数理解析研究所講究録
947巻 1996年 213-222



# 遺伝アルゴリズムによる制約付きマルコフ決定過程の解法

# A Solving Method of a MDP with Constraint by Genetic Algorithm

平山克己 1)・河合　一 2)

Katsumi Hirayama 1) · Hajime Kawai 2)

1) 鳥取大学 工学研究科 社会システム工学専攻
1)Course in Engineering of Social Development ,Tottori Univ
2) 鳥取大学 工学部
2)Department of Social System Engineering ,Tottori Univ

We consider discreate time Markov decsion process (MDP) with finite state space, finite action space and two kinds of immediate rewards. The problem is to maximize time average reward generated by on reward stream. subject to that the other reward is not smaller than a prescribed value. The problem is analyzed in the range of pure stationary policies. MDP with one optimality criterion and no constraint can be silved by usual policy improvement method. MDP with one reward constraint can be solved by linear programming, in the range of mixed policies. On the other hand, however, when we restrict the policies to pure polices the problem is some conbinatrial problem, for which any solving method has not been discovered. In this paper, we propose an approach applying Genetic Algorithm in order to carry on a search process effectively and to obtain a near optimal pure stationary policy. A numerical example is given to exmine the effeciency of the approach proposed here.

## 1　はじめに

本論文では、有限状態空間、有限決定空間、及び2種類の直接利得を持つ離散時間マルコフ決定過程(Malkov Decision Process：略してMDP)を取り扱い、一方の利得から生じる時間平均利得をある与えられた値以上に保持する純定常政策の中で，他方の利得から生じる時間平均利得を最大にする政策を定める制約付きＭＤＰ問題 [1] について考える。

一制約を持つＭＤＰは，既に Beulter and Ross [2] により混合戦略の範囲で考察され，最適政策は，せいぜい二つの純政策の混合政策により与えられることが示されている。ただし，混合政策の下では各決定を毎期確率的に選択することになり，管理上面倒な点が多く，純政策の範囲内で最適政策を求めることは現実的な意味で重要な問題であると思われる。しかし，純政策に限定すると、組合せ的問題となり、厳密解の導出が非常に困難となる。そこで、本研究では制約付マルコフ決定過程に対して、遺伝アルゴリズムに政策改良法を加味した新しいアプローチを提案する。

## 2　制約つきＭＤＰ

はじめに以下の記号を定義する。

$I=\{0,1,\cdots,N\}$：状態空間　　$p_{ij}^k$：状態$i$で決定$k$を選択したときの推移確率

$a_i^k, b_i^k$：状態$i$で決定$k$を選択したときに生じる直接利得

$S$：純政策の集合，すなわち $S=k_0\times k_1\times\cdots\times k_N$　　$s$：純政策，すなわち，$s\in S$

ここですべての純政策に対し，マルコフ決定過程は完全エルゴティクであるとする。すなわち，定常分布を持つ。
$g(s)$：政策$s$を採用したときの直接利得$a_i^k$から生じる時間平均利得
$h(s)$：政策$s$を採用したときの直接利得$b_i^k$から生じる時間平均利得
$\pi^s=(\pi_0^s,\cdots,\pi_N^s)$：政策$s$を採用したときの定常分布

なお，表現の簡潔化のため，$p_{ij}^s,a_i^s,b_i^s$をそれぞれ，政策$s$採用したときの推移確率および，状態$i$における直接利得を表すとする。$g(s),h(s)$はそれぞれ，$g(s),v_i(s)$および$h(s),w_i(s)$ $(i\in I)$を未知数とする次の連立方程式

$$\begin{cases} g(s)+v_i(s) = a_i^s+\sum_j p_{ij}^s v_j(s) \quad i=1,\cdots,N \\ v_0(s)=0 \end{cases} \tag{1}$$

$$\begin{cases} h(s)+w_i(s) = b_i^s+\sum_j p_{ij}^s w_j(s) \quad i=1,\cdots,N \\ v_0(s)=0 \end{cases} \tag{2}$$

の解として与えられる。あるいは，定常分布を用い

$$g(s) = \sum_i \pi_i^s a_j^s \tag{3}$$

$$h(s) = \sum_i \pi_i^s b_j^s \tag{4}$$

$$\pi_j = \sum_i \pi_i^s p_{ij}^s, \quad j\in I \tag{5}$$

$$\sum_i \pi_i = 1 \tag{6}$$

で与えられる。以上の記号を用いると我々の問題は次のように表現される。

Object

$$\max_s g(s) \tag{7}$$

Subject to.

