 T が 10 以上である場合は「if 単打 then 2 ベース進塁」、10 未満である場合は「if 単打 then 1 ベース進塁」とした。

*1 アウトカウント 0 から 3, 各塁のランナーの有無の組み合わせ状態集合 $S = \{s_i | s_0, s_1, s_2, \dots, s_{24}\}$

3. 打順 k の要求機能分析法

「得点する」というタスクを直接得点に結びつく機能と間接的に得点に結びつく機能に分類する。前者を「塁に出る働き」と「走者を返す働き」とし、後者を「非得点圏 (a) / 得点圏 (b)」での打撃確率 H の相違」と「進塁規則」とした。これらを打順 k に要求される機能とし、評価項目 $\ell = (\ell \mid O^a, O^b, S^a, S^b, T)$ と記述する。各評価要素 ℓ は、非得点圏 (a) / 得点圏 (b) での出塁率 O 、長打率 S 、盗塁数 T であり、各々に対する選手の各能力を非得点圏 (a) / 得点圏 (b) での出塁力、長打力、走塁力とよぶ。

打順 k の要求機能分析は、

1. 打順 k に要求される各評価要素 ℓ の定量化
 \Rightarrow 要求値 D_k^ℓ の算出
 2. 打順 k が得意とする各評価要素 ℓ の定量化
 \Rightarrow 特長値 U_k^m ($m = D^{O^a}, D^{O^b}, D^{S^a}, D^{S^b}, D^T$) の算出
- の2段階から構成される。

3.1 要求値 D_k^ℓ 算出アルゴリズム

- A. 重回帰分析に基づく方法 (以後, MR 法)
- B. 状態の分析に基づく方法 (以後, SA 法)

の2つの方法を提案する。以下に、各方法の変数を定義し、算出法の概略を図2に示す。

- D_k^ℓ : 打順 k における各評価要素 ℓ の要求値 ($0 \leq D_k^\ell \leq 100$)
- $\sum_\ell D_k^\ell$: 打順 k の重要度 ($0 \leq \sum_\ell D_k^\ell \leq 500$)
- $V(s_i)$: s_i の状態価値 ($0 \leq V(s_i) \leq 1$)
- $d_{s_i}^\ell$: s_i における各評価要素 ℓ の要求値 ($0 \leq d_{s_i}^\ell \leq 100$)
- $p(s_i, k)$: s_i で k 番打者が打席に立つ確率 ($0 \leq p(s_i, k) \leq 1$)
- e^ℓ : 各評価要素 ℓ の相対的重要度

1. 1~9 番打者の打撃確率 H , 進塁規則 (盗塁数 T) にしたがって RL 値を算出
2. 目的変数を RL 値, 説明変数を 1~9 番打者の各評価要素 ℓ の値とした 45 項目からなる重回帰式 $\rightarrow D_k^\ell$

A) Multi linear Regression based Method (MR 法)

1. 状態 s_i に要請される選手の順位付け★1
 非得点圏 (a) / 得点圏 (b) の各々に対して実行
 - (a) 属性を出塁力, 長打力, 走塁力とし, 水準値は各々, 1~4, 1~4, 1~2 と設定
 - (b) 順位付けを行う選手群の打撃確率 H , 進塁規則 (盗塁数 T) を設定
 - (c) (b) で設定した選手の打撃確率 H を a_j ($j = 1, 2, \dots$) とし, 状態 s_i に要請される選手の順位付け法を適用
2. 手順 1 の結果を用いて s_i ($i = 1, 2, \dots, 24$) に対してファジィ・コンジョイント分析を適用
3. $d_{s_i}^\ell$ の算出
 - (a) $V(s_i)$ の決定★2
 - (b) 手順 2 で得る s_i の各属性寄与率を $V(s_i)$ に応じて線形変換 $\rightarrow d_{s_i}^\ell$
4. D_k^ℓ の算出
 - (a) $p(s_i, k)$ の決定★3
 - (b) 打順 k ($k = 1, 2, \dots, 9$) に対して, $\sum_i \{p(s_i, k) \times d_{s_i}^\ell\}$ を計算し, 正規化
 - (c) e^ℓ の算出★4
 - (d) (b) で得る各値を e^ℓ に応じて線形変換 $\rightarrow D_k^\ell$

B) States Analysis based Method (SA 法)

