知能と複雑系 113－7
(1998. 7. 24)

例からの学習による野球監督の先発選手決定問題

山口 陽三*, 乾 伸雄**, 野瀬 隆**, ビピン インドルキア**, 小谷 善行**
* 東京農工大学大学院工学研究科電子情報工学専攻情報工学講座
** 東京農工大学工学部情報コミュニケーション工学科

〒184-8588 東京都小金井市中町 2-24-16
東京農工大学工学部情報コミュニケーション工学科
TEL: 042-387-4608　E-mail: yozo@cc.tuat.ac.jp

あらまし 近年、大規模なデータから計算機に自律的に学習や知識獲得を行わせる研究が盛んである。本稿では、エキスパートシステムの実用例の一つとして野球の監督を題材とし、例からの学習によってその実現を試みた。我々は野球の選手起用は、選手の守備・打撃・走塁の能力を考慮することによって行われると考え、それぞれの能力による選手個々人の「貢献度」を定義し、問題の定式化を行った。そして貢献度を予測する重回帰式を求め、野球シミュレータ上で実験を行った。その結果、我々の方法で先発選手を決定した監督は、ランダムに先発選手を決定する監督に対して 57 % の勝率を収めた。

キーワード　エキスパートシステム, 野球, 学習

The Starting Member Decision Problem of Baseball by Learning from Instances

Yozo Yamaguchi*, Nobuo Inui**, Takashi Nose**, Bipin Indurkhya**, Yoshiyuki Kotani**
* Department of Computer Science, Graduate School of Technology,
Tokyo University of Agriculture and Technology
** Department of Information and Communication, Faculty of Technology,
Tokyo University of Agriculture and Technology

2-24-16 Naka-machi, Koganei, Tokyo, 〒184-8588
Department of Information and Communication, Faculty of Technology,
Tokyo University of Agriculture and Technology
TEL: 042-387-4608　E-mail: yozo@cc.tuat.ac.jp

Abstract Research work dealing with knowledge acquisition as well as learning from large databases has been popular nowadays. This paper proposes a practical baseball manager expert system to select the starting members by learning from instances. The system calculates the estimation of defense, batting and running abilities and finally selects appropriate members for each game. We then defined each member's expected activity value called the degree of achievement. By using the multiple regression function of statistical analysis, we tested the degree of achievement on our baseball simulator. The result concluded that the baseball manager by this method recorded 57 % of winning against a baseball manager system which randomly selects the starting members.

key words　expert system, baseball, learning

1　はじめに

人間がある問題を解決しようとする場合に、あるときは規則や公理に基づいて推論し、またあるときは過去の経験を生かして最適と思われる解を選択したりすると思われる。置かれた状況によって人間は様々な方法によって問題解決を行っていると思われるが、それらの作業は一般的に、複数の要素が複雑に絡み合って行われており、その手順や規則を実際に分析することは難しい。なぜなら、人間自身もある問題解決を行う場合の手順を自分たちでもよくわかっていないからである。

一方で、計算機に大量のデータを与え、その中から意味のある知識を抽出しようとする研究や、与えられたデータから計算機に自律的に学習・知識獲得を行わせようとする研究が近年盛んになされている（文献 [1][2][3] 等参照)。その背景には計算機のメモリや CPU の性能の向上により、大規模なデータを実時間で処理できる環境が整いつつあることが挙げられよう。また、エキスパートシステムの研究分野においても人間の手順を計算機に忠実に行わせるだけでなく、計算機が自律的にその手順を学習していく方法は、エキスパートシステムの発展の一つの可能性として有効であると考えられる。

本稿では、エキスパートシステムの実用例の一つとして野球の監督を題材とし、野球監督のエキスパートシステムの設計及び実現を試みる。野球には、その試合で起きた出来事を忠実に紙面に表現するスコアブックという記録媒体があり、それによって表現されている情報は数多い。そこで我々は、野球の試合の情報を計算機内で表現するデータ構造を提案し（文献 [5])、その形式によって野球の試合の保存を行い、計算機による学習を行った。具体的には、野球の選手起用を考える上で必要となってくる、期待される「選手の活躍の度合」を重回帰式で表現し、そして最適と思われる個々の選手の組合せを出力することを考えた。

