DD
2002
68

発見的探索手法への階乗進数表現の適用

平 成 14 年 度

渡 瀬 一 紀

概 要

従来、組織的に解くことが難しいと思われていた問題に対しても、コンピュータの計算能力を利用したさまざまな発見的な探索方法が開発されてきている。遺伝的アルゴリズムと呼ばれる、生物の進化を模した探索方法もそのような方法のひとつである。遺伝的アルゴリズムを使って近似解を求めようとする場合には、解候補を染色体と呼ばれる形式で表現する必要がある。解こうとする問題によって染色体表現は異なってくるが、従来は遺伝子として0と1を使うものや、順列を使う方法などが多く研究されてきた。しかし、染色体表現によっては探索能力に差が見られることも多く、どのような染色体表現を採用するかは重要な問題である。

本研究では、階乗進数を使って染色体を表現する方法を提案している。n 桁の階乗進数を使って $n!$ 個の数を表現することができる。したがって、まず階乗進数表現と順列を対応させることが考えられる。さらに、階乗進数を要素間の結合を表すものと解釈することもできる。このため階乗進数表現を集合分割問題へ適用することも可能である。階乗進数は順列と一対一に対応し、また交叉による致死遺伝子を抑制することも可能である。また、ある種の順序制約を容易に表現できる。また、集合分割問題においては、群の数を固定することなく探索を行うことも、逆に群の数を固定して探索を行うこともできる。このように階乗進数による染色体表現は、好ましいいくつかの性質を備えているので、階乗進数表現の有効性を評価することを本研究の目的とした。

本研究ではまず、解が順列として表現できる巡回セールスマン問題やフローショップスケジューリング問題などに適用して、階乗進数表現の有効性を検討した。特にフローショップスケジューリング問題において、従来の方法に比べて高い探索能力を持つことを、数値実験を通して明らかにした。

また本研究では、集合分割の応用として、分割表からの規則抽出を試みた。説明変数や階級の数が多くなると、作成し得る分割表の数は膨大なものとなる。そこで本研究では、分割表の探索に遺伝的アルゴリズムを利用したが、様々な分割表を表現する方法として階乗進数表現を使った。そして古典的な問題であるあやめの識別問題に適用し、有効な少数の規則を抽出することができた。

さらに、多変数でありかつあいまいな回答を含む減価償却方法選択問題へ適用し、提案方法の有効性について検討した。ここでは説明変数として26変数が設定されている。すべての変数を含む分割表を考えることは現実的ではないので、変数増減法を使って分割表の選択を行った。このような減価償却方法選択問題へ適用した場合にも、会計的知見と一致する少数の規則を得ることができた。

目 次

第1章　発見的探索と階乗進数 **4**

1.1 発見的探索方法 4

1.2 階乗進数表現と本研究の目的 6

第2章　階乗進数による順列の表現 **8**

2.1 階乗進数と順列の対応 8

2.2 巡回セールスマン問題への適用 10

2.3 フローショップスケジューリング問題への適用 13

2.3.1 NEH法 13

2.3.2 遺伝的アルゴリズムへの適用の例示 14

2.3.3 遺伝的アルゴリズムの数値実験 15

2.3.4 テストデータ 16

2.3.5 遺伝的アルゴリズムにおける数値実験結果 17

2.3.6 情報量を用いた結果の検討 18

2.3.7 他の探索方法との比較 21

2.3.8 結論 22

2.4 順序制約を持つフローショップスケジューリングへの応用 22

2.4.1 階乗進数による制約の表現 23

2.4.2 制約つき順列探索の数値実験 25

2.4.3 結論 28

2.5 因果モデリングへの応用 28

2.5.1 因果モデリングにおける課題 29

2.5.2 階乗進数による順序関係の表現 31

2.5.3 順列生成の例示 35

2.5.4 可能な順列の数え上げ 36

2.5.5 条件を満たす順列の探索 37

2.5.6 結論 38

第3章　階乗進数と集合分割 **39**

3.1 階乗進数による集合分割の表現 39

3.2 視覚的クラスタリング 41

3.3 木の生成 43

3.4 ヒストグラムモデル 44

3.4.1 さまざまな度数分布 45

3.4.2 情報量規準と度数分布の要約 46

3.4.3 遺伝的アルゴリズムを使ったヒストグラムの選択 46

3.4.4 到着分布への応用 47

3.4.5 長崎の観光地人気調査結果への応用 48
3.4.6 結論 50

第4章 分割表からの規則抽出 **52**
4.1 情報量規準による分割表の選択 52
4.2 過度要約 55
4.3 あやめの識別問題 57
4.4 変数の置き換え 58
4.5 分割表の区分選択と遺伝的アルゴリズム 59
4.6 変数の増加 60
4.7 変数の減少 63
4.8 タブーサーチ 64
4.9 規則抽出の手順 65
4.10 決定木との比較 66
4.11 結論 67

第5章 大規模な問題の例としての減価償却問題への応用 **68**
5.1 減価償却方法の性質の違いから導かれる仮説とその検証 68
5.1.1 減価償却方法 68
5.1.2 減価償却方法選択に関する調査の概要 69
5.1.3 調査項目 69
5.1.4 調査結果の概要 70
5.1.5 未上場企業に対する考慮事項調査 73
5.1.6 仮説の検定と統計的方法 74
5.1.7 償却速度の違いに起因する仮説 76
5.1.8 償却計算の基礎の違いに起因する仮説 80
5.1.9 償却可能範囲額と償却方法の採用 82
5.1.10 償却の単位と償却方法 82
5.2 減価償却方法選択のファジィ推論 83
5.2.1 ファジィ推論とファジィ規則 84
5.2.2 ニューラルネットワークによる主観の評価 85
5.2.3 簡略化ファジィ推論 89
5.3 経営環境に関する大規模な調査からの償却方法選択に関する規則の抽出 92
5.3.1 経営環境に関する調査の概要 92
5.3.2 1994年調査から得られた規則 94
5.3.3 1997年調査から得られた規則 97
5.3.4 2000年調査から得られた規則 98
5.4 結論 98

第6章 まとめ **100**

付録A 1991年および1992年のアンケート調査項目 **104**

付録B 経営環境に関する調査 **106**

第1章　発見的探索と階乗進数

従来、組織的に解くことが難しいと思われていた問題に対しても、コンピュータの計算能力を利用したさまざまな発見的な探索方法が開発されてきている。このような方法として、遺伝的アルゴリズム (Genetic Algorithm:GA)、シミュレーテッドアニーリング (Simulated Annealing:SA)、タブーサーチ (Tabu Search:TS) などがある。これらは、巡回セールスマン問題などの組み合わせ最適化問題においても広く用いられてきた。本研究の目的のひとつは、アンケートなどを通して得られる分割表から、意味のある少数の規則を得ることであるが、これも分割表に含まれる階級数が増えるにしたがって、可能な組合せが指数的に増加する性質を持っている。この規則抽出の過程において、主に遺伝的アルゴリズムを利用している。そしてその染色体表現として、階乗進数表現を採用している。そこで本章では、階乗進数およびその遺伝的アルゴリズムにおける利用について述べる。

1.1　発見的探索方法

巡回セールスマン問題などのように、基本的にはすべての場合を列挙しなければ最適解が求まらないような組み合わせ最適化問題において、近年、メタヒューリステックス [48] やモダンヒューリステックス [68] と呼ばれる発見的探索方法の適用が盛んに研究されてきている。発見的探索方法としては、遺伝的アルゴリズム (Genetic Algorithm:GA)、シミュレーテッドアニーリング (Simulated Annealing:SA) やタブーサーチ (Tabu Search:TS) などが代表的なものであろう。

遺伝的アルゴリズムは、生物の進化を模したアルゴリズムを使って、よりよい解を求めようとするもので、Holland[45] によって導入された手法である。シミュレーテッドアニーリングは、ある確率で現在の解よりも劣る解も受け入れることによって、局所解からの脱出を図るものである [32][33]。タブーサーチ [14][94][76] では、新しい近傍解へ移る際にタブーリストを参照する。タブーリストに含まれている近傍解への移動を禁止することによって局所解からの脱出を図っている。

本研究では、これらの方法の中で主に遺伝的アルゴリズムを利用している。これは遺伝的アルゴリズムの適用範囲が非常に広く [74]、また他の方法をその探索過程の中に含ませることも可能であるからである。遺伝的アルゴリズムは、選択、交叉、突然変異などの多様な操作を備えており、その一般的な手順は次のようにまとめることができる [28]。

1. 初期集団の生成
2. 終了条件が満たされるまでの繰り返し
   (a) 適応度の評価
   (b) 選択
   (c) 交叉
   (d) 突然変異

以下、手順にそって、遺伝的アルゴリズムの概略を述べる。まず、あらかじめ定められた数の

親を乱数を使って生成する。これらの親は染色体と呼ばれるもので表現される。これらの染色体は、考えている問題の現実世界での解候補を複数の遺伝子を使って表現したものである。例えば、ナップザック問題では、詰め込むか (1)、詰め込まないのか (0) といった遺伝子を使って表現可能である。また、巡回セールスマン問題では、可能な巡回路を都市番号からなる順列として染色体を構成することができる。このように現実の問題をどのような染色体として表現するかは遺伝的アルゴリズムの実行に先立って定めておく必要がある。また、集団の大きさをどのくらいにするかということもあらかじめ決めておかなければならない。

親の集団が構成されたら、それぞれの親の適合度を求める。巡回セールスマン問題では、その染色体が表す巡回路の長さをもとに適合度を定める。適合度はより好ましいものが高くなるように設定されることが多い。適合度が求められたら、その適合度を元に一組の親を選択する。この親をもとに交叉によって子を生成することとなるが、優秀な親（適合度の高い親）ほど選択される確率を高くする。このような操作を行うことによって、優秀な親ほど多くの子を生成したり、次の世代に残ることが可能となる。選択の方法としては、ルーレット戦略、ランク戦略などが広く用いられている。ルーレット戦略とは、その適合度に比例するように選択確率を定める方法である。適合度の似通った集団となった場合には、確率に違いが見られなくなるため、スケーリングもあわせて行われることが多い。また、最小化問題へ適応するためには、適合度を如何に定めるかという問題が残る。ランク戦略は、適合度そのものではなく、適合度の順位によって選択確率を定めるものである。このため、最小化問題への適用も容易に行うことができる。

交叉とは、それぞれの親から、その染色体の一部を引き継ぐ形で新しい染色体（子）を生成する手続きである。通常、一組の親から、子は２つ生成される。交叉の方法としては、染色体の切断点を１カ所定め、その前後をそれぞれの子に引き継ぐ１点交叉、切断点を２カ所設定する２点交叉、交叉時にマスクをかけていずれの親の要素を引き継ぐか決定する一様交叉などがある。

突然変異は染色体の一部を強制的に変更する手続きである。巡回路を順列で表す巡回セールスマン問題においては、特定の都市の訪問順を変更するという形で突然変異が施される場合が多い。これをシフト演算と呼んでいる。

選択、交叉、突然変異という操作によって、得られた新しい集団が次の世代の親となり、ふたたびこれらの操作の対象となる。このような世代交代を繰り返しながら、よりよい解を求めようとする点に、遺伝的アルゴリズムの特質がある。

シミュレーテッドアニーリングやタブーサーチでも突然変異に似た操作を使って、新たな解候補を作っていく。したがって、遺伝的アルゴリズムの特徴的な操作として、交叉を考えることができるであろう。

染色体を都市番号の順列として表現し、巡回セールスマン問題に遺伝的アルゴリズムを適用した場合を考えてみよう。一組の親に対して切断点を１カ所定め、切断点の前半と後半をそれぞれの子に引き継いだものとしよう。このようにして作られた子の染色体には、同じ都市番号が複数含まれる可能性がある。このような染色体は巡回路を表すものとは考えられない。これが致死遺伝子と呼ばれるものである。飯間、三宮 [19][20] は、スケジューリング問題において、交叉によって致死遺伝子が多数作られるような場合には探索効率が落ちることを確かめている。巡回セールスマン問題において順列として解候補を表現した場合に、交叉によって致死遺伝子が生じることを避けるためにさまざまな方法が開発されてきた [12][13][95][44]。そしてどのような方法をとるかによって、探索能力に大きな差が生じることも確認されている [11][73]。したがって、致死遺伝子の生じない染色体表現、交叉の方法を検討することは重要な課題である。

さらに、親に含まれる好ましい形質をどのように子に伝えるかということも重要な課題である。親が獲得した資質を適切に子に伝えることができなければ、効果的な探索が期待できないからで

ある。巡回セールスマン問題においても、形質を遺伝させるための工夫がなされてきた。山村ら [95] はサブツアー交換交叉を提案し、良好な結果を得ている。さらに、前川ら [44] は枝交換交叉を使うことを提案している。これら形質の遺伝に考慮した方法は、単に致死遺伝子を抑制した交叉に比べて、探索能力が高いことが確かめられている [73]。

このように、交叉という操作を行う際には、致死遺伝子の抑制、形質の保持など探索効率を高める工夫ををどのように実現するかが重要な課題であることがわかる。本研究では、これらの課題を解決する目的で、階乗進数による染色体表現を検討した。次章以降でみるように、染色体を階乗進数を使って表現したときには、順序制約がある順列の探索においても、致死遺伝子を容易に抑制しうる場合がある。また、集合分割問題では、親に共通する結合関係を維持することも可能である。

1.2 階乗進数表現と本研究の目的

本研究において、n 桁の階乗進数 $C_nC_{n-1}\ldots C_2C_1$ は、数

$$(n-1)!C_n+(n-2)!C_{n-1}+\ldots+1!C_2+0!C_1$$

を表すものとする。ここで、各桁の係数 C_j は $0 \leqq C_j < j (j=n,\ldots,1)$ の範囲の整数である。したがって常に $C_1=0$ であるので最後の項は不要であるが、説明の便宜上 n 個の項からなる階乗進数を扱う。

たとえば、4 桁の階乗進数で表すことのできる最大の数は、

$$3!\times 3+2!\times 2+1!\times 1+0!\times 0=4!-1$$

である。したがって、0 も含めて $4!$ 個の場合を表現することができる。

前節でみたように、遺伝的アルゴリズムを利用しようとする場合、どのような染色体表現や交叉の方法を取るかによって、探索能力に違いが生じる。本研究もさまざまな問題に対して遺伝的アルゴリズムの適用を試みているが、その染色体表現として階乗進数表現を採用した。

n 桁の階乗進数を使って、$n!$ 通りの場合を表現できることから階乗進数の利用としてまず考えられるのは順列表現への応用である。解が順列として表されるものとしては巡回セールスマン問題やフローショップスケジューリングがある。

一方、順列以外にも階乗進数と対応しうるものがある。本研究では階乗進数を使って要素間の結合関係を表す方法を検討した。この方法を用いることによって、集合分割問題への階乗進数表現の適用が可能となる。特に分割表のそれぞれの階級を要素とする集合を考えると、分割表からの規則抽出は、集合分割問題の一種と考えることができる。

このようにさまざまな問題に対して、階乗進数表現を適用することができる。そこで、本研究では、フローショップスケジューリング問題や集合分割問題などを中心として、遺伝的アルゴリズムにおける階乗進数表現の適用可能性と、その有効性を検証することを目的とした。特に、規則抽出では、多変数でありかつ曖昧な要素を持つ減価償却方法選択問題へ応用し、得られる知見の妥当性を検証することを目的とした。

まず第 2 章では、階乗進数と順列を対応させている。このため、巡回セールスマン問題や、フローショップスケジューリング問題への適用が可能となった。第 2 章では、これらの問題に対する階乗進数表現の有効性を、数値実験を通して検証する。階乗進数表現を採用した場合には、一様交叉を採用したとしても、交叉によって致死遺伝子が生じることを抑制しうる。さらに、ある種の順序制約は容易に表現することができる。このため制約を満足しなくなった場合に必要とされ

る修正や罰金などの操作が不要となる。第2章では、特にフローショップスケジューリング問題を中心に、階乗進数表現の有効性を検討している。

さらに本研究では階乗進数を使った集合分割についても検討した。n 個の異なる要素を k 個の群に分ける組み合わせの数は次式で示される第2種のスターリング数と呼ばれるものとなる [50]。

$$_nS_k = \sum_{i=0}^{k}(-1)^i\frac{(k-i)^n}{i!(k-i)!}$$

さらに、可能なすべての k に関する分割数の合計はベル数と呼ばれている。n 個の要素に対するベル数 $B(n)$ は第2種のスターリング数を使って次のように表すことができる [50]。

$$B(n) = \sum_{k=1}^{n} {}_nS_k$$

したがって、要素の数 n が増加するにしたがってすべての場合を列挙することは極めて困難となってくる。

第3章では集合分割に対する階乗進数表現のいくつかの適用例について見ることとする。ここでは、視覚的クラスタリング、CG における木の生成、ヒストグラムモデルへの適用を検討した。また特に本研究では、集合分割によって規則の抽出を試みている。この考え方を Fisher[9] のあやめの識別問題に適用した結果を第4章にまとめた。ここでは、規則抽出の方法として知られている決定木 [64][65] との比較も行っている。

第5章では、階乗進数表現を使った規則抽出の減価償却方法選択問題への適用を検討した。この問題は、多変数かつ大規模な問題という特徴を持っている。また、主観にもとづく観測データを分析対象としているために、曖昧な要素を含んでいる。このような例に対して、分割表からの規則抽出が有効であるか、また、変数の選択を如何に行うべきかを検討した。

ところで、Reeves[68] は発見的 (heuristic) という言葉を次のように定義している。
“ヒューリステックとは、よい (最適値に近い) 解を、実用的な計算コストで探索する技法である。このとき、実行可能性や最適性を保証することは不可能でもよく、多くの場合、ある実行可能解が最適値にどのくらい近いものかを明確にしなくてもよいことにする。”

この定義からもわかるように、発見的探索方法は必ずしも最適解に到達することを保証するものではない。実用的な計算コストで満足する近似解が得られればよいという考えが背景にある。したがって、各方法の優劣はいきおい数値実験に頼ることになる。実際、各方法の能力を比較するために、また改良の有無を検証するためにさまざまな数値実験が行われてきた。そこで本研究でも、特に順列と対応させた場合には、階乗進数表現の有効性を数値実験を通して検証するという方法をとった。

ところで、規則抽出問題などにおいて、発見的探索方法の有効性はどのように判断すべきであろうか。巡回セールスマン問題などのように明確な基準があるのであろうか。規則を適用したときの判別力、予測力も有効な指標であろう。しかし、規則抽出の目的が、予測や判別ではなく、事象の解釈、説明である場合には、単なる判別力ではその有効性を測ることができないであろう。求められるものは、解釈や説明の妥当性であろう。つまり、減価償却方法選択問題においては、得られた規則は会計的な知識とも合致する必要がある。このため、階乗進数表現の減価償却方法選択問題への適用では、仮説検定の結果から得られる知見との整合性についても検討した。

第2章　階乗進数による順列の表現

前章で述べたように、n 桁の階乗進数を使って、$n!$ 個の数を表すことができる。したがって、まず解が順列として表現される問題への階乗進数表現の適用を試みた。

2.1　階乗進数と順列の対応

階乗進数を使って、順列をもれなく数え上げる方法は古くから提案されている [60][98]。しかし、そこでは、階乗進数は単なるカウンタとして利用されているのみである。そこでまず、階乗進数と順列の対応から検討する。そのために、順列に含まれる要素の相対的な位置関係を次のように定義する。

定義　後位置

n 個の異なる要素からなる特定の順列（以下、基準順列と呼ぶ）を考える。要素の位置を先頭から順に位置 0、位置 1、...、 位置 $n-1$ とする。基準順列において、位置 p の要素 x_p と、位置 q の要素 x_q をとったとき、$p < q$ ならば要素 x_q は要素 x_p より後位置にあるという。

順列から階乗進数への変換を考える。異なる n 個の要素からなる基準順列において、位置 $k(k = 0, \ldots, n-1)$ の要素を x_k と表す。また n 桁の階乗進数を $C_n C_{n-1} \ldots C_2 C_1$ と表す。基準順列に含まれる n 個の要素からなる任意の順列が与えられたとき、対応する階乗進数の係数 $C_j (j = n, \ldots, 1)$ を次の規則を用いて決定する。

規則 1（順列から階乗進数への変換)

1) 与えられた任意の順列において、要素 x_k までの部分順列を取り出す。

2) 部分順列の中で、要素 x_k より後位置にある要素の数を階乗進数の係数 C_{n-k} とする。

たとえば、大きさ $4(n=4)$ の順列 $bdac$ を 4 桁の階乗進数に変換することを考えよう。基準順列は仮に昇順の順列 $abcd$ とする。基準順列の最初の要素から順に対応する係数の決定を行ってみよう。この場合には位置 0 の要素 a に対応する係数から決定されていくことになる。

順列 $bdac$ から要素 a までの部分順列 bda をとる。基準順列は $abcd$ なので、部分順列の中で要素 a より後位置にあるものの数は 2、つまり $C_4 = 2$ とする。

要素 b までの部分順列 b の中で、要素 b より後位置にあるものの数は 0、つまり、$C_3 = 0$ とする。

要素 c までの部分順列 $bdac$ の中で、要素 c より後位置にあるものの数は 1、つまり、$C_2 = 1$ とする。

要素 d までの部分順列 bd の中で、要素 d より後位置にあるものの数は 0、つまり、$C_1 = 0$ とする。

この結果、順列 $bdac$ に対応する階乗進数は 2010 となる。つまり、順列 $bdac$ は 2010 とコード化されることになる。

次に n 桁の階乗進数を大きさ n の順列に変換する規則を定める。異なる n 個の要素からなる基

準順列 $x_k(k=0,\ldots,n-1)$ と n 桁の階乗進数 $C_nC_{n-1}\ldots C_2C_1$ が与えられたものとする。この時、n 桁の階乗進数から大きさ n の順列を次の規則を用いて作成する。

規則 2 (階乗進数から順列への変換)

1) 大きさ n の配列を考え、左から順に位置 0、位置 1、...、位置 $n-1$ とする。

2)$k=0$ とする。

3) 要素 x_k を位置 C_{n-k} に格納する。

4)k の値を 1 増やす。

5)$k=n-1$ ならば終了。そうでなければ、大きさ n の配列において、すでに要素が格納されている場所を除いて、左から順に位置 0、位置 1、...、位置 $n-k-1$ とする。

6) 手順 3) から 5) までを繰り返す。

たとえば、4 桁の階乗進数 2010 から大きさ 4 の順列を次の手順で生成する。前と同じく基準順列は、昇順の順列 $abcd$ とする。基準順列の最初の要素から順に順列内での位置を決定していく。

$C_4=2$ であるから、要素 a を位置 2 へ格納する。そして、要素 a を格納した場所を除いて位置番号を付け直す。

$C_3=0$ であるから、要素 b を位置 0 に格納するとともに位置番号を付け直す。

$C_2=1$ であるので、要素 c を位置 1 に格納するとともに位置番号を付け直す。

$C_1=0$ であるので、要素 d を位置 0 に格納する。

この結果、階乗進数 2010 から順列 bdac が生成される (図 2.1)。

変数 a の割り当て

		a	

変数 b の割り当て

b		a	

変数 c の割り当て

b		a	c

変数 d の割り当て

b	d	a	c

図 2.1: 階乗進数から順列への変換

規則 1 および規則 2 により、順列から階乗進数、階乗進数から順列への変換を行う場合に、階乗進数と順列が 1 対 1 に対応することを示す。1 対 1 対応を示す過程で, 完備性、健全性と呼ばれる性質も証明する。

(完備性) 規則 1 によって、大きさ n の任意の順列を、n 桁の階乗進数に対応させることができる。

(証明) 考えている順列の任意の要素を x とする。x の基準順列での位置を $k(k=0,\ldots,n-1)$ とする。要素 x の階乗進数の係数を C_{n-k} とする。任意の順列から得られた x までの部分順列の中で、要素 x より後位置にあるものの数は高々$n-k-1$ である。つまり、$C_{n-k}<n-k$ であるので、階乗進数の係数の条件を満たす。

(健全性) 規則 2 によって、n 桁の任意の階乗進数を、大きさ n の順列に対応させることができる。

(証明) まず、$n=1$ の場合を考えよう。階乗進数は C_1 と表される。階乗進数の係数の条件か

ら $C_1 = 0$。つまり、大きさ1の箱 (位置番号は0) の配列に要素 x_0 を格納して順列を得る。大きさ n の順列が階乗進数からすでに構成されているものとする。このとき新たに $n+1$ 桁目を加えた $n+1$ 桁の階乗進数を考える。また、基準順列において位置0に新たな要素を加え、従来からの n 個の要素の位置番号をそれぞれ1つずつ増やす。さらに、位置0から位置 n までを持った大きさ $n+1$ 個の配列を用意する。階乗進数を $C_n C_{n-1} \ldots C_3 C_2 C_1$ と表す。階乗進数の定義から、$C_{n+1} < n+1$ であるから、基準順列の最初に追加された要素 x_0 を位置 C_{n+1} に格納することができる。要素 x_0 を格納した場所を除いて、再び左から順に位置0、...、位置 $n-1$ とする。これらの位置に大きさ n の順列を格納することによって、大きさ $n+1$ の順列を作ることができる。

(一意性) 規則1および規則2によって、順列と階乗進数を一意に対応させることができる。

(証明) 異なる2つの階乗進数 $C_n C_{n-1} \ldots C_3 C_2 C_1$ と $C'_n C'_{n-1} \ldots C'_3 C'_2 C'_1$ を考える. 階乗進数を辞書式に並べたとき、$C_n C_{n-1} \ldots C_3 C_2 C_1 > C'_n C'_{n-1} \ldots C'_3 C'_2 C'_1$ とする。このとき、$C_i > C'_i, C_j = C'_j (n \geqq j > i)$ となるような i が存在する。基準順列内の要素 $x_k (k = 0, \ldots, n-i-1)$ が格納される場所は等しいので、これらの要素を除いた大きさ i の配列を考える。$C_i > C'_i$ であるから、要素 x_{n-i} の格納場所は異なる。つまり、異なる階乗進数から同一の順列が生成されることはない。

また、2つの異なる順列を考える。この2つの順列の階乗進数表現をそれぞれ $C_n C_{n-1} \ldots C_3 C_2 C_1$ と $C'_n C'_{n-1} \ldots C'_3 C'_2 C'_1$ とする。このとき、基準順列の要素 $x_k (k = 0, \ldots, n-1)$ の2つの順列での格納場所を順に調べる。2つの順列は異なるのであるから、$C_{n-k} \neq C'_{n-k}$ となるような要素 x_k が少なくともひとつは存在するはずである。したがって、2つの階乗進数が等しくなることはない。

(一対一対応) 順列と階乗進数は一対一に対応する。完備性から任意の順列を階乗進数に対応させることができる。さらに、健全性から任意の階乗進数を順列に対応させることができる (逆写像の存在)。また、異なる順列や階乗進数から、同一の階乗進数や順列が生成されることはない (一意性)。したがって、順列と階乗進数は一対一に対応する。

ここで、順序表現と呼ばれる Grefenstette、Gopal、Rosmaita and Gucht[13] が提案した方法との相違点を見ておこう。順序表現では順列生成時に、残されている要素のうちの何番目のものを使うかという情報を染色体に記憶させる。したがって、遺伝子がとり得る最大値は、大きさ n の順列を扱う場合には、先頭から順に、$n-1, n-2, \ldots, 2, 1, 0$ となる。したがって、順序表現も階乗進数を使った表現であるとみることができる。異なる点は、階乗進数と順列の対応規則にある。順序表現が、生成される順列の最初の要素から順に定めていく方式なのに比べて、本研究で用いた方法は、基準順列の最初の要素から順に挿入すべき位置を定めていく。つまり、順序表現における染色体は生成される順列の出現順に関する情報を持つのに対して、本研究で用いた対応では基準順列の要素が持つ位置情報を表しているものと考えることができる。このため、本研究で採用した方法は、順序表現に対して位置表現ということもできる。対応規則が異なるために、同じ階乗進数から異なる順列が生成されることもある。上の例で用いた階乗進数2010から順序表現の規則によって順列を生成すると *cadb* となる (図2.2)。

2.2 巡回セールスマン問題への適用

巡回セールスマン問題やフローショップスケジューリング問題など、その解が順列として表されるものは多い。このような問題に対して、近年遺伝的アルゴリズムなどメタヒューリスティックスと呼ばれる方法を用いた探索が試みられてきた。巡回セールスマン問題に遺伝的アルゴリズムを適用しようとする場合には、交叉によって生じる致死遺伝子を抑制することが必要であった。こ

2番目の変数の割り当て

c			

0番目の変数の割り当て

c	a		

1番目の変数の割り当て

c	a	d	

0番目の変数の割り当て

c	a	d	b

図 2.2: 順序表現による変換

の問題を解決するために、Grefenstette、Gopal、Rosmaita and Gucht[13] は、順序表現を使った順列の表現を提案した。また、順列をそのまま染色体として使う場合にも、交叉の方法を工夫することによって、致死遺伝子を抑制することができる。例えば、Goldberg and Lingle[12] は部分写像交叉（PMX）と呼ばれる交叉方法を提案し、致死遺伝子が抑制できることを示した。

本研究では、階乗進数を使って順列を表現する方法を提案している。また、本研究では、発見的探索方法として主に遺伝的アルゴリズムを利用している。そこで、本節では、階乗進数表現の巡回セールスマン問題への適用と、遺伝的アルゴリズムの実行について述べる。

本研究では以下の手順で遺伝的アルゴリズムを実行している。

1. 親の初期集団 (大きさを $SIZE$ とする）を作成する。集団は適合度の高い順に並び替えておく。
2. 集団の中から NEW 組の親のペアを作る。親の選択はランクによる。
3. 交叉によって子を生成する。交叉は一様交叉を採用した。
4. 親の集団の中で最も適合度の高いものを除いた、$SIZE-1$ 個の親と、新しく生成された $2\times NEW$ 個の子に対して突然変異を行う。突然変異は、乱数で選ばれた遺伝子座の値 (階乗進数の係数) を対立遺伝子におきかえることによって行った。このとき、突然変異前と等しい値が選ばれることも許容した。
5. 突然変異を行った後で、$SIZE+2\times NEW$ 個の染色体の適合度を求める。これらの染色体を適合度の高い順に並び替え、上位 $SIZE$ 個を次の世代に残す。
6. あらかじめ定められた回数だけ、上記 2. から 5. までの繰り返す。

この手順を巡回セールスマン問題へ適用した。対象としたものは、玄、辻村、高村、程 [11] によって示された例である。ここでは距離行列が明示されており、追実験が容易であったからである。また、10 都市、20 都市、30 都市の例が示されているが、この中では最も規模の大きい 30 都市の例を使って数値実験を行うこととした。実験では、親の集団の大きさ $SIZE$ は 30、交叉を行うペアの数 NEW は 15、繰り返し回数は 5000 とした。また、異なる初期集団を作成し、合計 20 回の実験を行った。また、玄、辻村、高村、程 [11] では、巡回路を順列で表現したときに、致死遺伝子を抑制する交叉として開発された PMX(partially mapped crossover),OX(order crossover),CX(cycle crossover) の比較を行っているので、本研究では、順序表現 [13]、2 点順序交叉 [47] との比較実験を行うこととした。ここで、2 点順序交叉とは、切断点を 2ヵ所設定し、その外側の遺伝子はそのまま子に引き継ぐ。中央部は残りの要素をもう一方の親からその出現順に受け継ぐという操作で

ある。20 回の実験を行って得られた最良の巡回路をその距離が短い順に並べて図示した (図 2.3)。

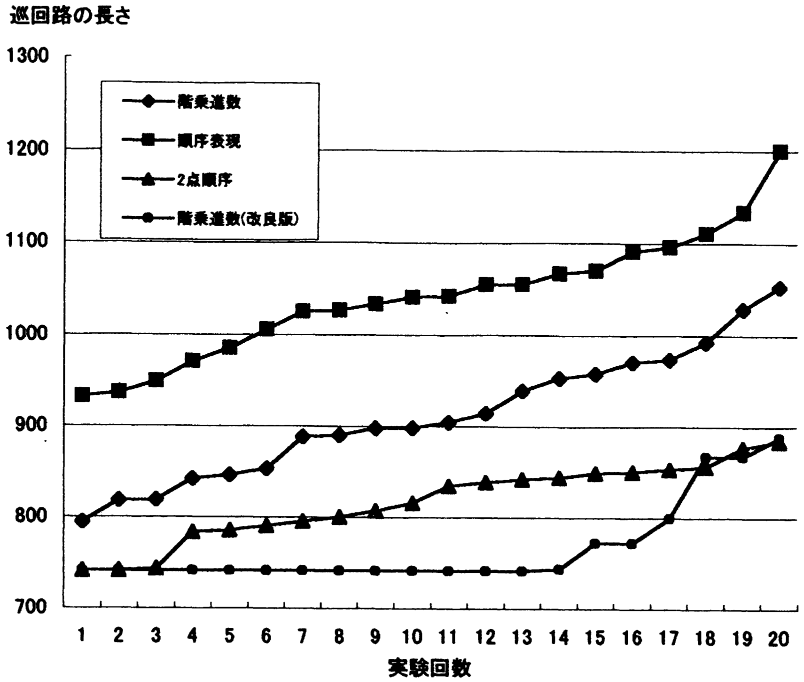


図 2.3: 30 都市の巡回セールスマン問題

数値実験では、階乗進数表現を採用した場合の最小巡回路長は 794 であった。交叉確率や突然変異率の設定が異なるので単純な比較はできないが、玄、辻村、高村、程 [11] の実験では、PMX から得られた最小巡回路が 1597、OX から得られた最小巡回路が 1593、CX から得られた最小巡回路が 944 であるので、これらの交叉方法に比べると良い結果を得ていると考えていいだろう。また順序表現を採用した場合の最小巡回路長が 932 であったことから、同じ階乗進数表現を使うとしても、本研究で採用した順列との対応の方が優れていることがわかる。

しかし一方、2 点順序交叉に比べると明かに探索能力が劣る。これは、階乗進数表現が致死遺伝子を抑制することには成功したものの、形質保持などを考慮していないからだと思われる。2 点順序交叉はその操作からある程度親に含まれる順序関係を保存する効果があると思われることから、階乗進数表現に比べて良い結果を得たものと考えることができる。

さらに、山村、小野、小林 [95] は親の形質を保持するような交叉方法を提案した。さらに前川、玉置、喜多、西川 [44] はマクロ遺伝子を使って、親に含まれる接続関係を保持しようとした。これらの方法は、従来の致死遺伝子の抑制のみを考慮した交叉方法に比べて、良好な探索結果を示している [73]。しかし、このような改良は、巡回セールスマン問題固有の性質を利用している。このため、他の問題への応用が困難な場合もある。

階乗進数を使って順列を表現した場合には、基準順列として任意の順列を選ぶことができる。そこで、次のような操作によって基準順列を作ることとした。任意の都市からスタートして最も近い距離にある都市をつないでいく。一般にはすべての都市が結ばれることなく、いずれかの都市に戻る。このようにして得られた一群の都市を基準順列の最初から順に配置していく。残された都市についても同様の操作を行って基準順列に格納する。このような操作によって、近い距離に

あるものが基準順列でも近い位置に格納されるので、親の初期集団を作る際にも、階乗進数の係数値を比較的小さい値 (0 または 1) とした。係数の値が小さければ生成された順列においても、前の要素と比較的近い位置に割り当てられる可能性が高くなるからである。このように基準順列と初期集団の係数値に改良を加えたものを階乗進数表現の改良版と呼ぶことにする。図 2.3 には、この方法によった場合の実験結果もあわせて示している。

図 2.3 から、階乗進数表現の改良版は他の方法に比べて良い結果を得ていることがわかる。この数値実験を通して、問題によって基準順列の構成を変えたり、また必要に応じて階乗進数の係数に条件をつけるという方法によって、探索能力が向上することがわかった。

2.3 フローショップスケジューリング問題への適用

フローショップスケジューリング問題も解が順列として表現されるため、巡回セールスマン問題のために開発された様々な交叉方法を適用することができる。村田、石渕、田中 [47] は、いくつかの交叉方法を比較して、2 点順序交叉が優れていることを示している。しかし、ここでは遺伝的アルゴリズムを単独で用いた場合には、シミュレーティッドアニーリングやタブーサーチよりも探索能力が低いこともあわせて報告されている。このため、遺伝的アルゴリズムの中にシミュレーティッドアニーリングの探索を取り入れることによって解の改善を試みている。また、巡回セールスマン問題において有効であった形質を遺伝させる方法を、フローショップスケジューリング問題に適用することも考えられるが、形質を保存する方法では、新しい解が生成されにくいという問題点も指摘されている [96]。

一方、Johonson[26] の研究以来、n 仕事 m 機械のフローショップスケジューリング問題に対して様々な方法が提案されてきた。これらの中には、Compbell、Dudek and Smith(CDS)[7]、Dannerbring(DES)[8] や Nawaz、Enscore and Ham(NEH)[51] など構成的ヒューリスティックスと呼ばれるいくつかの方法がある。Turner and Booth[83] は数値実験を通してこれら 3 つの方法 (CDS、DES、NEH) を比較した。そして、これらの中では NEH がもっとも優れていることを示した。

このようにフローショップスケジューリングの近似解法として、NEH などの構成的ヒューリスティックスと呼ばれるものと、遺伝的アルゴリズムなどのメタヒューリスティックスと呼ばれる方法が研究されてきた。しかし、これら 2 つの近似解法を組み合わせることはこれまでほとんど行われてこなかったと思われる。このため、本節では、構成的なヒューリスティックスである NEH の探索方法を、遺伝的アルゴリズムの中に取り入れ、探索能力の改善を図ることを目的とした。この目的のために、順列を階乗進数を使って表現することとした。

階乗進数による順列表現は順列との 1 対 1 対応も保証されるので、染色体表現における冗長性もない。また、交叉によって致死遺伝子が生じることがないので、1 点交叉や 2 点交叉のみならず、一様交叉も適用できる。さらに、順列と階乗進数を対応させるときに使われる順列（これを基準順列と呼ぶことにする）を工夫することによって、NEH の考え方を遺伝的アルゴリズムの探索過程に無理なく導入できる。これらの理由により、本研究では、順列を階乗進数として表現することとした。また、従来の方法との比較を行うために数値実験を行うこととした。

2.3.1 NEH 法

本研究では、同じ順序で m 機械において処理される n 仕事が与えられたときに、総完了時間 (makespan) 最小を目的とするフローショップスケジューリング問題を考える。このフローショッ

プスケジューリング問題に対して、NEH 法は、高速に優秀な近似解を生成する方法として知られている。NEH 法の概略は以下に示すように極めて単純なものである。

1) それぞれの仕事について、機械ごとの作業時間の合計を求める。

2) 総作業時間の大きい順に仕事を並べ替える。

3) 総作業時間の大きい 2 つの仕事を取り出す。これら 2 つの作業のみを考えたときの最適な作業順を決定する。これは 2 つの可能な順序に対する総完了時間を比較することによって求められる。ここで決定された 2 つの仕事の相対的な順序は以下の手順においても保存される。

4)3 番目に総作業時間の大きい仕事を取り出す。先に決定された 2 つの仕事の適当な位置に挿入することを考える。先に決定された 2 つの仕事の相対的な順序関係は維持するので、可能な順列は 3 通りである。この 3 通りすべてについて、総完了時間を計算し、総完了時間が最小となる挿入位置を決定する。