図 2: 要求値 D_k^ℓ 算出アルゴリズム

以下, 図 2 の SA 法における ★1 ~ ★4 について説明する。まず, ★1 の部分, 状態 s_i に要請される選手の順位付けについて説明する。以下に記号と変数を定義し, 状態 s_i に要請される選手の順位付け法の概略を図 3 に示す。

- 状態集合 $S = \{s_i \mid s_0, s_1, s_2, \dots, s_{24}\}$
- s : 現状態
- s' : 次状態
- a_j : 選手 j の打撃確率 H
- r : 総得点
- r' : 各状態での得点
- $Q(s_i, a_j)$: s_i において a_j を選択する価値
- エピソード: ノーアウト走者なし状態 s_1 からスリーアウト状態 s_0 までの 1 イニングの期間

1. すべての s_i, a_j に対して, $Q(s_i, a_j)$ を設定
2. 各エピソードに対して繰り返し:
 $s = s_1, r = 0$
 - (a) エピソードの各ステップに対して繰り返し:
 $Q(s_i, a_j)$ に従って, s_i での a_j を決定
 行動 a_j を取り, r', s' を観測
 $r + = r', s = s'$
 $s = s_0$ ならば繰り返しを終了
 エピソードに含まれる $\forall (s_i, a_j)$ に対して,
 $Q(s_i, a_j) + = r$
3. s_j ($i = 1, 2, \dots, 24$) での $Q(s_i, a_j)$ ($j = 1, 2, \dots$) を比較して, 降順に選手 j を順位付け

図 3: 状態 s_i に要請される選手の順位付け法

直交配列法により得られた 16 選手 (表 1) に対して適用した結果のうち, s_1 に要請される選手の順位を図 4 に示す。横軸が選手 j ($j = 1, 2, \dots, 16$), 縦軸が $p(s_1, a_j) = Q(s_1, a_j) / \sum_j Q(s_1, a_j)$ を示し, 各選手の縦軸上の数字はその選手の順位を表す。

表 1: 16 選手の各水準値

選手 j	出塁力	長打力	走塁力
1	1	1	1
2	2	2	2
3	3	3	1
4	4	4	2
5	1	2	2
6	2	1	1
7	3	4	2
8	4	3	1
9	1	3	1
10	2	4	2
11	3	1	1
12	4	2	2
13	1	4	2
14	2	3	1
15	3	2	2
16	4	1	1

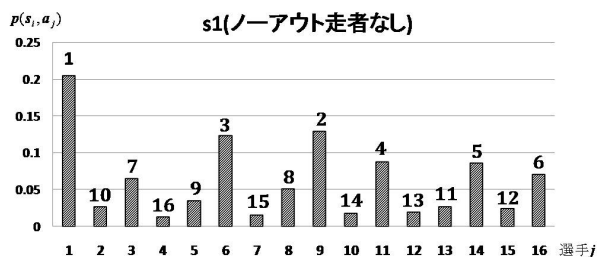


図 4: 状態 s_1 に要請される選手の順位

次に、★2の部分、 s_i の状態価値 $V(s_i)$ の決定について説明する。図5は2007年度セントラルリーグ(以後、セ・リーグ)の各チームの s_i ($i = 1, 2, \dots, 24$)からスリーアウト状態 s_0 までの期待得点値とその平均を示している。この図で各チームのグラフの形状が互いに類似することから、 s_8 ($\arg_{s_i} \max V(s_i)$)の状態価値 $V(s_8)=1$ とし、他の状態価値 $V(s_i)$ ($i = 1, 2, \dots, 24, i \neq 8$)は、 $V(s_8)$ に応じて線形変換し、状態価値 $V(s_i)$ ($i = 1, 2, \dots, 24$)を決定した。各チームの $V(s_i)$ とその平均を図6に示す。