以下、第 2 節では問題の定式化を述べ、第 3 節では、期待される、「選手が活躍する度合」の定義を行い、第 4 節でその活躍の度合を予測する重回帰式を求める。そして第 5 節では実験結果を示し、第 6 節で考察を述べる。

2　問題の定式化

野球というゲームには、攻撃と守備という二つの立場がある。それらを交互に担当していくことで試合が進んでいくわけだが、試合中の作戦に関しては、監督は自軍が現在どちらの立場に置かれているかを認識し、その上でその状況において最適と思われる作戦を考える。ところが、監督の試合前の作業である先発選手決定問題を考える場合、攻撃と守備、どちらの立場に立った場合にも試合に勝つという目的を果たす可能性が高いと思われる解を、試合が始まる前に決定しなければならない。その際には攻撃と守備の両方の立場に立つことを考えなければならないので、本稿では自軍の選手の能力として「守備に対する能力」「打撃に対する能力」「走塁に対する能力」の三つを考える。各選手のこれらの能力を考えた上で、また、チーム全体としてのこれらの能力も考えて、その試合において最適と思われる先発選手を決定する必要がある。ただ、投手は試合の勝敗を左右する、とても大きな役割を担うので、その選抜については他のポジションとは異なった処理を行う。本研究では、まず投手をできる全選手の中で、投球に関して期待される活躍の度合が最も高い選手を先発投手として選び、続いて残りの八つのポジションについての先発選手を選抜するという手順をとっている。

上記の「守備に対する能力」「打撃に対する能力」「走塁に対する能力」を、ここでは「守備貢献度」「打撃貢献度」「走塁貢献度」と名づけることとする。それらの定義は第 3 節で行う。

2.1 先発選手選抜問題

この問題は、10 ～ 20 人くらいのベンチ入り選手の中から、先発として出場する 9 人を選抜する問題である。単純に考えれば上記の三つの貢献度が高い選手から 9 人を選抜すればよいが、この問題の場合はポジションの制約がある。つまり、一つのポジションには一人の選手しか守らせることはできないし、一人の選手に複数のポジションを守らせることはできない。

また、ポジションによってその難しさのようなものは変わってくる。単純に考えれば、難しいポジションに守備貢献度の高い選手を配置するのがよい。そこで、ポジション別の守備機会数を例から求めた。ここで言う守備機会は、野球規則においての守備機会とは多少異なる。ここでは、アウトに関与した回数と失策数、そしてアウトに関与しなくても打球を処理した回数の和を守備機会数としてカウントした。30 試合分の例（文献 [9]）からの結果を表 1 に示す。

結局この問題は次のように考えられる。

$$
\begin{aligned}
I &= \{player_{\text{投手}}, player_{\text{捕手}}, \cdots, player_{\text{右翼手}}\} \\
X &= w_1 \times \sum_{pos=\text{投手}, pl \in I}^{pos=\text{右翼手}} \text{守備貢献度}_{pl} \times \text{ポジション別守備機会割合}_{pos} \\
&+ w_2 \times \sum_{pl \in I} \text{打撃貢献度}_{pl} + w_3 \times \sum_{pl \in I} \text{走塁貢献度}_{pl}
\end{aligned}
$$