5) すべての仕事について、手順 4) により挿入位置を決定する。

2.3.2 遺伝的アルゴリズムへの適用の例示

前節で、順列を階乗進数を使って表現する方法を述べたが、これを遺伝的アルゴリズムを使った探索に適用する。階乗進数と順列を対応させるために必要な基準順列は任意の順列を採用できる。そこで、総作業時間の小さい順に n 個の仕事を並べた時に得られる順列を基準順列として採用する。図 2.4 に示した 5 仕事 4 機械の仮設例では、基準順列は $acbde$ となる。

次に、係数 $C_j(j = n, \ldots, 1)$ が、条件 $0 \leqq C_j < j$ を満たすように、乱数を使って階乗進数 $C_n C_{n-1} \ldots C_2 C_1$ を生成する。この操作を繰り返して図 2.5 に示した 4 つの階乗進数を作ったものとしよう。これが親の初期集団となる。また、2.1 節の規則 2 を使って順列に変換し、総完了時間 (makespan) を計算する。図 2.5 には総完了時間および順列も合わせて示している。なお、染色体 (階乗進数) は適合度の大きい順 (総完了時間の小さい順) に並び替えている。

基準順列 $acbde$ の最後の 2 つの仕事 d、e に注目しよう。基準順列は総作業時間の小さい順に並び替えているので、これらの仕事が総作業時間の大きい 2 つの仕事である。これらの仕事に対する階乗進数の係数は、最も総作業時間の大きい仕事 e については常に 0、次に総作業時間の大きい仕事 d については、0 または 1 のみをとりうることがわかる。とりうる数の範囲が限定されていることから、乱数を使って階乗進数を作ったとしても、可能な組み合わせのすべてを生成しうることが予想される。実際、図 2.5 の初期集団においても、可能な 2 つの組み合わせ 0,1 と 0,0 が出現している。また、3 番目に作業時間が大きい仕事 b まで考えた場合でも、可能な 6 通りの組み合わせのうち 3 通りが出現していることがわかる。このように、昇順の基準順列を採用し、階乗進数を使って順列を表現した場合には、総作業時間の大きいものに対する相対的な位置関係の探索はきめ細かく行うことができることがわかる。

	機械 1	機械 2	機械 3	機械 4	総作業時間
仕事 a	41	38	18	95	192
仕事 b	49	85	38	43	215
仕事 c	72	84	40	18	214
仕事 d	78	15	96	61	250
仕事 e	6	93	64	88	251

図 2.4: 5 仕事 4 機械の仮設例

	染色体						順列				
$no.1$	0	3	1	1	0	516	a	e	b	d	c
$no.2$	1	0	2	1	0	551	c	a	e	d	b
$no.3$	0	0	1	1	0	570	a	c	e	b	d
$n0.4$	4	2	0	0	0	623	b	d	c	e	a

図 2.5: 親の初期集団

交叉の処理を行うためには、1 組の親を選ぶ必要がある。交叉させる親をルーレット方式により選ぶ方法も広く行われているが、この方法は最小化問題には適用しづらく、スケーリングの必要もある。そこで本研究では、ランクをもとに親の選択を行うことにした。つまり、集団の大きさを $SIZE$ としたときに、$i(i = 0, \ldots, SIZE - 1)$ 番目に適合度の高い染色体が選ばれる確率 $p(i)$ を、

$$p(i) = (SIZE - i)/(\frac{1}{2} \times SIZE \times (SIZE + 1))$$

とした。

順列を階乗進数を使って表現することによる利点の 1 つは、交叉によって致死遺伝子が生じないことである。ふたつの階乗進数 $C_n C_{n-1} \ldots C_2 C_1$ と $C'_n C'_{n-1} \ldots C'_2 C'_1$ に対して、交叉位置を p とする 1 点交叉によって新たに子 $C_n C_{n-1} \ldots C_p C'_{p-1} \ldots C'_2 C'_1$ を生成したものとしよう。$C_j < j, C'_j < j (j = n, \ldots, 1)$ であるので、作られた子も階乗進数の係数の条件を満たす。つまり、交叉によって致死遺伝子は生じない。2 点交叉、一様交叉によっても致死遺伝子が生じないことは明らかである。本研究では、多様な解を生成し初期収束を避けるために一様交叉を採用した。

突然変異も、係数の値を変化させることによって実現できる。それぞれの染色体について、突然変異の処理を行う遺伝子座を乱数を使って決める。仮に、先頭から数えて $k(k = 0, \ldots, n - 1)$ 番目の遺伝子座が突然変異の対象になったものとしよう。k 番目の遺伝子座の値は、階乗進数の係数の条件から、$n - k$ 未満の値でなければならない。そこで、0 から $n - k - 1$ までの乱数を作り新しい遺伝子とする。順列生成の規則 2 から、階乗進数の係数を変化させることによって、要素 x_k の位置を変えることができる。しかし、親の集団の中で最も優秀なものに対しては突然変異の操作を施さないこととした。これは、エリートを保存するためである。

突然変異を行った後で、階乗進数を順列に変換し、総完了時間 (makespan) を計算した。そして、総完了時間 (makespan) の小さい順に並び替え、必要な数の染色体を次の親として残した。

2.3.3 遺伝的アルゴリズムの数値実験

提案方法の探索能力を検証する目的で、数値実験を行った。本研究では、仕事の数 n として、20、50、100、200 の 4 通りを設定した。また、機械の数 m も 5、10、20 の 4 通りを設定した。したがって、仕事数と機械数の可能な組み合わせは 12 通りである。比較の対象は、次の 4 つである。

1)NEH

古典的なヒューリスティックスである NEH の考え方を提案方法は取り入れている。提案方法がどの程度 NEH の解を改善する効果があるかどうか確かめるために数値実験に加えた。

2)1 点順序交叉 ("1 点" と略記)

Reeves[67] が採用した交叉方法である。染色体は順列を使って表現し、突然変異はシフト操作である。なお、Reeves は初期集団の中に、NEH で得られた解を含めているが、探索能力を比較するために、初期集団はすべて乱数を使って生成した。

3) 中央部順序交叉 ("中央" と略記)

フローショップスケジューリングを扱った村田、石渕、田中 [47] の数値実験において、遺伝的アルゴリズムの交叉処理としては最も優れた結果を示しているものである。1点順序交叉と同じく、染色体は順列を使って表現し、突然変異はシフト操作である。交叉位置を2ヶ所設定し、両端部はそのまま一方の親からそのまま引き継ぐ。中央部は他の親から、その出現順に引き継ぐ交叉方法である。なお、次の方法と同じく2点順序交叉の範疇に入るものなので、区別する目的で中央部順序交叉と呼ぶことにした。

4) 両端部順序交叉 ("両端" と略記)

順序制約付き順列の遺伝的アルゴリズムを使った探索において、飯間・三宮 [19][20] は2点順序交叉を採用している。これは、切断点を2ヶ所設定する、切断点の間の遺伝子は一方の親からそのまま引き継ぐ、切断点の左側はもう一方の親からその出現順に引き継ぐ、切断点の右側もやはりもう一方の親から引き継ぐが、右から見て同じ順になるように割り当てるという操作によって達成される。このような2点順序交叉を仮に両端部順序交叉と呼ぶことにしよう。突然変異はやはりシフト操作である。

5) 順序表現 ("順序" と略記)

Grefenstette、Gopal、Rosmaita and Gucht[13] が提案した染色体表現である。これも階乗進数表現のひとつと考えられることから、比較の対象に加えた。交叉は一様交叉を採用した。また、突然変異は提案方法と同じく特定の遺伝子座の値を条件を満たす数におきかえることによって行った。

6) 提案方法 ("提案" と略記)

前節で述べたように、染色体は階乗進数を使って表現する。階乗進数と順列の対応は、2.1 節の規則1および2による。また、交叉は一様交叉を採用した。

これらの方法のうち、2)、3)、4)、5)、6) が遺伝的アルゴリズムによって解を求めようとするものである。交叉と突然変異の操作部分以外は同じ手順によった。以下に概略を示す。

集団の大きさ $SIZE$ は n、つまり仕事の数と同じ大きさとした。初期集団作るために、$SIZE$ 個の階乗進数を乱数を使って生成した。5)、6) では階乗進数をそのまま利用し、2)、3) および 4) では、順列に変換した。したがって、それぞれの方法の数値実験において、初期集団は同一のものからスタートしている。得られた初期集団は適合度の高い順（総完了時間の短い順）に並び替える。

前述のように、交叉はそれぞれの方法で異なる。交叉はランクをもとに行う。$SIZE/2$ 組の親を作り、$SIZE$ の子を生成した。交叉の結果得られた $SIZE$ の子と、親の集団の中で最も適合度の高いものを除いた $SIZE + (SIZE - 1)$ 個の染色体について突然変異を施す。2)、3) および 4) ではシフト操作によって突然変異を実行した。また、5) および 6) では乱数によって選ばれた遺伝子座の階乗進数の係数を他の値と置き換えることによって突然変異を実行した。

交叉と当然変異の操作を行った後で、親と子の合計 $2 \times SIZE$ 個の染色体を適合度の高い順に並び替える。そして、上位 $SIZE$ の染色体を次の親として残した。世代交代数はおおむね解が安定すると思われる 1000 世代まで行った。

2.3.4 テストデータ

過去の数値実験、例えば Osman and potts[61] などにおいては、各仕事の各機械での作業時間を1から100までの乱数で与えている。したがって、本研究でも1から100までの一様乱数を使って作業時間を決定した仮設例を設けることとした。1から100までの一様乱数を使って作られたものを $typeA$ の仮設例と呼ぶことにする。

しかし一方、、現実には作業時間がランダムになることはむしろまれであるという指摘がある。Rinnooy Kan[69] は、作業時間に見られる 2 つのパターンを想定している。1 つは、機械によって平均作業時間が異なるものであり、もう 1 つは仕事によって機械毎の作業時間が異なる傾向をもつ場合である。このうち前者の場合には、特定の機械での作業時間が全体の完了時間を規定するため解の探索は容易であるが、後者は良い近似解を求めることがより困難であることがわかっている [67]。したがって、本研究でも仕事によって平均作業時間が異なる仮設例を設定することにした。Rinnooy Kan[69] に倣って、各仕事の各機械での処理時間を次の範囲の一様乱数で定める。なお、ε_i は仕事毎に 1、2、3、4、5 の中からランダムに選ばれた数である。

$$20\varepsilon_i + 1 : 20\varepsilon_i + 20$$

このようにして作られた仮設例を *typeB* の仮設例と呼ぶことにする。

2.3.5 遺伝的アルゴリズムにおける数値実験結果

各アルゴリズムの優劣を判定するために、2 つの指標を計算した。1 つ目の指標は、30 個の数値例に対して何回他のアルゴリズムに比べて良い解を得ることができたかを数えたものである。なお、同じ完了時間となる場合があるので、合計は必ずしも 30 と一致しない。2 つ目の指標は、一対比較の結果得られる t 値である。一対比較は、一対の実験 (本研究の場合にはある特定の数値例に対する実験) から得られる観測結果の差がある値 (本研究では 0) と等しいと考えることができるかを統計的に確認する方法であり、検定手順は次のようにまとめられる。

*Step*1 2 つの統計量の差の母平均を μ とし、帰無仮説 H_0 を次のように設定する。

$$H_0 : \mu = 0$$

*Step*2 対立仮説 H_1 を次のように設定する。

$$H_1 : \mu \neq 0$$

*Step*3 大きさ n(本研究では 30) の標本についての差 $x = x_1 - x_2$ の平均を $\overline{x}$、標本分散を s^2 とするとき、次の検定統計量 t を計算する。

$$t = \overline{x}/(s/\sqrt{n})$$

*Step*4 帰無仮説 H_0 のもとでは、検定統計量 t は自由度 $n-1$ の t 分布に従うことを利用して検定を行う。

本研究では、有意水準を両側 1 %に設定した。また標本数はいずれの検定においても 30 であるので、棄却域は

$$t < -2.750 \text{ または } t > 2.750$$

となる。表には、検定結果もあわせて示している。提案方法が、比較の対象となる方法に比べて優れていると思われる場合には△、劣っていると思われる場合には▼を記入している。

表 2.1 に、作業時間として 1 から 100 までの一様乱数を使った type A のテストデータに対する実験結果を示した。提案方法は、仕事数や機械数によらず NEH の解を改良していることがわかる。さらに、同じ階乗進数表現である順序表現に比べてもすべての場合において優れていることがわかった。また、仕事数や機械数が大きい (n=100、m=20 または n=200、m=20) 場合には、順序交叉を用いた 3 方法すべてに対しても、提案方法が優れていることがわかった。

表 2.1: type A のデータに対して最良解を見つけた回数

n	m	NEH	1 点	中央	両端	順序	提案
20	5	7(△)	29( )	23( )	26( )	11(△)	26
20	10	1(△)	14(▼)	17( )	10( )	1(△)	5
20	20	0(△)	6( )	8( )	6( )	0(△)	11
50	5	9(△)	26( )	26( )	27( )	9(△)	24
50	10	0(△)	8( )	12( )	4( )	0(△)	13
50	20	0(△)	9( )	9( )	1(△)	0(△)	13
100	5	16( )	28( )	27( )	28( )	8(△)	25
100	10	0(△)	13(▼)	15(▼)	8( )	0(△)	7
100	20	0(△)	2(△)	6(△)	0(△)	0(△)	22
200	5	18( )	29( )	30( )	28( )	10(△)	26
200	10	1(△)	13( )	15( )	8( )	0(△)	7
200	20	0(△)	1(△)	4(△)	0(△)	0(△)	26

仕事によって作業時間の平均を変えた type B の仮設例に対する実験結果を表 2.2 に示した。表 2.2 から明らかなように、提案方法はほとんどすべての場合において、他の方法より優れた結果を示している。特に、仕事数が 50 以上であり、機械数が 10 以上の規模の問題においては、提案方法の優位性は顕著である。最良解を見つけた回数、統計的な検定結果の両方において、提案方法は他の方法に比べて優れている。

乱数を使ってテストデータを与える type A の場合、それぞれの作業時間が互いに相殺して、総作業時間の差は小さくなる傾向がある。これは、NEH の考え方を取り入れた提案方法にとって、不利な状況である。一方、type B では、それぞれの仕事によって作業時間の平均が異なるので、このような状況は起こりにくい。このことが、type A に比べて、type B において、提案方法がより優れた結果を示すことができた原因だと思われる。

一方、type A のテストデータも含めて、規模の大きい問題 (n=100、m=20 または n=200、m=20) において、提案方法が優れた結果を示している。これらのことから、提案方法はそれぞれの仕事によって作業時間の平均が異なる場合や、比較的規模の大きい問題において有効であるといえるだろう。

2.3.6 情報量を用いた結果の検討

数値実験の結果から、提案方法は従来の方法より優れていることがわかった。本節では、その原因を情報量をもとに検討する。

n 桁の階乗進数を $C_nC_{n-1}\ldots C_2C_1$ とするとき、2.1 節の仮定から、係数 C_j は条件 $0 \leqq C_j < j(j=n,\ldots,1)$ を満たす。いま係数 C_2 に注目しよう。条件から C_2 は 0 もしくは 1 の値をとる。C_1 は常に 0 であるから、$C_2=1, C_1=0$ と $C_2=0, C_1=0$ はそれぞれ総作業時間が大きい 2 つの作業の可能な順列を表す。初期集団において、係数値は乱数を使って定めているので、1 と 0 はほぼ同数含まれているはずである。つまり、係数 C_2 のエントロピーは 1 に近い値となっているはずである。一方、探索の結果、係数 C_2 のエントロピーが減少したものとしよう。これは可能な 2 つの順列のうち特定のものに収束したことを意味している。つまり、係数のエントロピーを計算す

表 2.2: type B のデータに対して最良解を見つけた回数

n	m	NEH	1点	中央	両端	順序	提案
20	5	14(△)	29( )	28( )	28( )	14(△)	28
20	10	12(△)	25( )	27( )	28( )	8(△)	27
20	20	12(△)	26( )	26( )	28( )	5(△)	23
50	5	2(△)	5(△)	10(△)	4(△)	0(△)	24
50	10	0(△)	5(△)	9( )	9(△)	0(△)	15
50	20	0(△)	2(△)	12( )	5( )	0(△)	16
100	5	0(△)	0(△)	4(△)	0(△)	0(△)	27
100	10	0(△)	0(△)	3(△)	0(△)	0(△)	29
100	20	0(△)	0(△)	7(△)	1(△)	0(△)	22
200	5	10( )	11(△)	24( )	5(△)	0(△)	15
200	10	0(△)	0(△)	3(△)	0(△)	0(△)	27
200	20	0(△)	0(△)	2(△)	0(△)	0(△)	28

ることによって、特定の順列に収束した度合いを測ることができるであろう。

エントロピーは他の係数についてのも計算することができる。係数 C_j は j 個の値をとりうる。大きさ n の親の集団においてそれぞれ n_j 個含まれていたとすると、係数 C_j のエントロピー $E(C_j)$ は次式で計算することができる。

$$E(C_j) = -\Sigma\ (n_j/n)log_2(n_j/n)$$

係数 C_2 の場合と同じように、係数のエントロピーを計算することによって、可能な順列のうち特定の順列に収束した度合いを測ることができるであろう。そこで、提案方法が特に優れていると思われる n=100、m=20 と n=200、m=20 の場合について、係数のエントロピーを計算した。対象とした係数は、$C_2, C_5, C_{10}, C_{15}, C_{20}, C_{25}, C_{30}$ の 7 つである。なお、染色体を順列、もしくは順序表現を使って表す方法では、第 1 節の規則 1 を使って階乗進数に変換し、エントロピーを計算した。結果を表 2.3 に示した。

係数 C_2 のエントロピーの初期値はいずれの場合も 1 に近い。1000 世代の探索の後では、いずれの方法を見ても小さくなっている。このことから、総作業時間の大きな 2 つの仕事の相対的な順序は一方に収束していることがわかる。他の係数についてもエントロピーが小さくなる傾向にあるが、その度合いは遺伝的アルゴリズムの設計によって異なる。係数 C_2 や C_5 等では、順序交叉を用いた場合のほうが、提案方法よりも小さな値となっている場合があるが、C_{15} 以上になると明らかに提案方法の方が小さい。つまり、提案方法によれば、総作業時間の大きい仕事のみならず、総作業時間の小さな仕事に対しても相対的な順序関係が探索とともに収束しつつあることがわかる。このことは可能な順序のうち、少数の順序に固定されていっていることを表す。これは、総作業時間の大きい順に作業順を固定していくという NEH の特徴と一致する。

遺伝的アルゴリズムは大域的探索に優れているが、局所探索能力は低いと考えられている。このため、村田、石渕、田中 [47] は、遺伝的アルゴリズムによる大域的探索を行った後で、シミュレーティッドアニーリングを使った局所探索を行う方法を提案している。一方、本研究では、階乗進数を使って順列を表現することによって、NEH の特徴である深さを優先した探索を行うことが可能となった。深さ優先探索は総作業時間の大きい仕事の相対的な位置関係を固定した局所探索

表 2.3: 探索後の階乗進数のエントロピー

係数	C_2	C_5	C_{10}	C_{15}	C_{20}	C_{25}	C_{30}
初期値	0.99	2.29	3.24	3.81	4.19	4.45	4.69
1 点	0.24	0.61	0.93	1.47	1.65	1.60	1.90
中央	0.24	0.59	0.95	1.26	1.44	1.72	1.66
両端	0.21	0.71	1.10	1.28	1.84	1.78	2.02
順序	0.22	0.62	0.93	1.52	1.41	1.98	1.84
提案	0.27	0.48	0.34	0.44	0.54	0.50	0.46

係数	C_2	C_5	C_{10}	C_{15}	C_{20}	C_{25}	C_{30}
初期値	0.99	2.29	3.24	3.82	4.18	4.45	4.69
1 点	0.03	0.16	0.22	0.36	0.83	1.08	1.49
中央	0.03	0.18	0.37	0.39	0.82	1.08	1.42
両端	0.08	0.17	0.33	0.57	1.01	1.34	1.77
順序	0.08	0.25	0.39	0.52	0.91	1.14	1.14
提案	0.09	0.19	0.23	0.19	0.26	0.35	0.50

係数	C_2	C_5	C_{10}	C_{15}	C_{20}	C_{25}	C_{30}
初期値	1.00	2.31	3.28	3.86	4.26	4.55	4.79
1 点	0.20	0.58	0.95	1.32	1.34	1.83	1.61
中央	0.25	0.54	1.12	1.21	1.35	1.67	1.78
両端	0.16	0.68	1.08	1.20	1.46	1.73	1.76
順序	0.30	0.85	1.27	1.16	1.72	1.82	1.99
提案	0.43	0.49	0.66	0.57	0.58	0.57	0.56

係数	C_2	C_5	C_{10}	C_{15}	C_{20}	C_{25}	C_{30}
初期値	1.00	2.31	3.29	3.86	4.26	4.55	4.79
1 点	0.04	0.11	0.24	0.50	0.53	0.75	0.72
中央	0.02	0.11	0.32	0.38	0.54	0.67	0.59
両端	0.15	0.35	0.71	0.74	0.84	1.29	0.98
順序	0.13	0.17	0.43	0.64	0.71	0.94	0.95
提案	0.15	0.17	0.22	0.10	0.14	0.19	0.19

であると考えることができる。実験から確かめられたように、世代交代が進むにつれて、各変数のエントロピーは減少しており、深さ優先の傾向が強くなっている。つまり、提案方法は大域的探索から次第に、総作業時間の大きい仕事の相対的な順序関係を固定した局所探索に移行する探索アルゴリズムであると考えることができる。これが従来の方法に比べてよい解を得ることができた原因であると思われる。

2.3.7 他の探索方法との比較

本節では、シミュレーティッドアニーリングやタブーサーチとの探索能力の比較を行う。

1) シミュレーティッドアニーリング (SA)

本研究では,Osman and Potts[61] の方法を参考にした。シミュレーティッドアニーリングでは,たとえ現在より悪い解であってもある確率で採用することによって局所解からの脱却を図ろうとするものである。Osman and Potts[61] では,探索回数Kを次式で設定することを提案している。

$$K = max\{3300ln\ n + 7500ln\ m - 18250, 2000\}$$

ここで、nは仕事数であり、mは機械数である。したがって,100 仕事 20 機械の場合には約 20,000 回の探索となるが,遺伝的アルゴリズムとの比較を行うために処理時間をそろえた。*typeA* と *typeB* の問題では遺伝的アルゴリズムの処理時間に若干差が見られるが,それぞれの処理時間におおむね等しくなるように,試行錯誤的に探索回数を増やした。*typeA* の問題については,500,000 回,*typeB* の問題に対しては探索回数を 555,000 回とすることでおおむね等しい処理時間となった。

2) タブーサーチ

タブーサーチは、Taillard[76] の方法を参考にした。タブーサーチでは,タブーリストと呼ばれるものを使って,再び同じ解に戻ることを抑制している。本研究では,タブーリストの大きさは 20 とし,総完了時間をリストの内容とした。

Taillard は、近傍解を作るときにシフト操作を採用している。そして,ある要素を順列内の可能なすべての位置に挿入したときの総完了時間を $O(mn)$ で計算できる方法を提案している。したがって短時間の間に相当数の解を探索できる。本研究では,すべての要素についてすべての可能な位置に挿入した場合の総完了時間を計算することとし,これを 1 世代とした。したがって仕事数が 100 の場合には 1 世代で 10,000 個の解を探索することになり,10 世代で探索回数は 100,000 となる。シミュレーティッドアニーリングの場合と同様に処理時間を揃えることとした。*typeA* の問題の場合には,1,300 世代 (探索回数 13,000,000), また *typeB* の問題の場合には 1,700 世代 (探索回数 1,700,000) でおおむね等しい処理時間となった。

各方法の優位性を検証するために,30 回の試行中何回最もよい解を得たかを数えた。また統計的に検証するために t 分布を使った一対比較を行った。一対比較の検定統計量は,大きさ n の標本についての差 $x = x_1 - x_2$ の平均を $\overline{x}$, 標本分散を s^2 とするとき,

$$t = \overline{x}/(s/\sqrt{n})$$

となる。本研究では,有意水準を両側 5 %に設定している。標本数はいずれの検定においても 30 であるので,t 値の棄却域は $t < -2.045$ または $t > 2.045$ となる。

作業時間を 1 から 100 までの一様乱数で設定した *typeA* の仮説例では,シミュレーティッドアニーリングが最も良い解を見つけた回数は多かった。タブーサーチがそれについでいる。遺伝的アルゴリズムは提案方法も含めて一度も最も良い解を見つけることができなかった。しかしながら,仕事によって処理時間の長短に傾向を持たせた *typeB* の仮設例では,提案方法が良い結果を示している。30 回の数値例のうち 24 回は最も良い解を見つけた。

表 2.4: タブーサーチ (TS) やシミュレーティッドアニーリング (SA) との比較

typeA	探索回数	最良解回数	平均処理時間 (秒)
NEH	−	0	−
1 点順序	100,000	0	36.1
2 点順序	100,000	0	36.7
TS	13,000,000	9	40.7
SA	500,000	22	37.5
提案方法	100,000	0	36.6
typeB	探索回数	最良解回数	平均処理時間 (秒)
NEH	−	0	−
1 点順序	100,000	0	40.9
2 点順序	100,000	5	40.1
TS	17,000,000	0	43.0
SA	555,000	2	40.1
提案方法	100,000	24	40.1

表 3 に示した t 検定の結果からも同様のことがいえる。*typeA* の仮設例において最も良い解を見つけられたのはシミュレーティッドアニーリングである。他の方法に対して統計的に有意な結果が得られた。1 点順序交叉と 2 点順序交叉の間には有意な差が見られない。提案方法も 1 点順序交叉や 2 点順序交叉に対しては優れていることが確かめられたが, シミュレーティッドアニーリングやタブーサーチに比べると悪い。また, シミュレーティッドアニーリングはタブーサーチより優れているという統計的な検定結果が得られた。しかし,*typeB* の仮設例に対しては提案方法が最も優れているという結果が得られた。他のすべての方法に対して統計的に有意な差が見られた。

2.3.8 結論

本研究では、順列表現に階乗進数を使い、フローショップスケジューリング問題への適用を検討した。ここで、基準順列として、合計作業時間の昇順の順列を採用することによって、作業時間合計の大きい仕事相互の相対的位置関係を密に探索する遺伝的アルゴリズムを設計することができた。この考え方は、構成的ヒューリスティックスとして知られている NEH と共通するものである。つまり、遺伝的アルゴリズムによる探索に、NEH の考えを導入することが可能となった。また、階乗進数表現の染色体の交叉からは致死遺伝子が生じないことを利用して、一様交叉を採用した。これは、少ない世代交代のうちに集団が局所解に収束することを回避する効果があるものと思われる。これらのことから、従来提案されてきた遺伝的アルゴリズムに比べ、よい解を得ることができたものと思われる。

2.4 順序制約を持つフローショップスケジューリングへの応用

順列を階乗進数を使って表現することにより、ある種の順序制約を容易に表すことができる。本節では、階乗進数表現の順序制約つき順列の探索への応用を検討する。

表 2.5: 一対比較の結果

typeA	*NEH*	1 点順序	2 点順序	*TS*	*SA*
1 点順序	▼				
2 点順序	▼	−			
TS	▼	▼	▼		
SA	▼	▼	▼	▼	
提案方法	▼	▼	▼	△	△

typeB	*NEH*	1 点順序	2 点順序	*TS*	*SA*
1 点順序	▼				
2 点順序	▼	▼			
TS	▼	−	△		
SA	▼	▼	−	▼	
提案方法	▼	▼	▼	▼	▼

△上の方法が左の方法より平均が小さい (優れている)
▼左の方法が上の方法より平均が小さい (優れている)
有意水準両側 5 % (片側 2.5 %)

2.4.1 階乗進数による制約の表現

本研究では、あるグループに属する仕事は他のあるグループに属する仕事より先に処理されなければならないといういわば直列的な制約を考える。これはある特定の顧客からある期間内に受注した仕事は、その後に受注した仕事より先に処理しなければならないような場合に生じる。一般には、その他多くの顧客からの仕事も処理していることが考えられるので、順序制約が付与されているものと、順序制約がない仕事を同時に考慮してスケジューリングを行う必要がある。このような制約を持つものの 1 例を図 2.6 に示した。

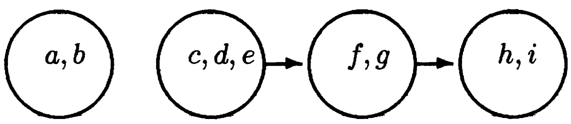


図 2.6: 先行ー後続関係の仮設例

図 2.6 において、仕事 c, d, e は仕事 f, g に、また仕事 f, g は仕事 h, i に先立って処理されなければならないことを表している。一方、仕事 a, b や仕事 c, d, e の間の順序制約はない。図 2.6 の制約を満たす順列の数は、

$$\frac{9!}{7!} \times 3! \times 2! \times 2! = 1728$$

である。このような制約を持つ順列を階乗進数を使って表現してみよう。

最初に基準順列を作る。基準順列は、まず制約のないものを最初に並べ、その後に直列的な制約を持つものを先行する順に並べて作る。図2.6の例に対しては、例えば $abcdefghi$ や $badcegfih$ などが基準順列として採用しうる順列である。ここでは、順列 $abcdefghi$ を基準順列としよう。

階乗進数表現の規則から、基準順列において j 番目の要素 x_j の係数 C_{n-j} がとり得る数は、$0 \leqq C_j < j$ である。従って、要素 a に対応する係数 C_9 のとり得る値は9未満の数である。仮に係数 C_9 の値として4をとったものとしよう。同様に、要素 b に対応する係数 C_8 も8未満の任意の数を取ることができる。仮に $C_8 = 5$ としたものとしよう。

次に要素 c の係数 C_7 を考える。階乗進数の条件から、C_7 は7未満の数となるが、順序制約があるために、いくつかの値をとることができない。要素 c は後続する要素として、要素 f、g, h, i を持っている。このため、これらの格納場所を確保しつつ要素 c の位置を定めなければならない。階乗進数の係数は、空位置の何番目に格納するかを定めるものであった。従って、後続変数の数だけ上限の値は小さくなる (図2.7)。つまり、直列的な制約がある場合には、階乗進数の各係数が満たすべき条件を次のようにまとめることができる。

変数 a の割り当て ($C_9 = 4, 0 \leqq C_8 < 8$)

				a				

変数 b の割り当て ($C_8 = 5, 0 \leqq C_8 < 8$)

				a		b		

変数 c の割り当て ($C_7 < 3$, × には配置できない)

			×	a	×	b	×	×

図 2.7: 変数の配置

(階乗進数の係数の条件)

基準順列において、j 番目の要素 x_j が fw_j 個の後続変数を持つ場合には、係数 C_j がとり得る数は、

$$0 \leqq C_j < j - fw_j$$

でなければならない。

この条件をもとに各係数の範囲を定めることができる。結果を表2.6に示した。階乗進数452000010は表2.6の条件を満たしている。また、これを順列に変換して $decfagbih$ を得ることができるが、これは図2.6の制約条件を満たしている。

さらに表2.6より、係数の可能な組み合わせは、

$$9 \times 8 \times 3! \times 2! \times 2! = 1728$$

となることがわかるが、これは制約を満たす順列の数と一致する。異なる階乗進数からは異なる順列が生成されることをあわせて考えると、階乗進数と制約を満たす順列は1対1に対応することがわかる。

階乗進数の係数が条件 $0 \leqq C_j < j - fw_j$ を満たしているものとしよう。この場合に交叉を行ったとしても、新たに作られる階乗進数の係数はやはり条件 $0 \leqq C_j < j - fw_j$ を満たしている。従って、階乗進数から得られる順列は順序制約を満たしていることになる。つまり、直列的な制約

表 2.6: 階乗進数の条件

係数	C_9	C_8	C_7	C_6	C_5	C_4	C_3	C_2	C_1
上限	8	7	2	1	0	1	0	1	0

の場合には、階乗進数を使って順列を表現することによって、同じ要素が複数現れる順列や、順序制約を満たさないものを抑制することができる。これは、遺伝的アルゴリズムを使った探索において、解を修正したり、罰金を与える必要がないことを意味している。

2.4.2 制約つき順列探索の数値実験

順列を使って染色体を表現した場合に、交叉において同じ要素を複数含む致死遺伝子は抑制できても、制約を満たさない順列を生成してしまう可能性がある。これに対処する方法として、1) 制約を満たすように順列を修正する方法と、2) 制約を満たさない程度に応じてペナルティを課す方法が検討されてきた [19][20]。本研究では、前者を修正法、後者を罰金法と呼ぶことにする。これらの方法の具体的手順を以下に示す。なお、これらの作業は染色体を順列を使って表現した場合に必要となるのであり、提案方法のように階乗進数を使って染色体を表現した場合には必要がないものである。

1) 修正法

遺伝的アルゴリズムを使った探索の過程で、順序制約を満たさない染色体が現れた場合に、制約を満たすように順列を一部手直しして完了時間を求める方法である。本研究では、順列の最初の要素から順に他の要素と比較を行って、制約を満たさない場合には置き換えることによって修正を行った。したがって、仕事数を n とした場合、$O(n^2)$ の計算量が必要である。なお、元の順列を修正後の順列に置き換えるという方法も考えられるが、飯間・三宮 [19][20] の研究を参考に、このような置き換えは行わないこととした。

2) 罰金法

探索過程で制約を満たさないものが現れた場合には、適合度の計算において何らかのペナルティを与えようとするものである。つまり、適合度 (完了時間) を次式を使って計算する。

$$\text{適合度} = z + \theta G$$

ここで、z は順序制約を考慮せずに求めた場合の完了時間であり、G は順序制約を満たさない個数 (要素の組み合わせの数) である。これは、順列の最初の要素から順に調べていって、順序制約を満たしていない場合を数えることによって求めた。したがって、計算量は仕事数を n とした場合、$O(n^2)$ である。罰金法の最大の問題点は、係数 θ をいくらぐらいに定めるかという点である。試行的な実験から、総完了時間の変動はせいぜい数 10 程度であったので、罰金のパラメータを 10 刻みで設定した。具体的には、0、$10, 20, 30, 40, 50, 60, 70, 80, 90, 100, \infty$ の 12 通りに係数を設定して、遺伝的アルゴリズムを実行した。そして、得られた解が最も良い場合を採用した。したがって、1 つの問題を解くために 12 回の実験を必要とする。

本研究では、次の 4 つのタイプの問題を対象に数値実験を行った。これらは、問題の規模、制約の強弱による違いを見るために設定したものである。4 つの仮設例を図 2.8 に示した。なお、図中の数は、仕事の番号を表す。基準順列はこの順番、つまり昇順に構成されているものとする。

最初のタイプの仮設例 *type* 1 では、仕事の数が 30 であり、うち 10 の仕事については制約はないものとする。また、10 個の仕事は残りの 10 個の仕事より先に処理されなければならないものと

type 1　30 仕事　制約なしの割合は 1/3

$\boxed{1\sim10}$　$\boxed{11\sim20} \rightarrow \boxed{21\sim30}$

type 2　50 仕事　制約なしの割合は 2/5

$\boxed{1\sim20}$　$\boxed{21\sim30} \rightarrow \boxed{31\sim40} \rightarrow \boxed{41\sim50}$

type 3　100 仕事　制約なしの割合は 7/10

$\boxed{1\sim70}$　$\boxed{71\sim80} \rightarrow \boxed{81\sim90} \rightarrow \boxed{90\sim100}$

type 4　100 仕事　制約なしの割合は 3/10

$\boxed{1\sim30}$　$\boxed{31\sim40} \rightarrow \boxed{41\sim50} \ldots \boxed{81\sim90} \rightarrow \boxed{91\sim100}$

図 2.8: 4 タイプの仮設例

する。*type* 2 の仮設例では仕事の数が 50 である。うち 20 については順序制約が付されていない。他の 30 の仕事は、これらの中では最初に処理されなければならない 10 個の仕事、ついで処理されなければならない 10 個の仕事そしてこれら 20 個の仕事の後に処理されなければならない 10 個の仕事から構成されているものとする。

type 3 と *type* 4 の仮設例はともに仕事の数が 100 であり、前者は制約が弱く 70 %の仕事については順序制約がないものとする。後者は逆に 70 %の仕事については順序制約が付されているものとする。いずれの場合においても、制約が付されている仕事は 10 個ずつのグループを構成し、直列的な制約が付されているものとする。

比較の対象となるアルゴリズムは次の 5 つである。

1) 両端部順序交叉+修正法
2) 両端部順序交叉+罰金法
3) 中央部順序交叉+修正法
4) 中央部順序交叉+罰金法
5) 提案方法

1) から 4) までのアルゴリズムでは、順列を使って染色体を表す。また、突然変異はシフト操作である。5) の提案方法では、階乗進数を使って染色体を表す。また、突然変異は係数を変化させることによって行う。

これらのアルゴリズムをうえで述べた 4 つのタイプの問題に適用した。このとき、それぞれのタイプ毎に 30 の異なる数値例を作った。各仕事の各機械での処理時間は 1 から 100 までの一様乱数を使って定めている。また, グループに含まれる仕事は総作業時間の小さい順に並び替えている。表 2.6 の係数の条件を満たすように乱数を使って階乗進数を作った。提案方法ではこれをそのまま親の初期集団として採用する。その他の方法においては、これを順列に変換したものを初期集団とした。実験結果を表 2.7 にまとめた。

各アルゴリズムの優劣を判定するために、3 つの指標を計算した。1 つ目の指標は、30 個の数値例に対して何回他のアルゴリズムに比べて良い解を得ることができたかを数えたものである。なお、同じ完了時間となる場合があるので、合計は必ずしも 30 と一致しない。2 つ目の指標は、一対比較の結果得られる t 値である。棄却域は両側 1 %に設定した。したがって、t 値が 2.750 より大きいとき提案方法から得られた解が有意に優れていると判断される。3 つ目の指標は計算時間である。実験では、 *Dell powerEdge* 300 *pentium* Ⅲ 600MHz *cpu* を用い、OS は $Windows$ 95、

表 2.7: 数値実験の結果

type 1	両端部順序交叉		中央部順序交叉		提案方法
	修正	罰金	修正	罰金	
最良回数 (回)	5	2	4	11	11
t 値	2.35	0.60	0.43	-1.85	—
平均時間 (秒)	2.43	31.50	2.37	17.97	1.37

type 2	両端部順序交叉		中央部順序交叉		提案方法
	修正	罰金	修正	罰金	
最良回数 (回)	1	0	6	0	23
t 値	7.45	12.32	3.83	8.96	—
平均時間 (秒)	9.637	125.17	9.50	71.50	4.60

type 3	両端部順序交叉		中央部順序交叉		提案方法
	修正	罰金	修正	罰金	
最良回数 (回)	0	0	5	0	25
t 値	11.77	15.06	6.02	9.94	—
平均時間 (秒)	53.20	810.27	51.60	373.23	20.97

type 4	両端部順序交叉		中央部順序交叉		提案方法
	修正	罰金	修正	罰金	
最良回数 (回)	0	0	1	0	29
t 値	16.85	35.70	8.02	38.30	—
平均時間 (秒)	77.73	669.90	68.37	379.73	20.63

注) 平均時間 : 30 仮設例の 1 例あたりの平均処理時間

プログラミング言語は C を使った。表 2.7 に示した値は 1 数値例あたりの計算時間 (単位秒) である。罰金法の場合には、12 種類のペナルティについて計算した場合の合計計算時間である。