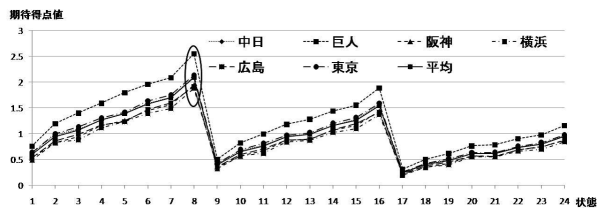


図5: 各チームの s_i から s_0 までの期待得点値とその平均

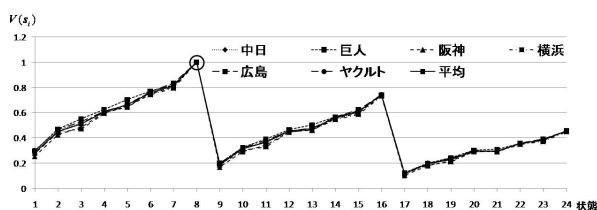


図6: 各チームの $V(s_i)$ とその平均

そして、★3の部分、 s_i で k 番打者が打席に立つ確率 $p(s_i, k)$ の決定について説明する。2007年度セ・リーグのRI値上位9名、下位9名、セ・リーグ各チームのラインナップを用いて $p(s_i, k)$ を算出し、選手の打撃確率 H と進塁規則(盗塁数 T)が大きく異なるラインナップにおいても $p(s_i, k)$ に有意な差はみられない。図7~図9に1, 2, 3番打者が s_i で打席に立つ確率 $P(s_i, k)$ ($k = 1, 2, 3$)を示す。

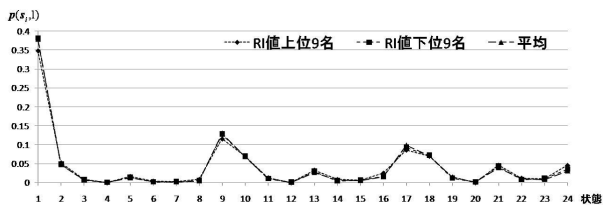


図7: 1番打者が状態 s_i で打席に立つ確率 $p(s_i, 1)$

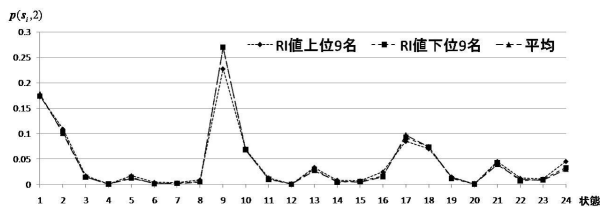


図8: 2番打者が状態 s_i で打席に立つ確率 $p(s_i, 2)$

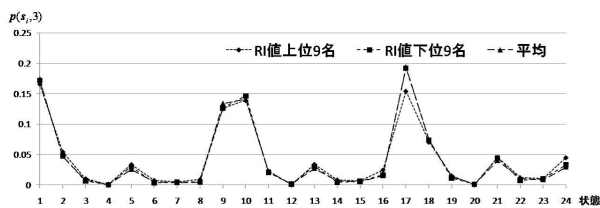


図9: 3番打者が状態 s_i で打席に立つ確率 $p(s_i, 3)$

最後に、★4の部分、各評価要素 l の相対的重要度 e^l の算出について説明する。重回帰分析によって得られるRI値に対する各評価要素 l の相対的重要度 e^l の算出方法を図10に示す。

また、この算出方法によって得られた評価項目 $l = (l^a, l^b, S^a, S^b, T)$ の各評価要素 l の相対的重要度 e^l を表2に示す。

1. 任意の回数 (a), (b) を繰り返し
 - (a) 選手 j の打撃確率 H と進塁規則(盗塁数 T)を設定
 - (b) RI値を算出
2. 評価要素 l の相対的重要度 e^l の算出
 目的変数をRI値,
 説明変数を選手 j の l の各値として重回帰分析 $\rightarrow e^l$

図10: 各評価要素 l の相対的重要度 e^l 算出法

表2: 各評価要素 l の相対的重要度 e^l

l	相対的重要度 e^l
O^a	1148
O^b	1070
S^a	499
S^b	691
T	502

3.2 特長値 U_k^m 算出アルゴリズム

3.1節で算出した要求値 $m=(m | D^{O^a}, D^{O^b}, D^{S^a}, D^{S^b}, D^T)$ の要素間の関係が打順ごとにより異なっていることに着目する。全打順の要求値 m を相対比較することによって、打順 k で必要とされる各評価要素 l の定量化(特長値 U_k^m の算出)を行う。分析には経営効率分析法(DEA)の1種であるSuperCCRモデル[Kaoru 02]を用いる。

DEAは事業体(政府、学校、軍隊)などの意思決定主体(Decision Making Unit: DMU)の効率性を相対的に評価する手法である。効率性は一般に入力と出力の商で与えられるが、入力と出力に複数の要素がある場合には、各要素の重要性をあらかじめパラメータが決定できない。これに対し、DEAはある対象の効率性を最大にするようにパラメータを選択して、主体の得意分野を評価する分析手法である。野球選手の能力の相対評価[廣津 04]にも用いられている。ここでは、DMUを全打順として適用する。算出手順の概略を図11に示す。