のときに可能な I の組合せの中で最大の X を与える 9 人の選手の組を選び出す問題

なお、各貢献度はその平均・分散が異なるので、平均 50、分散 10 の偏差値得点に標準化した。

2.2 打順決定問題

先発選手選抜問題によって 9 人の選手の組が選び出されたら、その 9 人の選手をある順番（打順）に並べる必要がある。これについては制約はなく、どの選手がどの打順に入ってもいいが、攻撃の最大の目標である、「多く得点すること」を果たす可能性が高い順番に並べるのがよい。そこで「打撃貢献度」「走塁貢献度」の、全体に対する打順別の割合を例から求めた。つまり、実際の試合においてどの打順にどの程度の貢献度を持つ選手が並べられているか、を調査した。30 試合分の例（文献 [9]）からの集計結果を表 2 に示す。打撃貢献度の定義については 第 3 節で述べる。また、第 3 節で述べるが、走塁の貢献度を厳密に表現できるような値が野球の一般の記録の上ではあまりないので、ここでは盗塁数を対象とした。

結局この問題は次のように考えられる。

$$
\begin{aligned}
J &= \{player_{1\,\text{番}}, player_{2\,\text{番}}, \cdots, player_{9\,\text{番}}\} \\
Y &= w_4 \times \sum_{dajun=1\,\text{番}, pl \in J}^{dajun=9\,\text{番}} \text{打撃貢献度}_{pl} \times \text{打順別貢献度}_{dajun} \\
&+ w_5 \times \sum_{dajun=1\,\text{番}, pl \in J}^{dajun=9\,\text{番}} \text{走塁貢献度}_{pl} \times \text{打順別盗塁割合}_{dajun}
\end{aligned}
$$

のときに可能な J の組合せの中で最大の Y を与える 9 人の選手の順番を決める問題

なお、各貢献度はその平均・分散が異なるので、平均 50、分散 10 の偏差値得点に標準化した。

表 1: ポジション別の守備機会数 (総和)

ポジション	守備機会数 (%)	
投手	236	(8.4 %)
捕手	396	(14.1 %)
一塁手	655	(23.3 %)
二塁手	298	(10.6 %)
三塁手	194	(6.9 %)
遊撃手	351	(12.5 %)
左翼手	204	(7.3 %)
中堅手	269	(9.6 %)
右翼手	207	(7.4 %)

表 2: 打順別の打撃貢献度と盗塁数 (総和)

打順	打撃貢献度 (%)	盗塁数 (%)
1	14.951 (11.3 %)	31(22.1 %)
2	14.717 (11.1 %)	29(20.7 %)
3	14.928 (11.2 %)	22(15.7 %)
4	15.243 (11.5 %)	16(11.4 %)
5	14.787 (11.1 %)	4(2.1 %)
6	15.239 (11.5 %)	6(4.3 %)
7	14.275 (10.7 %)	13(9.3 %)
8	14.508 (10.9 %)	6(4.3 %)
9	14.226 (10.7 %)	13(9.3 %)

3 貢献度の定義

本節では各種貢献度の定義を行う。ここで言う貢献度は、ある選手が対象とする試合においてどの程度の活躍をすることができるか、を表す値で、「守備貢献度」「打撃貢献度」「走塁貢献度」「投球貢献度」の四つを定義する。

3.1 守備貢献度

選手個人の守備率と呼ばれる値をそのまま適用する。守備率とは、守備についている選手がアウトをとることができたと考えられる機会の中で失策を犯す割合を、1 から引いたものである。[0, 1] の範囲の値をとり、高い方が失策の確率が低い、つまり守備に関して貢献が期待されると考えられる。

3.2 走塁貢献度

野球の一般的な記録においては、走塁に着目した項目は少ない。盗塁、得点、残塁等が挙げられるが (その定義については文献 [10] を参照)、得点や残塁といった項目は必ずしも選手の走塁の能力を表現しているとは言えない。後続の打者の能力によるところが大きいからである。そこで本稿では、盗塁を取り上げ、盗塁成功率をもって走塁貢献度とすることにした。一般に盗塁の成功率が高い選手は、足が速かったり走塁が上手であることが多い。そういう意味でも、盗塁成功率はある程度走塁貢献度を表すにふさわしいと考えることができよう。これも [0, 1] の範囲の値をとり、高い方が貢献が期待されると考えられる。