仕事数 30 で、制約なしの仕事の割合を 1/3 とした *type* 1 の問題の場合、最良解を得ることのできた回数は、中央部順序交叉の罰金法と提案方法が 11 回と最も多かった。しかし、 t 検定の結果からは有意差は認められず、規模の小さな問題においては、優劣が見られなかった。

仕事数が 50 で、制約が 3 グループ間に設定されている *type* 2 の仮設例に対する実験結果からは、提案方法が他の方法よりも優れていることを確認することができる。つまり、30 回の試行中 23 回は、提案方法が最も優れた解を得ている。次いで良い結果を表しているものは中央部順序交叉の修正法であるが、回数は 6 回に留まっている。t 値を見てもすべての比較において正の値をとっており、かつ、2.750 よりも大きい。つまり、いずれの方法と比較しても提案方法から得られる解の総完了時間の平均は小さいということができる。

仕事数が 100 である *type* 3 の仮設例に対する数値実験からも提案方法の有効性を確認することができる。全体の 30 %の仕事に対して制約が付されている *type* 3 の仮設例に対する実験からは、30 回中 25 回において提案方法が最も良い解を得ている。中央部順序交叉の修正法が 5 回であり、

両端部順序交叉が最も良い結果を得た場合は一度もなかった。t 値をみてもすべて 2.750 より大きい値となっている。このことからも提案方法が良い解を得ていることがわかる。

全体の 7 割の仕事に直列的な制約が付いている *type* 4 に対する実験結果では、提案方法の優位性はさらに顕著に現れている。30 回の試行中 29 回において提案方法が最も良い解を得ている。t 値からも、提案方法から得られた解が、他の方法から得られた解の総完了時間よりも短いことがわかる。これらの実験を通して、規模が大きくなったり、また、制約が付されている仕事の割合が高いほど提案方法は優れた結果を示していることがわかった。

処理時間を比較してみよう。仕事数が比較的少ない 30 の場合には、修正法と提案方法の計算時間の差はそれほど顕著ではない。1 つの数値例に対する実行時間は 2 秒ないし 3 秒程度である。しかし、仕事数が増えるにしたがって計算時間の差は大きくなる。例えば、仕事数 100 の場合には、提案方法の計算時間が 20 秒程度であるのに対して、順序交叉を用いた場合の修正法は、おおむね 2 倍から 4 倍の計算時間を必要としている。罰金法の場合にはさらに多くの計算時間を必要としている。これは、順列を用いて染色体を構成した場合には、順序制約を満たさないものを数えたり、修正する必要があるためだと考えられる。階乗進数を用いた場合には、致死遺伝子が抑制されるのでこのような操作は不必要である。このことが計算時間の差の原因であると考えられる。つまり、計算時間の面からみてても提案方法が優れていることがわかる。なお、両端部順序交叉が中央部順序交叉に比べて計算時間が長いのは、子を作成する場合に一方の親の順番に遺伝子を並び替える必要が 2 回あるためだと思われる。

2.4.3 結論

本研究ではいわば直列的な順序制約があるフローショップスケジューリング問題を対象に、染色体を順列を使って表現する方法と階乗進数を使って表現する方法との比較を行った。階乗進数を使った場合には、同じ要素が複数現れるもののみならず、順序制約を満たさないものも自動的に排除できる。このため、計算時間が短くてすむという利点が予想されたが、実験を通してこのことを確認した。さらに、仕事数が 50、100 と増えた場合には、得られた解の総完了時間 (makespan) も他の方法より短いことがわかった。

順序制約を満たす順列は、可能な順列の部分集合となる。このため、染色体を順列を使って表現した場合には、交叉や突然変異によって致死遺伝子が生じないようにすることは困難である。飯間、三宮 [19][20] は、連続して投入できないという、本研究とは異なった制約条件を扱っているが、順列表現を用いたために致死遺伝子の発生を抑制することはできなかった。本研究で扱ったいわば直列的な制約を持つ場合でも、順列を使って染色体を表現した場合には致死遺伝子を抑制することは困難であろう。

2.5 因果モデリングへの応用

相関係数から因果のネットワークを構成しようとする試みは古くからある。特に Blalock が変数間に因果のパスを設定すべきかどうかということに関する操作的な方法を提案したことによって、因果モデリングは大きな前進をみた [6]。しかしながら、この方法も得られた相関係数からたちどころに因果モデルを作るというものではなく、思考実験を繰り返しながらよりよいモデルを得ようとするものである。その後、Blalock の方法を改善する試みがなされているが [77]、前提条件が厳しすぎるなど、改善の余地が残されている。　因果モデルを構成しようとするときの第 1 の問題点は、変数の数が多くなるにつれて想定しうるモデルが急激に増加することである。した

がってすべてのモデルについて比較検討することは不可能である。さらに、もうひとつのより深刻な問題は、得られた相関係数から因果のモデルが一意に定まらないことである。したがってモデルを特定するためには、変数間の先行ー後続関係に関する先験的な知識を必要とする。つまり、モデル探索も変数間の順序関係を考慮しつつ行わなければならない。本研究では、これら2つの問題を解決するために、階乗進数を使った順列表現とモデル探索について検討した。

2.5.1 因果モデリングにおける課題

Blalock の方法は、与えられた相関行列から因果モデルを構成しようとするものである。いま、Blalock が示した都市化の指数 (x_1) 以下 5 変数からなるモデルを考える。それぞれの変数間の相関係数の値は表 2.9 のようになっている。

表 2.8: Blalock の例における相関係数

変数	x_1	x_2	x_3	x_4	x_5
都市化の指数 (x_1)	1.000	− 0.389	0.670	0.264	0.736
非白人の割合 (x_2)		1.000	0.067	− 0.531	− 0.440
白人の所得 (x_3)			1.000	0.042	0.598
非白人の教育程度 (x_4)				1.000	0.386
非白人の所得 (x_5)					1.000

Blalock の方法では、まず候補となる因果モデルを設定する。次に因果のパスが設定されていない変数間の相関係数の値をモデルから推定し、これと観測値を比較する．両者の差が大きければモデルを修正することによってよりよいモデルを得ようとするものである。このような操作の結果、Blalock は図 2.9 を得ている。

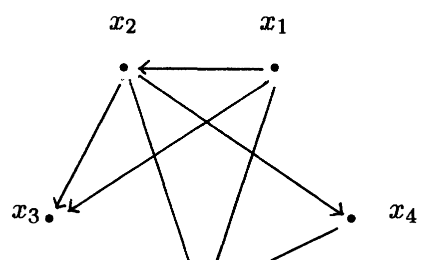


図 2.9: Blalock の因果モデル

ここで何らかの理由によって、5 変数間に次のような因果の向きが設定できるものと仮定しよう。

$$x_1 \rightarrow x_2 \rightarrow x_3 \rightarrow x_4 \rightarrow x_5$$

このとき、すべての因果のパスが設定された完全逐次モデルは次のように表される。

$$
\begin{aligned}
x_2 &= p_{12}x_1 + p_{u2}r_{u2} && (2.1)\\
x_3 &= p_{13}x_1 + p_{23}x_2 + p_{u3}r_{u3} && (2.2)\\
x_4 &= p_{14}x_1 + p_{24}x_2 + p_{34}x_3 + p_{u4}r_{u4} && (2.3)\\
x_5 &= p_{15}x_1 + p_{25}x_2 + p_{35}x_3 + p_{45}x_4 + p_{u5}r_{u5} && (2.4)\\
& && (2.5)
\end{aligned}
$$

ここで、p_{ij} はパス係数である。また、各変数 $x_i(i = 1, 5)$ および誤差 r_{ui} は、平均 0、標準偏差 1 となるように正規化されているものとする。また、構造方程式中の誤差変数が、同じ式の中の説明変数と無相関であることが仮定できるならば、通常の回帰分析の方法によってパス係数を求めることができる。もちろん完全逐次モデルには、不要なパスが含まれている可能性がある。Blalock は思考実験によってそれらのパスを取り除こうとした。構造方程式をそれぞれ独立な回帰式と考えると、それぞれの式は考えられるすべての変数を含んだモデルである。これらの中には従属変数を説明する上で不要なものも含まれている可能性がある。そこで、これらの変数を変数増減法を使って取り除くことを試みた。その結果次のような方程式系を得ることができる。

$$
\begin{aligned}
x_2 &= -0.39x_1 && (2.6)\\
x_3 &= 0.82x_1 + 0.39x_2 && (2.7)\\
x_4 &= -0.53x_2 && (2.8)\\
x_5 &= 0.35x_1 - 0.25x_2 + 0.38x_3 + 0.14x_4 && (2.9)\\
& && (2.10)
\end{aligned}
$$

これを図示すると図 2.10 のようになる。

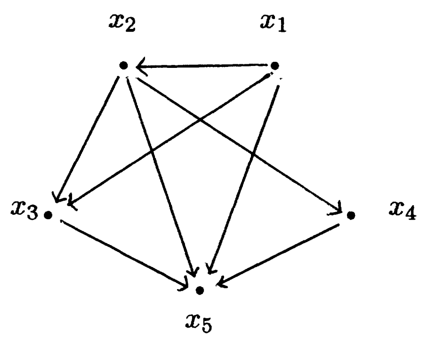


図 2.10: 変数増減の結果得られた因果モデル

得られたモデルを Blalock の結果と比較してみよう。Blalock の結果との違いは、x_3 から x_5 へのパスが設定されている点のみである。しかしながら、Blalock においても変数 x_3 から変数 x_5 へのパスの必要性が次の段階での検討課題として挙げられている。変数増減法による方法は、両者の間にパスを設定すべきであることを示している。つまり、完全逐次モデルからの変数増減法によるモデル推定によって、Blalock の方法と同じ結果を得ることができると考えられるであろう。

一方、図 2.10 において直接のパスが引かれていない、変数 x_1 と x_4、変数 x_3 と x_4 の間の相関係数の値を、モデルから推定することができる。表 2.9 に推定結果と観測値を示した。両者の乖離が小さいほど、得られた因果モデルは観測結果に適合していることになる。

表 2.9: 相関係数の推定結果

相関係数	推定値 1	観測値
r_{12}	0.206	0.264
r_{34}	− 0.038	0.042

モデルの適合度に関する総合的な指標は提案されていないが、仮に最大の乖離度でモデルの適合度を表すとすれば、図 2.10 で示される因果モデルの適合度は 0.042 − (− 0.038)=0.08 となる。ところで、得られたモデルは、

$$x_1 \rightarrow x_2 \rightarrow x_3 \rightarrow x_4 \rightarrow x_5$$

という先行ー後続関係を前提としていた。したがって、順列 $x_1x_2x_3x_4x_5$ の適合度が 0.08 であると解釈することが可能である。仮に、他の先行—後続関係が仮定されるならば、異なる因果モデルが導かれることは容易に想像できる。現実の観測結果にもっとも適合するモデルを探索するためには、変数の数を n として $n!$ 個の完全逐次モデルに対して上記の検討を行わなければならない。

モデルを探索する上でもうひとつの、より深刻な問題は、相関係数から因果モデルが一意に定まらないということである。例えば、偏相関係数 $r_{12.3} = 0$ という関係からは、次の 2 つのモデルを考えることができる。

$$x_1 \leftarrow \quad x_3 \rightarrow \quad x_2 \qquad\qquad x_1 \rightarrow \quad x_3 \rightarrow \quad x_2$$

図 2.11: $r_{12.3} = 0$ から得られるモデル

このことはやみくもに探索しても、解釈が困難な意味のないモデルを選んでしまう危険性があることを示唆している。これを避けるためには、発生の時間順などから先験的にわかっている順序関係を取り入れたうえで探索を行う必要がある。これはまた、順序関係に制約条件がついた順列の探索と同義である。制約条件がついた順列の探索方法に関するいくつかの提案がなされているが、より一般的な制約を扱うことができるように、本研究では階乗進数による順列表現を検討した。

2.5.2 階乗進数による順序関係の表現

基準順列に含まれる複数の要素を考える。これらの要素の中から 1 つの要素を選んだとき、他の要素がすべてこの要素より後位置にあるならば、この要素は先端に位置するという。たとえば、5 個の要素からなる順列を考える。基準順列として、降順の順列 $edcba$ を考えたものとしよう。このとき、要素 b は要素 c より後位置にあるという。また、要素 a、b、c、d の中では、要素 d が先端に位置するという。

先行ー後続関係は、有向グラフを使って表すことができる。本研究ではサイクルを含まないグラフを考える。また、自明なパスは除かれているものとする。自明なパスとは、そのパスがなくても先行ー後続関係が明らかなもので、そのパスの向きを逆にした場合にサイクルを生じるものをいう。また、要素 x から要素 y へ至る同一方向のパスの経路が存在する場合に、要素 x を要素 y の先行要素、要素 y を要素 x の後続要素と呼ぶことにする。特に、要素 x から要素 y へ直接のパスが設定されている場合に要素 x を要素 y の親、要素 y を要素 x の子と呼ぶことにする。

いま、任意の要素 x の順列内における位置を決定したいものとしよう。要素 x より先行する要素がある場合には、生成された順列においても、それらの要素より要素 x は後に配置されなければならない。また、後続する要素がある場合には、それらを格納する場所が確保されるように要素 x の配置を決めなければならない。

階乗進数と順列を対応させるだけならば、任意の順列を基準順列として採用できる。しかし、順序制約を満たす順列を生成しようとするときには、基準順列も適切なものを選ぶ必要がある。基準順列の候補はいわゆるトポロジカルソート [16] を使って得ることができる。したがって、基準順列において先行する要素は後続する要素の前に配置されているものとする。この性質に加えて、基準順列は次の性質も併せ持っているものと仮定する。

（基準順列に関する仮定）

基準順列に含まれる任意の要素を選び、その要素を x とする。要素 x が親を持つ場合に、それらの要素の中で任意の要素を $p(x)$ と表す。もし基準順列において、要素 $p(x)$ と要素 x の間に要素が配置されている場合には、それらの要素は次のいずれかであるものとする。

1) 要素 $p(x)$ の後続要素である。

2) 要素 x の先行要素である。

なお、自明なパスを除いているので、1) と 2) の 2 つの条件を満たす要素は存在しない。このとき、任意の要素の位置を決める手順を次のようにまとめることができる。

規則 3(順列の生成アルゴリズム)

任意の要素 x_q に対応する階乗進数の係数を C_{n-q} とする。また、要素 x_q に後続する要素の数を fw_q とする。

１． 後続する要素の数から、

$$0 \leqq C_{n-q} < n-q-fw_q$$

である。要素 x_q に先行する要素がない場合には、この条件を満たすように C_{n-q} を定めればよい。

２． 先行する要素がある場合

要素 x_q の親の中で、基準順列において先端に位置するものを x_p とする。また、基準順列において要素 x_p と要素 x_q の間に 0 個以上のの要素があるものとする。これらの要素を x_r とする。このとき、

2－1　$pnt = C_{n-p}$ とする

2－2　$C_{n-r} < pnt$ ならば $pnt = pnt - 1$

2－3　$C_{n-r} \geqq pnt$ でかつ x_r が x_q の親ならば $pnt = C_{n-r}$

2－4　すべての x_r について 2－2 および 2－3 を繰り返す

この結果求められる pnt に対して、

$$pnt \leqq C_{n-q} < n-q-fw_q$$

を満たすように C_{n-q} を定める。

この規則を満たす階乗進数を生成し得ることを以下に示す。大きさ n の基準順列で位置 q にある要素を $x_q(q=0,\ldots,n-1)$, 要素 x_q に対応する階乗進数の係数を C_{n-q} と表す。上のアルゴリズムによって順列を生成するためには、

$$pnt \leqq C_{n-q} < (n-q) - fw_q \tag{2.11}$$

となるような C_{n-q} が得られることが必要である。ここで、pnt は先行要素の位置から決まる下限値、また fw_q は要素 x_q の後続要素の数である。つまり、任意の要素に関して、

$$pnt < (n-q) - fw_q \tag{2.12}$$

が成り立つことを示す必要がある。

配置可能性を数学的帰納法を使って証明する。基準順列が n 個の要素から構成されているものとする。このとき、先頭の要素が配置可能であることは明らかである。そこで、すでに q 個の要素が順列内に割り当てられているものと仮定し、$q+1$ 個目の要素 (基準順列内の位置番号は q) の配置を考えているものとする。この要素を x_q とする。

要素 x_q の後続要素はすべて基準順列において要素 x_q より後位置に配置されているから、後続要素の数 fw_q は高々$n-q-1$ である。つまり、常に $(n-q)-fw_q>0$ である。これより、要素 x_q が先行要素を持たない場合には、$0 \leqq C_{n-q} < (n-q) - fw_q$ を満たす C_{n-q} を常に得ることができる。したがって要素 x_q が先行要素を持つ場合のみを考えればよい。

要素 x_q の親の中で、基準順列において先端に位置するものを x_p とする。要素 x_p に対応する階乗進数の係数を C_{n-p}、後続要素の数を fw_q とする。いくつかの場合に分けて証明を示す。

1) 基準順列において要素 x_p の直後に要素 x_q が配置されている場合。

手順から、$pnt = C_{n-p}$ とする. また、要素 x_p は配置可能であるという帰納法の仮定から、$C_{n-p} < (n-p) - fw_p$。要素 x_q は要素 x_p の子であるから、要素 x_q の持つ後続要素の数は、要素 x_p の後続要素の数より少ない。つまり、$fw_q \leqq fw_p - 1$ である。

また、基準順列において要素 x_p は位置 p に、要素 x_q は位置 q に格納されているので、$p=q+1$ である。したがって、

$$\begin{aligned} (n-q) - fwq &\geqq (n-p-1) - fw_p - 1 \\ &= (n-p) - fw_p \\ &> C_{n-p} \\ &= pnt \end{aligned} \tag{2.13}$$

となり、C_{n-q} を定めることができる。

2) 基準順列において要素 x_p と要素 x_q の間に要素が配置されている場合。

基準順列に関する仮定から、これらの要素は、要素 x_p の後続要素か、要素 x_q の先行要素かのいずれかである。まず、これらの要素がすべて要素 x_p の後続要素である場合を考える。

2 − 1) 基準順列において、要素 x_p と要素 x_q の間に配置されている要素がすべて要素 x_p の後続要素である場合。

基準順列において、要素 x_p と要素 x_q の間に配置されている要素を $x_r(r=p+1,\ldots,q-1)$、対応する階乗進数の係数を C_{n-r} と表す。生成される順列においても、要素 x_p の後続要素 x_{r} はすべて要素 x_p の後に配置される。つまり、$pnt = C_{n-p}$ としたとき、$C_{n-r} \geqq pnt$ であるが、仮定から x_{r} はいずれも要素 x_q の先行要素ではないので、pnt の値は更新されず C_{n-p} のままである。

基準順列において要素 x_p と要素 x_q の間に配置されている要素の数は $q-p-1$ である。要素 x_q および要素 x_{r} はすべて要素 x_p の後続要素であるから、

$$fw_q \leqq fw_p - (q-p-1) - 1 \tag{2.14}$$

である。さらに、要素 x_p がすでに配置可能であるという帰納法の前提から、

$$C_{n-p} < (n-p) - fw_p \tag{2.15}$$

である。したがって、

$$\begin{aligned} (n-q) - fw_q &\geqq (n-q) - fw_p - (q-p-1) - 1 \\ &= (n-p) - fw_p \\ &> Cn - p \\ &= pnt \end{aligned} \tag{2.16}$$

となり、C_{n-q} を定めることができる。

2 − 2) 基準順列において、要素 x_q に先行する要素が、要素 x_p と要素 x_q の間に少なくとも 1 個配置されている場合。

要素 x_q の親の中で、生成された順列において最後尾に配置された要素を $x_s (p < s < q)$ とする。したがって、上の規則 3(2 − 3) から要素 x_s を格納した時点で $pnt = C_{n-s}$ と更新される。

基準順列において、要素 x_s と要素 x_q の間には $q - s - 1$ 個の要素が配置されている。要素 x_s は要素 x_q の親であるから、基準順列に関する仮定から、これらの要素は要素 x_s の後続要素であるか、要素 x_q の先行要素である。このうち、要素 x_q の先行要素の数を m (≧ 0)、要素 x_s の後続要素の数を $q - s - 1 - m$ とする。

要素 x_s はすでに割当可能であるという帰納法の仮定から、

$$C_{n-s} = pnt < (n-s) - fw_s \tag{2.17}$$

である。さらに、要素 x_q は要素 x_s の後続要素であるから、

$$fw_q \leqq fw_s - (q-s-1-m) - 1 \tag{2.18}$$

である。また仮定から、要素 x_s は要素 x_q の親の中で最後尾に配置された要素であったので、生成された順列において、要素 x_q のm個の先行要素はすべて要素 x_s より前に配置される。したがって、これらの要素が順列内に配置されるたびに pnt の値は 1 つずつ減じられて、要素 x_q の配置を考えるときには、

$$pnt = C_{n-s} - m \tag{2.19}$$

となっているはずである。これらのことから、

$$\begin{aligned} & (n-q) - fw_q \\ \geqq\ & (n-q) - fw_s + (q-s-1-m) + 1 \\ =\ & (n-s) - fw_s - m \\ >\ & C_{n-s} - m \\ =\ & pnt \end{aligned} \tag{2.20}$$

となる。

2.5.3 順列生成の例示

再びBlalock[6] の例を考えよう。先に見たようにこれは5要素を含むモデルである。いま仮に、都市化の指数 (x_1) は非白人の割合 (x_2) と非白人の教育程度 (x_3) に先行する変数であり、非白人の割合 (x_2) は白人の所得 (x_3) に影響を及ぼし、日白人の教育程度 (x_3) は非白人の所得 (x_5) に影響を及ぼすということが先験的にわかっているものとしよう。このような先行ー後続関係を図示すると図 2.12 のようになる。モデルが規定しているのは、要素 x_1 は要素 x_2 と要素 x_3 に先行し、また、要素 x_2 が要素 x_3 に先行し、要素 x_4 が要素 x_5 に先行するということだけである。要素 x_2 と要素 x_4 の間の順序関係や、要素 x_3 と要素 x_5 の間の順序関係はなんら規定していない。なお、図 2.12 が規定している順序関係は、2.5.1 で想定した順序関係 $x_1 \rightarrow x_2 \rightarrow x_3 \rightarrow x_4 \rightarrow x_5$ を含んでいることに注意しよう。このような変数間の部分的な先行ー後続関係を満たしながら、観測された相関係数とできるだけ矛盾しない因果モデルを探索することを考える。

$$
\begin{tikzcd}
 & x_2 \arrow[r] & x_3 \\
x_1 \arrow[ur] \arrow[dr] & & \\
 & x_4 \arrow[r] & x_5
\end{tikzcd}
$$

図 2.12: 先行ー後続関係の仮設例

図 2.12 の先行ー後続関係から、まず要素 x_1 が基準順列の先頭に配置される。x_1 に対して子の関係にあるのは x_2 と x_4 であるが、いずれも x_1 のみを親として持つので基準順列に配置することができる。仮に x_4 を配置したものとしよう。次に配置し得るものは、x_2 と x_5 であるが、2.5.2 の基準順列の仮定を満たすためには、x_5 を配置する必要がある。ついで x_2、x_3 の順に基準順列に配置される。この結果、基準順列として、$x_1x_4x_5x_2x_3$ を得ることができる。この基準順列が 2.5.2 の基準順列に関する仮定を満たしていることは容易に確かめることができる。

順列生成においては、基準順列の最初の要素から配置すべき場所を決定していく。図 2.12 に対して基準順列として $x_1x_4x_5x_2x_3$ を採用したので、まず要素 x_1 の配置場所を決定する。

要素 x_1 に対応する階乗進数の係数は C_5 と表される。C_5 は階乗進数において、4!の係数に相当する。したがって、$C_6 < 5$。また、要素 x_1 は4個の後続要素をもつ。要素 x_1 に先行する要素はない。したがって、2.5.2 の規則 3(1.) から、$0 \leqq C_5 < 1$ となる。つまり $C_5 = 0$。このとき、x_1 は位置0に挿入される。

次に、要素 x_4 の配置場所を決める。要素 x_4 に対応する階乗進数の係数は $C_4(< 4)$ と表される。

$$C_5 = 0 \quad \begin{array}{|c|c|c|c|c|}\hline x_1 & - & - & - & - \\ \hline\end{array}$$

図 2.13: 要素 x_1 に対する係数の決定

後続要素の数は 1 である。さらに、要素 x_1 を先行要素としてもつ。基準順列において、要素 x_1 との間に他の要素はない ($pnt = C_5$、つまり $pnt = 0$)。したがって、2.5.2 の規則 3(2.) から、$0 \leqq C_5 < 3$ となる。仮に $C_5 = 1$ としよう。

$$C_4 = 1 \quad \begin{array}{|c|c|c|c|c|}\hline x_1 & - & x_4 & - & - \\ \hline\end{array}$$

図 2.14: 要素 x_4 に対する係数の決定

同様に要素 x_5 の係数 C_3 を決める。また、後続要素の数は 0 である。したがって、まず $C_3 < 3$。要素 x_5 は先行要素として要素 x_4 をもつ。$pnt = C_4$、つまり $pnt = 1$ とする。この結果、2.5.2 の規則 3(2.) から、$1 \leqq C_3 < 3$。仮に $C_3 = 1$ としよう。

$$C_3 = 1 \quad \begin{array}{|c|c|c|c|c|}\hline x_1 & - & x_4 & x_5 & - \\ \hline\end{array}$$

図 2.15: 要素 x_5 に対する係数の決定

次に要素 x_2 の配置場所を決める。要素 x_2 に対応する階乗進数の係数は $C_2(< 2)$ である。また、後続要素の数は 1 である。先行要素として x_1 を持つ。したがって、$pnt = C_5 = 0$ である。基準順列において要素 x_1 と要素 x_2 の間に要素 x_4 と要素 x_5 が配置されている。これらの要素に対応する階乗進数の係数はともに pnt より大きい。また、いずれの要素も x_2 の親ではない。したがって、pnt の値を更新する必要はない。つまり 2.5.2 の規則 3(2.) から、$0 \leqq C_2 < 1$ つまり $C_2 = 0$ である。

最後に、変数 x_3 の配置場所を定める。階乗進数と順列の対応規則から、$C_1 = 0$ とならなければならない。これを確かめてみよう。

後続要素はないから、まず $C_1 < 1$。また、親として、x_2 を持つ。したがってまず 2.5.2 の規則 3(2 − 1) から、$pnt = C_2$、つまり $pnt = 0$ とする。この結果、$0 \leqq C_1 < 1$ つまり $C_1 = 0$ となる。

この結果階乗進数として 01100 が得られる。これを順列に変換し、$x_1x_2x_4x_5x_3$ を得る。この順列が規定されている先行ー後続関係を満たしていることは容易に確かめることができる。

2.5.4 可能な順列の数え上げ

順序関係が規定されている場合には、条件を満たす順列の数は、可能な順列のごく一部になることが容易に予想される。条件を満たすものだけを階乗進数を使って数え上げる。

前節の検討から、$C_5 = 0$。さらに要素 x_4 に対応する階乗進数の係数 C_4 は 0 から 2 までの値をとり得ることがわかった。したがって、$C_4 = 0, C_4 = 1, C_4 = 2$ の 3 つの場合に分けて考えることができる。このような作業を繰り返すことによって可能なすべての場合を数え上げることができる。

規定された順序関係から、制約を満たす順列が少数となることが予想される場合には、階乗進数を使ってすべての場合を数え上げ、問題に応じた適合度を計算することによって、総当り探索

$C_2 = 0$ $\boxed{x_1 \mid x_2 \mid x_4 \mid x_5 \mid -}$

図 2.16: 要素 x_2 に対する係数の決定

$C_1 = 0$ $\boxed{a \mid d \mid c \mid b \mid f \mid e}$

図 2.17: 要素 f に対する係数の決定

が可能となる。それぞれの階乗進数に対応する順列と、2.5.1 で定義した適合度 (最大乖離度) を表 2.10 に示した。

表 2.10: 制約を満たす順列の数え上げ

C_5	C_4	C_3	C_2	C_1	順列	最大乖離度
0	0	0	0	0	$x_1x_4x_5x_2x_3$	0.689
0	0	1	0	0	$x_1x_4x_2x_5x_3$	0.696
0	0	2	0	0	$x_1x_4x_2x_3x_5$	0.816
0	1	1	0	0	$x_1x_2x_4x_5x_3$	0.616
0	1	2	0	0	$x_1x_2x_4x_3x_5$	0.463
0	2	2	0	0	$x_1x_2x_3x_4x_5$	0.078

表 2.10 から、可能な順列は 6 通りであることがわかる。それぞれの順列に対して、完全逐次モデルを想定し、変数増減法を使って不要なパスを除去する。得られたモデルで因果のパスが設定されていない変数間の相関係数を推定し観測値と比較する。その差の最大値をモデルの適合度とする。したがって、この場合には適合度の小さいものが望ましいモデルとなる。表 2.10 から、順列として、$x_1x_2x_3x_4x_5$ をとたもの、つまり Blalock が想定していたと思われる順序関係から得られるモデルが明かに優れていることを確認することができる。

2.5.5 条件を満たす順列の探索

順序制約を加味しても、総当り探索が事実上不可能な場合には、ヒューリステックな方法を採用する必要がある。SA などと同様に、階乗進数を使った制約条件付き探索でも、新たな近傍解を作る必要があるが、これは階乗進数の係数を変化させることによって達成可能である。

2.5.3 に示した順列 $x_1x_2x_4x_5x_3$ を作る過程で、それぞれの係数が満たすべき条件は表 2.11 にまとめられる。

新たな解を作るために、階乗進数の係数の中の任意のものを変化させる。例えば、係数 C_3 は 1 または 2 の値を取ることができるので、$C_3 = 2$ としてみよう。この結果、要素 x_5 に後続する変数に対する階乗進数の係数のとり得る範囲が変わってくるが、これも 2.5.2 の方法を用いて決めることができる。

条件のうち、上限値はその変数が持つ後続要素の数によってのみ規定される。つまり、先行要素の配置が変化しても、上限値は変わらない。したがって、下限値のみに注目すればよい。さらに、下限値が変化する可能性がある変数は、変化させた変数 (今の場合は変数 x_5) に後続する変数である。つまり、一部分の修正のみで別のモデルを得ることができる。

表 2.11: 階乗進数の条件

係数	C_5	C_4	C_3	C_2	C_1
上限	0	2	2	0	0
下限	0	0	1	0	0
採用	0	1	1	0	0

親の階乗進数の係数が変化したことによって、子の下限値が変化した場合でも、採用されている値がその条件を満たしているならば、子の係数を変化させる必要はない。これらのことから、ある変数の階乗進数の係数が変化した場合でも、修正は部分的な範囲で収まることが期待できる。

2.5.6 結論

因果モデルを作成するにあたっては、先験的に得られている順序関係を考慮しながら、モデルの探索を行う必要がある。直列制約を持つものを階乗進数で表現する方法を前節で検討したが、より複雑な制約条件を持つものも階乗進数で表現可能であることがわかった。先験的な順序関係が、2.5.2（規準順列に関する仮定）を満たすものであるならば、階乗進数を使った数え上げ、探索が可能であることがわかった。一方、順序制約を満たす階乗進数から他の階乗進数へ変換することが可能であることもわかったが、遺伝的アルゴリズムなどへ応用するまでには至らなかった。遺伝的アルゴリズムなどの発見的探索手法への応用は今後の課題としたい。

第3章　階乗進数と集合分割

分析対象をいくつかのグループに分類するという局面にしばしばわれわれは直面する。この目的のために従来から用いられてきたものに、クラスター分析がある。クラスター分析には階層的クラスタリングと非階層的クラスタリングがある。さらに、階層的クラスタリングには、最近隣法、最遠隣法など様々な方法がある [97]。階層的クラスタリングは、クラスターの生成過程を図示できるという長所がある反面、クラスターの形成をどこで停止すべきかという問題がある。また、最近隣法や最遠燐法など群の間の距離の定義によって、異なる結果が得られることがある。特に最近隣法によった場合、いわゆる鎖効果によって、直感とは異なるクラスターが形成されることもある。一方、非階層的クラスタリングの代表的な方法に k - $mean$ 法と呼ばれるものがある [46]。しかしこれにはあらかじめ群の数を指定しなければならないという問題がある。

われわれが提案しようとする方法も非階層的クラスタリングに属するものである。しかし、われわれが直面する問題では、あらかじめ群の数がわからないものが多い。この場合、最適な分割を、群の数を変化させながら探索しなければならない。つまり様々な群の数を許容しながら、最適な分割を探索する必要がある。したがって、k - $mean$ 法の適用は好ましくない。そこで、本研究では、階乗進数を使って、集合の分割を表現することとした。階乗進数を使うことによって、対象となる要素をいくつかの連結成分からなる森、つまりいくつかのグループとして表現することができる。また、階乗進数の係数のとり得る値を定めることによって、特定の2要素が同一群に属するか、または異なる群とすべきかという条件を容易に指定することができる。われわれはまず階乗進数表現を視覚的クラスタリング問題に適用した。さらに、階乗進数の係数が正という条件を付与することによって、すべての要素が連結している木に対応させることもできる。このことから、階乗進数表現を使い、与えられた形状を持つ木の探索を行うことができた。また、集合分割の特殊な応用例としてヒストグラムモデル [89] などへの適用について検討する。

3.1　階乗進数による集合分割の表現

階乗進数が与えられたとき、次の変換規則によって森へ変換する。

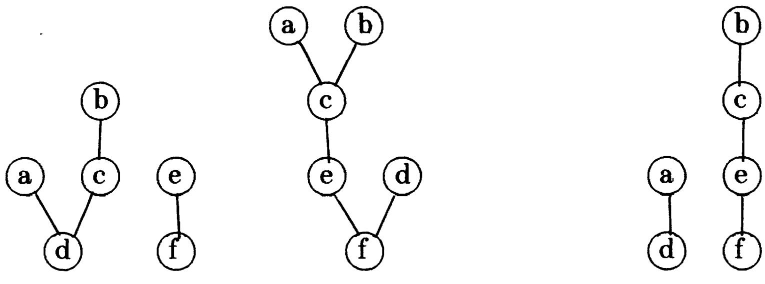
規則4　階乗進数から森への変換規則

異なる n 個の要素からなる配列 $x_k(k=0,\cdots,n-1)$ を考える。これを基準要素列と呼ぶことにする。また、n 桁の階乗進数を $C_nC_{n-1}\cdots C_2C_1$ と表す。このとき、基準要素列の要素 x_k は、要素 $x_{k+C_{n-k}}$ の要素と結びついているものと解釈する。

いま仮に基準要素列を、(a,b,c,d,e,f) としよう。また、図3.1の親1に示す階乗進数311010が与えられたものとしよう。基準要素列の最初の要素 x_0 は a であり対応する階乗進数の係数は3である。したがって、要素 a は x_{0+3}、つまり d と結びついているものと解釈する。同様に2番目の要素 $x_1=b$ に対応する階乗進数の係数値は1であるので、要素 b は要素 x_{1+1}、つまり要素 c と結びついていると解釈する。このような操作を繰り返すことによって、図3.2aに示すような木を生成させることができる。

基準要素列	a	b	c	d	e	f
親 1	3	1	1	0	1	0
親 2	1	1	2	2	1	0
子	3	1	2	0	1	0

図 3.1: 森を生成する階乗進数の例示



(a) 親 1　(b) 親 2　(c) 子

図 3.2: 階乗進数から作られた森

また同様の操作を行うことによって、図 3.1 の親 2 の階乗進数からは図 3.2b に示した森を生成することができる。図 3.2a に示した森は 2 つの連結成分からなっている。一方、図 3.2b の森ではどの 2 点も通路で結ばれているので、連結グラフである。生成された森がいくつの連結成分を持つかは、階乗進数の係数に含まれる 0 の数によって決まる。図 3.1 の親 1 に示した階乗進数には 0 が 2 個含まれている。このため、生成される森も 2 つの部分に分かれる。これに対して、図 3.1 の親 2 に示した階乗進数は最後の要素のみが 0 で、他の要素は正である。この結果、連結グラフが生成される。このことから、階乗進数の最後の要素を除いて、値が正となるような条件を与えた場合には、木を生成しうることとがわかる。また、係数の値として 0 を認めた場合には、全要素を、含まれる 0 の数だけの群に分割していることがわかる。

ところで図 3.1 の子 1 に示した階乗進数は、親 1 および親 2 の階乗進数から、いわゆる一様交叉によって生成したものである。子 1 に対応する木を図 3.2c に示した。図 3.2a、図 3.2b の木と比べるとかなり異なっているように見える。しかし、仔細に観察すると 3 つの木は共通した性質を持っていることもわかる。つまり、いずれの木においても 2 組の要素、$(b,c)(e,f)$ は結びついている。これは、親 1 および親 2 の階乗進数において、要素 b と要素 e に対応する係数値がともに等しいことに起因している。両親の係数値が等しいときには、一様交叉を行ったとしても同じ値が子に引き継がれる。また、階乗進数の係数値はどの要素と連結しているかを示すものなので、両親の係数値が等しい場合には、その係数に対応する連結関係は子に伝えられる。このことは遺伝的アルゴリズムによって探索を行った場合、要素間の結びつきという親の形質が子に遺伝することを表している。

3.2 視覚的クラスタリング

分析対象が、空間上でいくつかのかたまりとして視覚的に判然と認められるような場合を考える。このような場合、人間は視覚的にグループを認識する。これを視覚的クラスタリングと呼んでいる [58]。本節の目的は、人間が視覚的に行っているクラスタリングを自動生成するアルゴリズムを、階乗進数を使って実現することである。

岡田、加藤、小沢 [58] は、視覚的クラスタリングを行うために、総合距離尺度と呼ばれるものを導入した。これはある分割が与えられたときに、クラスター内距離とクラスター外距離を求め、その加重平均として定義される。なお、距離の計算に先立って、正規化という処理を行う。これは、個体間の最大距離 d_{max} を基準として、$d_{max}=1$、$0 \leqq d_{ij} \leqq 1$ となるような変換を意味する。ここで、i、j は任意の 2 つの個体を、d_{ij} はその 2 つの個体間の距離を表す。このとき、クラスター内距離和 D_I、クラスター外距離和 D_O は次のように表される。

$$D_I = \sum_{k=1}^{g} \sum_{i \in C_k} \sum_{j \in C_k} d_{ij}$$

$$D_O = \sum_{k=1}^{g} \sum_{i \in C_k} \sum_{j \notin C_k} (d_{max} - d_{ij})$$

ここで、C_k は k 番目のクラスター、g は群の数を表す。また、クラスター内距離和 D_I、クラスター外距離和 D_O を α $(0 \leqq \alpha \leqq 1)$ を使って加重平均したものが総合距離尺度である。

$$D = \alpha\ D_I + (1-\alpha) D_O$$

岡田、加藤、小沢 [58] は総合距離尺度が最小となる分割を試行錯誤的に求めている。まず、全体が 1 つの群に含まれている状態を考え、これを初期分割とする。そして、個体を順に 1 つずつ選び、それを独立した群、または他の群に含めた場合の総合距離尺度を計算する。そして最も総合距離尺度が小さくなるような分割を次の分割として採用する。このような操作を何度か行っても総合距離尺度に変化が見られなくなったとき探索を終了する。また総合距離尺度には未知のパラメータ α が含まれているので、さまざまな α に対して探索を行うことになる。