- u_k^m : 打順 k における m に対する重み
 - y_k^m : 打順 k の m の値(D_k^l)を標準化した値
 - $U_k^m = u_k^m y_k^m$: 打順 k の m の特長値
1. DMU: 全打順
 入力項目: 全打順に対して入力1.0
 出力項目: y_k^m ($m = D^{O^a}, D^{O^b}, D^{S^a}, D^{S^b}, D^T$)
 2. 打順 k ($k = 1, 2, \dots, 9$)に対して繰り返し:
 - (a) 以下のLP問題の解を算出
 分析する打順を α とすると
 目的関数: $\max \theta_\alpha = \sum_m u_\alpha^m y_\alpha^m$
 制約式: $\sum_m u_k^m y_k^m \leq 1$ ($k \neq \alpha, k = 1, 2, \dots, 9$)
 $u_k^m \geq 0$
- 算出した仮想出力値 $u_k^m y_k^m$ を特長値 U_k^m へ

図11: 特長値 U_k^m 算出アルゴリズム

4. 実験と考察

提案手法の有用性を 2007 年度セ・リーグ各チームのスターティングメンバー 48 名 (投手を除く 8 名 × 6 チーム) のデータを用い、検証する。MR 法, SA 法で算出した全打順の要求値 D_k^ℓ の 5 角形レーダを図 12, 図 13 に示す。中央の数字は打順 k の重要度 $\sum_\ell D_k^\ell$ を表す。また、打順 k の評価要素 ℓ の特長値 U_k^m を表 3 に示す。上段が MR 法, 下段が SA 法で算出した D_k^ℓ を基に算出した特長値 U_k^m である。

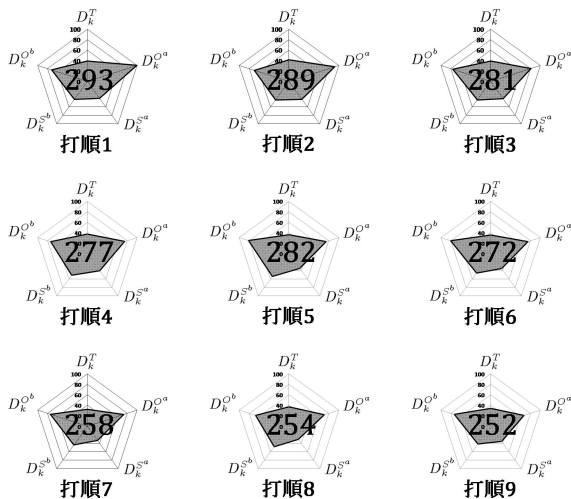


図 12: 要求値 D_k^ℓ by MR 法

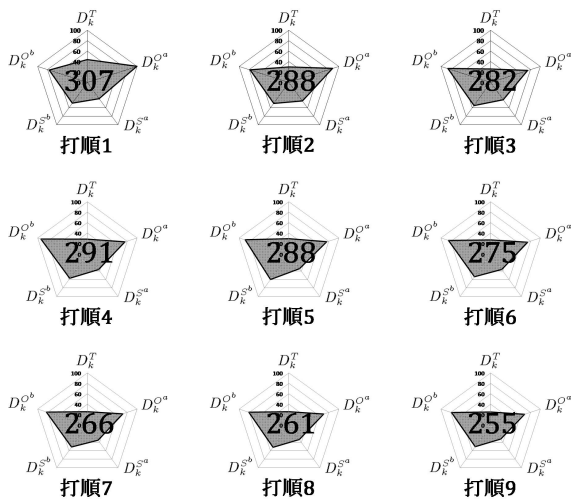


図 13: 要求値 D_k^ℓ by SA 法

表 3: 特長値 U_k^m (上段 MR 法, 下段 SA 法)

$m \backslash k$	1	2	3	4	5	6	7	8	9
D^{O^a}	1.2	0.1	0.0	0.0	0.0	0.0	0.2	0.0	0.0
D^{O^b}	0.0	0.2	0.0	0.0	0.0	0.1	0.0	0.0	0.0
D^{S^a}	0.2	0.0	0.4	0.0	0.8	0.7	0.4	0.0	0.3
D^{S^b}	0.0	0.0	0.0	1.3	0.0	0.1	0.2	0.0	0.1
D^T	0.0	1.2	0.7	0.1	0.0	0.2	0.2	0.1	0.0
D^{O^a}	0.0	0.0	0.0	0.5	0.9	0.0	0.0	0.3	0.0
D^{O^b}	0.0	0.0	0.4	0.0	1.3	0.3	0.0	0.2	0.1
D^T	0.0	1.3	0.1	0.0	0.0	0.0	0.0	0.4	0.0
D^{S^a}	4.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