3.3 打撃貢献度

選手が 1 回の打席での打撃でチームにとってどれだけの貢献をしたかを表す。1 回の打撃が終了すれば、アウトカウントや走者位置といった、野球の試合の状況に変化が起こる。ここではアウトカウントと走者位置の組を作ってそれぞれの場合について、その状況から攻撃終了までの間に k 点以上入る確率 (得点確率) を求めておき、得点確率を上昇させる状況に遷移させるような打撃にプラスの評価を与え、得点確率を下降させる状況に遷移させるような打撃にマイナスの評価を与える。得点確率の表の一部を表 3 に示す。

表 3: 得点確率表

状況	$k=1$	$k=2$	$\cdots$
0 死 0 塁	0.33	0.18	$\cdots$
0 死 1 塁	0.47	0.24	$\cdots$
0 死 1.2 塁	0.63	0.40	$\cdots$
⋮	⋮	⋮	⋮
1 死 満塁	0.76	0.46	$\cdots$
2 死 満塁	0.46	0.26	$\cdots$

ある打者の 1 回の打撃によって k 点入った場合の打撃貢献度は

$$\text{打撃貢献度}(\text{状況}_{\text{打撃前}}, \text{状況}_{\text{打撃後}}) = \begin{cases} 1.0 - \text{得点確率}(k, \text{状況}_{\text{打撃前}}) & (k \geq 1) \\ \text{得点確率}(1, \text{状況}_{\text{打撃後}}) - \text{得点確率}(1, \text{状況}_{\text{打撃前}}) & (k = 0) \end{cases}$$

と定義した。つまり打者の打撃によって得点が入らなかった場合でも 1 点以上入る確率（純粋に、得点できる確率）を上昇させた打撃についてはプラスの評価が与えられる。

この計算によると打撃貢献度は、$(-1, 1)$ の範囲の値をとる。そこで便宜上、これを $(0, 1)$ の範囲に線形写像した。

3.4 投球貢献度

投手がある試合においてどれだけチームに貢献したかを計るのは難しい。野球規則によって投手に与えられる記録には「勝利投手」「敗戦投手」「セーブ」といったものが挙げられるが、これらはある規則に従って便宜上つけられるもので、その投球内容の評価にまで至っていないことが多い。

その中でも防御率と呼ばれる値がある。この値は、ある投手の責任においての失点が、9 イニングあたりどれだけあるかを示すもので、比較的投手の力量や投球内容を示す値にはなる。イニングは一つのアウトも 1/3 イニングと数えられるので、防御率は平均の投球回数（イニング）が多い投手についても少ない投手についても平等の評価を行うことができる値でもある。また、野球の攻撃における最大の目的が「多くの得点を記録すること」にあるならば守備における最大の目的は「失点を最小限にとどめること」であり、失点をもとに計算される防御率という値は、投手のチームに対する貢献度を表すにふさわしいものであると考えられる。よって、ここでは投手の防御率をもって貢献度とする。防御率は投手の責任による失点の割合であるから、少ない方がチームにとっての貢献度が高いと考えられる。

4 貢献度を求める重回帰式

ある選手のある試合において、期待される活躍の度合（貢献度）を第 3 節で定義した。本節では、個々の選手の持つデータから、この貢献度を予測する重回帰式を求める。

4.1 守備貢献度・走塁貢献度

第 3 節で述べた通り、次のようにする

$$
\begin{aligned}
\text{守備貢献度} &= \text{守備率} \\
\text{走塁貢献度} &= \text{盗塁成功率}
\end{aligned}
$$