総合距離尺度を用いることによって、全体が 1 つのものから、それぞれの個体が独立した群であるものまでを統一的に扱うことができる。このような長所がある反面、以下のような問題点も残されている。第 1 には、探索アルゴリズムが局所解に陥りやすいということである。そこで本研究では、大域的な探索手法を導入するため、遺伝的アルゴリズムを利用することとした。

2 番目のより深刻な問題は、パラメータ α を合理的に決定する基準がないことである。クラスター内距離とクラスター外距離のいずれに重点をおくかは問題によって異なるであろう。クラス内距離を重視すれば (α の値が大きければ)、それぞれの個体は独立した群として認識されやすくなり、逆にクラスター外距離を重視すれば (α の値が小さければ) 全体は 1 つの群として認識されやすくなる。したがって、α の値によっては、分割数が異なる結果となる。岡田、加藤、小沢 [58] は、α を 0.001 刻みで変化させ、最も広範囲の α に対して得られる分割数を重視する方法を提案している。しかし、基本的には個体数だけの分割数が考えられるので、大量のデータを扱う場合には、何らかの方法で無意味は分割を抑制したり、逆に明らかに異なる群とすべき個体が同一のグループに入ることを抑制する必要がある。

図 3.4 に示したものは、岡田、加藤、小沢 [58] がアルゴリズムの検証に使った 2 次元の点パターンである。人間がクラスタリングを行うとすれば、おそらく対象を 3 つの群に分けるであろう。岡

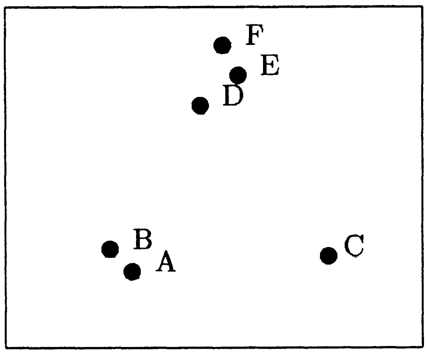


図 3.3: 6 点のクラスター問題

田、加藤、小沢 [58] は、α を 0.001 刻みで変化させ、得られる分割を観測した。その結果、最も広範囲の α（α =0.273〜0.777) に対して、期待される分割を得ている。探索回数に対する比率は 50.5 %である。しかし、全体が 1 つのものから、それぞれの個体が独立した群となるものまで他の 5 通りの分割すべてが得られており、それらの合計もやはり 50 %に近い [58]。

われわれが対象を分割しようとする場合、個体数よりも少ないいくつかの群に分けることを考えているだろう。つまり、全体が 1 つの群となることや、それぞれが独立した群となることを期待しているのではないだろう。この点からいえば、最も離れた 2 点が同一グループになることや、最も近い 2 点が異なる群になることは不自然である。そこで、このような分割を抑制することを考える。表 3.1 は、図 3.4 の点パターンから個体間の距離を測定したものである。

表 3.1: 要素間の距離

	A	B	C	D	E	F
A	0.000	0.424	2.608	2.377	2.953	3.231
B	0.424	0.000	2.902	2.247	2.860	3.089
C	2.608	2.902	0.000	2.625	2.683	3.130
D	2.377	2.247	2.625	0.000	0.640	0.000
E	2.953	2.860	2.683	0.640	0.000	0.447
F	3.231	3.089	3.130	0.854	0.447	0.000

グループの構成を行う前に要素間の結合に関して、いくつかの制約を設けることを考える。ここでは、視覚的な判断からいくつかの条件を設定してみよう。これらの点の中で最も離れているものは、(A, F) であることがわかる。次に離れているものは、(C, F) であり、(A, C) の距離も大きいことから、これら 3 点が同じ群に含まれないように調整することを考える。さらに、最も近いものは (A, B) であり、次に近いものは (E, F) であるので、これら 2 組の要素は同じ群に属するものとして探索を行うことを考える。

これらの条件を満たしつつ遺伝的アルゴリズムによる探索が行えるように、(D, E, B, C, A, F) を基準要素列として定める。異なる群としたい要素を基準要素列の最後に並べる。他の要素は任意の位置に置く。前節に定めた階乗進数から森への変換規則において、基準要素列の要素 $x_k(k = 0$、$\cdots$、$n-1)$ に対応する階乗進数の係数 C_{n-k} は、要素 x_k がそこから数えて何番目の要素と結合しているかを表すのであった。そこで、$C_{n-k} = 0$ とすれば要素 x_k が基準要素列において後に配

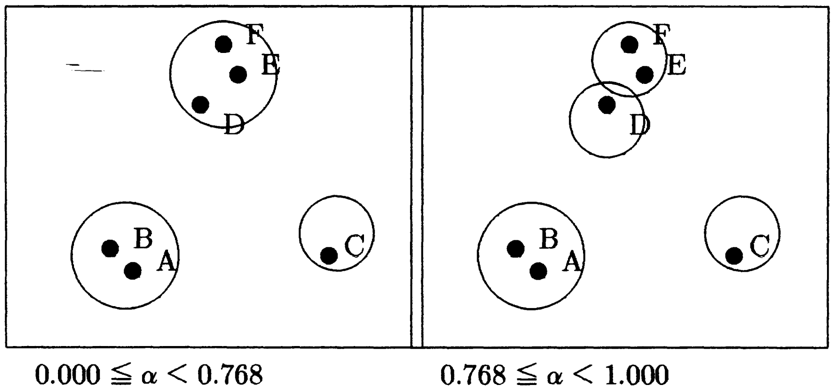


図 3.4: 階乗進数による探索結果

置されている要素と結合することはない。つまり、同一の群となることはない。いま、要素 C、A、F はそれぞれ別の群となるようにしたいので、これらの要素に対応する階乗進数の係数は 0 とおく。

また、要素 E は要素 F と結合するようにしたい。このためには E に対応する階乗進数の係数 C_5 を 4 とすればよい。この結果、要素 E は常に要素 F と結合することになる。同様に、要素 B に対応する階乗進数の係数 C_4 を 2 とすれば、要素 B は常に要素 A と結合することになる。この結果、階乗進数が満たすべき条件は次のようになる。なお、ここで - は階乗進数の係数の条件を満たす任意の値を取り得ることを示す。

表 3.2: 階乗進数の範囲

基準要素列	D	E	B	C	A	F
係数の範囲	-	4	2	0	0	0

階乗進数の係数の範囲から、可能な組合せは 6 通りにすぎないことがわかる。したがって、すべての場合について、総合距離尺度を計算することができる。なお、岡田、加藤、小沢 [58] にならって、α の値は 0 から 1 まで 0.0001 刻みで行った。その結果、次の 2 つの分割のみが得られた。要素の結合について、条件をつけた場合には、可能な群の数が 3 の場合と 4 の場合の 2 通りしか現れていない。岡田、加藤、小沢 [58] の実験においては、6 通りの分割が現れていたが、提案方法によると、すべてが 1 つのグループになったり、個々の要素が独立した群になるなどの場合を排除しうることを確認できた。これは、階乗進数の係数に条件を付したためであり、より規模が大きな問題にあっても適用できるものである。

3.3 木の生成

階乗進数の最後の係数 C_1 のみを 0 とし、他の要素は正とした場合には、全連結な木が得られる。これを利用して、あらかじめ与えられた形状の木の生成を試みた。従来、コンピュータグラフィックスにおいては、自然な木を生成するために、フラクタルや L システムを利用してきた。これら

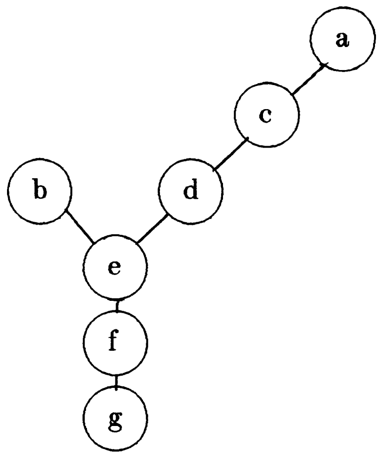


図 3.5: 階乗進数と木

の手法は、再帰的な方法を使って、現実的な木の生成を行うことに成功している。しかし、最終的にどのような形になるかは、あらかじめ予想することが難しいという問題がある。

基準要素列を (a,b,c,d,e,f,g) としよう。要素 g に対応する階乗進数の係数のみが 0 であり、他の要素に対応する係数は正とする。例えば階乗進数 2311110 はこの条件を満たす。これを図 3.5 にはこれから生成される木も合わせて示している。

図 3.5 から、要素 g が根になっており、要素 a と要素 b が葉になっていることがわかる。その他の点はいずれも 2 つ以上の連結している点を持っている（以後このような点を本研究では節と呼ぶことにする）。いま、要素 (a,b) は常に葉にしたいものとしよう。このとき、要素 (a,b) に対応する階乗進数の係数 (C_7,C_6) を次の条件を満たすように定めれば要素 (a,b,c) はいずれも葉とすることができる。

$$2 \leqq C_7 < 7,\ 1 \leqq C_6 < 6$$

要素 a に対応する階乗進数の係数 C_7 は常に 2 以上であるので、要素 b と結合することはない。つまり、要素 b は葉となる。このように葉としたい要素を基準要素列の最初に並べ、それらが相互に結合しないように階乗進数の係数値を使って調整することができる。そして、これらの要素にあらかじめ目標となる座標値を設定しておく。この座標値にできるだけ近くなるような階乗進数を探索することによって、あらかじめ定められた形状の木を生成することができる。

3.4 ヒストグラムモデル

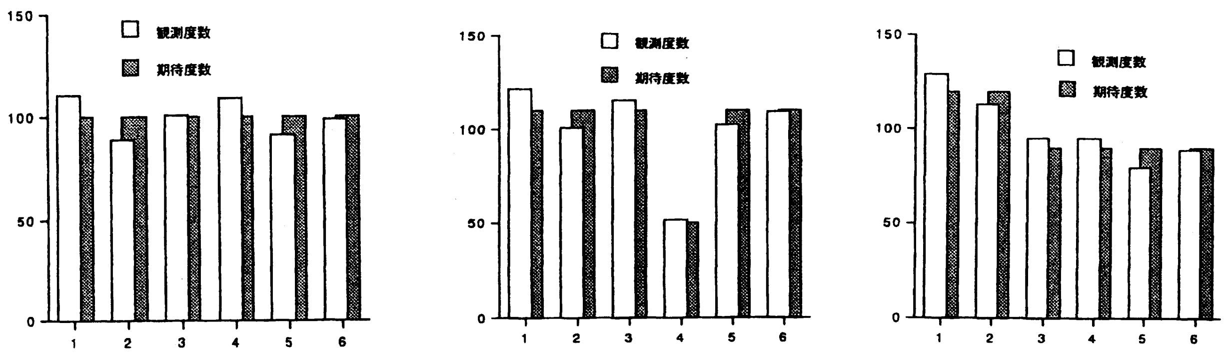
品質管理などにおいて、観測や実験から得られたデータを度数分布の形にまとめることはよく行われることである。結果の概略を見るだけでなく、ＡＢＣ分析などの基礎資料として用いられることも多い。しかし、一般には観測結果には誤差が含まれるために、得られた度数分布が真の分布を表すことは希である。この結果、観測される度数分布から定性的な知識を得ることが困難な場合も生じてくる。本節の目的は、観測された度数分布を簡略化することによって、特徴の把握を容易にしようとするものである。

3.4.1 さまざまな度数分布

まず1個のサイコロを考えよう。サイコロの目は等確率に出現するものとする。このサイコロを600回投げて、目のでた数を記録したとしよう。期待度数はそれぞれの目について100であるが、観測される度数にはばらつきが見られることが普通である。仮に図3.6(1)に示されるような結果が得られたものとしよう。それぞれの期待度数が等しいことから、このような度数分布を平均型の度数分布と呼ぶことにする。

この観測結果から知りたいことは、1の目のでる確率が0.185であり、2の目のでる確率が0.148であるといったことではない。知りたいことはあくまで各々の目のでる確率は等しいということである。つまり、知りたいことは観測度数の背後に隠れている分布、つまり期待度数からなる分布である。

また、別のサイコロを考えよう。このサイコロは目4の出現確率のみが他の目の出現確率より小さいものとする。このように、凹凸のある度数分布は連続的な量をいくつかの階級に区分したときに現れる一般的な分布である。このように凹凸のある分布を凹凸型と呼ぶことにしよう。図3.6(2)に600回のサイコロ投げを行った時の観測度数と、期待度数を示した。図3.6(2)では目4の出現確率が他の目の出現確率より特に低く設定しているので、強凹凸型と呼ぶことにする。もちろん、強い、弱いは相対的なものであり、便宜的な名前である。これから述べるように取り扱い、探索方法に違いはない。



平均型の度数分布　　強凹凸型の度数分布　　階段型の度数分布

図 3.6: 様々な度数分布

本研究では、さらにもうひとつ別の分布モデルを考える。ＡＢＣ分析のように、属性に関する度数分布を扱う場合には、度数の大きい順に並び替えて階段型の度数分布に変形することがある。このような例は、サイコロの目のでる確率は等確率ではなく、1と2の目のでる確率が高いという場合に対応させることができる。前と同じく600回のサイコロ投げの実験を行って図3.6(3)のような観測結果を得たものとしよう。1と2の目のでる確率が他の目の出現確率より大きくなっているので、階段型の度数分布と呼ぶことにしよう。ここでも、知りたいことは1の目がでる確率は0.215であり、2の目のでる確率は0.188であるというものではないであろう。出現確率の大きい1と2、およびそれ以外のグループに分かれるというものであろう。

つまり、本節での目的は、図3.6のような観測結果が得られたときに、出現確率が等しいと思われる事象をまとめることによって、それぞれ背後にあると思われる度数分布を推定することであ

る。そして、出現確率が等しいと思われる事象をまとめて、よりばらつきの少ない度数分布を作成することを度数分布の要約と呼ぶことにする。

3.4.2 情報量規準と度数分布の要約

赤池の情報量規準（AIC）を使って、度数分布の異なる階級をまとめた方が良いのか、それとも別々の出現確率を持った階級であると考えた方が良いのかということを判定する方法が提案されている [70]。度数分布の要約の場合、AIC は次式で計算される。

$$AIC = -2\sum\{(\text{観測度数})log(\text{出現確率の推定値})\} + 2(\text{階級数} - 1)$$

図の結果から AIC を求めてみよう。観測結果は、それぞれの事象がそれぞれの出現確率を持った多項分布から得られたものであると考えることができる。図の観測結果は、すべての目の出る確率が異なったさいころから得られたものであると考えてみよう (モデル 1a)。それぞれの目の出る出現確率の推定値は最尤法によって求めることができるが、それは観測度数を合計で除した値と等しくなる。表にはその結果も示している。この出現確率の推定値を使って AIC を求めると 2156.196 となる。

一方、図の観測結果はすべての事象の出現確率が等しい多項分布から得られたものと考えると (モデル 1b)、それぞれの事象の出現確率は等しいものとなり、その推定値は最尤法により 0.167 となる。この出現確率の推定値を使って AIC を計算すると 2150.111 となる。

AIC によるモデル選択は、AIC が小さいほうのモデルをより好ましいモデルとして選ぶ。したがって、図の観測結果に対しては、すべての事象の出現確率は等しいと考えるモデルの方が、より好ましいモデルとして選択される。

3.4.3 遺伝的アルゴリズムを使ったヒストグラムの選択

情報量規準を用いて、より好ましいモデルを選択しようとするとき、階級数（本研究では事象の数）が増加するにしたがって、考えることのできるモデルの数は急速に増加していくので、効率的にモデルを選択する方法が必要になってくる。このため、遺伝的アルゴリズムを利用する方法を検討する。

遺伝的アルゴリズムを使って探索を行うためには、様々に要約された度数分布を染色体として表現する必要がある。本研究では、階乗進数を使って度数分布を表現することとしたが、離れた階級をまとめることを認めるかどうかで、次の 2 つの場合を想定することができるであろう。

最初は、事象の順番に特に意味がない場合である。サイコロモデルもこのような例と考えることができる。図 3.6 の (1) や (2) の例がこれに該当する。このような場合には、離れた位置にある階級の結びつきも許容する必要がある。したがって、n 桁の階乗進数 $C_nC_{n-1}\cdots C_2C_1$ の係数の条件は、

度数分布要約の条件 (1)　$0 \leqq C_j < j$

となる。この条件によって、任意の階級は他の任意の階級と結びつくことが可能である。例えば、図 3.6(2) のように目 4 の出現確率のみが他の目の出現確率より小さいと考えるモデルは階乗進数を使って、112010 と表現することが可能である。

一方、ABC 分析などのように、事象が出現確率の大きい順などに並び替えられている場合もある。時系列的に観測された結果を度数分布に表したような場合も事象の出現順は意味を持つ。サイコロモデルでは図 3.6(3) の場合がこれに該当する。このような場合の度数分布の要約は隣り合ったものをまとめるべきかどうかということが問題となる。つまり、離れた階級を結びつけることを考える必要はない。したがって、このような場合において、階乗進数の係数が満たすべき条件は次のように記述できるであろう。

度数分布要約の条件 (2)　$0 \leqq C_j < 1$

すべての階乗進数の係数は 0 または 1 の値のみをとる。そして 1 の値を取る場合には、当該階級は次の階級とまとめられることを表す。

3.4.4 到着分布への応用

階級の並びに意味がある場合、つまり階乗進数の係数の条件が $0 \leqq C_j < j(j = n, \cdots, 1)$ で与えられる場合を考える。例としてを到着分布の要約を考える。図 3.7 は待ち行列のシミュレーションを行う目的で、本学食堂の券売機へ到着した人数を観測したものである。観測日は 1988 年 11 月 15 日であり、午前 11 時から午後 1 時半までの昼食時間帯に測定を行った。現在は複数の券売機が設置してあるが、当時は 1 台のみであった。この券売機への到着状況を 1 分単位に記録した。1 人で測定したので、実際には 1 分間の到着人数を測定し、測定結果を次の 1 分間の間に記録するという方法によっている。したがって、実際の到着人数は、観測された数のおおむね 2 倍程度になるものと予想されるが、分析では生の観測結果を用いる。

このような観測結果から、客の到着分布を求めることを考える。食堂の到着にみられるように、単位時間内の到着人数が時間とともに変化する場合がある。したがって、それぞれの時間帯で異なる平均到着率を求める必要があるが、観測された結果を直接用いると変動が大きくなる恐れがある。このため観測結果から平均到着率が等しいと考えてもよい範囲を求める必要が生じる。

また、最大の階級数はデータ数を n として、階級数は $2\sqrt{n}-1$ に抑えるべきであるとされるので [70]、観測された結果を 6 分毎にまとめた度数分布を分析対象とした。

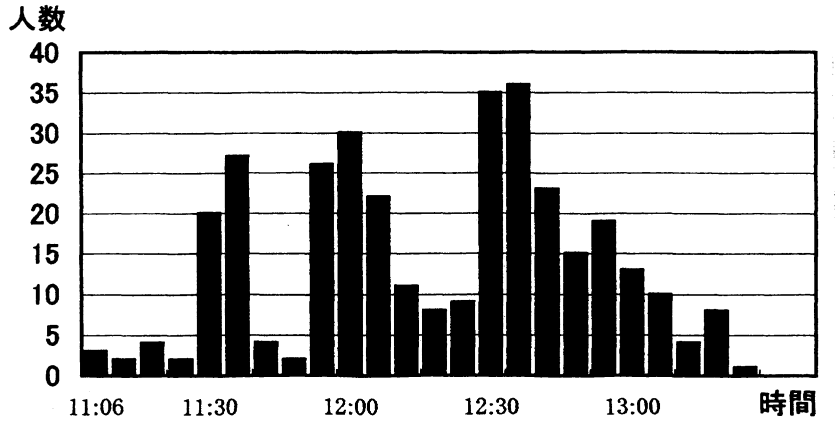


図 3.7: 食堂への客の到着

11 時 6 分から 13 時 30 分の間に 334 人の客の到着があった。図 3.7 には 24 の階級が含まれてい

るので、24 桁の階乗進数を用意する。それぞれの係数がとりうる値は 1 または 0 である (ただし、C_1 は常に 0)。親の集団は 30 とした。交叉は 15 組に対して行い、ランクによって親の選択を行った。突然変異は取りうる数を乱数を使って発生させ、新しい係数値とした。世代交代数は 1000 である。要約後の度数分布を 3.8 に示した。

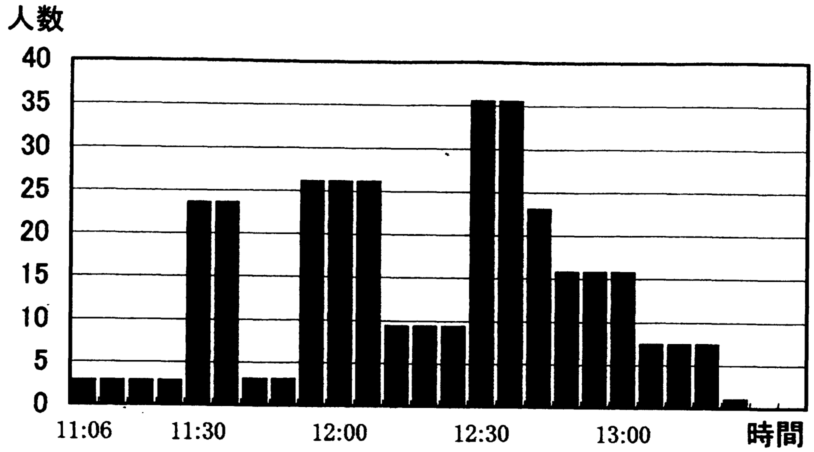


図 3.8: 券売機への到着分布の推定

図 3.8 から、単位時間当たりの到着人数が多い山の部分が 3ヵ所あることがわかる。最初の山は 11 時 30 分からであるが、これは本学事務職の昼休みがこの時間から始まるためだと思われる。次の山は 12 時ぐらいにみられるが、これは大学の 2 時限目が終了する時刻であり、次の 12 時 30 分からの山は、付属高校が昼休みに入るためと思われる。このように、推定された度数分布は、容易に解釈可能であり、券売機への到着分布の特徴をよく表しているものと思われる。また、事務職と高校の場合には、それぞれ休み時間の開始時刻からピークが始まっているのに対して、大学生の場合には、休み時間の少し前からピークが始まることもわかる。これらの分析から、推定された度数分布を検討することによって、特徴をより的確に把握することができるものと思われる。

3.4.5 長崎の観光地人気調査結果への応用

ABC 分析のように、出現度数の大きい順に並び替えられた度数分布の例として、長崎の観光地の人気調査結果を考える。長崎の観光地を 6ヵ所を選び、学生を対象に 1996 年 6 月に人気調査を行った。調査では、6ヵ所のうち行ってみたいところまたはもう一度行ってみたいところを 1 つだけ選んでもらった。回収総数は 234 通であり、うち有効回答数は 208 通であった。調査結果を表 3.3 に併せて示している。なお、入学後の学生生活の長短によって、1、2 年生の回答結果と 3、4 年生の回答結果を、それぞれ低学年および高学年と区別して表している。全体として、ハウステンボスの人気が高そうであるが、それ以外の観光地については、低学年と高学年では差がみられるようなので別個に分析を行った。

この例では階級数は 6 である。したがって 6 桁の階乗進数を用意すればよい。親の集団の大きさは 30 とし、毎世代 15 組も交叉を行った。世代交代数は 100 である。

また、観光地の人気にどのような傾向が見られるかを知りたいので、度数分布の要約のみではなく、その信頼性を評価することを考えた。本研究では、リサンプリング [62] によって信頼性を評価することとした。リサンプリングでは、得られた標本を母集団と考え、そこから乱数を使っ

表 3.3: 長崎の観光地の人気調査結果（単位：人）

観光地	低学年	高学年	合計
平戸生月	5	20	25
佐世保九十九島	14	5	19
ハウステンボス	83	35	118
浦上平和公園	2	3	5
グラバー園オランダ坂	11	3	14
雲仙島原	12	15	27
合計	127	81	208

て同数の標本を新たに作り、各種の統計量の計算を行う。本研究の場合、低学年の場合 127、高学年の場合 81 の標本から乱数を使って同数の標本を選ぶ。乱数を使っているので、複数選択されるものもあれば、一度も選択されないものもある。したがって、度数はその都度異なった数となることが予想される。このような作業を何回も繰り返して注目している統計量、例えば平均や分散の変動などを調べる。本研究の場合、信頼性を評価したいのは、平均や分散などではなく、分布の形そのものである。

図 3.4 に 1,000 回のリサンプリグを行った結果を示した。観測結果を人気順に並び替え階段型のデータとして分析した。低学年の場合には 16 パターン、高学年の場合には 22 パターンが現れたが、図 3.4 には、低学年と高学年それぞれの場合において出現回数の多い 8 パターンを表示している。なお、低学年の場合の基準順列は左から順に、ハウステンボス、佐世保九十九島、雲仙島原、グラバー園オランダ坂、平戸生月、浦上平和公園であり、高学年の基準順列は左から順に、ハウステンボス、平戸生月、雲仙島原、佐世保九十九島、グラバー園オランダ坂、浦上平和公園である。

すべての観光地の人気に差がないとするパターンは染色体形式で (111110) と表現される。リサンプリングの結果、(111110) のパターンは低学年、高学年とも 1 度も現れなかった。このことから、観光地の人気には差があることがわかる。逆に、すべての観光地の人気に差があるとするパターンは (000000) と表現されるが、低学年において 6 回、高学年において 11 回しか現れなかった。これらのことから、6ヵ所の観光地は人気の面から幾つかのグループに分かれることが推定できる。

リサンプリングの結果、最も頻繁に現れたパターンは、低学年では (011010) であり、これは学生の人気が高い順に、{ハウステンボス}、{佐世保九十九島、雲仙島原、グラバー園オランダ坂}、{平戸生月、浦上平和公園} の 3 グループに分かれることを意味している。それぞれのグループ内では人気に差がないことを表している。例えば、低学年では雲仙島原とグラバー園オランダ坂の間に人気の差はないが、雲仙島原と平戸生月では人気に差があることを示している。

高学年では {ハウステンボス}、{平戸生月、雲仙島原}、{佐世保九十九島、グラバー園オランダ坂、浦上平和公園} の 3 つに分けられるというパターンが最も頻繁に現れている。。低学年では知名度が要因となっているものと推定されるが、高学年では、温泉や歴史、ドライブコースなどの要因が重視されているものと思われる。

観測結果からは、低学年、高学年ともにハウステンボスの人気が高いと思われるが、ハウステンボスの人気が、他の観光地より高いとするパターンは、ともに (0****0) と表現される。ここで、*は 1 または 0 のいずれでもよいことを表している。低学年では、リサンプリングから得られた 1,000 のパターンすべてが (0****0) の形式であった。これに対して高学年では、(0****0) と

表 3.4: 観光地人気調査結果からのリサンプリング

染色体	回数	染色体	回数
0 1 1 0 1 0	356	0 1 0 1 1 0	346
0 1 1 0 0 0	146	0 1 0 0 1 0	178
0 1 0 1 0 0	102	0 0 0 1 1 0	130
0 1 1 1 0 0	101	0 1 0 1 0 0	90
0 0 1 0 1 0	85	0 0 0 0 1 0	54
0 0 1 1 0 0	69	0 0 0 1 0 0	39
0 1 0 1 1 0	36	1 0 0 1 1 0	33
0 1 0 0 1 0	34	1 0 0 0 1 0	31
低学年の場合		高学年の場合	

なったパターンは 911 回であった。このことから、高学年では、ハウステンボスの人気はやはり高いものも、他の観光地、特に平戸の人気と同程度と考えられる場合も十分想定し得ることがわかった。高学年で、さらに雲仙まで同程度の人気であると考えるパターンは (11***0) と表現されるが、この出現回数は 3 回にとどまっている。このことから、雲仙島原までハウステンボスと同程度の人気だと考えることには無理があることがわかる。

低学年と高学年の回答結果で大きく異なっているのは、佐世保と平戸の人気である。低学年では佐世保の人気が高いが、高学年では平戸の人気が高い。人気に差があるかどうかをリサンプリングの結果から分析してみよう。

まず、低学年において、佐世保が平戸より人気が高いとするパターンは (*+++**) と表現される。ここで+++は、いずれかの遺伝子座において 0 がたつことを意味している。1,000 回のリサンプリング中、896 回がこのパターンであった。一方、高学年において平戸が佐世保より人気が高いとするパターンも (*++***) と表現できる。1,000 回のリサンプリング中 998 回がこのパターンであった。これらの分析から、低学年と高学年では、佐世保と平戸の人気に逆転が起きていることを、相当程度の確信をもって推定することができる。

3.4.6 結論

本節では、観測された度数分布から、その背後にある真の分布を推定するために、遺伝的アルゴリズムと情報量規準を用いた方法を検討した。食堂の券売機への到着分布を対象として分析した場合には、生の観測結果よりも解釈が容易になることもわかった。

本節の後半では、得られた分布の信頼性を評価するために、リサンプリングによるシミュレー

ションを検討した。長崎の観光地の人気調査結果を対象に分析を行い、低学年と高学年では人気に差があることがわかった。また、いくつかの遺伝子の値を任意としたり（本文＊）、連続する遺伝子座の中のいずれかに特定の値をとる（本文＋）ものなどを考えることによって、さまざまなパターンの出現確率も容易に計算できることがわかった。

第4章　分割表からの規則抽出

近年、観測されたデータから知識を得ようとする研究が活発になされてきている。工学、医学などにおいては、これらの主な目的は予測にあるように思われる。しかし一方、説明を行う目的で知識を得たい場合もある。次章では、どのような企業がどのような会計方法を選択しているのかを記述する少数の規則の抽出を検討している。このような場合には、知識を得る目的は予測ではないであろう。

しばしば規則は決定木を用いて表される。ID3[64]、C4.5[65]、See5/C5.0[66] などが効果的に決定木を構成する方法として提案されている。しかし、決定木から知識を得ようとする場合には、根の変数が変わると得られる知識が大きく変わってしまう危険性がある。このことは予測が目的の場合にはそれほど大きな問題とはならない。しかし、解釈が目的の場合には重大な問題である。さらに決定木では同程度の説明力がある変数があった場合に、一方をモデルに含めない傾向がある。このことも予測が目的である場合には問題とならないが、解釈が目的の場合には、同程度の効果をもった変数も含めていた方が、解釈が容易になる場合もある。一方、遺伝的アルゴリズムを使って知識を獲得しようとする研究もなされている。しかしながら探索空間が莫大なものとなったり[24]、[52]、またあるものは、あらかじめ規則の数を定めておく必要がある [49]。

本章では集合分割を使った知識獲得のための異なる方法を提案する。本研究ではいくつかの選択肢から選ばれたアンケート調査結果を分析対象とし、観測結果を分割表にまとめることができるものと仮定する。したがって、ID3 とおなじく離散的な値をとる変数を考える。本研究でも、ID3と同じく分割表の階級をいくつかのグループに分けることによって少数の知識を得ようとするものであるが、ID3 が対象を細分化しながら規則を記述していくのに対して、本研究では分割表の階級を融合することによっていくつかのグループを構成する方法をとった。

4.1　情報量規準による分割表の選択

問題としている属性 (被説明変数) の値を予測・推定するために、いくつかの説明変数を設定してアンケート調査を実施したものとする。説明変数も名義的または順序的な尺度の場合には、アンケート調査結果はクロス集計され分割表の形にまとめられる。一般に説明変数は多数設定される場合が多いから、被説明変数を予測・推定する目的でさまざまな分割表を作成することができる。

仮にある属性によって 2 群 A、B に分けられる観測対象があったとしよう。そして、いずれの群に属するかを予測・推定するために、2 つの説明変数 x、y を設定して調査を実施したものとする。2 つの変数ともに big または $small$ を値としてとるものする。このとき、群を判別する目的でさまざまな分割表を作成することができる。表 4.1 は説明変数 x による分割表であり、表 4.2 は説明変数 y による分割表である。

また、説明変数として両方の変数をとった分割表を作成することもできる。これを表 4.3 に示す。

さらに、表 4.3 の分割表において、いくつかの階級をまとめて表 4.4 に示すような分割表を作成することも可能である。表 4.4 の分割表は変数 y の値が $small$ である場合に、変数 x をまとめた分割表となっている。

表 4.1: 説明変数 x による分割表

変数 x	群 A	群 B	合計
big	73	6	79
small	50	42	92
合計	123	48	171

表 4.2: 説明変数 y による分割表

変数 y	群 A	群 B	合計
big	33	41	74
small	90	7	97
合計	123	48	171

表 4.1 から表 4.4 までみてきたように、群を予測する目的で、さまざまな分割表を作成することができる。このため、どのような変数を説明変数として選ぶか、また、階級の一部を要約すべきかどうかの選択が必要となる。前者が説明変数選択の問題であり、後者が区分選択の問題である。

変数選択と区分選択を統一的に行うために，赤池の情報量規準 (AIC) を利用する方法が提案されている [70]、[71]。分割表の選択における AIC は次式で計算される。

$$AIC = (-2)\sum_{i,j} n(i,j) log \frac{n \cdot n(i,j)}{n(i) \cdot n(j)} + 2(K_i - 1)(K_j - 1)$$

ここで、i は分割表の階級を、j は群を表す。したがって、$n(i,j)$ は観測度数を示す。また、$n(i), n(j)$ はそれぞれ周辺分布を表し、n はデータの個数を表す。また K_i は分割表に含まれる階級の数を表し、K_j は群の数を表す。

表 4.4 の分割表に対して AIC を計算すると次のようになる。

$$\begin{aligned} AIC(Table4.4) &= (-2)(18 log \frac{18 \cdot 171}{123 \cdot 20} + 2 log \frac{2 \cdot 171}{48 \cdot 20} + 15 log \frac{15 \cdot 171}{123 \cdot 54} \\ &+ 39 log \frac{39 \cdot 171}{48 \cdot 54} + 90 log \frac{90 \cdot 171}{123 \cdot 97} + 7 log \frac{7 \cdot 171}{48 \cdot 97}) \\ &+ 2(3-1)(2-1) \\ &= -71.92 \end{aligned}$$

それぞれの分割表に対しても同様の方法で AIC を計算することができる。

$$AIC(Table4.1) = -31.71$$

$$AIC(Table4.2) = -49.01$$

$$AIC(Table4.3) = -69.96$$

表 4.3: 説明変数 x と y による分割表

変数 x	変数 y	群 A	群 B	合計
big	*big*	18	2	20
big	*small*	55	4	59
small	*big*	15	39	54
small	*small*	35	3	38
合計		123	48	171

表 4.4: いくつかの階級をまとめた分割表

変数 x	変数 y	群 A	群 B	合計
big	*big*	18	2	20
small	*big*	15	39	54
	small	90	7	97
合計		123	48	171

AIC を用いた分割表の選択では、最も AIC の値が小さい分割表が選ばれる。この結果、表 4.4 の分割表が、被説明変数の値を予測・推定するうえで最も有効であると判断される。表 4.4 から，変数 y が *big* であるか *small* であるかということが群の判別には重要であること、さらに変数 y が *big* である場合には，変数 x の値が *big* であるか *small* であるかという要素を加えることによって予測能力が増すことがわかる。

一方、表 4.4 の分割表から、推論規則を記述すると次のようになるであろう。

規則 1　if x is big and y is big then group A

規則 2　if x is small and y is big then group B

規則 3　if y is small then group A

ところで、説明変数の数や区分数が増加するにしたがって、作りうる分割表の数は急激に増加する。従って、説明変数の数や区分数が増加したときにも、分割表の選択が行える方法を開発することが課題となる。

表 4.3 に示されるような最高次の分割表から、階級を要約した分割表を作成しようとした場合、その組合わせの数はベル数として知られている集合の分割数となる [50]。原始カテゴリー数 4 の場合には組合わせの数は 15 であるが、カテゴリー数が 8 となると 4,140 となり、カテゴリー数が 27 の場合には 5.46×10^{20} となる。このように原始カテゴリー数が増加するにしたがって組合せの数は急激に増加するので、すべての場合について AIC を計算し、比較することは不可能になってくる。そこで本研究では、遺伝的アルゴリズムを利用して分割表の探索を行うこととした。遺伝的アルゴリズムを使った探索を行うことによって、階級数の増加に対応することができる。

4.2 過度要約

情報量規準を用いて規則を得ようとする場合のもうひとつの問題が過度要約である。過度要約とは分割表に含まれる階級が過度にまとめられた結果、意味のない規則を得ることをいう。仮設例を使って説明する。

表 4.5 は前と同じく変数 x、y を含んだ分割表である。ただし、それぞれの変数は big、$middle$、$small$ の 3 つの値をとるものとする。この 2 変数を使って 2 群を識別する規則を抽出することを考える。

表 4.5: 2 変数からなる分割表

変数 x	変数 y	群 A	群 B	合計
big	big	1	3	4
big	$middle$	0	2	2
big	$small$	10	1	11
$middle$	big	2	12	14
$middle$	$middle$	1	9	10
$middle$	$small$	2	0	2
$small$	big	5	0	5
$small$	$middle$	9	1	10
$small$	$small$	2	7	9
合計		32	35	67

表 4.5 の分割表をもとに表 4.6 に示すような要約された分割表を作成することができる。

表 4.6: 過度に要約された分割表

変数 x	変数 y	群 A	群 B	合計
big $middle$ big $small$ $middle$	big big $middle$ $small$ $middle$	6	33	39
$middle$ $small$ $small$ big	$small$ big $middle$ $small$	26	2	28
合計		32	35	67

この分割表では、原始分割表の階級が 2 つに要約されている。要約後の第 1 カテゴリには群 B に属する観測結果が、第 2 カテゴリには群 A に属する観測結果が多く含まれていることから、群の分別も達成されていることがわかる。しかしながら、要約後の 2 つのカテゴリから規則を記述すると、

if (x is big or middle or small) and (y is big or middle or small) then group B

if (x is big or middle or small) and (y is big or middle or small) then group A

となり、規則としては意味を成さないことがわかる。これは、要約された分割表の度数を計算するときに、最初の分割表の各階級の値を単純に合計しているためである。

そこで、最初の分割表が与えられ、要約後の分割表の度数を計算するときに、もともとの階級のそれぞれが要約後の分割表の階級のいずれに属すると考えられるかを調べ、2つ以上の階級に属すると考えられる場合には、便宜的に1/2や1/3ずつそれぞれの階級に属するとして度数の計算を行うことにした。

例えば、表4.6に示されるような分割表の規則の前件部は、2つの規則ともに、

if (x is big or middle or small) and (y is big or middle or small)

と表現される。最初に与えられた分割表の最初の階級は、前件部として、

if x is big and y is big

を持つと解釈できるので、いま考えている分割表の2つのカテゴリの前件部をともに満たすものと考えることができる。したがって、群Aの度数1の1/2、群Bの度数3の1/2をそれぞれを要約後の2つの階級の度数として再配分する。最初の分割表の9つの階級すべてについて同様の操作を行うと次のような分割表を得ることができる。

表 4.7: 再配分後の分割表

変数x	変数y	群A	群B	合計
big	*big*			
middle	*big*			
big	*middle*	16	17.5	33.5
small	*small*			
middle	*middle*			
middle	*small*			
small	*big*	16	17.5	33.5
small	*middle*			
big	*small*			
合計		32	35	67

表4.7の分割表は群の識別に適していないことは明らかである。つまり、最初の分割表のそれぞれの階級の度数を再配分することによって、見かけ上は分離されていたとしても、意味を持たない規則を示すような分割表を排することができる。