提案手法の問題点:

重回帰分析に基づく MR 法では、全打順を一括して決定することを前提としている。しかし、 α 番打者が固定された場合に残りの打順の要求値 D_k^ℓ ($k = 1, 2, \dots, 9, k \neq \alpha$) に影響を及ぼす。

例えば、1 番打者を走塁力のある選手 A で固定する場合と長打力のある選手 B で固定する場合とでは 2 ~ 9 番打者の要求値 D_k^ℓ ($k = 2, 3, \dots, 9$) は異なるはずである。

一方、SA 法も α 番打者が固定された場合には、 $V(s_i)$ と $p(s_i, k)$ を再計算しなければならない。これは、 $V(s_i)$ と $p(s_i, k)$ がラインナップを構成する全選手の組み合わせの平均値であることから、選手の選び方やデータの数が算出される D_k^ℓ ($k = 1, 2, \dots, 9$) に与える影響を考える必要があることを示している。

また SA 法の各手順において、状態 s_i に必要とされる選手の順位付け法に対しては、各状態で期待得点値を高める選手に報酬を与えるメカニズムのため、上位の選手は順位を明確に示すことができるが、下位の選手に対しては順位が曖昧になること、 s_i の状態価値 $V(s_i)$ の決定に対しては、 $V(s_i)$ を s_i からスリーアウト状態 s_0 までの期待得点値から決定したが、他にも s_i での期待得点値から決定することなどが考えられることがそれぞれの問題として挙げられる。

また、特長値 U_k^m は打順 k の特長値として算出したが、全打順を 1 つの DMU としたときの全打順の特長値との関係からの評価が必要である。

5. 結論

本稿では最適ラインナップ構成に向けて、要求値 D_k^ℓ と特長値 U_k^m の算出の 2 段階から成る打順 k の要求機能分析法を提案した。要求値 D_k^ℓ の算出においては、重回帰分析に基づく MR 法と各状態の分析に基づく SA 法の 2 つの手法を、特長値 U_k^m の算出においては、DEA を用いた分析法を紹介した。

本稿では扱っていないが、選手の能力分析、整合分析においても研究を進めている。現時点では、MR 法で算出した要求値 D_k^ℓ を用いた整合分析により得られるラインナップの方が、SA 法より高い得点力を示している。このことを踏まえ SA 法で算出される D_k^ℓ を MR 法に近づけるために、本稿で指摘した SA 法の各手順での問題点を解決していくことを今後の課題として挙げる。

参考文献

[菅原 08] 菅原 俊治: マルチエージェントパラダイムとネットワーク, 人工知能学会誌, 23 巻, 5 号 (2008)

[橋浦 07] 橋浦 弘明: ソフトウェア開発グループ演習のためのチーム編成の最適化支援, メディア教育研究, 第 3 巻, 2 号 (2007)

[水島 91] 水島 温夫: R & D における人材・組織管理への行動科学的アプローチ, 学術大会後援要旨集, 6, 19-26 (1991)

[岡部 02] 岡部 一光: 職務と人材のミスマッチ解消のためのデータマイニングに基づくコンピテンシーモデルの構築, 日本データベース学会, Vol.1, No.1 (2002)

[J. アルバート 04] J. アルバート, J. ベネット (後藤寿彦監修, 加藤貴昭訳): メジャーリーグの数理科学 <下>, シュプリンガー数学リーディングス, 第 2 巻 (2004)

[Cover 77] Cover, T.M. and Keilers, C.W.: An offensive earned-run average for baseball, Operations Research, 25, 729-740 (1977)

[Bukiet 97] Bukiet B, Harold E.R. and Palacios J.L.: A Markov Chain Approach to Baseball, Operations Research, 45, No.1, 14-23 (1997)

[廣津 04] 廣津 信義, 宮地 力: 野球チームのラインナップ選定のための数理的一手法-日本代表チームの選定を例として-, オペレーションズ・リサーチ, 49, 380-389 (2004)

[Kaoru 02] Kaoru Tone: A slacks-based measure of super-efficiency in data envelopment analysis, European Journal of Operational Research, 143, 32-41 (2002)