4.2 打撃貢献度

守備・走塁とは異なり、打撃にはその様々な側面を表すいろいろな値が考えられる。ここでは次の説明変数を用意し、これらが絡み合って貢献度の予測が行えると考えた。

「出塁率」「長打率」「犠打成功率」「盗塁成功率」「最近 3 試合の打率」

出塁率・長打率の定義については文献 [10] にその説明を譲る。犠打成功率・盗塁成功率は、いずれもそれが成功した回数を、企画した回数で割った値である。最近の打率は、ここでは仮に 3 試合とし、該当選手が試合に出場していなかった場合には、その選手の従来の打率が使われるとした。これによって最近の打撃の調子といったものを考慮に入れられる。ただ、1 試合前の打率と 3 試合前の打率とでは及ぼす影響が異なると考え、ここでは試合が遠ざかるごとに影響が弱まる、反比例のモデルを考え、反比例の関数を重みとしてかけた上で平均値をとることとした。なお、上記の説明変数は、長打率を除いては $[0,1]$ の範囲の値をとる。長打率は理論的には $[0,4]$ の範囲の値をとりうるが、$[0,1]$ の範囲に落ち着くことが圧倒的に多い。

30 試合分のデータ（文献 [9]）、のべ 502 人のデータから、表 4 の偏回帰係数を得た (重相関係数 0.355、寄与率 0.126、分散比 14.312)。

これよりある選手の打撃貢献度を予測する重回帰式は次のようになる。

$$
\begin{aligned}
\text{打撃貢献度} = 0.0958 &\times \text{出塁率} + 0.1206 \times \text{長打率} + 0.0240 \times \text{犠打成功率} \\
&- 0.0057 \times \text{盗塁成功率} - 0.0402 \times \text{最近の打率} + 0.4177
\end{aligned}
$$

表 4: 打撃貢献度の説明変数と偏回帰係数

説明変数	偏回帰係数
出塁率	0.0958
長打率	0.1206
犠打成功率	0.0240
盗塁成功率	−0.0057
最近の打率	−0.0402
定数	0.4177

表 5: 投球貢献度の説明変数と偏回帰係数

説明変数	偏回帰係数
被打率	45.9581
与四死球率	4.6232
奪三振率	−2.5808
平均イニング数	0.0808
最近の防御率	0.0694
最近の投球数	−0.0226
定数	−8.6395

4.3 投球貢献度

ある投手のある試合での投球の予測は、打撃同様に難しい。様々な要素の影響で実力通りの力が発揮されないのは打撃もそうだが、打撃は 9 人に 1 回の割合で打順が回ってくるのに対して投

球はその投手の任務中は 1 球 1 球その投手が投げなければならず、その一人の投手の調子や実力が試合に及ぼす影響は大きい。また、1 球を投げるごとに疲労が蓄積されると考えられ、以前の試合での疲労が影響する場合もある。それらのことを考えた上で、説明変数として最近 3 試合での防御率に加え、最近 4 日間での投球数も加えた。説明変数として次のものを用意した。

「被打率」「1 イニングあたりの与四死球数」「1 イニングあたりの奪三振数」
「1 試合あたりの平均イニング数」
「最近 3 試合での防御率」「最近 4 日間の試合での投球数」

被打率は、投手の、相手打者と対したときの相手打者の打率で [0, 1] の範囲の値をとる。打率の定義については 文献 [10] に譲る。与四死球は、投手が相手打者に与える四死球、奪三振は投手が相手打者から奪う三振で、いずれも 0 以上の数値をとりうる（奪三振数が獲得アウト数を超えることはめったにないが、ありえなくない）。最近 3 試合での防御率、最近 4 日間の投球数については打撃貢献度の場合と同様に、試合が遠ざかるごとに影響が弱まる、反比例のモデルを考え、反比例の関数を重みとしてかけた上で平均値をとった。防御率については該当投手が試合に出場していなかった場合には、その投手の従来の防御率が使われるとし、投球数については該当投手が試合に出場していなかった場合には、0 球の投球を行ったとした。

60 試合分のデータ（文献 [8] [9]）、のべ 202 人のデータから、表 5 の偏回帰係数を得た (重相関係数 0.343、寄与率 0.117、分散比 4.322)。