また、表4.5に示した分割表から、再配分計算を行って、表4.8の分割表を得ることができる。表4.8の分割表では、最初と最後のカテゴリに群Bに属する観測結果がまとめられており、2番目と3番目の階級が群Aの要因をあらわしているものと考えることができる。表4.8から得られる4つの規則を記述すれば次のようになるであろう。

if (x is big or middle) and (y is big or middle) then group B

if (x is big or middle) and (y is small) then group A

if (x is small) and (y is big or middle) then group A

if (x is small) and (y is small) then group B

表 4.8: 4 階級に要約された分割表

変数 x	変数 y	群 A	群 B	合計
big	*big*			
big	*middle*	4	26	30
middle	*big*			
middle	*middle*			
middle	*small*	12	1	13
big	*small*			
small	*big*	14	1	15
small	*middle*			
small	*small*	2	7	9
合計		32	35	67

これらの規則によって 67 個のサンプルを識別した場合の判別率は、

$$判別率 = \frac{26+12+14+7}{67} = 88\ \%$$

となる。したがって、規則としても有効であることがわかる。さらに、これらは互いに排反であるので、最初の分割表の度数を単純に合計した値とも等しくなっている。このように要約された分割表から、規則を記述しようとする場合には、再配分計算をおこなうことによって意味のない分割表を得ることを避けることができる。

4.3 あやめの識別問題

あやめの識別問題とは、Fisher[9] によって与えられた 4 変数による 3 群の識別問題である。ここでは、変数を x_1, x_2, x_3, x_4 と表し、3 群を g_1, g_2, g_3 と表すことにする。

Fisher の例では、変数は実数値であるが、変数の範囲を 3 等分し分割表を作成してみた。それぞれの変数の最大値を $xmax$, 最小値を $xmin$ とするとき、置き換えは次のように行っている。

$x > xmin + 2/3 \times (xmax - xmin)$ ならば *big*

$x > xmin + 1/3 \times (xmax - xmin)$ かつ $x < xmin + 2/3 \times (xmax - xmin)$ ならば *middle*

$x < xmin + 1/3 \times (xman - xmin)$ ならば *small*

変数の値を $big, middle, small$ とした時に得られる分割表は表 4.9 のようになる。

たとえば、すべての変数の値が *small* であるようなあやめは、群 1 に属するものが 1 個だけあったことを表は示している。

それぞれの変数について 3 区分しているので、可能な階級の数は 81 であるが、実際に度数が観測されたものは 20 階級である。本章の目的は、このような分割表が与えられたときに、群を識別するような小数の意味のある規則を得ることである。

表 4.9: あやめの識別問題の原始分割表

	変数の値				変数置き換え						
	x_1	x_2	x_3	x_4	x_1	x_2	x_3	x_4	g_1	g_2	g_3
1 :	*small*	*small*	*small*	*small*	2	2	2	2	1	0	0
2 :	*small*	*middle*	*small*	*small*	2	3	2	2	35	0	0
3 :	*small*	*big*	*small*	*small*	2	5	2	2	9	0	0
4 :	*middle*	*middle*	*small*	*small*	3	3	2	2	1	0	0
5 :	*middle*	*big*	*small*	*small*	3	5	2	2	4	0	0
6 :	*small*	*small*	*middle*	*meddle*	2	2	3	3	0	5	1
7 :	*small*	*middle*	*middle*	*middle*	2	3	3	3	0	1	0
8 :	*middle*	*small*	*middle*	*middle*	3	2	3	3	0	15	0
9 :	*middle*	*middle*	*middle*	*middle*	3	3	3	3	0	23	0
10 :	*big*	*middle*	*middle*	*middle*	5	3	3	3	0	3	0
11 :	*middle*	*small*	*big*	*big*	3	2	3	5	0	0	1
12 :	*middle*	*small*	*big*	*big*	3	2	5	5	0	0	6
13 :	*middle*	*middle*	*middle*	*big*	3	3	3	5	0	1	4
14 :	*middle*	*middle*	*big*	*big*	3	3	5	5	0	0	18
15 :	*big*	*small*	*big*	*big*	5	2	5	5	0	0	1
16 :	*big*	*middle*	*big*	*big*	5	3	5	5	0	0	13
17 :	*big*	*big*	*big*	*big*	5	5	5	5	0	0	2
18 :	*middle*	*small*	*big*	*middle*	3	2	5	3	0	1	2
19 :	*middle*	*middle*	*big*	*middle*	3	3	5	3	0	1	1
20 :	*big*	*middle*	*big*	*middle*	5	3	5	3	0	0	1

4.4 変数の置き換え

本研究では、分割表に含まれるいくつかの階級を要約することによって、より少数の規則を導くことを目的としている。たとえば、表 4.9 の最初の 2 つの階級がまとめられたものとすると、そこから得られる規則は次のように記述される。

if x_1 *is small and* x_2 *is small or middle and* x_3 *is small* x_4 *is small then ...*

このような表現には、"*small or middle*"などが含まれており、計算機処理には向かない。そこで、変数の値として、素数を使うことにした。たとえば、*small* に対して 2、*middle* に対して 3 を与えたものとしよう。このとき、"*small or middle*"を 2 と 3 の最小公倍数 6 を使って表現する。そうすることによって、上の規則は、

if x_1 *is* 2 *and* x_2 *is* 6 *and* x_3 *is* 2 *and* x_4 *is* 2 *then ...*

と表現される。このような変数の置き換えを行うと、分割表のそれぞれの階級がこの規則の前件部を満たすかどうかは簡単に判断することができる。つまり、任意の階級について、それぞれの変数の値で対象としている規則の前件部の値を割ったときに、余りが 0 となれば、その変数は規則の前件部に含まれる当該変数の条件を満たすと判断することができる。そして、すべての変数について余りが 0 であるならば、規則の前件部を満たすことになる。

このような置き換えを行うことの利点のひとつは、分割表の階級数を規則の数と等しくするこ

とができることである。このことによって、必要な記憶領域を節約でき、また処理速度の向上も期待できる。表 4.9 には、変数値を置き換えた結果も併せて示している。

変数置き換えの~~もう~~一つの利点は、禁止したい組み合わせを容易にコントロールできる点である。たとえば、"*small or big*" といった前件部の条件は、規則の解釈を困難にしてしまう。このような組み合わせも最小公倍数を使うことによって、容易に抑制することができる。"*small or big*" という組み合わせは、最小公倍数 10 を使って表現できるので、探索の過程で 10 となるような階級の分割が現れた場合には、十分大きい AIC とすることで、このような組み合わせを抑制することができる。

4.5 分割表の区分選択と遺伝的アルゴリズム

表 4.9 の分割表の階級を要約して少数の規則を抽出することを考える。表 4.9 は区分選択の対象となるもとの分割表と言う意味で、原始分割表と呼ぶことにする。原始分割表には 20 の階級が含まれている。階級を要約するということは、この 20 の階級をいくつかのグループに分割するということに他ならない。したがって、第 3 章で検討したように、階乗進数を使って集合の分割を表現することとした。原始分割表に対しては 20 桁の階乗進数を用意すればよい。

遺伝的アルゴリズムは以下の手順によった。

1) 原始分割表の階級数に等しい桁数をもつ階乗進数を使って染色体を表現する。

2) 親の集団の大きさ $SIZE$ は 30 とした。

3) 交叉はランクによった。親の集団を AIC の小さい順に並び替え、i 番目 (i=0,$\cdots$,29) の染色体の選択される確率 $pr(i)$ は次式によった。

$$pr(i) = (SIZE - i)/(SIZE \times (SIZE + 1)/2)$$

また、交叉は $SIZE/2$ 組について行った。したがって、新たに作られる子の染色体は SIZE 個である。

4) 親のうちもっとも適合度の高い (AIC が小さい) 染色体を除いた $SIZE - 1$ 個の染色体と、新たに作られた $SIZE$ 個の子の染色体に対して突然変異を行う。突然変異は、$SIZE \times 2 - 1$ 個の染色体それぞれについて、乱数を使って遺伝子座を定め、当該遺伝子座 (階乗進数の係数) が取りうる値と置き換えることによって行った。エリートを保存するために、親の中でもっとも適合度の高い染色体に対しては突然変異の操作を行わなかった。

5) 交叉と突然変異を行った後で、$SIZE \times 2$ 個の染色体の適合度を求める。適合度は、染色体が表す分割表から得られる AIC の値とした。このとき、解釈が困難となる選択肢の組み合わせとなるものについては、十分大きな値を AIC 値として、抑制することとした。$SIZE \times 2$ 個の染色値を適合度の高い順に並び替える。そして上位 $SIZE$ 個の染色体を親として残す。

6) 上記 3) から 5 までの操作を定められた回数 (100 世代) 行って、もっとも適合度の高い染色体を選ぶ。

あやめの識別問題に対して上記の方法を適用し、最も AIC が小さい分割表として、表 4.10 を得た。

要約後の分割表から規則を記述すると次のようになる。

規則 1　if (x_1 *is small or middle*) *and* (x_3 *is small*) *and* (x_4 *is small*) *then* g_1.

規則 2　if (x_2 *is small of middle*) *and* (x_3 *is middle*) *and* (x_4 *is middle*) *then* g_2.

表 4.10: 原始分割表から直接得られた分割表

規則 $No.$	x_1	x_2	x_3	x_4		g_1	g_2	g_3		合計
1	6	30	2	2	:	50	0	0	:	50
2	30	6	3	3	:	0	47	1	:	48
3	15	30	15	5	:	0	1	45	:	46
4	15	6	5	3	:	0	2	4	:	6
合計					:	50	50	50	:	150

規則 3　*if* (x_1 *is middle or big*) *and* (x_3 *is middle or big*) *and* (x_4 *is big*) *then* g_3.

規則 4　*if* (x_1 *is middle or big*) *and* (x_2 *is small or middle*) *and* (x_3 *is big*)

and (x_4 *is middle*) *then* g_3.

これらの規則の後件部は分割表の度数をもとに記述した。この 4 つの規則によって、150 サンプル中 146 サンプルは正しく判別できることがわかる。しかし、この方法には、以下のような問題点がある。

最初の問題点は、原始分割表に含まれる階級の数が多くなると原始分割表から直接規則を抽出することが困難となることである。特に第 5 章で扱うように多くの説明変数が設定されているときには、それらの変数をすべて含んだ分割表を考えることは困難になってくる。

第 2 の問題点は、すべての変数を含んだ原始分割表から規則の抽出を行っているので、規則に貢献していない変数が残る可能性があるということである。このこととも関連していると思われるが、規則の記述が長くなっている。そこで、これらの問題を解決するために変数増減法を検討することとした。

4.6 変数の増加

まず、あやめの識別問題で変数 x_1 のみからなる分割表を作成すると表 4.11 のようになる。変数 x_2, x_3, x_4 についても同様の検討を行うことができる。それぞれの変数について作成される原始分割表から遺伝的アルゴリズムを使って区分選択を行ったところ、すべての変数について、3 つの階級を持つ分割表から得られる AIC がもっとも小さな値となった。従って、それぞれの変数から得られる分割表は表 4.12、表 4.13、表 4.14 のようになる。

表 4.11: 変数 x_1 のみからなる分割表

	x_1	g_1	g_2	g_3	total	
no.1	2	45	6	1	52	
no.2	3	5	41	32	78	
no.3	5	0	3	17	20	
total		50	50	50	150	AIC=－119.88

表 4.12: 変数 x_2 のみからなる分割表

	x_2	g_1	g_2	g_3	total	
no.1	2	1	21	11	33	
no.2	3	36	29	37	102	
no.3	5	13	0	2	15	
total		50	50	50	150	AIC=－35.95

表 4.13: 変数 x_3 のみからなる分割表

	x_3	g_1	g_2	g_3	total	
no.1	2	50	0	0	50	
no.2	3	0	48	6	54	
no.3	5	0	2	44	46	
total		50	50	50	150	AIC=－264.51

表 4.11、表 4.12、表 4.13、表 4.14 に示したように、1 変数のみからなる分割表を作成した場合には、変数 x_4 を用いた分割表の AIC がもっとも小さな値となった。そこで、まず変数 x_4 を採用しよう。

次の問題は、変数 4 の次にどの変数を採用するかということである。仮に、変数 x_1 を追加して分割表を作成したものとしよう。このときに得られる原始分割表は表 4.15 のようになる。

変数 x_4 のみからなる分割表の階級数は 3 であり、変数 x_1 も 3 つのカテゴリーを持つので、可能な階級の数は 9 であるが、実際に度数を持つのは、表 4.15 に示すように 7 階級のみであった。この原始分割表からの区分選択は、7 桁の階乗進数を使って行うことができる。遺伝的アルゴリズムを使った探索の結果、表 4.16 の分割表が選択された。

変数 x_2 および変数 x_3 を新たに追加した原始分割表を作成し、遺伝的アルゴリズムを使って区分選択を行ったところ、それぞれ表 4.17、表 4.18 の分割表が、もっとも小さい AIC を与えるものとして選択された。

これらの分析から明らかなように、変数 x_4 対して、新たに変数 x_3 を追加することによって、AIC の値は－286.9 と改善される。従って、変数 x_3 を追加して得られる分割表が選択される。

変数 x_3 と x_4 からなる分割表にさらに新たな変数を追加することを考える。このとき、変数 x_3x_4 の可能なすべての階級を含む分割表ではなく、表 4.18 に得られた 4 階級からなる分割表に対して、

表 4.14: 変数 x_4 のみからなる分割表

	x_4	g_1	g_2	g_3	total	
no.1	2	50	0	0	50	
no.2	3	0	49	5	54	
no.3	5	0	1	45	46	
total		50	50	50	150	AIC=－275.69

表 4.15: 変数 x_1 と x_4 からなる分割表

	x_1	x_4	g_1	g_2	g_3	total	
no.1	2	2	45	0	0	45	
no.2	3	2	5	0	0	5	
no.3	2	3	0	6	1	7	
no.4	3	3	0	40	3	43	
no.5	5	3	0	3	1	4	
no.6	3	5	0	1	29	30	
no.7	5	5	0	0	16	16	
total			50	50	50	150	AIC=− 257.46

表 4.16: 変数 x_1 と x_4 からなる要約後分割表

	x_1	x_4	g_1	g_2	g_3	total	
no.1	6	2	50	0	0	50	
no.2	30	3	0	49	5	54	
no.3	15	5	0	1	45	46	
total			50	50	50	150	AIC=− 275.69

新たな変数の追加を試みる。もし、仮に x_3 と x_4 の可能なすべての組み合わせをもとに原始分割表を作ると、9 × 3、つまり 27 の階級を持つ分割表を対象とすることもあり得る。このような方法をとった場合、それぞれ 3 区分を持つ変数が 4 つ含まれた分割表は、81 の階級を持つ可能性があり、5 つ含む分割表は、243 の階級をもつ可能性もある。このことは、原始分割表に含まれる変数の数が増えるに従って、原始分割表の階級が急速に増えることとなり、やがては、すべての変数を含んだ分割表の階級数に近くなってしまう。この結果、変数増減法をとる理由がなくなってしまう。一方、先に得られた分割表の階級数は 4 であり、新たな変数を追加したとしても得られる分割表の階級数は高々 12 である。このように先に得られた分割表をもとに変数増加を行うことによって、原始分割表の階級数を抑制することが可能となる。

先に得られた変数 x_3 と x_4 からなる分割表に変数 x_1 を追加したときに得られる原始分割表は表 4.19 のようになった。変数 x_3 と x_4 からなる分割表のそれぞれの階級が、変数 x_1 によって細分化されていることがわかる。度数が 0 となった階級は除いているので、可能な 12 階級のうち、9 階級が出現している。この原始分割表からの区分選択は、9 桁の階乗進数を使って行うことができる。この結果、表 4.20 の分割表が選択された。

この結果、9 階級からなる分割表を 6 階級に要約した方が、AIC は小さくなることがわかった。しかしながら、変数 x_3 と x_4 からなる分割表から得られる AIC よりは大きいことから、新たに変数 x_1 を追加することは好ましいことではないこともわかる。変数 x_2 を追加した場合にも、同様の手順を踏むことができる。変数 x_2 を追加した原始分割表から得られる区分選択後の分割表は表 4.21 のようになった。

先に求められた x_3 と x_4 からなる分割表の AIC も − 286.91 であった。従って、AIC 値には差がないことになる。このような場合に、次のステップにはどちらの分割表をもとにすれば良いだろうか。本研究では、変数増減の過程において、後述のようにタブーサーチを用いている。この結

表 4.17: 変数 x_2 と x_4 からなる要約後分割表

	x_2	x_4	g_1	g_2	g_3	total	
no.1	30	2	50	0	0	50	
no.2	6	3	0	49	5	54	
no.3	30	5	0	1	45	46	
total			50	50	50	150	AIC=− 275.69

表 4.18: 変数 x_3 と x_4 からなる要約後分割表

	x_3	x_4	g_1	g_2	g_3	total	
no.1	2	2	50	0	0	50	
no.2	3	3	0	47	1	48	
no.3	5	3	0	2	4	6	
no.4	15	5	0	1	45	46	
total			50	50	50	150	AIC=− 286.91

果、変数 x_2, x_3, x_4 からなる分割表をもとに次のステップに移行する。ただし、最良解の更新は行わない。変数は徐々に増加する方法を取っているので、このことによっていたずらに多くの変数を含む分割表を抑制することができるものと考えられる。

4.7 変数の減少

一方、変数を減少させる手順はどのように考えたらよいだろうか。単純に変数を削除しても階級数は変化しない。他方、変数を削除することは各階級が表す前件部の条件が緩くなることを意味するから、再配分後の度数もまた変化する可能性がある。そこで、本研究では、次の手順で変数を減少させることとした。

変数減少の手順

1) 原始分割表に含まれている変数を 1 つ選ぶ。

2) その変数を削除する。

3) 削除後の原始分割表に対して区分選択を行う。

4) 減少させるべき変数を削除する。

表 4.18 によって、変数 x_3 と x_4 からなる分割表が得られている。これに対して、変数を増加させる方法は上に述べたとおりである。逆に表 4.18 から変数 x_3 を削除することを考える。そこでまず、表 4.18 から単純に変数 x_3 を削除した表 4.22 を作る。しかし変数を削除しただけでは度数は変化しない。このため、表 4.22 の分割表をもとに区分選択を行う。区分選択は、元データの個々の観測数を再配分することによって行う。再配分の結果、階級数 3 である表 4.23 の分割表が選択された。

表 4.23 から得られる AIC は− 275.69 であり、変数 x_3 と x_4 からなる表 4.18 から得られる AIC − 286.91 より大きい。このように変数を削除することによって、もとの分割表から得られる AIC より大きくなる場合には、当該変数の削除は行わないこととした。

表 4.19: 変数 x_1、x_3 と x_4 からなる分割表

	x_1	x_3	x_4	g_1	g_2	g_3	total	
no.1	2	2	2	45	0	0	45	
no.2	3	2	2	5	0	0	5	
no.3	2	3	3	0	6	1	7	
no.4	3	3	3	0	38	0	38	
no.5	5	3	3	0	3	0	3	
no.6	3	5	3	0	2	3	5	
no.7	5	5	3	0	0	1	1	
no.8	3	15	5	0	1	29	30	
no.9	5	15	5	0	0	16	16	
total				50	50	50	150	AIC=− 265.31

一方、階級を要約する過程で、一部の変数が"all"を表す数のみから構成される分割表が得られることがある。このような場合、当該変数は規則の構成に何ら寄与していないので、AIC の評価に係わらず強制的に削除することとした。なお、この操作によって変数の削除を行った場合には、次節タブーリストの更新は行わない。

4.8 タブーサーチ

ところで、表 4.23 の分割表は、最初の変数を選ぶときに採用した表 4.14 の分割表である。この分割表から次のステップに移行した場合には、再び変数 x_3 が選択されてしまい、意味のない繰り返しに陥る危険がある。これを避けるために、タブーサーチを利用することとした。タブーリストには、直近数回に追加または削除された変数を格納することとした。

直前のステップで求められた分割表において採用されていない変数の追加をまず試みる。AIC 値が最も小さくなる変数の追加を行うが、当該変数が最近削除されたためにタブーリストにある場合には、次に AIC 値が小さい変数を追加する。そして、当該変数をタブーリストに追加する。その後、採用されている変数の削除を試みる。削除することによっても、直前のステップで得られている分割表の AIC より小さい値となるものがあれば当該変数の削除を行う。このとき、タブー

表 4.20: 変数 x_1、x_3 と x_4 からなる要約後分割表

	x_1	x_3	x_4	g_1	g_2	g_3	total	
no.1	6	2	2	50	0	0	50	
no.2	2	3	3	0	6	1	7	
no.3	3	3	3	0	38	0	38	
no.4	5	3	3	0	3	0	3	
no.5	15	5	3	0	2	4	6	
no.6	15	15	5	0	1	45	46	
total				50	50	50	150	AIC=− 279.95

表 4.21: 変数 x_2、x_3 と x_4 からなる要約後分割表

	x_2	x_3	x_4	g_1	g_2	g_3	total
no.1	30	2	2	50	0	0	50
no.2	6	3	3	0	47	1	48
no.3	6	5	3	0	2	4	6
no.4	30	15	5	0	1	45	46
total				50	50	50	150

AIC=− 286.91

表 4.22: 表 4.18 から変数 x_3 を削除した分割表 (1)

	x_4	g_1	g_2	g_3	total
no.1	2	50	0	0	50
no.2	3	0	47	1	48
no.3	3	0	2	4	6
no.4	5	0	1	45	46
total		50	50	50	150

AIC=− 286.91

リストを走査して、当該変数が最近追加されたものであるならば、当該変数の削除は行わず、ついで小さい AIC 値となる変数の削除を検討する。削除された変数はタブーリストに追加し、タブーリストの更新を行う。追加や削除の操作を行った後に得られる分割表をもとに次のステップに移行する。

このことから、本研究で採用した変数増減法は、すべての変数に対して追加または削除の検討を行う過程を 1 ステップとするタブーサーチと見ることもできる。

4.9 規則抽出の手順

本研究で採用した、規則抽出の手順をまとめると次のようになる。なお、遺伝的アルゴリズムの実行手順は 4.5 に示した通りである。

1) 分割表への変数の追加

1.0　前のステップの原始分割表の AIC を保存する。

1.1　分割表に含まれていない変数を一つ選ぶ。

1.2　当該変数の加えた原始分割表を作成する。

1.3　遺伝的アルゴリズムを使って区分選択を行う。

1.4　分割表に含まれていないすべての変数について、1.1 から 1.3 までの処理を行い、タブーリストに含まれておらず、かつ区分選択後の AIC が最も小さくなる変数を選ぶ。

1.5　1.4 で選ばれた変数を追加し、区分選択を行った分割表を作成する。また、この変数をタブーリストに追加し、タブーリストの更新を行う。

2) 変数の削除

2.1　分割表に含まれている変数を一つ選ぶ。

2.2　当該変数を除いた分割表を作成する。

表 4.23: 表 4.18 から変数 x_3 を削除した分割表 (2)

	x_4	g_1	g_2	g_3	total	
no.1	2	50	0	0	50	
no.2	3	0	49	5	54	
no.3	5	0	1	45	46	
total		50	50	50	150	AIC=－275.69

2.3　遺伝的アルゴリズムを使って区分選択を行う。

2.4　分割表に含まれているすべての変数について、2.1 から 2.3 間の処理を行い、タブーリストに含まれておらず、かつ区分選択後の AIC が最も小さくなる変数を選ぶ。

2.5　2.4 で選ばれた変数を削除したときの区分選択後の AIC が、1.0 で保存した前のステップの AIC より小さければその変数を分割表から削除し、タブーリストに加え、タブーリストの更新を行う。もし、前のステップの AIC より大きければ、変数の削除は行わない。

3) 不要な変数の除去

すべての階級において、"all" を意味する値になっている変数があれば、その変数を分割表から削除する。このとき、タブーリストの更新は行わない。

4) あらかじめ定めれた回数、1) から 3) の処理を繰り返す。

あやめの識別問題に対して 10 ステップのタブーサーチ探索を行ったところ、変数 x_3, x_4 からなる表 4.18 の分割表が最良の分割表として選ばれた。この分割表から規則を抽出し、記述すると次のようになる。

これから得られる規則は、次のように記述される。

if (x_3 *is small*) *and* (x_4 *is small*) *then* g_1

if (x_3 *is middle*) *and* (x_4 *ismiddle*) *then* g_2

if (x_3 *is big*) *and* (x_4 *is middle*) *then* g_3

if (x_3 *is middle or big*)*and*(x_4 *is big*) *then* g_3

分割表の度数から明らかなように、これら 4 つの規則によって、150 例中 146 例は正しく識別することができる。

ところで、4.2.3 において原始分割表から直接求めた規則と比較してみよう。まず、両方の規則の判別力には差がないことがわかる。また、要素 x_3, x_4 に関する前件部が等しいこともわかる。一方、変数増減法を使って得られた規則には、変数 x_1, x_2 に関する記述が含まれておらず、より簡明な規則が得られていることがわかる。

4.10　決定木との比較

表 4.9 のような分割表が得られたときに、効果的に規則を抽出する方法として ID3[64]、C4.5[65]、See5/C5.0[66] などが知られている。決定木による規則抽出と比較するために、表 4.9 を See5 に適用してどのような規則が得られるか検討することとした。See5 および提案方法によって得られた規則を表 4.24 にまとめた。

表 4.24 から決定木によった場合よりも、提案方法によった場合のほうが得られる規則の記述が

表 4.24: See5 および提案方法から得られた規則

See5 からの規則

規則 *No.*	x_1	x_2	x_3	x_4	群
1	−	−	−	*small*	g_1
2	−	−	*middle*	*middle*	g_2
3	−	−	*big*	−	g_3
4	−	−	−	*big*	g_3

変数増減法によって得られた規則

規則 *No.*	x_1	x_2	x_3	x_4	群
1	−	−	*small*	*small*	g_1
2	−	−	*middle*	*middle*	g_2
3	−	−	*big*	*middle*	g_3
4	−	−	*middle or small*	*big*	g_3

詳細であることがわかる。特に、群 g_1 を識別するための規則 1 において顕著な違いが見られる。表 4.9 を仔細に眺めてみると、群 g_1 に属する 50 のサンプルの要素 x_3, x_4 の値はすべて *small* となっていることがわかる。したがって、予測目的ならば、いずれかの変数の値が (See5 は要素 x_4 を選んでいる) わかれば十分である。しかし、群 g_1 の特徴を捉え、解釈することが目的ならば、2 つの要素ともに規則に含まれていることが望ましい。この意味において、解釈目的ならば提案方法から得られる規則の方が望ましいと考えることができる。

4.11 結論

本章では、分割表としてまとめられる観測結果から、被説明変数に関する意味のある規則を抽出することを試みた。このためには様々な分割表を評価し、選択するという作業が必要になった。原始分割表に含まれる階級数が少なかったり、含まれる変数の数が少なければ 4.5 に示したように、原始分割表から直接規則を抽出する方法が有効であろう。しかし、変数が多く含まれている調査であったり、階級数が多くなることが予想される場合には、直接原始分割表から求めることは困難となってくるし、また得られた規則も冗長なものになる危険がある。

本研究では、分割表における区分選択は階級の分割と考え、階乗進数を使ってさまざまな分割表を表現した。また、大規模な問題にも対応できるように、変数増減法を検討した。この結果、直接原始分割表から求めた規則に比べ、より簡明で同等の判別力のある規則を得ることができた。一方、決定木から得られる規則に対しては、より詳細な規則となっていることがわかった。このことから、解釈目的で規則を抽出しようとする場合には、提案方法の方が好ましいと思われる。

第５章　大規模な問題の例としての減価償却問題への応用

本章では、大規模な分割表からの規則抽出の例として、減価償却選択問題を取り上げる。減価償却とは、費用配分の原則にもとづいて固定資産の取得価額をその耐用期間における各事業年度に配分する手続である。減価償却費の計上にあたっては耐用期間など見積もりの要素が含まれている。さらに減価償却方法としていくつかの代替的な方法が認められている。

一般に固定資産投資の増加に伴って減価償却費も増加する傾向にあり、したがって、期間利益に対する減価償却の影響も増大する。さらに、減価償却費は単に帳簿上計上される費用であり、現金支出を伴わないという特徴があるため、自己金融という面から資金政策とも密接に関係している。どのような償却方法を選択するかによって計上される減価償却費の額は異なるので、どのような償却方法を選択するかということは企業にとて重要な問題であり、償却方法の選択状況も企業によって違いがみられるところである。

減価償却の重要な目的は、正確な損益計算であり、利益に及ぼす影響を考慮しながら償却費を増減することは損益計算を歪めるものとして是認されない。しかしながら、企業における償却方法選択の状況は、必ずしも正確な期間損益計算という観点のみからでは説明できない状況にある。そこで本研究では、償却方法選択状況を説明する因果律を、アンケートなどを通して得られた観測結果から抽出し、さらに得られた因果律の妥当性について検証することを目的とした。

5.1　減価償却方法の性質の違いから導かれる仮説とその検証

5.1.1　減価償却方法

我が国にあっては、会計理論上もまた実務上も、定率法と定額法が代表的な償却方法である考えることができる。したがって、本研究では定率法と定額法の計算方法の違いに起因すると思われる償却方法選択要因を仮説の形にまとめることとした。定額法および定率法による償却費の計算式より、両方法の計算上の違いとして次の２点を指摘することができるであろう。

・定率法は定額法に比べて償却速度が速い。

定額法が毎期一定の償却費を計上するのに対し、定率法は償却の早期により多くの償却費を計上する。このため、投下資本の回収割合（減価償却累計の取得価額に対する割合）は、償却計算の基礎となった耐用期間が終了する以前の如何なる時点でみても、定率法による場合の方が高くなる。

・定率法は未償却残高をもとに償却費の計算を行う。

定額法は、取得価額をもとに計算するので、取得価額の記録が必要である。これに対して、定率法は未償却残高に一定の償却率を乗じて償却費を計算するので、取得価額の記録を必要としない。このことは、追加投資が頻繁に行われるような場合には、計算の簡略化につながる。

これらの計算上の違いをもとに仮説を設定した。本研究では、償却速度の違いによるもの５つと、償却計算のもとになる基礎の違いによると思われるもの２つを設定した。そしてこれらの仮

説を検討するためにアンケート調査を実施した。

5.1.2 減価償却方法選択に関する調査の概要

減価償却方法として、定率法と定額法のいずれの方法を採用するかによって、計上される期間利益は異なってくる。本研究ではどのような要因が償却方法選択に影響を与えているかということを実態調査を通して分析することとした。このため1991年より継続的に企業アンケートを実施してきた。収益性や資産重要性などに対する経営者の主観と償却方法選択との関連をみる目的で、1991年11月に製造業に属する上場企業、さらに1992年12月には製造業に属する未上場企業へのアンケート調査を行った。1994年からは、減価償却資産に関する質問のほか、たとえば市場集中度など広く経営環境に関する質問も加えてアンケートを実施してきた。この調査は1994年、1997年、2000年と3回実施してきている。これらのアンケート結果から、減価償却方法選択に関する規則を抽出することが本章の目的であるが、本節では1991年および1992年に行ったアンケートの概要を述べることとする。

1991年の上場企業、1992年の未上場企業に対するアンケートは、主に減価償却方法と減価償却資産に関する質問からなっていた。質問項目としては、減価償却方法選択状況のほか、資産重要性、経済的耐用年数に関することなど減価償却資産に関するものを設定した。上場企業と未上場企業に対する質問項目は同じものとした。これは上場企業と未上場企業によって違いがあるかどうかをみるためであった。

5.1.3 調査項目

製造業に属する上場企業に対するアンケート調査は1991年11月、経理部長宛の質問用紙の郵送、回答用はがきの回収により行った。調査対象は1217社であり、497社から回答を得ることができた（回収率40.8 %）。また、未上場企業に対する調査は1992年12月に1078社を対象に行った。調査は上場企業の場合と同じく経理部長宛の郵送と回答用はがきの回収によって行った。ここで対象となる企業は東洋経済社四季報「未上場企業版('92年下期)」に掲載されている製造業に属する企業とした。回収数は462通であり、回収率は43 %であった。1991年の上場企業、1992年の未上場企業に対するアンケートの内容は付録に示した。

1991年および1992年の調査において、共通の質問項目は次の10項目である。

質問(a) 減価償却方法の選択状況
質問(b) 総資産に占める有形固定資産の割合
質問(c) 資産除却時の除却損、売却損の発生状況
質問(d) 有形固定資産に対する資産的支出の大きさ
質問(e) 設備投資資金の借入依存度
質問(f) 収益性
質問(g) 収益性の変動
質問(h) 計算機の利用状況
質問(i) 償却計算の範囲
質問(j) 経済的耐用年数

これらの質問項目は、すべていくつかの選択肢の中から、最も当てはまると思われるものを選択するという方法によっている。つまり、経営者の主観を尋ねるという方法を取った。主観を

聞くという方法を取ったために、回答結果に対する曖昧さは残る。それにもかかわらず、主観を尋ねるという方法を取ったのは、どの償却方法を選択するかということは、結局は経営者の主観に依存するのではないかという考えによる。

5.1.4 調査結果の概要

ここでは、各調査項目の単純集計を見ていくことにする。本節で分析の対象とするものは、上場企業に対するアンケート結果であるが、比較のために未上場企業に対する調査結果もあわせて示す。なお未記入や不明分は除いているので、合計は回収数には必ずしも一致しない。各集計表では、比率の差の検定結果も併せて示している。2つの母比率の差の検定は、

帰無仮説 H_0 :2つの母比率は等しい

に対して、検定統計量

$$t = (\hat{p}_1 - \hat{p}_2) / \sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1} + \frac{1}{n_2})}$$

を計算することによって行うことができる。ここで、$\hat{p}_1, \hat{p}_2$ は標本比率であり、n_1, n_2 はそれぞれの標本数である。また、$\hat{p}$ は次式で求められる。

$$\hat{p} = (\hat{p}_1 n_1 + \hat{p}_2 n_2)/(n_1 + n_2)$$

帰無仮説のもとでは、検定統計量は標準正規分布に従うから、$|t| \geqq 1.64$ のとき危険率片側5％で帰無仮説を棄却できる。比率の差の検定から、該当する区分で上場企業の比率の値が大きいと思われる場合には△を、上場企業の方が小さいと思われる場合に▼を記入している。

減価償却の方法 (表 5.1)

最初の質問 (a) は定率法と定額法の選択状況である。この選択の違いを説明することが最終的な目標である。表の集計結果から、上場企業、未上場企業ともに定率法によっている企業が多いことがわかる。特に未上場企業の場合には圧倒的である。比率の差の検定からも未上場企業において定率法のみで償却を行っている企業の割合が高いことを確認することができる。逆に定率法と定額法を併用している企業の割合は上場企業の方が高い。上場企業、未上場企業ともに、定率法と定額法が半々という企業の割合は低い。このことから、企業においては明確な意志をもって償却方法を選択しているであろうことが推測される。すべて定額法によっている企業は上場企業4.6％、未上場企業3.5％であり、大きな差は見られない。

有形固定資産の重要性 (表 5.2)

総資産に占める有形固定資産の割合は未上場企業の方が高いと感じている企業が多い。総資産に占める割合が高いまたはやや高いと感じている企業は上場企業が25.9％であるのに対して未上場企業では34.8％に達している。逆にやや低いと感じている企業は上場企業の方が多い。

除却損や売却損の発生状況 (表 5.3)

除却損や売却損の発生状況には、差がないようである。ほとんどの企業は普通または少ないと感じている。上場企業未上場企業ともに除却損売却損の発生が多いまたはやや多いと感じている

表 5.1: 質問 (a)　償却方法の採用状況

償却方法		上場企業	未上場企業	検定結果
1	すべて定率法	370(74.1 %)	402(87.8 %)	▼
2	定率法による部分が多い	74(14.8 %)	31(6.8 %)	△
3	定率法と定額法が半々である	8(1.6 %)	4(0.9 %)	
4	定額法による部分が多い	24(4.8 %)	5(1.1 %)	△
5	すべて定額法によっている	23(4.6 %)	16(3.5 %)	
	合計	499(100.0 %)	458(100.0 %)	

(注) △ : 危険率 5 %で上場企業の比率が大きい
▼ : 危険率 5 %で上場企業の比率が小さい

表 5.2: 質問 (b)　有形固定資産の重要性

有形固定資産の割合		上場企業	未上場企業	検定結果
1	高い	43(8.6 %)	59(12.9 %)	▼
2	やや高い	86(17.3 %)	100(21.9 %)	▼
3	普通	182(36.5 %)	171(37.4 %)	
4	やや低い	136(27.3 %)	80(17.5 %)	△
5	低い	51(10.2 %)	47(10.3 %)	
	合計	498(100.0 %)	457(100.0 %)	

(注) △ : 危険率 5 %で上場企業の比率が大きい
▼ : 危険率 5 %で上場企業の比率が小さい

企業は 1 割前後である。

設備投資の大きさ (表 5.4)

有形固定資産に対する資本的支出も未上場企業の方が多いと感じている。比率の差の検定結果は質問 (b) 総資産に占める有形固定資産の割合に対する回答と同じものとなった。上場企業にあっては未上場企業に比べて資本的支出が少ないと感じている企業が多い。

借入依存度 (表 5.5)

上場企業、未上場企業ともに借入依存度が高いと回答した企業の割合が最も少ない。逆に借入依存度が低いと回答した企業の割合が最も高いという結果になった。上場企業と未上場企業の比較では借入依存度が高いまたはやや高いと回答した企業の割合に差が見られた。いずれの回答においても上場企業の割合が未上場企業の割合より高い。

収益性 (表 5.6)

表 5.3: 質問 (c)　除却損、売却損の発生

除却損、売却損の発生	上場企業	未上場企業	検定結果
1　多い	16(3.2 %)	15(3.3 %)	
2　やや多い	33(6.6 %)	38(8.3 %)	
3　普通	161(32.4 %)	151(33.0 %)	
4　やや少ない	136(27.4 %)	108(23.6 %)	
5　少ない	151(30.4 %)	145(31.7 %)	
合計	497(100.0 %)	457(100.0 %)	

表 5.4: 質問 (d)　有形固定資産に対する資本的支出

資本的支出	上場企業	未上場企業	検定結果
1　多い	17(3.4 %)	36(7.9 %)	▼
2　やや多い	74(14.9 %)	104(22.8 %)	▼
3　普通	225(45.2 %)	183(40.1 %)	
4　やや少ない	119(23.9 %)	81(17.8 %)	△
5　少ない	63(12.7 %)	52(11.4 %)	
合計	498(100.0 %)	456(100.0 %)	

(注) △：危険率 5 %で上場企業の比率が大きい
▼：危険率 5 %で上場企業の比率が小さい

未上場企業に対する調査が 1992 年上場企業に対する調査が 1991 年であったにもかかわらず未上場企業の方が業績に関しては楽観的である。未上場企業の方が最近の収益性は他の企業に比べて高いと感じている企業が多く逆に低いと感じている企業の割合は小さい。

収益性の変動 (表 5.7)

収益性の変動については普通と感じている企業が最も多く、大きいと感じている企業や小さいと感じている企業の割合は小さい。上場企業と未上場企業の間では収益性の変動が小さいと感じている企業の割合に差が見られた。未上場企業の方が利益の変動は小さいと感じている企業の割合が高い。

償却費の計算手段 (表 5.8)

上場企業未上場企業ともに償却費の計算は計算機によっている場合が圧倒的に多い。上場企業では 95.4 %、未上場企業では 90.8 %の企業が計算機処理を行っている。比率の差の検定から上場企業の計算機処理の割合が未上場企業における割合よりも高くなっていることがわかった。

償却の範囲 (表 5.9)

表 5.5: 質問 (e)　設備投資資金の借入依存度

借入依存度	上場企業	未上場企業	検定結果
1　高い	53(10.6 %)	65(14.2 %)	▼
2　やや高い	92(18.5 %)	106(23.1 %)	▼
3　普通	116(23.3 %)	104(22.7 %)	
4　やや低い	100(20.1 %)	75(16.4 %)	
5　低い	137(27.5 %)	108(23.6 %)	
合計	498(100.0 %)	458(100.0 %)	

(注) △ : 危険率 5 %で上場企業の比率が大きい
▼ : 危険率 5 %で上場企業の比率が小さい

表 5.6: 質問 (f)　収益性

収益性	上場企業	未上場企業	検定結果
1　高い	24(4.8 %)	49(10.7 %)	▼
2　やや高い	102(20.5 %)	97(21.2 %)	
3　普通	113(22.7 %)	118(25.8 %)	
4　やや低い	159(32.0 %)	129(28.2 %)	
5　低い	99(19.9 %)	64(14.0 %)	△
合計	497(100.0 %)	457(100.0 %)	

(注) △ : 危険率 5 %で上場企業の比率が大きい
▼ : 危険率 5 %で上場企業の比率が小さい

上場企業、未上場企業ともに個別償却を行っている企業の割合が最も高かった。上場企業で 63.1 %、未上場企業で 76.6 %が個別償却を行っていることがわかった。一括償却や総合償却などを行っている企業の割合は上場企業の方が高い。

経済的耐用年数 (表 5.10)

上場企業未上場企業間で経済的耐用年数に関する差はほとんど見られない。上場企業未上場企業双方において法定耐用年数の 7 から 8 割程度と回答した企業が過半数であった。次に多い回答が法定耐用年数と同じぐらいというもので上場企業で 27.9 %、未上場企業で 31.1 %であった。

5.1.5　未上場企業に対する考慮事項調査

1992 年の未上場企業に対する調査では減価償却方法を採用するにあたって考慮する事項を併せて尋ねた。以下の項目の中から減価償却方法選択にあたって考慮するものを選んでもらった (複数回答)。ここでの質問項目は上場企業が償却方法を変更した際に監査意見や脚注に記載されている変更理由を参考に設定した。また特に傍点の語をキーワードとして回答していただいた。結果を

表 5.7: 質問 (g)　収益性の変動

収益性の変動	上場企業	未上場企業	検定結果
1　大きい	28(5.6 %)	28(6.1 %)	
2　やや大きい	107(21.5 %)	112(24.5 %)	
3　普通	201(40.4 %)	168(36.8 %)	
4　やや小さい	128(25.7 %)	102(22.3 %)	
5　小さい	34(6.8 %)	47(10.3 %)	▼
合計	498(100.0 %)	457(100.0 %)	

(注) △ : 危険率 5 %で上場企業の比率が大きい
▼ : 危険率 5 %で上場企業の比率が小さい

表 5.8: 質問 (h)　償却費の計算手段

計算手段	上場企業	未上場企業	検定結果
1　計算機処理	474(95.4 %)	417(90.8 %)	△
2　帳簿上で計算	23(4.6 %)	42(9.2 %)	▼
合計	497(100.0 %)	459(100.0 %)	

(注) △ : 危険率 5 %で上場企業の比率が大きい
▼ : 危険率 5 %で上場企業の比率が小さい

表 5.11 に示した。

償却方法選択において最も考慮されるのは資金の早期回収であり 65 %近くの企業がこれを選んでいる。この理由によれば定率法を選択することに傾向になるであろう。未上場企業で定率法のみを採用している企業の割合が高い原因のひとつと考えられる。上場企業において償却方法変更の際に多くみられた費用配分の適正化も 50 %ほどの企業で選ばれている。これら以外で多いのは陳腐化の速さ (35.9 %)、技術革新の速さ (39.8 %)、設備の使用状況 (30.3 %) であった。前の 2 つは資金の早期回収にも通じ後の理由は費用配分の適正化とも通じる。

5.1.6　仮説の検定と統計的方法

本節では、19991 年の上場企業に対するアンケート結果を用いて、償却方法選択に関するいくつかの仮説検定を行う。仮説の検討を行う前に、本研究ではいわば被説明変数となる償却方法の採用状況をみておこう。償却方法の採用状況は質問 (a) でたずねている (表 5.1)。企業がどのような償却方法を採用しているかということは有価証券報告書からも知ることはできるが、複数の方法を採用している場合には、金額的な面からの重要性を知ることができない。そのため、質問項目を設定するにあたって金額面から判断していただくことにした。調査結果によると、定率法のみで償却を行っている企業が 370 社 (74 %) と圧倒的に多い。次に多いのが主として定率法によっている場合で 74 社 (15 %) である。半々であると答えた企業を含めても主として定額法によっていると思われる企業は 55 社 (11 %) と少ない。従って、本節の分析では、表 5.1 の「3　定率法と定額法による部分が半々である」以下をまとめて、「主に定額」として分析することとした。この結

表 5.9: 質問 (i)　償却資産の範囲

償却資産の範囲	上場企業	未上場企業	検定結果
1　おおむね個別償却	315(63.1 %)	351(76.6 %)	▼
2　類似資産の一括償却	57(11.4 %)	34(7.4 %)	△
3　総合償却	127(25.5 %)	73(15.9 %)	△
合計	497(100.0 %)	459(100.0 %)	

(注) △：危険率 5 %で上場企業の比率が大きい
▼：危険率 5 %で上場企業の比率が小さい

表 5.10: 質問 (j)　経済的耐用年数

法定耐用年数と比較して	上場企業	未上場企業	検定結果
1　半分以下	10(2.0 %)	8(1.8 %)	
2　半分ぐらい	79(15.8 %)	70(15.3 %)	
3　7〜8 割程度	271(54.3 %)	237(51.9 %)	
4　法定耐用年数と同じぐらい	142(31.1 %)	73(15.9 %)	
合計	499(100.0 %)	457(100.0 %)	

果、1991 年調査データに対しては、「定率のみ」で償却を行っている 370 社と、「主に定率」で償却を行っている 74 社、「主に定額」法で償却を行っている 55 社の比較分析となる。

本研究ではアンケート結果を分析対象とするために、分割表を作成し統計的な検討を行うことが多い。主な内容は、独立性の検定、調整残差による期待度数と観測度数の乖離の検討である。独立性の検討は χ^2 分布を用いて行うことができる。検定統計量は次式で表される。

$$\chi^2 = \sum_i \sum_j (f_{ij} - e_{ij})^2 / e_{ij} \tag{5.1}$$

ここで f_{ij}, e_{ij} はそれぞれ分割表を構成するセル ij の観測度数と期待度数である。χ^2 は分割表を構成する変数が互いに独立であるという帰無仮説のもとでは近似的に $\chi 2$ 乗分布に従う。従って、χ^2 を計算し、$\chi 2$ 乗分布表の値と比較することによって独立性の検定を行うことができる。本研究では、危険率 5 %で検定を行っている。帰無仮説が棄却される場合には、*を付して表すことにした。

各セルの観測度数と期待度数の差が有意であるかどうかは、次式で計算される調整残差を用いて検討することができる。ただし、n はデータ数、n_i, n_j はそれぞれ分割表の第 i 行または第 j 列の合計（周辺分布）を表す。

$$r_{ij} = (f_{ij} - e_{ij}) / \sqrt{e_{ij}(1 - \frac{n_i}{n})(1 - \frac{n_j}{n})} \tag{5.2}$$

調整残差は、帰無仮説

$H_0 : E(f_{ij}) = e_{ij}$

の下では平均 0、標準偏差 1 の標準正規分布にしたがうことが知られている [15]。したがって、検定統計量 r_{ij} を計算し標準正規分布表と比較することによって、観測度数と期待度数の差が有意で

表 5.11: 減価償却方法選択にあたって考慮する事項 (複数回答)

(イ)費用配分の適正化	231(50.0 %)
(ロ) 対象となる設備の使用状況	140(30.3 %)
(ハ) 毎期計上される償却費の平準化	128(27.7 %)
(ニ)事業環境が良いか悪いかということ	55(11.9 %)
(ホ)操業度の上昇、低下	63(13.6 %)
(ヘ)技術革新の速さ	184(39.8 %)
(ト)設備投資が大きいと考えられること	69(14.9 %)
(チ) 資金の早期回収による財務健全化	300(64.9 %)
(リ) 対象となる設備の陳腐化の速さ	166(35.9 %)
(ヌ) 巨額投資が完了したことによる負担増	71(15.4 %)
(ル) 事務の合理化	93(20.1 %)
(ヲ)修理費などの維持費用の発生状況	85(18.4 %)
合計	462(100.0 %)

あるかどうかを検定できる。本研究では、危険率を片側 5 %に設定した。観測度数が期待度数よりも大きいと考えられる時には△、観測度数が期待度数より小さいと考えられる時には▼を付けて表すこととした。

5.1.7 償却速度の違いに起因する仮説

収益性に関する仮説

定率法と定額法の選択に関して、最も一般的にいわれている仮説は次のものであろう。

(仮説 1) 定率法は定額法に比べて節税効果が高い。したがって、収益性が十分高い企業は定率法を採用する 。

定率法は定額法に比べて償却速度が速い。したがって、償却期間の早期に現金支出が節約できる。このため、定率法を採用した方が期間利益の現在価値は大きくなる。節税効果が大きいということはこれを意味している。この仮説を検証するために、償却方法の採用状況と質問 (f) との関連性を調べる。

質問 (f) は、収益性の良否をたずねたものである。ただ、収益性が良いか悪いかを直接たずねるかわりに財務比率の高低を、なおかつ数年間の大まかな傾向を聞いている。分析の際には、総資本経常利益率や売上高経常利益率が高いと答えた企業は収益性が高い企業、低いと答えた企業は収益性が低い企業と考えることにする。なお、分割表を作成する際には、「高い」または「やや高い」と回答した企業を「高い」、「低い」または「やや低い」と回答した企業を「低い」とした。これは、「やや高い」や「やや低い」という控えめな回答もそれぞれ「高い」または「低い」という経営者の主観を表しているものと考えることができるからである。このような選択肢のまとめは以降の質問に対しても行った。

表 5.12 は償却方法の採用状況と収益性によって分割表を作成したものである。χ^2 検定により、償却方法の採用状況と収益性が独立であるという帰無仮説は危険率 5 %で棄却することができる。また、調整残差の検定結果から、収益性が高い企業では定率法のみを採用する割合が高く、逆に

表 5.12: 償却方法と収益性

収益性	定率のみ	主に定率	主に定額	合計
高い	113(△)	6(▼)	7(▼)	126
普通	90(　)	19(　)	4(▼)	113
低い	166(▼)	48(△)	44(△)	258
合計	369	73	55	497

χ^2=36.99*

（注）＊：危険率5％で独立であるという帰無仮説を棄却
△：危険率5％で観測度数が期待度数より大きい
▼：危険率5％で観測度数が期待度数より小さい

収益性が低い企業では一部定額法によったり、主に定額法による割合が高くなっていることがわかる。この分析結果は収益性仮説を支持するものである。

償却費平準化に関する仮説

定率法により計算される償却費は毎期異なったものとなる。設備投資の頻度が少なければ計上される償却費の流れも予測がつきやすいし、有形固定資産の重要性が低ければ計上される償却費の変動もそれほど考慮しなくてもよいであろう。しかしながら、有形固定資産の重要性が高く設備投資が頻繁に行われるような企業が定率法によって償却を行うと、経過年数の異なる資産が混じりあうことによって、計上される償却費の時系列的な推移が複雑になりまた変動も大きくなるおそれがある。さらに大規模な設備投資が続く場合には、当初の償却費負担が過大となることもあり得る。これらのことは利益の変動をもたらし、企業評価を下げる原因ともなる。したがって次の仮説をたてることができる。

（仮説 2) 有形固定資産の重要性の高い企業や、設備投資が活発な企業では、償却費の変動を小さくするために定額法を採用する。

今回の調査では、設備投資が活発であるかどうかをみるために質問 (d) を、また、有形固定資産の重要性をきくために質問 (b) を設定していた。これらの質問も、財務的な数値ではなく、経営者の主観的な判断を尋ねた。財務的な数値ならば有価証券報告書から得られるということもあるが、経営者の判断が償却方法選択に直結しているのではないかという考えによる。

質問 (b) の結果から、有形固定資産の重要性は他社並みと考えている企業が 182 社 (36 %) と最も多いことがわかる。しかし、高いまたはやや高いと考えている企業があわせて 129 社 (26 %) にとどまっているのに対し、低いまたはやや低いと考えている企業は 187 社 (37 %) に達している。調査対象は製造業に限定しているので、37 %もの企業が有形固定資産の重要性は高くはないと考えているということは意外であった。

また質問 (d) の結果から、資本的支出の大きさは他企業並みと考えている企業が 225 社 (45 %) と最も多く、ついでやや少ないと考えている企業 119 社 (24 %) であることがわかる。しかしながら、資本的支出が少ないまたはやや少ないと感じている企業は合計 182 社 (36 %) であり、多いまたはやや多いと感じている企業 91 社 (18 %) の倍にも達している。このことは、先の質問 (b) でもみられたところであり、製造業に属する企業にあっても有形固定資産の重要性が漸次低下していることを示しているものと思われる。

表 5.13: 償却方法と資産の重要性

資産の重要性	定率のみ	主に定率	主に定額	合計
高い	81(▼)	23( )	25(△)	129
普通	140( )	29( )	13(▼)	182
低い	149(△)	21(▼)	17( )	187
合計 370	73	55	498	

χ^2=17.05*

(注) ＊：危険率 5 %で独立であるという帰無仮説を棄却
△：危険率 5 %で観測度数が期待度数より大きい
▼：危険率 5 %で観測度数が期待度数より小さい

表 5.14: 償却方法と設備投資の大きさ

資本的支出	定率のみ	主に定率	主に定額	合計
多い	57(▼)	14( )	20(△)	91
普通	176(△)	32( )	17(▼)	225
少ない	137( )	27( )	18( )	182
合計	370	73	55	498

χ^2=14.77*

(注) ＊：危険率 5 %で独立であるという帰無仮説を棄却
△：危険率 5 %で観測度数が期待度数より大きい
▼：危険率 5 %で観測度数が期待度数より小さい

表 5.13 および表 5.14 は、設備投資の大きさや有形固定資産の重要性と、償却方法の採用状況の関連性をみたものである。いずれの表においても償却方法との関連性がみられる。特に有形固定資産の重要性や設備投資が活発な企業では関連性が顕著である。調整残差の検討から、有形固定資産の重要性や設備投資が活発な企業では、定率法のみの採用は少なく、主に定額法による場合が多いことがわかる。この分析結果は仮説を支持するものである。

利益変動に関する仮説

先の仮説は、いわば費用の面から利益の変動を検討したものであるが、収益の面からも利益の変動を考えることができる。昭和 30 年代の海運業で定率法が多く採用されている理由として、収益の変動の大きさとそれに起因する投下資本の早期回収の可能性が指摘されている [72]。これを仮説の形にまとめると次のようになるであろう。

(仮説 3) 利益変動の激しい企業では、早期に投下資本の回収を図るために定率法を採用する。

この仮説を確かめるために設けた項目が質問 (g) である。ここでは収益性という言葉を使わずに利益率の変動の大きさをたずねた。これは、質問の意図を明確にするためでもあるが、収益以外例えば材料費の変動が大きい企業についても早期回収の必要性を考えることができるので、収益の変動に限定しないほうがよいと考えたためである。

利益変動の大きさと償却方法の採用状況をまとめたものが表 5.15 である。χ^2 検定の結果からは利益変動の大きさと償却方法選択の間に関連性はみられない。また、各区分ごとにみても調整残差に有意差はほとんどみられない。

表 5.15: 償却方法と利益変動の大きさ

利益の変動	定率のみ	主に定率	主に定額	合計
大きい	97( )	18( )	20( )	135
普通	148( )	36(△)	17( )	201
小さい	125( )	19( )	18( )	162
合計	370	73	55	498

χ^2=5.82

(注) △：危険率 5 %で観測度数が期待度数より大きい

経済的耐用年数に関する仮説

次の仮説は、除却損や売却損に関するものである。資産を使用の途中に除却したり売却したりすることはしばしばみられることである。特に使用の初期においては、原価に比して割合低い売却価格を強いられることもあり、定額法に比べて資産の低価評価になる定率法は保守主義会計を満足させるものとされる。これを仮説の形にまとめると次のようになるであろう。

(仮説 4) 償却計算の基礎となった耐用年数より経済的耐用年数が短くなるような企業では、除却損や売却損の発生を少なくするために定率法を採用する。

表 5.16: 償却方法と除却損・売却損の発生

除却損等の発生	定率のみ	主に定率	主に定額	合計
多い	34( )	6( )	9(△)	49
普通	121( )	24( )	16( )	161
少ない	215( )	42( )	30( )	287
合計	370	72	55	497

χ^2=3.02

(注) △：危険率 5 %で観測度数が期待度数より大きい

表 5.16 は償却方法と除却損や売却損の発生状況をまとめたものである。χ^2 検定を行っても有意差はみられなかった。したがって償却方法との関連性もいえないが、除却損や売却損には減価償却方法がすでに反映しているという仮説検証上の問題点がある。そこで経済的耐用年数の長短と償却方法の関連をみることにした。

質問 (j) から、経済的耐用年数が法定耐用年数と同じくらいだと感じている企業は 139 社 (28 %) にすぎない。多くの企業は、経済的耐用年数は法定耐用年数の 7〜8 割程度と考えている。この "感覚" と償却方法採用の関連性をみたものが表 5.17 である。表 5.17 からも償却方法と経済的耐用年数の間の関連性をみることはできない。経済的耐用年数が短い企業が定率法を採用する傾向にあるとはいえないので、仮説を支持する結果は得られなかった。

表 5.17: 償却方法と経済的耐用年数

経済的耐用年数	定率のみ	主に定率	主に定額	合計
半分位	69( )	14( )	6( )	89
7〜8 割	206( )	39( )	26( )	271
同じ位	95(▼)	21( )	23(△)	139
合計	370	74	55	499

χ^2=6.82

(注) △：危険率 5 %で観測度数が期待度数より大きい
▼：危険率 5 %で観測度数が期待度数より小さい

ところで表 5.17 から、経済的耐用年数が法定耐用年数と同じ位と感じている企業で、定額法を採用する傾向が強く、定率法を採用する傾向が弱いこととがわかる。設定した仮説は、経済的耐用年数が短い場合に対するものなので、このような傾向を説明することはできない。この原因については、後ほど検討することにする。

借入依存に関する仮説

仮説 5 は資金調達に関するものである。資金調達に関しては相反するふたつの仮説を設定することができる。

(仮説 5-1) 資金調達を借入金に頼っている企業は、借入の返済資金を確保するために定率法を採用する。

借入金の返済を行うためには現金が必要である。このため現金支出がより抑えられる定率法を採用するというものである。仮に、返済資金確保仮説と呼ぶことにしよう。

(仮説 5-2) 資金調達を借入金に頼っている企業は、財務外観を良くし、借入を容易にするために定額法を採用する。

定率法による場合、特に償却の早期には、多くの償却費を計上するために、定額法による場合より利益が少なくなる。この結果、追加資金を借り入れようとする場合に不利になるおそれがある。このため、借入依存の高い企業は定額法を採用するというものである。これを財務外観性仮説 [79] とよんでいる。

表 5.18 は借入依存度と償却方法採用状況をまとめたものである。両者の間には関連性がみられる。つまり、借入依存度の高い企業では、定額法を採用する傾向が強く、定率法を採用する傾向は弱い。これは、財務外観性仮説を支持するものとなっている。

5.1.8 償却計算の基礎の違いに起因する仮説

簡便方法選択に関する仮説

償却計算が未償却残高をもとに行えるという定率法の特徴は、償却計算の簡便化につながる。したがって、次のような仮説をたてることができる。

(仮説 6) 償却計算を簡略化しようとする企業は定率法を採用する。

表 5.19 は償却計算を計算機処理によっている場合と手作業によっている場合での償却方法採用

表 5.18: 償却方法と借入依存度

借入依存度	定率のみ	主に定率	主に定額	合計
高い	87(▼)	24( )	34(△)	145
普通	85( )	21( )	10( )	116
低い	198(△)	28(▼)	11(▼)	237
合計	370	73	55	498

χ^2=38.95*

(注) ＊：危険率５％で独立であるという帰無仮説を棄却
△：危険率５％で観測度数が期待度数より大きい
▼：危険率５％で観測度数が期待度数より小さい

状況をみたものである。仮説が正しければ、手作業によっている場合には定率法を採用する割合が高くなるはずであるが、χ^2 検定の結果からも有意差はみられず、観測度数と期待度数の間にも差はみられない。つまり、償却計算を手作業によっている場合でも、簡略化のために定率法を採用するという傾向はみられない。

表 5.19: 償却方法と償却計算の手段

償却計算の手段	定率のみ	主に定率	主に定額	合計
計算機処理	350( )	72( )	52( )	474
帳簿上で計算	18( )	2( )	3( )	23
合計	368	74	55	497

χ^2=0.76

自動修正能力に関する仮説

定率法の優れた点として、自動修正能力をあげる説がある [56][57]。自動修正能力とは、耐用年数の見積の誤りによる未償却残高や償却累計の誤差を、通常の償却計算を通して解消していく機能である。これも償却計算の基礎が未償却残高であることに起因している。自動修正能力は誤差が顕在化しにくい総合償却において特に必要な機能であるとされる。したがって次のような仮説をたてることができる。

(仮説 7) 総合償却を行っている企業では、耐用年数の誤りによる誤差を自動修正するために定率法を採用する。

表 5.20 は償却計算の単位と償却方法の採用状況をまとめたものである。なお、アンケートでは、種類や耐用年数の似通っている資産を一括にした償却計算を一括償却、異なる種類の資産をまとめたものを総合償却としている。

χ^2 検定の結果、償却計算の単位と償却方法の採用状況には関連性がみられる。しかしながら、仮説から予想されたものとは逆に、総合償却を行っている企業では、主に定額法による場合が多く、定率法のみで償却を行っている企業の割合は低い。この結果は仮説 7 からは導くことができず、後述のように他の要因を入れて考える必要がある。

表 5.20: 償却方法と総合償却

償却計算の範囲	定率のみ	主に定率	主に定額	合計
個別償却	245(△)	44( )	26(▼)	315
一括償却	43( )	9( )	5( )	57
総合償却	82(▼)	21( )	24(△)	127
合計	370	74	55	499

$\chi^2=12.19$

（注） ＊：危険率５％で独立であるという帰無仮説を棄却
△：危険率５％で観測度数が期待度数より大きい
▼：危険率５％で観測度数が期待度数より小さい

5.1.9 償却可能範囲額と償却方法の採用

以上の７つの仮説はアンケート調査を行う前にあらかじめ想定していたものである。検証の結果、仮説 1. 収益性仮説、仮説 2. 償却費平準化仮説および仮説 5-2. 財務外観性仮説は支持されることがわかった。しかし一方では、当初予想していなかった結果、つまり想定していた仮説からは説明できない結果も得られている。これらの点を検討しておこう。

最初は、仮説 4. 経済的耐用年数仮説に表れた有意差である。先に述べたように全体としてみた時には、経済的耐用年数の長短と償却方法の採用には関連性はみられない。しかし、調整残差をみてみると、経済的耐用年数が法定耐用年数と同じくらいと答えている企業において主に定額法によっている企業が多く、定率法のみによっている企業が少ないという結果が得られていた。

経済的耐用年数が法定耐用年数とほぼ同じであるならば、除却損や売却損の発生は償却方法によらない。したがって定率法のみによる企業が少なく、主に定額法による企業が多いということは仮説 4 からは説明できない。

経済的耐用年数が比較的長い企業で、定率法が少なく定額法が多い理由として、税法における 95％償却を考えることができるであろう。法人税法の規定によると、耐用年数が経過して帳簿価額が 10％になった後も、５％に達するまで償却を続けることができる（法人税施行令 61 ①）。定率法でも定額法でも耐用年数経過後には未償却残高が 10％になるように設計されているが、さらに続けて償却を行う場合には定率法による場合の方が償却の速度は遅くなってしまう。たとえば、耐用年数 10 年の資産を帳簿価額が５％になるまで償却するとすれば、定額法の場合には 11 年で済むが、定率法では 13 年を必要とする。さらに、耐用年数が 15 年の資産であれば、定額法の 16 年に対し定率法では 20 年を必要とする。一般には、定率法による場合の方が速く償却を行うことができる。しかしながら、経済的耐用年数が長く、耐用年数を超えて償却を行う場合には、定額法による場合の方が速く償却を完了することができる。これが、比較的経済的耐用年数が長い企業で、定額法を採用する割合が高く、逆に定率法を採用する割合が低い理由と考えられるだろう。

5.1.10 償却の単位と償却方法

自動修正仮説からは、総合償却を行っている企業では、耐用年数の誤差を自動修正するために定率法を採用する割合が高くなることが予想されていた。しかし、調査結果からは、個別償却を行っている企業に定率法を採用している割合が高く、総合償却を行っている企業では定額法を採用する割合が高いことがわかった。この原因を検討してみよう。

説明変数相互の関連性をみてみると、総合償却か個別償却かという償却計算の単位は資産の重要性との間に関連性が強いことがわかる。資産の重要性はまた償却方法の選択とも関連があった。そこで、資産の重要性を第3変数として導入する。また、分割表の検討から、資産の重要性が高い場合には、定額法を採用する傾向や総合償却が採用される傾向にあり、また、資産の重要性が低い場合には、定率法を採用する傾向や個別償却が採用されやすいという関連性が認められる。そこで、資産の重要性が高い企業のみ、および資産の重要性が低い企業のみに限定して償却計算の単位と償却方法の関係を見ることとした。

償却計算の単位と償却方法選択の間に見られた関係が、資産の重要性に影響を受けないものであるなら、資産の重要性が高い企業と低い企業に分けて分割表を作成した場合でも有意差は表れるはずである。表10に資産の重要性毎に分割表を作成した場合の検定結果を示している。検定の結果からは、いずれの指標で見ても独立であるという帰無仮説は棄却されない。このことから、償却計算の単位と償却方法選択の間に見られる関係は、資産の重要性によって引き起こされた疑似相関の可能性が強い。

表 5.21: 資産の重要性毎にみた償却計算の単位と償却方法の関係

資産の重要性が高い企業

	定率のみ	主に定率	主に定額	合計
個別償却	44	12	14	70
一括償却	13	3	1	17
総合償却	24	8	10	42
合計	81	23	25	129

$\chi^2=2.81$

資産の重要性が低い企業

	定率のみ	主に定率	主に定額	合計
個別償却	111	16	8	135
一括償却	15	1	4	20
総合償却	23	4	5	32
合計	149	21	17	187

$\chi^2=6.80$

5.2 減価償却方法選択のファジィ推論

本節の目的は、償却方法選択に関する規則を設定して、その規則の有効性を検証することである。定率法のみで償却を行っている企業と、定額法を採用している企業との間で、収益性の指標、例えば総資本経常利益率などの比較を行うと、定率法のみで償却を行っている企業の平均値の方が高いことは容易に確かめることができる[80]。本節で確かめようとするものは、逆の命題、つまりある条件を満たすならば、特定の償却方法を採用する傾向があるというものである。本節では、償却方法との関連性が見られた要因からなる分割表を作成し、これから検証のための規則を導いた。そしてこれらの規則の有効性を検証するためにファジィ推論を利用した。

5.2.1 ファジィ推論とファジィ規則

判別分析では、判別したい事象に関係していると思われる変数を設定し、それらの変数を使ってどの程度判別できたかを検証し、必要ならば説明変数の入れ替えを行うという手順を踏むことになる。しかしながら、データが変れば、選択される変数の組合わせも変化する場合があり、うまく判別できたとしても定性的な知識としてまとめることが困難である。

一方、推論の規則が明示的で理解しやすいものにファジィ推論がある。max-min 合成による場合には、どのルールによって結論が得られたかということも理解しやすい。しかし、ファジィ推論による場合には、どのようなルールを設定するか、また、ファジィ集合のメンバーシップ関数をどのように決定すべきかという問題が残る。本研究では、償却方法選択に関連するアンケート調査を行っていたので、ファジィ推論のためのルールをこのアンケート調査をもとに決定することとした。

前節で見たように、それぞれの仮説毎に独立性の検定を行ったところ, 収益性, 資産重要性, 借入依存度との間に償却方法選択との関連性が見られた。それぞれの分割表を再び表 5.22 に示した。なお, 表 5.22 には調整化残差の検定結果もあわせて示している。

表 5.22: 収益性、資産重要性、借入依存度と償却方法の選択

(a) 収益性と償却方法の選択

収益性	定率のみ	主に定率	主に定額	合計
高い	113(△)	6(▼)	7(▼)	126
普通	90()	19()	4(▼)	113
低い	166(▼)	48(△)	44(△)	258
合計	369	73	55	497

(b) 資産重要性と償却方法の選択

資産の重要性	定率のみ	主に定率	主に定額	合計
高い	81(▼)	23( )	25(△)	129
普通	140( )	29( )	13(▼)	182
低い	149(△)	21(▼)	17( )	187
合計 370	73	55	498	

(c) 借入依存度と償却方法の選択

借入依存度	定率のみ	主に定率	主に定額	合計
高い	87(▼)	24( )	34(△)	145
普通	85( )	21( )	10( )	116
低い	198(△)	28(▼)	11(▼)	237
合計	370	73	55	498

△：危険率 5 %で観測度数が期待度数より大きい
▼：危険率 5 %で観測度数が期待度数より小さい

表 5.22 から, 収益性が高いと感じている企業では, 定率法のみで償却を行う傾向が強いことがわかる。また, 資産の重要性や借入依存が高いと感じている企業では定額法を採用する傾向が強いことがわかる。一方, 収益性, 資産重要性, 借入依存度のそれぞれについて, 普通と回答した企業においては, 償却方法選択に傾向がほとんど見られなかった。従って、収益性、資産重要性、借入依存

度のすべてについて「高い」または「低い」と回答している企業のみを抽出して分割表を作成し直した。さらに、この過程において、「主に定率」と「主に定額」の2群をまとめて集計した。これは、これら2群の有意さに共通性が見られること、さらにはこれら2群を識別する知見が得られていないことによる。

表5.23には,収益性、資産重要性、借入依存度の3つの要因からなる分割表を示した。いずれかの分割表において普通と回答した企業を除いているので、合計は205社である。表5.23では、それぞれの項目について、高いまたは低いと回答した企業のみを集計したにもかかわらず、調整化残差の有意差は8階級中3階級に現れたにすぎない。

遺伝的アルゴリズムなどの発見的方法を使わなくても、有意差が現れた3階級から次の2つの規則を導くことができる。

規則1：収益性が高く、借入依存度が低い企業では、定率法のみで償却を行う。

規則2：収益性が低く、資産重要性が高く、借入依存度が高い企業では、定額法を採用する。しかしこれらの規則が適用できるのは、8階級中3階級、205社中106社に過ぎない。さらに、106社に適用して識別したとしても判別率は77％に過ぎない。この例のように群に含まれるデータ数に偏りがあり、また多くの階級が群を識別する能力を持っていないと考えられる場合には、分割表から得られた規則をそのまま観測結果に適用しても、部分的かつ不十分な識別結果しかえられない。しかし一方、得られた2つの規則には、高いや低いなどファジィ集合と考えることができる要素が含まれている。そこで、得られた2つの規則をファジィ推論を行うための規則と考えて、有効性を探ることとした。

表5.23: 償却方法選択状況の分割表

収益性	資産重要性	借入依存度	定率のみ	定額採用
高	高	高	5( )	2( )
高	高	低	16(△)	0(▼)
高	低	高	3( )	0( )
高	低	低	39(△)	4(▼)
低	高	高	20(▼)	27(△)
低	高	低	11( )	5( )
低	低	高	27( )	12( )
低	低	低	24( )	10( )
	合計		145	60

5.2.2 ニューラルネットワークによる主観の評価

ファジィ推論を行うためには、分析対象となるそれぞれの企業が、“収益性が高い”や“借入依存度が低い”というファジィ集合を満たす度合いを定めなければならない。収益性、資産重要性や借入依存度を測定するためには、既存の財務比率を利用することができる。しかし、多くの考えうる財務比率からどの比率を選ぶか、またどのようにして0から1までの数に変換するかという問題があるので、ニューラルネットワークの出力をそのままファジィ集合との一致度と考える方法をとった。これは、いくつかの財務比率を入力とするニューラルネットワークを組み、“高い”または“低い”というアンケートの調査結果を教師信号として学習させる。そして、得られたニューラルネットワークの出力をファジィ集合との一致度とする方法である。この方法をとることによっ

て、規則 1 または規則 2 が適用できない階級に含まれる企業の識別も行うことができ、さらにはいずれかの質問に“普通“と回答した企業にも規則を適用することができるようになる。しかし一方では、財務比率を得ることができる企業に対象が限定される。

収益性ひとつとってみても、これを測る目的で計算される財務比率は多数あるので、代表的と思われる財務比率を選ぶことにした。本研究では、通産省編「わが国企業の経営分析」に含まれている財務比率を参考に選んだ。収益性を測る比率としては、次の 3 つの比率を選んだ。ただし、減価償却方法の違いによる影響を除くため、償却前の値を使って計算している。また、財務比率を計算するにあたっては、コスモ証券編コスモ財務データ 1995 年版を利用している。

$$\text{総資本経常利益率} = \frac{\text{経常利益} + \text{有形固定資産減価償却費}}{\text{総資本} + \text{有形固定資産減価償却累計} + \text{割引譲渡手形}}$$

$$\text{売上高経常利益率} = \frac{\text{経常利益} + \text{有形固定資産減価償却費}}{\text{売上高}}$$

$$\text{粗付加価値率} = \frac{\text{粗付加価値合計}}{\text{売上高}}$$

財務データが得られた 226 社のうち、収益性が高いと回答した会社は 57 社、低いと回答した会社は 126 社、普通と回答した会社は 43 社であった。このうち、収益性が高い、もしくは低いと答えた会社 183 社について $Mann$ - $Whitney$ のW検定を行った。

$Mann$ - $Whitney$ のW統計量は、収益性が高いと回答した企業群と低いと回答した企業群のすべての財務比率の組合わせ $(x_i、x_j)$ のうち、$x_i > x_j$ となるものの個数を表す [10]。ここで、x_i は収益性が高いと答えた企業の財務比率、x_j は収益性が低いと答えた企業の財務比率を表す。したがって、W統計量が大きい時（または小さい時）に当該財務比率で 2 群はよく識別されていることになる。また、Wの平均と分散は次式で表され、正規分布に近似的にしたがうことから、通常の統計的検定を行うことができる [10]。

$$E(W) = mn/2$$

$$Var(W) = mn(m+n+1)/12$$

ただし、ここで m, n は 2 群のデータ数であり、収益性に関する検定の場合には、それぞれ 57 と 126 である。したがって 2 群の組合せの数は 7,182 通りとなる。表 5.24 に検定結果を示した。

表 5.24: 収益性比率について Mann-Whitney 検定

	組合せの数	W統計量	同%	検定結果
総資本経常利益率	7,182	6,547	91	*
売上高経常利益率	7,182	6,583	92	*
粗付加価値率	7,182	5,015	70	*
ニューラルネット出力	7,182	6,634	92	*

*は危険率片側 2.5 %で収益性が高いと回答した企業の値が大きい

表 5.24 の検定結果から、いずれの財務比率も収益性の高低を判断する指標として有効であることがわかる。しかしながら、複数の財務比率を用いることによって識別の程度がさらに向上することも予想されるし、また、ファジィ集合との一致度として利用するためには 0 から 1 までの

範囲に財務比率の値を変数変換する必要がある。そこで、3入力1出力の図5.1のようなニューラルネットを組んだ。教師信号は、当該企業が収益性が高いと回答している場合に0.9、低いと回答している場合に0.1とした。

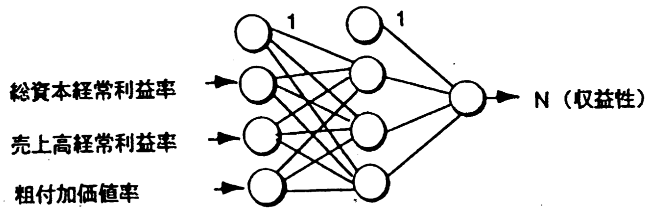


図 5.1: 収益性のニューラルネット

財務比率はそれぞれ値の範囲が異なる。そこで、財務データが得られた226社の平均と標準偏差を用いて標準化した。各企業の財務比率は、おおむね±3 σ 以内に納まることを期待して、標準化後の財務比率を k=1/3 のシグモイド関数に入力し、その出力値を図5.1に示されたニューラルネットの入力値としている。

収益性が高いまたは低いと回答している183サンプルに対する3,000回の学習の結果、2乗誤差の平均は0.0517であった。そして、学習後のニューラルネットと、各社の総資本経常利益率など3つの財務比率を使って計算した出力値に対して $Mann$ - $Whitney$ 検定を行った。表5.24にはその結果もあわせて示している。

ニューラルネットの出力値も、他の財務比率と同様に収益性の高低の識別に有効であることがわかる。さらに、単独の財務比率を用いた場合よりも、若干ではあるが、識別の程度が向上していることがわかる。また、ニューラルネットの出力であるので、当然のことではあるが、0から1までの範囲におさまっている。したがって、学習後のニューラルネットの出力値をもって、当該企業の収益性の高低を測る指標とすることにした。

借入依存度についても、同様の分析を行って図5.2に示されるニューラルネットを組んだ。 借入依存度については当初、借入金利子率も入力変数に加えていたが、借入依存度が高いまたは低いという感覚との関連性がみられなかったので削除した。

$$\text{借入社債比率} = \frac{\text{短期借入金＋長期借入金＋社債＋割引譲渡手形}}{\text{総資本＋有形固定資産減価償却累計＋割引譲渡手形}}$$

$$\text{金融費用負担率} = \frac{\text{金融費用}}{\text{売上高}}$$

財務数値が得られた266社のうち、借入依存度が高いまたは低いと回答した企業は合計175社であった。この175社の財務比率と借入依存度の感覚を学習データとして、3,000回の学習を行った。2乗誤差の平均は0.1156であった。また、収益性の場合と同じくそれぞれの財務比率を単独で用いるよりは、ニューラルネットの出力値による場合の方が、W検定量の値は大きかった。このことから、ニューラルネットの出力を用いた方が、借入依存度が高いまたは低いという感覚をより分離できるものと思われる。

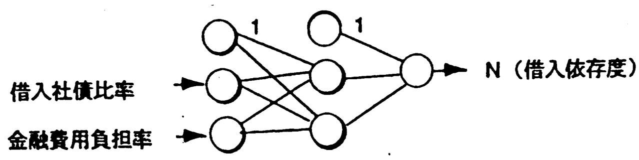


図 5.2: 借入依存度のニューラルネット

表 5.25: 借入依存度に関する比率について Mann-Whitney 検定

	組合せの数	W統計量	同%	検定結果
借入社債比率	7,350	5,911	80	*
金融費用負担率	7,350	5,670	77	*
ニューラルネット出力	7,350	5,937	81	*