これよりある投手の投球貢献度を予測する重回帰式は次のようになる。

$$
\begin{aligned}
\text{投球貢献度} = 45.9581 &\times \text{被打率} + 4.6232 \times \text{与四死球率} - 2.5808 \times \text{奪三振率} \\
&+ 0.0808 \times \text{平均イニング数} + 0.0694 \times \text{最近の防御率} \\
&- 0.0226 \times \text{最近の投球数} - 8.6395
\end{aligned}
$$

5 実験

実験は、自作の野球シミュレータ上で行った。そのシミュレータ上では、東京農工大学等の 8 大学の実データから作られたキャラクターが動く。各チームのベンチ入り選手数は異なるが、9～18 人である。開発環境は UNIX 上の C 言語である。

5.1 実験条件

今回の実験では、本稿で提案した手法で先発選手決定を行う監督とランダムに先発選手決定を行う監督とを対戦させた。ランダムな手法と言っても、投手を含めて個々の選手が守ることのできるポジション以外は守らない、という制約はつけた。その制約の中でどの選手を選抜するか、そして選抜した選手をどう打順として並べるか、というところの決定はランダムに行うこととした。

今回の実験では、第 2 節における重みは、$w_1 : w_2 : w_3 = 5 : 5 : 1, w_4 : w_5 = 2 : 1$ として実験を行った。また、対戦においては次の条件を設けた。

- 同じチームの同じデータを双方に与え、同じチームが対戦する。
- 先発選手を決定したあとは、試合中の選手交代は双方行わないとする。

- 攻撃の作戦は、両監督ともただ打者に打たせる作戦だけを使うとする。
- 延長戦は無制限で、決着がつくまで試合を行うとする。

5.2 実験結果

同じチーム同士を 200 試合ずつ対戦させるセットを、6 セット (6 チームについて) 行った。1200 試合の対戦の結果、次の結果が出た。

監督	勝利数	勝率 (%)
本稿で提案する手法	685	57.1
ランダムな手法	515	42.9

6 考察

重回帰分析によって選手の貢献度を予測することによって先発選手の決定を行う手法が、ランダムに先発選手を決定する手法よりは有効であった。今後、多くの試合を学習させていくことでより強い監督を実現できると思われる。また、今後の課題として、「守備」「打撃」「走塁」の最適なバランスを学習によって決めていくこと等が挙げられる。そして今回の手法を試合前の先発選手決定問題だけでなく、試合中においての選手交代問題にも適用できるかどうか、その可能性も探っていきたい。

参考文献

[1] 河野 浩之：「データベースからの知識発見の現状と動向」, 人工知能学会誌 Vol.12, No.4 (1997-7), pp.3-10, 1997

[2] 喜連川 優：「データマイニングにおける相関ルール抽出技法」, 人工知能学会誌 Vol.12, No.4 (1997-7), pp.19-26, 1997

[3] 沼尾 雅之, 清水 周一：「流通業におけるデータマイニング, 人工知能学会誌 Vol.12, No.4 (1997-7), pp.34-41, 1997

[4] 毛利 隆夫, 田中 英彦：「記憶に基づく推論と多変量解析との比較」, 人工知能学会第 32 回知識ベースシステム研究会資料, pp.56-63, 1995

[5] 山口 陽三, 乾 伸雄, 野瀬 隆, 小谷 善行：「野球のスコアブックのデータ構造の提案」, 人工知能学会第 8 回合同研究会資料, pp.77-82, 1997

[6] 有馬 哲, 石村 貞夫：「多変量解析のはなし」, 東京図書, 1987

[7] 東京大学教養学部統計学教室編：「統計学入門」, 東京大学出版会, 1991

[8] 東京新大学野球連盟 2 部 平成 9 年秋季リーグ戦公式記録, 1997

[9] 東京新大学野球連盟 2 部 平成 10 年春季リーグ戦公式記録, 1998

[10] 日本野球連盟他編纂：「野球規則」, 1994