*は危険率片側2.5 %で借入依存が高いと回答した企業の値が大きい

次に資産重要性の高低を判断するためのニューラルネットを組んだ。入力として用いた財務比率は次の3つである。

$$減価性有形固定資産比率 = \frac{有形固定資産 + 有形固定資産減価償却累計 - 土地}{総資本 + 有形固定資産減価償却累計 + 割引譲渡手形}$$

$$労働装備率 = \frac{有形固定資産 + 有形固定資産減価償却累計 - 建設仮勘定 - 土地}{期末従業員数}$$

$$有形固定資産対売上高 = \frac{有形固定資産 + 有形固定資産減価償却累計}{売上高}$$

資産重要性が高いまたは低いと回答した企業は合計145社であった。この145社を学習データとして、前と同じく3,000回の学習を行った。2乗誤差の平均は0.0911であった。なお、収益性や借入依存度と同様に、資産重要性が高いと回答している場合に0.9、低いと回答している場合に0.1をそれぞれ教師データとしている。また、財務比率を単独で用いるより、ニューラルネットの出力値を用いた方が、W検定量も大きくなった。したがって、収益性や借入依存度と同様に、ニューラルネットの出力を用いた方が、資産重要性が高いまたは低いという感覚をより分離できるものと思われる。

それぞれの学習後のニューラルネットを使って、収益性、資産重要性、借入依存度のすべてについて高いまたは低いと回答している100社を対象に主観のニューラルネットワークによる判別をおこなった。“高い”と回答している場合には0.9、“低い”と回答している場合には0.1として学習しているので、ニューラルネットの出力が0.5以上のとき“高い”、0.5未満のとき“低い”と判断することとした。結果を表5.27にまとめている。

収益性に対する回答結果を学習したニューラルネットは“収益性が高い”と回答した企業31社のうち27社を正しく識別している。同様に“収益性が低い”と回答した69社のうち64社を正し

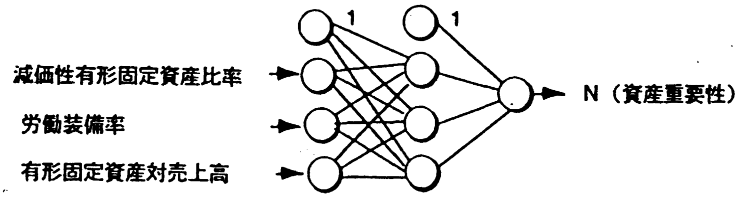


図 5.3: 資産重要性のニューラルネット

表 5.26: 資産重要性に関する比率について Mann-Whitney 検定

	組合せの数	W統計量	同%	検定結果
減価性有形固定資産比率	5,046	4,266	85	*
労働装備率	5,046	3,928	78	*
有形固定資産対売上高	5,046	4,214	84	*
ニューラルネット出力	5,046	4,380	87	*

*は危険率片側 2.5 %で資産重要性が高いと回答した企業の値が大きい

く識別している。識別率は 91 %である。また、資産重要性に関する回答結果を学習したニューラルネットの識別率は 82 %、借入依存度を学習したニューラルネットの識別率は 81 %であった。それぞれのニューラルネットがいわば主観を識別しようとするものである点を考慮すると、ある程度の識別率は確保できたものと考えていいだろう。したがって、これらニューラルネットワークの出力値をそれぞれ、“収益性が高い”、“資産需要性が高い”、“借入依存度が高い”というファジィ集合との一致度とみなすこととした。逆に、1 から出力値を引いたものを、“収益性が低い”、“資産重要性が低い”、“借入依存度が低い”というファジィ集合との一致度を表すものと解釈した。

5.2.3 簡略化ファジィ推論

本研究では、後件部を実数値で表現した簡略化ファジィ推論を用いる。入力を $x_j(j=1,\cdots,m)$ とし、出力を y とすると、簡略化ファジィ推論の規則 $R_i(i=1,\cdots,n)$ は次のように表される。

$$R_i : if\ x_1\ is\ A_1 \ldots x_m\ is\ A_m\ then\ y_i\ is\ w_i$$

ここで、A_j はファジィ集合、w_i は実数値である。また i はルール番号である。ファジィ推論の推論結果 z は次式で与えられる。

$$\mu_i = \prod_{j=1}^{m} \mu_{A_{ij}}(x_j)$$

$$z = \frac{\sum \mu_i w_i}{\sum \mu_i}$$

表 5.27: 要因ごとの判別結果

(*a*) 収益性に関する判別結果

	判別結果		
回答	収益性高い	収益性低い	合計
収益性高い	27	4	31
収益性低い	5	64	69
合計	32	68	100

(*b*) 資産重要性に関する判別結果

	判別結果		
回答	資産重要性高い	資産重要性低い	合計
資産重要性高い	35	7	42
資産重要性低い	11	47	58
合計	46	54	100

(*c*) 借入依存度に関する判別結果

	判別結果		
回答	借入依存度高い	借入依存度低い	合計
借入依存度高い	41	7	48
借入依存度低い	12	40	52
合計	53	47	100

後件部の実数値はデルタルールと呼ばれる方法によって求めることができる。修正係数を0.1に設定し、1000回の学習の結果、$w_1 = \mathbf{1.15}$、$w_2 = \mathbf{0.50}$ を得ることができた。

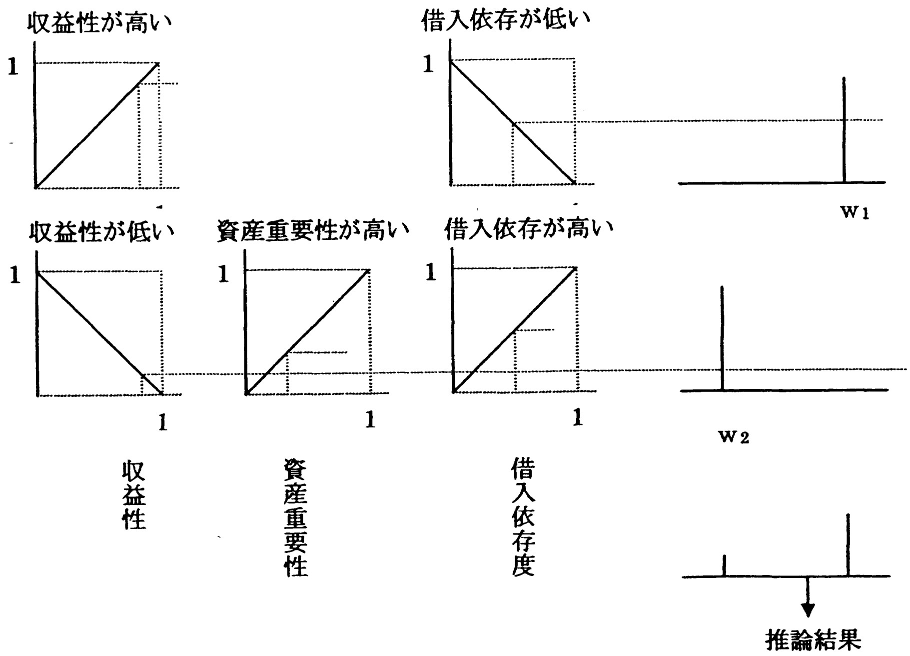


図 5.4: 2つの規則によるファジィ推論

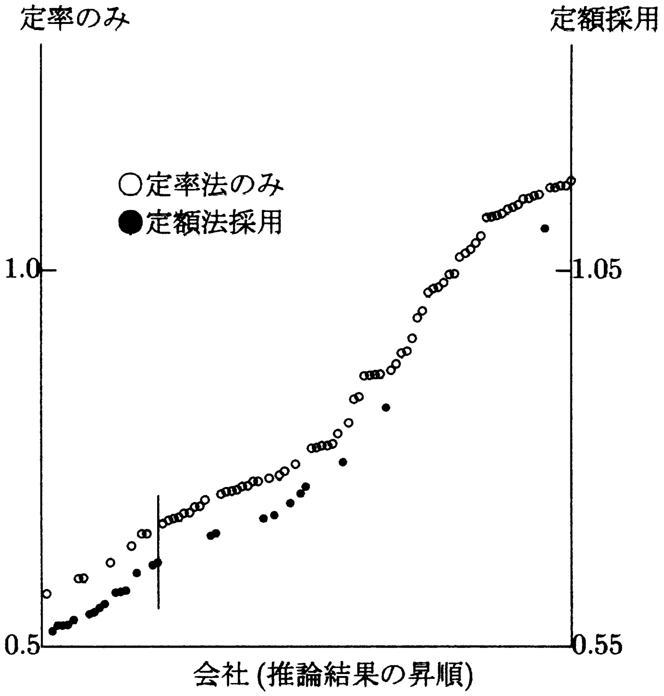


図 5.5: 2 つの規則による判別結果

表 5.28: 償却方法選択の判別結果

	判別結果		
回答	定率法のみ	定額法採用	合計
定率法のみ	68	7	75
定額法採用	10	15	25
合計	78	22	100

財務数値が得られる 100 社について、償却方法選択状況を、得られた規則をもとに推定した。それぞれの会社の出力を小さい順に並べ、標本における誤判別が最小となる点をカットオフポイントに設定した。本研究の場合、出力が 0.66 以下ならば定額法を採用する企業、それ以上ならば定率法のみで償却を行う企業と判別されることとなった (図 5.5)。実際の選択状況と判別状況を表 5.28 に示した。この表 5.28 から、2 つの規則のみでも 83 %の企業を識別できることがわかった。

前章でみたあやめの識別問題とは異なり、分割表に含まれる階級には群を識別する能力がないと思われるものが含まれていた。このため、分割表から得られた規則を観測結果に直接適用しても、部分的な判別しかできない。分割表からの規則抽出の有効性を確認するために、ファジィ推論に応用して判別力を確認した。この結果、直接規則を適用できないサンプルに対しても判別を行うことができるようになった。そして 83 %の企業は正しく判別できた。このような分析を通して、分割法から得られる規則が汎用性を持っていると考えることもできるであろう。

5.3 経営環境に関する大規模な調査からの償却方法選択に関する規則の抽出

本章前半では、あらかじめ設定した仮説をもとにアンケートを実施し、償却方法選択との関連性が見られた変数からなる分割表を作成した。そしてその分割表から少数の規則を抽出した。したがって、得られる規則も容易に解釈できるものであった。しかし、あらかじめいくつかの仮説が設定されるとは限らない。このような場合には、関連があるかどうかわからない変数も含めて分割表の作成、規則の抽出を行わなければならない。われわれは第 4 章において、変数増減法を使った規則抽出について検討した。この方法を減価償却方法選択問題へ適用してみる。

5.3.1 経営環境に関する調査の概要

企業の置かれている状況を広く調査する目的で、上場企業等を対象に、1994 年、1997 年、2000 年の 3 回にわたってアンケート調査を行った。調査項目は減価償却方法選択状況の他、29 項目 (1994 年調査は 26 項目) を設定した。調査項目は表 5.29 の通りである。選択肢の後の数は、第 4 章で述べたようにそれぞれの選択肢を表す素数である。なお、詳細な調査項目および単純集計結果は付録に示した。

1994 年の調査は、1994 年 11 月に実施した。金融、電力、ガスを除く全上場企業 1,969 社を対象に行い、うち 655 社から回答を得た。回収率は 33.3 %である。1997 年の調査は 1997 年 8 月に実施した。対象は株式公開企業のうち銀行証券・保険業を除く 2,937 社であり、有効回答総数は 833 社であった。回収率は 28.4 %である。2000 年調査は 2000 年 12 月に実施した。対象は、日本国内証券市場で資金調達を行っている一般事業業種 (銀行金融・保険・証券業を除く) の 3,265 社である。有効回答数は 664 社、回収率は 20.3 %であった。1994 年調査では、上場企業に限っていたが、その後店頭公開企業なども含めた。また、いずれの調査でも質問用紙を経理部長宛てに発送し、同封はがきに回答を記入し返送していただくという方法を取っている。

表 5.29: アンケートに含まれる説明変数

項目名	選択肢
(1) 輸出割合	大きい (2)、普通 (3)、小さい (5)
(2) 収益性	高い (2)、普通 (3)、低い (5)
(3) 利益変動	大きい (2)、普通 (3)、小さい (5)
(4) 棚卸資産回転	長い (2)、普通 (3)、短い (5)
(5) 材料価格変動	大きい (2)、普通 (3)、小さい (5)
(6) 輸入依存	高い (2)、普通 (3)、低い (5)
(7) 現金対掛売り	現金売りが多い (2)、半々(3)、掛売りが多い (5)
(8) 売掛回収	長い (2)、普通 (3)、短い (5)
(9) 借入依存	大きい (2)、普通 (3)、小さい (5)
(10) 主力銀行	大きい (2)、普通 (3)、小さい (5)
(11) 社長出自	創業者 (2)、生え抜き (3)、外部人材 (5)
(12) 製品販売先	少数の販売先 (2)、多岐に渡る (3)
(13) 仕入先分布	少数の仕入先 (2)、多岐に渡る (3)
(14) 役員賞与	大きい (2)、普通 (3)、小さい (5)
(15) 設備投資効果	概ね予想通り (2)、予想とはずれる (3)
(16) 設備投資期間	長い (2)、普通 (3)、短い (5)
(17) 利益水準	高い (2)、普通 (3)、低い (5)
(18) 有固割合	高い (2)、普通 (3)、低い (5)
(19) 自己資本比率	高い (2)、普通 (3)、低い (5)
(20) 付加価値率	高い (2)、普通 (3)、低い (5)
(21) 研究開発	高い (2)、普通 (3)、低い (5)
(22) 販売費	高い (2)、普通 (3)、低い (5)
(23) 資金余裕	大きい (2)、普通 (3)、小さい (5)
(24) 損益分岐点	高い (2)、普通 (3)、低い (5)
(25) 市場環境	成熟期 (2)、市場拡大期 (3)、萌芽的時期 (5)
(26) 市場集中	高い (2)、普通 (3)、低い (5)
(27) 多角化	大きい (2)、進めたい (3)、必要ない (5)
(28) 社長専門	営業 (2)、技術 (3)、経理 (5)、総務 (7)
(29) 耐用年数	半分以下 (2)、半分位 (3)、7〜8 割程度 (5)、法定程度 (7)

本研究で実施したアンケート調査では、減価償却方法の選択状況に関する選択肢として、1) 定率法のみで償却、2) 一部定額法採用、3) 主に定額法で償却を設定していた。しかし、1994 年度および 1997 年度の調査では、定率法にみによって償却を実施している企業が圧倒的であり, 一部に定額法を採用している企業や主に定額法で償却を行っている企業は少ない。このため、1994 年調査結果と 1997 年調査結果を分析する際には、一部定額法採用と主に定額法で行っている企業をまとめて、定額法採用企業として分析することとした。したがって、1994 年および 1997 年調査結果に対して、定率法のみで償却を行っている企業と、定額法を採用している企業を識別する規則抽出を行う。

平成 10 年 3 月の税法改正で、平成 10 年 4 月 1 日以降取得された建物の償却方法は、定額法の

みによることとされた(それ以前に取得された建物については定額法または定率法の選択)。このため、2000年調査に対しては、一部定額法を採用している企業において, 建物の償却方法に関する税法改正の影響があると思われることから、定率法のみで償却を行っている企業と、主に定額法で償却を行っている企業との比較分析を行うこととした。

表 5.30: 減価償却方法の選択状況

減価償却方法	1994年調査	1997年調査	2000年調査
1　定率法のみ	470(72.1 %)	640(77.0 %)	155(23.4 %)
2　一部定額法	121(18.6 %)	136(16.4 %)	430(65.0 %)
3　主に定額法	61(9.4 %)	55(6.7 %)	77(11.6 %)
合計	652(100.0 %)	831(100.0 %)	662(100.0 %)

1994, 1997各調査において、減価償却選択状況について回答のなかったアンケート結果は、研究の目的に鑑み検討の対象外とした。さらに、得られた規則の妥当性を検証する目的で、アンケート結果を、分析用と検証用に2分した。このとき、アンケート結果を会社コード（1994年調査では業種コード）を使って並び替え、交互に分析用と検証用に区分した。これは、分析用と検証用で業種の偏りがないようにするためである。1994, 1997年の各調査において、分析用と検証用のデータ数は次のようになった。

表 5.31: 分析と検証用データ数

調査年度	分析用	検証用	計
1994年	326	326	652
1997年	416	415	831
2000年	116	116	332

2000年調査において分析用と検証用のデータ数が少ないのは、先に述べたように、一部定額法を採用していると回答している企業を除いているためである。

それぞれの調査において、分析用データから規則の抽出を行い、検証用データを使って規則の有効性を検証することとした。

5.3.2　1994年調査から得られた規則

規則抽出は、4.5および4.9の方法によっているが、まとめると次のような手順となる。

1) 分割表への変数の追加

1.0　前のステップの原始分割表のAICを保存する。

1.1　分割表に含まれていない変数を一つ選ぶ。

1.2　当該変数の加えた原始分割表を作成する。

1.3　遺伝的アルゴリズムを使って区分選択を行う。

1.4　分割表に含まれていないすべての変数について、1.1から1.3までの処理を行い、タブーリストに含まれておらず、かつ区分選択後のAICが最も小さくなる変数を選ぶ。

1.5　1.4で選ばれた変数を追加し、区分選択を行った分割表を作成する。また、この変数をタ

ブーリストに追加し、タブーリストの更新を行う。
2) 変数の削除

2.1　分割表に含まれている変数を一つ選ぶ。

2.2　当該変数を除いた分割表を作成する。

2.3　遺伝的アルゴリズムを使って区分選択を行う。

2.4　分割表に含まれているすべての変数について、2.1から2.3間の処理を行い、タブーリストに含まれておらず、かつ区分選択後のAICが最も小さくなる変数を選ぶ。

2.5　2.4で選ばれた変数を削除したときの区分選択後のAICが、1.0で保存した前のステップのAICより小さければその変数を分割表から削除し、タブーリストに加え、タブーリストの更新を行う。もし、前のステップのAICより大きければ、変数の削除は行わない。

3) 不要な変数の除去

すべての階級において、"all"を意味する値になっている変数があれば、その変数を分割表から削除する。このとき、タブーリストの更新は行わない。

4) あらかじめ定めれた回数、1) から3)の処理を繰り返す。

遺伝的アルゴリズムは以下の手順によった。

1) 原始分割表の階級数に等しい桁数をもつ階乗進数を使って染色体を表現する。

2) 親の集団の大きさ$SIZE$は30とした。

3) 交叉はランクによった。親の集団をAICの小さい順に並び替え、i番目(i=0,⋯,29)の染色体の選択される確率$pr(i)$は次式によった。

$$pr(i) = (SIZE - i)/(SIZE \times (SIZE + 1)/2)$$

また、交叉は$SIZE/2$組について行った。したがって、新たに作られる子の染色体はSIZE個である。

4) 親のうちもっとも適合度の高い(AICが小さい) 染色体を除いた$SIZE - 1$個の染色体と、新たに作られた$SIZE$個の子の染色体に対して突然変異を行う。突然変異は、$SIZE \times 2 - 1$個の染色体それぞれについて、乱数を使って遺伝子座を定め、当該遺伝子座(階乗進数の係数) が取りうる値と置き換えることによって行った。エリートを保存するために、親の中でもっとも適合度の高い染色体に対しては突然変異の操作を行わなかった。

5) 交叉と突然変異を行った後で、$SIZE \times 2$個の染色体の適合度を求める。適合度は、染色体が表す分割表から得られるAICの値とした。このとき、解釈が困難となる選択肢の組み合わせとなるものについては、十分大きな値をAIC値として、抑制することとした。$SIZE \times 2$個の染色値を適合度の高い順に並び替える。そして上位$SIZE$個の染色体を親として残す。

6) 上記3) から5までの操作を定められた回数(100世代)行って、もっとも適合度の高い染色体を選ぶ。

1994年アンケート結果からは、階級数10の分割表が選ばれた。結果を表5.32に示す。

表5.32に含まれている変数は14変数である。得られた階級は10であるので、規則数も10となる。表中の値は選択肢を表す素数の最小公倍数である。"all"を意味する最小公倍数となったものは－としている。規則抽出のために使用した分析データでは、10階級中9階級において調整残差に有意差が見られている。定率法のみの選択を意味する規則が5、定額法を採用することを意味する規則が4個である。表5.32には、10個の規則を検証データに適用した結果も併せて示している。得られる10階級の度数のうち、5階級において調整残差の有意差が見られた。うち4階級に現れた有意差は分析データと同じである。従って、この4個の規則を1994年調査結果から得られ

表 5.32: 1994 年結果から得られた分割表

1994年	2	3	6	7	9	10	11	12	14	15	18	19	23	25	分析データ 定率	分析データ 定額	検証データ 定率	検証データ 定額
1 :	15	—	—	—	5	15	—	—	—	2	—	—	6	—	16(△)	0(▼)	12	4
2 :	2	3	15	5	—	—	6	—	5	—	3	2	—	6	0(▼)	2(△)	1	0
3 :	2	—	—	—	—	—	2	—	—	—	—	—	—	—	30(△)	0(▼)	24(△)	3(▼)
4 :	2	—	—	—	—	—	15	—	—	—	—	—	—	—	28(△)	3(▼)	28(△)	4(▼)
5 :	15	—	15	—	—	—	—	2	15	—	—	—	—	2	12(△)	1(▼)	12	6
6 :	15	—	—	—	—	—	2	—	—	—	—	6	—	—	25(△)	5(▼)	26(△)	6(▼)
7 :	15	—	—	—	—	—	15	—	—	—	—	6	—	—	37(▼)	28(△)	51	23
8 :	15	—	—	—	—	—	2	—	—	—	—	5	—	—	10(▼)	10(△)	12	6
9 :	15	—	—	—	—	—	3	—	—	—	—	5	—	—	4(▼)	13(△)	8(▼)	15(△)
10 :	15	—	—	—	—	—	5	—	—	—	—	5	—	—	14	6	7(▼)	11(△)
														total	175	67	180	78

た規則として抜き出すこととした。さらに 4 個の規則のみを抜き出すことによって、変数の数は 14 から 3 に減少する。なぜなら、その他の変数は、これら 4 規則の前件部に含まれないからである。表 5.33 に変数名、選択肢の内容等をまとめた。

表 5.33: 1994 年結果から得られた 4 規則

		収益性	社長出自	自己資本比率	分析データ 定率	分析データ 定額	検証データ 定率	検証データ 定額
no.	3 :	高い	創業者	—	30(△)	0(▼)	24(△)	3(▼)
no.	4 :	高い	生え抜き、外部人材	—	28(△)	3(▼)	28(△)	4(▼)
no.	6 :	普通、低い	創業者	高い、普通	25(△)	5(▼)	26(△)	6(▼)
no.	9 :	普通、低い	生え抜き	低い	4(▼)	13(△)	8(▼)	15(△)

ところで、最初の規則と 2 番目の規則には、社長出自が「創業者」、「生え抜き、外部人材」が含まれている。収益性はともに高い」なので、これら 2 規則はまとめることができる。この結果、最終的には 3 規則を得ることができる。これらの規則を記述すると次のようになる。

規則　1　収益性が高い企業は定率法のみで償却する割合が高い。

規則　2　収益性が普通または低くても、社長が創業者であり、自己資本比率が高いまたは普通ならば、定率法のみで償却する割合が高い。

規則　3　収益性が普通または低く、社長が生え抜きであり、自己資本比率が低ければ、定額法を採用する割合が高い。

償却方法の選択と収益性が関連していることは第2章でも確認したことである。また、自己資本比率が高ければ借入依存度も低くなるであろう。一方、オーナー社長であるかどうかということと決算政策とは密接に関係していることが報告されている [75]。つまり、定率法選択を意味する規則1および規則2は妥当な規則であると思われる。規則3は規則1および2と逆のものとなっている。従って、定額法採用を意味する規則として妥当であると思われる。

5.3.3 1997年調査から得られた規則

1997年調査からも同様の方法でいくつかの規則を抽出した。1997年アンケートから得られる分割表を表5.34に示した。

表5.34: 1997年結果から得られた分割表

1997年調査								分析データ		検証データ	
		11	12	13	18	19	27	定率	定額	定率	定額
no.	1 :	6	2	−	2	6	−	6(▼)	5(△)	12(△)	2
no.	2 :	6	3	−	2	6	−	34	8	34	9
no.	3 :	5	−	−	2	6	−	10(▼)	10(△)	6(▼)	8(△)
no.	4 :	6	−	−	15	6	6	110(△)	12(▼)	102(△)	24(▼)
no.	5 :	5	−	−	15	6	6	36(▼)	14	35	14
no.	6 :	−	−	−	15	6	5	73(△)	1(▼)	63(△)	11(▼)
no.	7 :	−	−	2	−	5	−	11(▼)	16(△)	17	7
no.	8 :	−	−	3	−	5	−	36(▼)	22(△)	37(▼)	23(△)
							total	316	88	306	98

得られた分割表は6変数を含み、8階級からなっている。分析データから得られた分割表では、8階級中7階級で調整残差において有意差が見られた。従って、分析データを「定率のみ」と「定額採用」にうまく区分していると見ることができる。一方、検証用データに8規則を適用したときに得られる分割表においても、8階級中4階級において分析データと等しい検定結果となっている。そこで、これら4階級から得られる規則を抜き出すこととした。

4つの規則には、4つの変数が含まれている。社長出自、自己資本比率は1994年調査からの規則においても含まれていた変数である。一方、仕入先分布、有形固定資産割合、多角化などの変数は1994年規則には含まれていない。調整残差の検定結果から、2番目の規則と3番目の規則が定率法のみで償却を行っている企業を識別するための規則である。自己資本比率と有形固定資産割合 (資産重要性) に注目してみよう。ともに、自己資本比率は高いまたは低く、資産重要性は普通または低いである。第2章でもみたように、これらの性質は定率法のみで償却を行っている企業に見られたものである。従って、これら2規則は第2章で得られた知見と矛盾するものではないと考えられる。

一方、最初の規則は自己資本比率が高いまたは普通でも、有形固定資産割合が高く、社長が外部人材であるならば、定額法を採用する傾向にあることを示している。第2章でも資産重要性が

表 5.35: 1997 年結果から得られた 4 規則

	社長出自	仕入先分布	有形固定資産割合	自己資本比率	多角化	分析データ 定率	分析データ 定額	検証データ 定率	検証データ 定額
no. 3 :	外部人材	ー	高い	高い、普通	ー	10(▼)	10(△)	6(▼)	8(△)
no. 4 :	創業者、生え抜き	ー	普通、低い	高い、普通	大きい、進めたい	110(△)	12(▼)	102(△)	24(▼)
no. 6 :	ー	ー	普通、低い	高い、普通	必要ない	73(△)	1(▼)	63(△)	11(▼)
no. 8 :	ー	多岐に渡る	ー	低い	ー	36(▼)	22(△)	37(▼)	23(△)

高い企業では定額法採用の傾向が見られた。従って、第 2 章で得られた結果と矛盾するものではない。

最後の規則は、自己資本比率が低く、仕入先が多岐に渡る企業では、定額法を採用する傾向にあることを示している。自己資本比率が低いことは借入依存度が高いことと通じる。従って、第 2 章で得られた結論と矛盾しない。しかし、仕入先分布という直ちに解釈ができない変数も残っている。これは 2 番目の規則と 3 番目の規則に含まれる多角化も同様である。これらの変数の必要性、解釈は今後の検討課題としたい。

5.3.4 2000 年調査から得られた規則

最後に 2000 年調査結果から得られた分割表を 5.36 に示した。含まれる変数は 5 個であり、階級数は 7 である。分析データでは 7 階級中 4 階級に有意差が表れている。

1994 年、1997 年調査結果と同様に検証用データに対して 7 規則を適用して得られる分割表も併せて表 5.36 に示している。しかし、調整残差において見られる有意差は分析データのものとまったく異なる。これは得られた規則の信頼性が極めて低いこと示している。本節の冒頭でも述べたように、平成 10 年に税法改正が行われ、建物の減価償却方法は定額法のみとなった。このことが影響しているとも考えられる。一方、不況の長期化などによって、償却方法選択の誘因が変化している可能性も否定できない。これらの検討も今後の課題としたい。

5.4 結論

本章では、減価償却方法を対象として、その選択状況を説明する少数の規則を得る方法を検討してきた。特に、アンケートから得られる分割表をもとに、規則抽出の方法を検討した。5.1 では、減価償却方法の性質の違いから導かれるいくつかの仮説に対して統計的な検討を行った。そして、収益性、資産重要性、借入依存度が償却方法の選択と関連していることをみた。これら 3 要因か

表 5.36: 2000 年結果から得られた分割表

2000年調査	8	17	19	20	28	分析データ 定率	分析データ 定額	検証データ 定率	検証データ 定額
no. 1 :	2	15	−	−	105 :	12(△)	0(▼)	8(△)	2
no. 2 :	15	5	2	−	3 :	4(△)	0(▼)	4(▼)	5
no. 3 :	15	5	15	−	3 :	1(▼)	10(△)	2	1
no. 4 :	−	6	−	2	− :	16	7	18(△)	2(▼)
no. 5 :	−	6	−	15	− :	8(▼)	10(△)	15	7
no. 6 :	−	5	−	2	70 :	14(△)	0(▼)	8(△)	2
no. 7 :	−	5	−	15	70 :	23	11	21(▼)	20(△)
					total :	77	38	76	39

らなる分割表を作ることによって、償却方法選択状況を説明するための 2 つの規則を得た。

しかし、分割表に含まれるすべての階級が規則として有効なわけではない。実際、5.1 の分析でも、8 階級中 3 階級が意味を持つ階級と考えられた。このため、少数の階級から得られた規則がどの程度有効であるか確認する必要が生じる。収益性、資産重要性、借入依存度はすべて財務比率を使って測定することができる。このことから、得られた規則を推論規則とするファジィ推論を行うこととした。「収益性が高い」などのファジィ集合との一致度はニューラルネットワークを組んで求めた。そして簡略化ファジィ推論の結果、83 %は正しく識別できた。

一方、分割表から直ちに少数の規則を得ることが困難な場合も多い。特に変数の数が増加すると作成しうる分割表の数も膨大なものとなる。本研究でも、1994,1997,2000 年に行った実態調査では、減価償却方法選択状況の説明変数として 26 ないし 29 項目を設定していた。すべての変数を含んだ分割表は階級に含まれる度数が小さくなるという欠点があり、可能なすべての分割表を検討することは実際的には不可能である。このため本研究では遺伝的アルゴリズムを使ってモデルの探索を行うこととした。モデルの評価基準としては AIC を使い、さまざまな分割表を階乗進数を使って表現し、少数の規則を得ることができた。多くの変数を使っているために、仕入れ先の分布や多角化への取り組みなど、規則に含まれているすべての変数について妥当性を確認することはできなかった。しかし、収益性や自己資本比率など 5.1 で得られた知見と合致する変数も含まれていることを確認することもできた。規則に含まれる変数すべてについて、定量化することができないので、5.2 のような分析を行うことはできなかったが、会計的知見とも矛盾しない規則が得られたことから、多変数かつ大量のデータがある場合にも、分割表から規則を抽出するという提案方法は有効であると考えてもいいであろう。

第6章　まとめ

遺伝的アルゴリズムを利用しようとする場合、どのような染色体表現や交叉の方法を取るかによって、探索能力に違いが生じる。本研究では染色体表現として階乗進数表現を採用し、さまざまな問題に対して適用を試みた。階乗進数表現の有効性を検証することが本研究の目的であった。

第2章では、階乗進数と順列を対応させた。階乗進数と順列の対応規則から、階乗進数と順列は一対一に対応する。また、一般的な性質として、交叉によって致死遺伝子が生じることを抑制することができる。このため、多点交叉にも適用できることがわかった。一方、本研究で用いた順列との対応規則においては、いわば非対称性を指摘することができる。基準順列の後に配置された要素ほど、取りうる係数の範囲が狭くなる。このことを利用して、大域的探索から局所的探索への移行を実現することが可能となった。実際、フローショップスケジューリングにおいて、作業時間の合計が大きいものを基準順列の後に配置することによって、他の方法に比べて良い探索結果を得ることができた。これは、作業時間の合計が大きいものから順に相対的な順序を定めていくという深さを優先するNEH法の考えかたを取り入れることができたためと考えられる。

また、本研究で採用した順列との対応規則を使って、ある種の順序制約を容易に表現できることがわかった。特に、本研究で直列的な順序制約と呼んだものに対しては、制約条件を満たす順列と階乗進数を一対一に対応させることができた。この結果、解の修正や罰金の導入などの必要がなくなり、短い探索時間でより良い解を探索することができた。さらに複雑な順序制約を持つものの例として、因果モデリングを取り上げた。ここでも、ある種の順序制約を満たす階乗進数を生成し得ることがわかったが、遺伝的アルゴリズムへの実装は今後の課題として残されている。

第3章以降は、階乗進数を使った集合分割の表現について検討した。特に、解釈を目的とした分割表からの規則抽出を中心に分析した。アンケートでは、被説明変数を識別するために、多くの説明変数が設定される場合が多い。これらの説明変数の中には、被説明変数との関連性が明らかでないものも含まれているであろう。そこで本研究では、多変数大規模な調査結果からの規則抽出の例として、経営環境に関するアンケート調査結果を取り上げた。このアンケート調査では、経営環境に関する多数の質問と減価償却方法の選択状況に関する質問を設定していた。減価償却方法選択のように経営者の主観が反映するものでは、説明変数を使ってそのすべての行動を規定することは不可能であろう。得られた規則も部分的にしか適用できないものであった。しかし、得られた規則は会計的知見とも合致するものであったことから、本研究で採用した規則抽出の方法は妥当であると判断しても良いと思われる。

本研究では、原始分割表に含まれる階級を要約する場合に階乗進数表現を利用した。ここでは、階乗進数表現の利点について考えてみることにしよう。

遺伝的アルゴリズムなどの発見的方法を使って要素をいくつかのグループに分けようとする場合には、非階層的なクラスタリング手法の方が扱いやすいであろう。例えばn個の要素からなる集合を、m群に分けようとするときには, それぞれの要素に1からmまでの数値を与えた染色体を作り、あらかじめ定めらた基準にもとづいて探索を行えば良い。しかしながら、この方法では、群の数が固定されているという欠点がある。もし、考え得る数の群を容認しようとすれば、$m = n$として、探索を開始しなければならない。

また、仮に群の数を m 個にしたかったとしても、探索の過程で、特定の群番号が染色体から除かれてしまうということもあり得るので，群の数が m 未満になる可能性が高い。もし、群の数を m に保とうとすれば、これらは致死遺伝子として扱わざるを得ない。さらに、この方法の欠点は、それぞれの要素に割り当てられた郡番号が意味を持つということである。ある染色体では、要素1と要素2にそれぞれ群番号1が割り当てられていたとしよう。また、別の染色体では、同じく要素1と要素2に群番号2が割り当てられていたとしよう。これら2つの染色体はともに、要素1と要素2が同じ群であることを表しているにもかかわらず、群番号が異なるために、交叉によって結合が切れる可能性がある。つまり、この方法では，両親に共通にみられる形質を適切に子に伝えることができない。このような事態をさけるためと思われるが、加藤、小沢 [27] は染色体表現の統一や優性遺伝多点交叉と呼ばれる複雑な処理を行っている。

本研究で利用した階乗進数表現、さらには、かつてあやめの識別問題に対して利用した順列表現 [92] では、群の数をあらかじめ定めておく必要もなく、両親に共通にもたれている要素間の結びつきの保存も可能である。また、群の数を与えられた数に保つことも容易である。つまり、群番号を使って染色体を表現したときの欠点を、階乗進数表現では回避することができる。

さらに、階乗進数表現ではいくつかの特徴がある。そのひとつは突然変異の操作である。突然変異では、ある要素の値を置き換えた。階乗進数の解釈から、このことは、当該要素を根とする木をまとめて他の群につけかえることを意味している。これは、遺伝的プログラミングにおける交叉の一種と考えることができる。

また、階乗進数表現の非対称性から、基準順列における要素の並べ方を工夫することによって、探索能力を向上させることができる。われわれが扱っている分割表からの規則抽出では、群を識別する能力があると思われる階級が重要な階級である。他方、全体の分布と差が認められないものは、群の識別にはあまり役立たないだろう。したがって、本研究と異なって階級数が多い場合には、群に関して平均的な度数となっている階級は、基準順列の最初に，逆に、分布が平均と異なるものを基準順列の最後に配置することによって、探索能力を改善することが期待できる。また、あらかじめ要素を複数の群に分けておいて、それぞれの群を他の群と違う特色を持ついくつかのグループに要約することもできる。これらの操作は、集合分割を階乗進数を使って表現した場合には容易に実現できるものである。

判別力の点から、提案方法を評価してみよう。従来、判別のためには判別関数を求めることが多かった。また、ニューラルネットワークを直接用いることもできる。しかし、高い判別力を獲得したとしても判別関数やニューラルネットワークの重みを解釈することが困難なために、その結果から知識 (規則) を得ることは困難であった。一方、ファジィ推論では規則そのものを使うために、満足すべき判別力が得られたならば、直ちに有効な知識を獲得することができる。したがって、本研究では、推論のための規則をまず抽出するという方法を取った。このような方法を取った場合、結果としての判別力は他の方法と比べてどうであろうか。5.2 の減価償却方法選択に関する判別を対象に比較を行ってみよう。

表 6.1(学習用 100 社) には、収益性、資産重要性、借入依存度のすべてにおいて「高い」または「低い」と回答した企業のうち、財務比率が得られた 100 社を対象に、ニューラルネットワーク、線形判別関数、提案方法の判別力を比較したものである。5.2 でも示したように、2 つの規則を使ってファジィ推論を行った場合の判別率は 83 %であった。ここで、ニューラルネットワークによる判別とは、基礎となった財務比率 8 個を使ってニューラルネットワークを組み、償却方法選択状況を教師信号として学習させ、その出力値によって最も誤判別が小さくなるように分類したものである。線形判別関数も同様に 8 個の財務比率を直接使った。このとき償却方法選択状況はダミー変数として与えている。やはり予測値をもとに並び替え、最も誤判別が小さくなる点をカットオ

フポイントとして選んだ。

表 6.1: 判別率の比較

方法	学習用 (100 社) 「高い」または「低い」	検証用 (126 社) 「普通」	計 (226 社)
ニューラルネットワーク	87 %	72 %	79 %
線形判別関数	85 %	71 %	77 %
ファジィ推論 (提案方法)	83 %	76 %	79 %

表 6.1 からわかるように、学習用 100 社に対する判別率は、ニューラルネットワークによった場合が最も高く 87 %であった。線形判別関数を使った場合は 85 %であり、提案方法による判別率はすでに述べたように 83 %である。合計 100 社であることを考えると、2 ないし 4 社誤判別が高いということになる。

一方、いずれの場合にも得られた判別モデルを使って、いずれかの項目に対して「普通」と回答した企業の判別を行うことができる。いずれかの項目に「普通」と回答した企業は合計 126 社であった。これらの判別結果も表 6.1 に示している。ここでは、ファジィ推論を使った提案方法が最もよい結果となっている。ニューラルネットワークを使った場合が 72 %、線形判別関数を使った場合が 71 %であるのに対し、ファジィ推論を使った提案方法では 79 %の企業を正しく識別している。この 126 社はいわば検証用データと考えることもできるので、このデータに対して高い判別率を得ることができたことはモデルの有効性を確認するうえで重要であろう。この結果、全体 226 社に対しても提案方法の判別率は 79 %となり、他の方法に比べて遜色はない。このことから、本研究で検討してきた分割表からの規則抽出は有効な方法であると考えていいと思われる。

最後に今後の課題について述べてみたい。階乗進数と順列の対応では、数としての階乗進数の可能性について検討してみたい。順列表現と異なり、階乗進数では四則をはじめさまざまな演算を定義できる。このような操作が順列にどのような効果をもたらすかという点について調べてみたい。たとえば補数をとるという操作によって、順列は逆順となることが予想されている。このような操作を探索に取り入れれることで、探索能力に差が生じるかどうか興味があるところである。また、階乗進数による集合分割においても、グループテクノロジー (GT) や画像処理における領域の認識などへの適用が考えられる。このような応用も今後の課題として残されている。

謝辞

減価償却方法選択などの決算行動に関する研究は、1980年代以降、慶應義塾大学理工学部管理工学科高橋吉之助研究室において精力的に続けられてきた。本研究もその一環として、実証的な分析を通して因果律の抽出を試みたものである。当初は多変量解析などの統計的方法を適用していた。しかし、1990年代になって、アンケート結果から推論規則を獲得することに重点を置くようになった。この間、一貫して故高橋吉之助先生からご指導をいただいた。アンケート結果という不確かさが残る対象を選ぶときにも、数々のご助言や励ましをいただいた。アンケート結果を使った最初の学会発表が、高橋先生との最後の共同研究となった。改めてご冥福をお祈りするとともに、心より謝意を表したい。

規則抽出における遺伝的アルゴリズムの適用にあたって、従来あまり用いられてこなかった階乗進数を利用している。この階乗進数の利用・適用については、慶応義塾大学理工学部管理工学科西野寿一先生、高橋正子先生から多くのご助言をいただいた。研究会を通したご指導がなければ本論文を完成させることは不可能であったろう。ここに記して謝意を表したい。また、本論文をまとめるにあたっては、予備審査および本審査の過程において、慶應義塾大学理工学部中村善太郎先生、篠崎信雄先生、森雅夫先生、萩原将文先生、そして慶應義塾大学理工学部管理工学科の諸先生から貴重なご助言をいただいた。ここに記して謝意を表したい。

また、学生時代から現在に至るまで高橋研究室の研究会は続いている。この研究会での議論が本研究を進める上での大きな基礎となっている。研究会のメンバーである佃純誠先生 (武蔵工業大学)、松村義之先生、高橋正子先生 (慶応義塾大学)、黒川行治先生 (慶應義塾大学)、江島夏実先生 (コンピュータ教育工学研究所) の諸先生からは有益な数々のご助言をいただいた。ここに記して謝意を表したい。

付 録A　1991年および1992年のアンケート調査項目

減価償却方法の選択状況および財務環境についておたずねします。以下の質問に対して、選択肢の中からもっとも該当すると思われるものを選び、その選択肢番号を返信用はがきの回答欄にご記入ください。この調査は、減価償却方法の採用状況と財務環境の関連性を統計的に分析することを目的としています。お聞きしております内容はいずれも主観的なものですので、経営者の方々の主観的な、または経験にもとづくご感想をお知らせいただければ幸いです。

(a) 金額的にみて、有形固定資産の減価償却費計上は次のいずれに該当しますか。なお、生産高比例法など他の償却方法による部分は除いてお考えください。

1　すべて定率法によっている

2　定率法による部分の方が多い。

3　定率法と定額法による部分が半々である。

4　定額法による部分が多い。

5　すべて定額法によっている。

(b) 他の製造業に属する企業に比べて、総資産に占める有形固定資産の割合は高い方だと思われますか。それとも低い方だと思われますか。

1　高い　2　やや高い　3　普通　4　やや低い　5　低い

(c) 他の製造業に属する企業に比べて、資産除却時の除却損、売却損は多い方だと思われますか。それとも少ない方だと思われますか。

1　多い　2　やや多い　3　普通　4　やや少ない　5　少ない

(d) 他の製造業に属する企業に比べて、有形固定資産に対する資本的支出は多い方だと思われますか。それとも少ない方だと思われますか。

1　多い　2　やや多い　3　普通　4　やや少ない　5　少ない

(e) 他の製造業に属する企業に比べて、設備投資資金を借り入れなどの外部資本に依存する割合は高い方だと思われますか。それとも低い方だと思われますか。

1　高い　2　やや高い　3　普通　4　やや低い　5　低い

(f) ここ数年間の傾向を大まかに見た時に、他の製造業に属する企業に比べて、総資本経常利益率や売上高経常利益率などの数値は高い方だと思われますか。それとも低い方だと思われますか。

1　高い　2　やや高い　3　普通　4　やや低い　5　低い

(g) 他の製造業に属する企業に比べて、総資本経常利益率や売上高経常利益率などの数値の時系列的な変動は大きい方だと思われますか。それとも小さい方だと思われますか。

1　大きい　2　やや大きい　3　普通　4　やや小さい　5　小さい

(h) 減価償却費の計算における計算機の利用状況についておたずねします。償却費の計算は次のいずれに該当しますか。

1　計算機で機械的に計算する部分が多い

2　帳簿上で各資産の償却費を計算する割合が高い

(i) 機械装置などの設備の償却についておたずねします。設備の償却は個別償却による部分が多

いでしょうか。それとも総合償却による部分が多いでしょうか。次の3つの中でもっとも代表的なものと思われるものをお選びください。

1　おおむね個別の資産を対象とした個別償却計算

2　種類や耐用年数の似通っている資産を一括にした償却計算

3　異なる種類の資産をもまとめた総合償却

(j) 主要生産設備の耐用年数についておたずねします。技術革新、陳腐化や不適応などを考慮した経済的耐用年数は、法定耐用年数に比べてどの程度の割合だと思われますか。

1　半分以下

2　半分ぐらい

3　7～8割程度

4　法定耐用年数と同じぐらい

付 録B　経営環境に関する調査

本節では 1994 年以降行った経営環境に関する調査の概要を示す。

質問（減価償却方法）金額的にみて、有形固定資産の減価償却方法は次のいずれに該当しますか。なお、定率法、定額法以外の償却方法による部分は除いてお考えください。

表 B.1: 減価償却方法の選択状況

減価償却方法	1994 年調査	1997 年調査	2000 年調査
1　定率法のみ	470(72.1 %)	640(77.0 %)	155(23.4 %)
2　一部定額法	121(18.6 %)	136(16.4 %)	430(65.0 %)
3　主に定額法	61(9.4 %)	55(6.7 %)	77(11.6 %)
合計	652(100.0 %)	831(100.0 %)	662(100.0 %)

質問 (1) 売上高のうち，輸出部分の占める割合は，他社に比べて大きい方だと思われますか。それとも小さい方だと思われますか。

表 B.2: 質問 (1) 売上に占める輸出の割合

(1) 輸出依存度	1994 年調査	1997 年調査	2000 年調査
1　大きい	37(6.6 %)	31(4.3 %)	38(5.9 %)
2　やや大きい	40(7.2 %)	61(8.4 %)	31(4.8 %)
3　普通	97(17.4 %)	93(12.8 %)	87(13.6 %)
4　やや小さい	65(11.6 %)	102(14.0 %)	75(11.7 %)
5　小さい	320(57.2 %)	442(60.6 %)	410(64.0 %)
合計	559(100.0 %)	729(100.0 %)	641(100.0 %)

質問 (2) ここ数年間の傾向を大まかに見た時に，総資本経常利益率や売上高経常利益率などの数値は，他社に比べて高い方だと思われますか。低い方だと思われますか。

表 B.3: 質問 (2) 収益性

(2) 収益性	1994 年調査	1997 年調査	2000 年調査
1　高い	62(9.5 %)	90(10.8 %)	68(10.3 %)
2　やや高い	101(15.5 %)	160(19.2 %)	118(17.8 %)
3　普通	148(22.7 %)	199(23.9 %)	160(24.1 %)
4　やや低い	184(28.3 %)	222(26.7 %)	180(27.1 %)
5　低い	156(24.0 %)	161(19.4 %)	137(20.7 %)
合計	651(100.0 %)	832(100.0 %)	663(100.0 %)

質問 (3) 他社に比べて，総資本利益率や売上高経常利益率などの数値の時系列的な変動は大きい方だと思われますか。

表 B.4: 質問 (3) 利益変動

(3) 利益変動	1994 年調査	1997 年調査	2000 年調査
1　大きい	57(8.8 %)	53(6.4 %)	33(5.0 %)
2　やや大きい	149(22.9 %)	201(24.1 %)	163(24.6 %)
3　普通	270(41.5 %)	350(42.0 %)	273(41.2 %)
4　やや小さい	98(15.1 %)	131(15.7 %)	126(19.0 %)
5　小さい	77(11.8 %)	98(11.8 %)	68(10.3 %)
合計	651(100.0 %)	833(100.0 %)	663(100.0 %)

質問 (4) 棚卸資産の回転期間は他社に比べて長い方だと思われますか。

表 B.5: 質問 (4) 棚卸資産の回転期間

(4) 棚卸資産回転期間	1994 年調査	1997 年調査	2000 年調査
1　長い	46(7.3 %)	45(5.5 %)	41(6.2 %)
2　やや長い	142(22.4 %)	187(22.9 %)	176(26.6 %)
3　普通	289(45.6 %)	359(44.0 %)	271(41.0 %)
4　やや短い	96(15.1 %)	130(16.0 %)	86(13.0 %)
5　短い	61(9.6 %)	94(11.5 %)	87(13.2 %)
合計	634(100.0 %)	815(100.0 %)	661(100.0 %)

質問 (5) 材料や仕入商品の価格変動は大きい方だと思われますか。

表 B.6: 質問 (5) 材料などの価格変動

(5) 仕入価格変動	1994 年調査	1997 年調査	2000 年調査
1　大きい	22(3.5 %)	33(4.0 %)	17(2.6 %)
2　やや大きい	101(16.1 %)	124(15.0 %)	116(17.5 %)
3　普通	351(55.9 %)	421(51.1 %)	316(47.7 %)
4　やや小さい	99(15.8 %)	140(17.0 %)	109(16.4 %)
5　小さい	55(8.8 %)	106(12.9 %)	105(15.8 %)
合計	628(100.0 %)	824(100.0 %)	663(100.0 %)

質問 (6) 材料や仕入商品を輸入に依存する割合は高い方だと思われますか。

表 B.7: 質問 (6) 材料などの輸入依存

(6) 輸入依存度	1994 年調査	1997 年調査	2000 年調査
1　高い	41(6.9 %)	48(6.2 %)	52(8.0 %)
2　やや高い	48(8.1 %)	76(9.8 %)	56(8.6 %)
3　普通	116(19.6 %)	128(16.6 %)	89(13.7 %)
4　やや低い	85(14.4 %)	125(16.2 %)	101(15.6 %)
5　低い	302(51.0 %)	396(51.2 %)	351(54.1 %)
合計	592(100.0 %)	773(100.0 %)	649(100.0 %)

質問 (7)　現金売りと掛け売りの割合は次のいずれに該当しますか。

表 B.8: 質問 (7) 現金売りと掛売の割合

(7) 販売方法	1994 年調査	1997 年調査	2000 年調査
1　現金売りが多い	134(20.7 %)	167(20.3 %)	139(21.0 %)
2　半々である	39(6.0 %)	56(6.8 %)	34(5.1 %)
3　掛売りの部分が多い	475(73.3 %)	600(72.9 %)	489(73.9 %)
合計	648(100.0 %)	823(100.0 %)	662(100.0 %)

質問 (8)　売掛金の回収期間は他社に比べて長いと思われますか。

表 B.9: 質問 (8) 売掛金の回収期間

(8) 売掛金の回収期間	1994 年調査	1997 年調査	2000 年調査
1　長い	30(4.6 %)	32(3.9 %)	29(4.4 %)
2　やや長い	118(18.2 %)	150(18.1 %)	125(18.9 %)
3　普通	368(56.8 %)	473(57.0 %)	347(52.5 %)
4　やや短い	73(11.3 %)	79(9.5 %)	81(12.3 %)
5　短い	59(9.1 %)	96(11.6 %)	79(12.0 %)
合計	648(100.0 %)	830(100.0 %)	661(100.0 %)

質問 (9)　設備投資などの資金を借入に依存する度合いは，他社に比べて大きい方だと思われますか。

表 B.10: 質問 (9) 設備投資資金の借入依存度

(9) 借入依存度	1994 年調査	1997 年調査	2000 年調査
1　大きい	57(9.0 %)	58(7.0 %)	60(9.1 %)
2　やや大きい	102(16.2 %)	123(14.9 %)	105(15.9 %)
3　普通	162(25.7 %)	191(23.1 %)	153(23.1 %)
4　やや小さい	98(15.5 %)	129(15.6 %)	92(13.9 %)
5　小さい	212(33.6 %)	325(39.3 %)	252(38.1 %)
合計	631(100.0 %)	826(100.0 %)	662(100.0 %)

質問 (10)　いわゆる主力銀行など特定の銀行からの借入が，借入総額に対する割合は大きい方だと思われますか。

表 B.11: 質問 (10) 主力銀行の比率

(10) 特定銀行からの借入割合	1994 年調査	1997 年調査	2000 年調査
1　大きい	103(16.7 %)	136(16.8 %)	103(15.7 %)
2　やや大きい	134(21.8 %)	162(20.0 %)	168(25.6 %)
3　普通	216(35.1 %)	251(31.0 %)	184(28.1 %)
4　やや小さい	57(9.3 %)	78(9.6 %)	51(7.8 %)
5　小さい	105(17.1 %)	183(22.6 %)	149(22.7 %)
合計	615(100.0 %)	810(100.0 %)	655(100.0 %)

質問 (11)　貴社の社長の出自としてもっとも近いと思われるものをひとつお選びください。

表 B.12: 質問 (11) 社長の出自

(11) 社長の出自	1994 年調査	1997 年調査	2000 年調査
1　創業者	219(33.7 %)	350(42.1 %)	290(43.7 %)
2　社内	224(34.5 %)	256(30.8 %)	213(32.1 %)
3　外部人材	206(31.7 %)	226(27.2 %)	161(24.2 %)
合計	649(100.0 %)	832(100.0 %)	664(100.0 %)

質問 (12)　貴社の製商品の販売先やサービスの提供先は次のいずれに該当しますか。

表 B.13: 質問 (12) 製商品などの販売先

(12) 販売先	1994 年調査	1997 年調査	2000 年調査
1　主に少数の得意先	177(27.2 %)	217(26.1 %)	194(29.3 %)
2　販売先は多岐	473(72.8 %)	613(73.9 %)	468(70.7 %)
合計	650(100.0 %)	830(100.0 %)	662(100.0 %)

質問 (13)　材料，商品など貴社の仕入先は次のいずれに該当しますか。

表 B.14: 質問 (13) 材料などの仕入先

(13) 材料等の仕入先	1994 年調査	1997 年調査	2000 年調査
1　少数の仕入先	185(28.9 %)	251(30.4 %)	236(35.8 %)
2　数量金額ともに多岐	455(71.1 %)	576(69.6 %)	424(64.2 %)
合計	640(100.0 %)	827(100.0 %)	660(100.0 %)

質問 (14)　役員賞与の大きさが，利益額と連動する度合いは大きい方だと思われますか。

表 B.15: 質問 (14) 役員賞与の利益連動

(14) 役員賞与の利益連動	1994 年調査	1997 年調査	2000 年調査
1　大きい	37(5.9 %)	42(5.2 %)	44(6.7 %)
2　やや大きい	36(5.7 %)	57(7.0 %)	59(8.9 %)
3　普通	310(49.1 %)	369(45.3 %)	284(43.0 %)
4　やや小さい	94(14.9 %)	138(16.9 %)	110(16.6 %)
5　小さい	155(24.5 %)	209(25.6 %)	164(24.8 %)
合計	632(100.0 %)	815(100.0 %)	661(100.0 %)

質問 (15)　大規模な設備投資の効果が，当初予想された効果より大きく異なることがありますか。

表 B.16: 質問 (15) 設備投資の効果

(15) 設備投資の効果	1994 年調査	1997 年調査	2000 年調査
1　おおむね予想通り	457(74.7 %)	596(74.4 %)	440(67.9 %)
2　大きくずれることもある	155(25.3 %)	205(25.6 %)	208(32.1 %)
合計	612(100.0 %)	801(100.0 %)	648(100.0 %)

質問 (16)　大規模な設備投資の効果が現れるまでの期間は，他社に比べて長い方だと思われますか。

表 B.17: 質問 (16) 投資効果が現れるまでの期間

(16) 投資効果が現れるまで	1994 年調査	1997 年調査	2000 年調査
1　長い	29(4.7 %)	23(2.9 %)	21(3.2 %)
2　やや長い	82(13.4 %)	119(14.9 %)	100(15.4 %)
3　普通	404(65.8 %)	487(61.1 %)	400(61.4 %)
4　やや短い	59(9.6 %)	98(12.3 %)	90(13.8 %)
5　短い	40(6.5 %)	70(8.8 %)	40(6.1 %)
合計	614(100.0 %)	797(100.0 %)	651(100.0 %)

質問 (17)　ここ数年の傾向を大まかに見た時に，利益水準はどのような位置にあると思われていますか。目標とする利益水準に比べて高いと思われますか，低いと思われますか。

表 B.18: 質問 (17) 最近の利益水準

(17) 利益水準	1994 年調査	1997 年調査	2000 年調査
1　高い	14(2.1 %)	22(2.6 %)	23(3.5 %)
2　やや高い	36(5.5 %)	83(10.0 %)	54(8.1 %)
3　普通	125(19.2 %)	184(22.1 %)	147(22.1 %)
4　やや低い	236(36.2 %)	326(39.1 %)	235(35.4 %)
5　低い	241(37.0 %)	218(26.2 %)	205(30.9 %)
合計	652(100.0 %)	833(100.0 %)	664(100.0 %)

質問 (18)　他社に比べて，総資産に占める有形固定資産の割合は高い方だと思われますか。

表 B.19: 質問 (18) 有形固定資産の割合

(18) 有形固定資産の割合	1994 年調査	1997 年調査	2000 年調査
1　高い	32(4.9 %)	44(5.3 %)	68(10.2 %)
2　やや高い	132(20.3 %)	181(21.8 %)	125(18.8 %)
3　普通	263(40.4 %)	342(41.1 %)	263(39.6 %)
4　やや低い	150(23.0 %)	168(20.2 %)	125(18.8 %)
5　低い	74(11.4 %)	97(11.7 %)	83(12.5 %)
合計	651(100.0 %)	832(100.0 %)	664(100.0 %)

質問 (19)　他社に比べて，自己資本比率は高い方だと思われますか。

表 B.20: 質問 (19) 自己資本比率

(19) 自己資本比率	1994 年調査	1997 年調査	2000 年調査
1　高い	157(24.1 %)	205(24.6 %)	175(26.4 %)
2　やや高い	170(26.1 %)	252(30.3 %)	179(27.0 %)
3　普通	137(21.0 %)	207(24.8 %)	147(22.1 %)
4　やや低い	112(17.2 %)	108(13.0 %)	98(14.8 %)
5　低い	76(11.7 %)	61(7.3 %)	65(9.8 %)
合計	652(100.0 %)	833(100.0 %)	664(100.0 %)

質問 (20)　貴社の主要製商品やサービスの付加価値率は，他社に比べて高い方だと思われますか。

表 B.21: 質問 (20) 付加価値率

(20) 付加価値率	1994 年調査	1997 年調査	2000 年調査
1　高い	42(6.5 %)	54(6.5 %)	49(7.4 %)
2　やや高い	142(22.0 %)	219(26.5 %)	175(26.4 %)
3　普通	270(41.9 %)	351(42.4 %)	264(39.8 %)
4　やや低い	127(19.7 %)	163(19.7 %)	145(21.8 %)
5　低い	64(9.9 %)	40(4.8 %)	31(4.7 %)
合計	645(100.0 %)	827(100.0 %)	664(100.0 %)

質問 (21)　他社に比べて，売上高に対する研究開発投資の割合は高い方だと思われますか。

表 B.22: 質問 (21) 売上に対する研究開発費の割合

(21) 研究開発投資	1994 年調査	1997 年調査	2000 年調査
1　高い	15(2.5 %)	29(3.7 %)	22(3.3 %)
2　やや高い	74(12.2 %)	84(10.6 %)	73(11.1 %)
3　普通	188(31.0 %)	245(30.9 %)	196(29.8 %)
4　やや低い	162(26.7 %)	198(25.0 %)	164(25.0 %)
5　低い	168(27.7 %)	236(29.8 %)	202(30.7 %)
合計	607(100.0 %)	792(100.0 %)	657(100.0 %)

質問 (22)　他社に比べて，売上高に対する販売促進費の割合は高い方だと思われますか。

表 B.23: 質問 (22) 売上に対する販売促進費の割合

(22) 販売促進費の割合	1994 年調査	1997 年調査	2000 年調査
1　高い	9(1.4 %)	21(2.5 %)	15(2.3 %)
2　やや高い	57(9.0 %)	98(11.9 %)	88(13.3 %)
3　普通	283(44.6 %)	348(42.1 %)	256(38.6 %)
4　やや低い	157(24.7 %)	214(25.9 %)	162(24.4 %)
5　低い	129(20.3 %)	146(17.7 %)	142(21.4 %)
合計	635(100.0 %)	827(100.0 %)	663(100.0 %)

質問 (23)　他社に比べて，資金的な余裕は大きい方だと思われますか。

表 B.24: 質問 (23) 資金的な余裕

(23) 資金の余裕	1994 年調査	1997 年調査	2000 年調査
1　大きい	100(15.4 %)	167(20.1 %)	127(19.1 %)
2　やや大きい	146(22.4 %)	215(25.8 %)	193(29.1 %)
3　普通	266(40.9 %)	298(35.8 %)	219(33.0 %)
4　やや小さい	76(11.7 %)	93(11.2 %)	71(10.7 %)
5　小さい	63(9.7 %)	59(7.1 %)	54(8.1 %)
合計	651(100.0 %)	832(100.0 %)	664(100.0 %)

質問 (24)　他社に比べて，売上高に対する損益分岐点の位置（損益分岐点比率）は高い方だと思われますか。

表 B.25: 質問 (24) 損益分岐点比率

(24) 損益分岐点比率	1994 年調査	1997 年調査	2000 年調査
1　高い	98(15.1 %)	90(10.8 %)	59(8.9 %)
2　やや高い	183(28.2 %)	213(25.6 %)	198(29.8 %)
3　普通	228(35.1 %)	354(42.5 %)	265(39.9 %)
4　やや低い	99(15.3 %)	135(16.2 %)	117(17.6 %)
5　低い	41(6.3 %)	40(4.8 %)	25(3.8 %)
合計	649(100.0 %)	832(100.0 %)	664(100.0 %)

質問 (25)　貴社の主要製商品・事業分野の市場は，次のいずれにもっとも該当すると思われますか。

表 B.26: 質問 (25) 主要製商品の市場

(25) 市場の状況	1994 年調査	1997 年調査	2000 年調査
1　成熟期	523(82.2 %)	635(77.1 %)	487(73.6 %)
2　市場拡大期	41(6.4 %)	168(20.4 %)	154(23.3 %)
3　萌芽的な時期	72(11.3 %)	21(2.5 %)	21(3.2 %)
合計	636(100.0 %)	824(100.0 %)	662(100.0 %)

質問 (26)　貴社の主要製商品・事業分野の市場集中度（概ね上位 8 社程度の市場占有率合計の値）は高い方だと思われますか。低い方だと思われますか。

表 B.27: 質問 (26) 主要製商品の市場集中度

(26) 市場集中度	1994 年調査	1997 年調査	2000 年調査
1　高い	207(32.8 %)	265(32.4 %)	192(29.0 %)
2　やや高い	186(29.5 %)	246(30.1 %)	209(31.5 %)
3　普通	98(15.5 %)	169(20.7 %)	136(20.5 %)
4　やや低い	52(8.2 %)	60(7.3 %)	65(9.8 %)
5　低い	88(13.9 %)	78(9.5 %)	61(9.2 %)
合計	631(100.0 %)	818(100.0 %)	663(100.0 %)

質問 (27)　貴社の多角化の度合いは次のいずれに該当すると思われますか。

表 B.28: 質問 (27) 多角化の度合い

(27)　多角化の度合い	1994 年調査	1997 年調査	2000 年調査
1　大きい	-	102(12.5 %)	79(11.9 %)
2　今後進めたい	-	481(58.8 %)	337(50.8 %)
3　多角化の必要はない	-	235(28.7 %)	248(37.3 %)
合計	-	818(100.0 %)	664(100.0 %)

質問 (28)　貴社の社長の専門分野としてもっとも近いと思われるものをひとつお選びください。

表 B.29: 質問 (28) 社長の専門分野

(28) 社長の専門分野	1994 年調査	1997 年調査	2000 年調査
1　営業	-	424(51.3 %)	340(51.5 %)
2　技術	-	241(29.2 %)	198(30.0 %)
3　経理	-	83(10.0 %)	62(9.4 %)
4　総務	-	78(9.4 %)	60(9.1 %)
合計	-	826(100.0 %)	660(100.0 %)

質問 (29)　主要生産設備の耐用年数についておたずねします。技術革新，陳腐化や不適応などを考慮した経済的耐用年数は，法定耐用年数にくらべてどの程度の割合だと思われますか。

表 B.30: 質問 (29) 経済的耐用年数

(29) 経済的耐用年数	1994 年調査	1997 年調査	2000 年調査
1　法定耐用年数の半分以下	-	17(2.1 %)	28(4.3 %)
2　法定耐用年数の半分ぐらい	-	126(15.9 %)	115(17.7 %)
3　法定耐用年数の 7～8 割程度	-	391(49.4 %)	275(42.3 %)
4　法定耐用年数と同じぐらい	-	257(32.5 %)	232(35.7 %)
合計	-	791(100.0 %)	650(100.0 %)

参考文献

[1] 青井和夫監、直井優編：社会調査の基礎、サイエンス社、1983

[2] 新井益太郎：減価償却の理論、同文館、1979

[3] 朝日会計社編：資金調達の実務、中央経済社、1984

[4] 新井清光：継続性の変更と正当な理由、企業会計、Vol.26、No.10、pp.30-35,1974

[5] 新井益太郎：減価償却の理論、同文館、1980

[6] Blalock,H.M.:Correlation and Causality:The Multivariate Case,Social Forces,No.39,pp.246-251,1961

[7] Campbell,H.G., R.A.Dudek and M.L.Smith :A Heuristic Algorithm for the n Job m Machine Sequencing Problem, Management Science, Vol.16,No.10,pp.630-637,1970

[8] Dannenbring, D.G.:An Evaluation of Flow Shop Sequencing Heuristics,Management Science,Vol.23,No.11,pp1174-1182,1977

[9] Fisher,R.A.:The Use of Multiple Measurements in Taxonomic Problems,Annals of Eugenics,7,pp.179-188,1936

[10] 石井恵一、堀素夫共訳：工科系のための統計概論（原著 Guttman,T. and S.S.Wilks:Introductory Engineering Statics)、培風館、1971

[11] 玄光男、辻村康寛、高村和央、程潤偉：GA による巡回セールスマン問題の解法の数値実験、足利工業大学研究集録、第 20 号、pp.401-407、1994

[12] Goldberg, D. E. and R. Lingle,Jr.:Alleles, Loci, and the Traveling Salesman Problem, Proc. 1st ICGA,1985

[13] Grefenstette, J., R. Gopal, B. Rosmaita and D. Van Gucht:Genetic Algorithms for the Traveling Salesman Problem,Proc.1st ICGA,1985

[14] Glover,F.,:Tabu search - Part Ⅰ,ORSA Journal on Computing 1/3,pp.190-206,1989

[15] Haberman,S.J.:Analysis of Qualitative Data,Academic Press,1978

[16] 萩原宏、西原誠一：データ構造とプログラミング技法、オーム社、1987

[17] 広瀬弘忠訳：因果分析法 (原著：Asher,H.B.:Causal Modeling)、朝倉書店、1983

[18] 伊庭斉志：遺伝的プログラミング、東京電機大学出版局、1996

[19] 飯間等, 三宮信夫「扱いにくい制約のあるスケジューリング問題に対するGAの挙動」第20回システムシンポジウム講演論文集,pp239-244,1994年

[20] 飯間等, 三宮信夫「遺伝アルゴリズムの動作に対する致死遺伝子の影響」計測自動制御学会論文集,Vol.31,No.5,pp569-576,1995年

[21] 飯野利夫：発生主義会計の利益平準化作用、会計ジャーナル、3巻増刊号、pp.40-49 、1971

[22] 飯野利夫：限定意見の意義と役割、企業会計、Vol.26、No.10、1974

[23] 井上良二：経営者会計行動と企業戦略、会計、Vol.134、No.2、34-47、1988

[24] 石渕久生、野崎賢、山本直久、田中英夫：遺伝的手法によるファジィーIF-Thenルールの選択、電気情報通信学会A、Vol.J76-A,No.10,pp.1465-1473,1993

[25] 市橋秀友、渡辺俊彦：簡略ファジィ推論を用いたファジィモデルによる学習型制御、日本ファジィ学会誌、Vol.2、No.3、pp.429-437、1990

[26] Jhonson, S.M. :Optimal Two- and Three-stage Production Schedules with Setup Times Included, Naval research logistics quarterly, Vol.1,No.1,pp61-68,1954

[27] 加藤常員、小沢一雅：遺伝的アルゴリズムを用いたクラスタリング、電子情報通信学会技術報告、PRU95-147,pp.19-24,1995

[28] 北野　宏明編：遺伝的アルゴリズム、産業図書、1993

[29] 木内佳市：『正当な理由』の解釈について、企業会計、Vol.27、No.8、pp.43-44、1975

[30] 木内佳市：減価償却論、同文館、1977

[31] 木村和三郎：新版減価償却論、森山書店、1981

[32] Kirkpatrick,S.,C.D.Gelatt,Jr.,M.P.Vecchi:Optimization by Simulated Annealing,SCENCE,Vol.220,pp.671-680,1983

[33] 喜多　一：シミュレーテッドアニーリング、日本ファジィ学会誌、Vol.9,No.6,pp.870-875,1997

[34] 日下部與一：監査基準逐条詳解、中央経済社、1972

[35] 日下部與一：新会計監査詳説、中央経済社、1981

[36] 国村道雄：わが国企業の減価償却政策、産業経理、Vol.43、No.2、pp.79-89、1983

[37] 国村道雄：減価償却・資産評価と会計政策－実証－、企業会計、Vol.38、No.7、pp.24-33、1986

[38] 熊野實夫：会計方針の変更と継続性の原則、企業会計、Vol.34、No.8、pp.79-84、1982

[39] 香村光雄：会計上の利益平準化の分析、大分大学経済論集、Vol.27、No.5、pp.26-61、1975

[40] 香村光雄：利益平準化とその影響－会計上の利益平準化の分析(続)-、大分大学経済論集、Vol.28、No.1、pp.38-58、1976

[41] 香村光雄：わが国企業における会計方法選択の一分析（一)、会計、Vol.127、No.2、pp.51-62、1985

[42] 香村光雄：わが国企業における会計方法選択の一分析（二）、会計、Vol.127、No.3、pp.70-82、1985

[43] 黒川行治、高橋正子、渡瀬一紀：日本企業の経営環境と決算行動に関する意識調査、三田商学研究、Vol.41,No.2,pp.87-107,1998

[44] 前川景示, 玉置久, 喜多一, 西川禮一：遺伝アルゴリズムによる巡回セールスマン問題の一解法、計測自動制御学会論文集、Vol.31,No.5,pp598-605,1995

[45] 皆川雅章、三上貞芳、横井浩史、高取則彦、鈴木恵二、川上敬共訳：遺伝的アルゴリズムの理論 (原著 Holland,J.H.:Adaptataion in Natural and Arti□cial Systems)、森北出版、1999

[46] 宮本定明：クラスター分析入門、森北出版、1999

[47] 村田忠彦, 石渕久生, 田中英夫：遺伝的アルゴリズムによるフローショップ・スケジューリングと多目的最適化問題への応用, 計測自動制御学会論文集,Vol.31,No.5,pp583-590,1995

[48] 室田一雄編：離散構造とアルゴリズムIV、近代科学社、1995

[49] 中島智晴、村田忠彦、石渕久生：クラシファイシステムによる言語識別ルールの獲得、日本経営工学会誌、Vol.47,No.3,pp.199-206,1996

[50] 成嶋弘：数え上げ組合せ論入門、日本評論社、1996

[51] Nawaz,M., E.E.Enscore,Jr. and I.Ham:A Heuristic Algorithm for the m-Machine, n-Job Flow-shop Sequencing Problem, OMEGA,Vol.11,No.1,pp91-95,1983

[52] 野村博義、若見昇：遺伝アルゴリズムによるファジィ推論ルールの決定法、電子情報通信学会論文誌A、Vol.J77-A,No.9,pp.1241-1249、1994

[53] 野崎賢、石渕久生、田中英夫：学習型ファジィ識別システムによる知識獲得、日本経営甲学会誌、Vol.45,No.5,pp.430-441、1994

[54] 日本公認会計士協会 25 年史編纂委員会：会計・監査資料、日本公認会計士協会、1977

[55] 野村博義、林勲、若見昇：デルタルールによるファジィ推論の自動チューニング手法と障害物回避への応用、日本ファジィ学会誌、Vol.4、No.2、pp.379-388、1992

[56] 沼田嘉穂：減価償却の知識、日本経済新聞社、1981

[57] 沼田嘉穂：減価償却の理論と実務、同文館、1982

[58] 岡田勇二、加藤常員、小沢一雅：総合距離尺度によるクラスタ分割 (CLAIM 法)、情報処理学会論文誌、Vol.35,No.11,pp.2291-2298

[59] 奥野忠一、久米均、芳賀敏郎、吉澤正：多変量解析法、日科技連、1971

[60] 奥村晴彦：アルゴリズム事典、技術評論社、1991

[61] Osman,I and C.Potts:Simulated Annealing for Permutation Flow-Shop Scheduling,OMEGA,Vol.17,No.6,pp551-557,1989

[62] 汪金芳、大内俊二、景平、田栗正章：ブートストラップ法、行動計量学、Vol.19,No.2,pp.50-81,1992

[63] 小澤康人：減価償却方法選択の基準、税経通信、Vol.39、No.14、pp.34-39、1984

[64] Quinlan,J.R.:Intorduction of Decision Trees,Machine Learning,Vol.1,pp.81-106,1986

[65] Quinlan,J.R.:C4.5:Programs for Machine Learning,Morgan Kaufmann Publishiers,1993

[66] Quinlan,J.R.:http://www.rulequest.com

[67] Reeves, C.R.:A Genetic Algorithm for Flowshop Sequencing,Computers Ops. Res.,Vol.22,No.1,pp5-13,1995

[68] Reeves,C.R. 編　横山隆一、奈良宏一、佐藤晴夫、鈴木昭男、荻本和彦、陳洛南訳：モダンヒューリステックス、日刊工業新聞社、1997

[69] Rinnooy Kan,A.H.G., Machine Scheduling Problems:Classification,Complexity and Computations. Marinus Nijjhoff The Hague,1976

[70] 坂元慶行、石黒真木夫、北川源四郎：情報量統計学、共立出版、1983

[71] 坂元憲行：カテゴリカルデータのモデル分析、共立出版、1985

[72] 佐波宣平：海運再建整備法における「減価償却不足の解消」、経済論叢、Vol.91,No.6,pp.1-25,1963

[73] 三宮信夫, 喜多一, 玉置久, 岩本貴司「遺伝アルゴリズムと最適化」, 朝倉書店,1998

[74] 坂和正敏、田中雅博：遺伝的アルゴリズム、朝倉書店、1995

[75] Smith,E.D.:The Effect of the Separation of Ownership from Control on Accounting Policy Decisions, The Accounting Review,Vol.51,No.4,pp.707-723,1976

[76] Taillard, E.,:Some Efficient Heuristic Methods for the Flow Shop Sequencing Problem,European Journal of Operational Research 47,pp65-74,1990

[77] Takagi,H.:An Algorithm for Causal Modeling in Epidemiological Studies,Behaviormetrika,No.8,pp.41-55,1980

[78] Takahasi,K.,Y.Kurokawa and K.Watase:Corporate Bankuruptcy Prediction in Japan,Journal of Banking and Finance,No.8,pp.229-247,1984

[79] 高橋吉之助、黒川行治、高橋正子：研究開発費の会計手続きの選択にみる決算行動、会計、pp.807-822,1991

[80] 高橋吉之助、江島夏実、渡瀬一紀、高橋正子、黒川行治：企業の決算行動の科学、中央経済社、1994

[81] 高橋正子、渡瀬一紀、黒川行治：日本企業の経営環境と決算行動および環境会計に関する意識調査、三田商学研究、Vol.44,No.2,pp.93-111,2001

[82] 武田昌輔編著：コンメンタール法人税法、第一法規

[83] Turner, S. and D.Booth:Comparison of Heuristics for Flow Shop Sequencing,OMEGA,Vol.15,No.1,pp75-85,1987

[84] 馬野元秀、岡本宏隆、鳩野逸生、田村坦之、河内二三夫、梅津祐久、木下淳一:ID3 に基づくファジィ・ルールの抽出とその推論法、9th Fuzzy System Symposium,pp.857-860,1993

[85] 渡瀬一紀：未上場企業の減価償却方法選択要因に関する調査報告、長崎総合科学大学紀要、Vol.35,No.1,pp.127-136,1994

[86] 渡瀬一紀、橘啓八郎：VDT 作業の健康障害に対する対策とその効果、OA 学会論集「情報系」、No.4.pp.63-72,1994

[87] 渡瀬一紀：会計政策と多角化に関する調査報告、長崎総合科学大学地域科学研究所紀要「地域論叢」、No.13,pp.29-34,1995

[88] 渡瀬一紀：情報量規準による償却方法選択要因の分析、経営情報学会誌、Vol.4,No.3,pp.15-30,1995

[89] 渡瀬一紀：遺伝的アルゴリズムを使った度数分布の要約とその信頼性の評価、長崎総合科学大学紀要、Vol.38,No.1,pp.1-10,1997

[90] 渡瀬一紀：ニューラルネットワークによる償却方法選択の判別分析、長崎県立大学論集、Vol.30,No.3,pp.341-358,1997

[91] 渡瀬一紀：階乗進数を使った因果モデルの獲得, 電子情報通信学会技術報告,AI99-19,pp71-78,1999

[92] Watase,K.,:Knowledge Acquisition from the Contingency Table, 日本ファジィ学会誌、Vol.12,No.2,pp.286-294,2000

[93] 渡瀬一紀：階乗進数を使った順列表現とフローショップスケジューリング問題への応用、電子情報通信学会誌Ｄ－Ｉ、Vol.J85-D-Ｉ,No.5,pp.411-423,2002

[94] Widmer,M., and A.Hertz:A New Heuristic Method for the Flow Shop Sequencing Problem,Europian Journal of Operational Research,No.41,pp.186-193,1989

[95] 山村雅幸、小野貴久, 小林重信：形質の遺伝を重視した遺伝的アルゴリズムに基づく巡回セールスマン問題の解法, 人工知能学会誌、Vol.7,No.6,pp1049-1059,1992

[96] 柳浦睦憲, 茨木俊秀：順序問題における遺伝的交叉法に対する一考察, 電学論（C）,6 月号,pp117-127,1994

[97] 安田三郎、海野道郎：社会統計学、丸善、1977

[98] 米田信夫：コンピュータくんの順列生成,bit,Vol.1,No.3,pp249-252,1969