$$h(s)\geq\alpha \quad s\in S \tag{8}$$

### 2.1 混合政策と純政策

本研究では触れていないが、例えば図1のように制約付きMDPを混合政策の範囲で考えると、理論的に厳密解が得られることが示されている[1]。しかし、混合政策は決定を確定的には選ばず、確率的に選ぶ政策である。したがって、意思決定者にとっては純政策の範囲で考える方が現実的であり、取り扱い易いと考えられる。

また、図1のように混合政策では端点を結ぶ直線と時間平均利得hの制約値$\alpha$が交わる点が最適解となる。しかし、純政策に限定すると実行可能解は離散的な点上に存在し、時間平均利得bの制約値$\alpha$によっては最適解は端点を結ぶ直線上にあるとは限らず、組合せ的問題となっている。そのため、理論的に厳密解を得る手段が現在では存在しない。

## 3 遺伝アルゴリズムの概要

遺伝アルゴリズムは生物進化の法則と遺伝のメカニズムを工学的に取り入れ最適化アルゴリズムとして構成したものである[3]。近年はモダンヒューリステックス[4]として注目されている。

### 3.1 遺伝アルゴリズムの概念

生物の各個体は、それぞれ固有の染色体を持ち、染色体は遺伝子の配列で構成されている。同様に、本手法も染色体と遺伝子に対応する決定と政策から構成され、以下のステップにより繰り返し計算を行なう[6]。

step1
世代を $t=0$ とする。$M$ 個の個体（政策）をランダムに生成して、初期個体群 $X(0)$ 、

$$X(0)=\{x_1(0),x_2(0),\cdots,x_M(0)\}$$

を設定する。（但し、各遺伝子は $1\sim K$ の10進数表示。）

step2
各個体の適応度（目的関数の値）を決める。この適応度に依存した一定のルールで個体の淘汰を行なう。（ルーレット戦略、エリート保存戦略、ランク戦略）

step3
一定の確率で交叉、突然変異を行い、新しい個体を生成。子は親と置き変わり新しい世代 $X(t+1)$ 、

$$X(t+1)=\{x_1(t+1),x_2(t+1),\cdots,x_M(t+1)\}$$

が形成される。

step4
終了条件により終了もしくは $t=t+1$ としてstep2へ戻る。

このアルゴリズムの主要部分は、適応度設定と適応度の高い個体を残す手続き、および新しい個体を生成する手続きである[7]。すなわち、淘汰が探索の方向を決定するハンドルとなり、交叉や突然変異が探索を進めるエンジンとなっている。これらの手続きが有効に働く時、遺伝アルゴリズムは効力を発揮する。

### 3.2 遺伝アルゴリズムの適用法

この節では、制約付きマルコフ決定過程の遺伝アルゴリズムへの導入、各パラメータの設定、及び設定した3ケースの適応度について説明する。前節における記号列で表される個体 $x(t)$ がマルコフ決定過程における純政策 $S$ にあたり、遺伝子の長さは状態数 $N+1$ となる。また、遺伝子座 $i$ は状態 $i$ に対応し、そこに入る遺伝子が、状態 $i$ での決定 $k_i$ である。以下に、本研究における遺伝アルゴリズムの適用手順について述べる。

現個体群を $U$ とし、対象とする個体（政策）を2章で定義した $s$ とする。まず、(5)、(6)より $\pi_j^s$ を求め、(3)、(4)より $g(s)$、$h(s)$ を計算し、表現型 $(h^s,g^s)$ とする。主な、パラメータを以下に示す。

$$\text{個体（政策）}s\in U\text{の表現型}:(h^s,g^s)$$
$$\text{個体の長さ（状態数）}:N+1$$
$$\text{個体数}:M$$
$$\text{個体}s(\in U)\text{の適応度}:F^s$$

淘汰は探索した最良解を破壊しないようにエリート保存戦略を用いた。

また、適応度については次の3つのケースを設定し、数値実験を行った。

**＜CASE1＞政策改良法を考慮しない場合**
ＣＡＳＥ１はＧＡだけの探索で、時間平均利得 $h$ の制約値$\alpha$を満たさない個体に対しては、ペナルティーとして適応度を0にし、次世代の遺伝子として継承しないようにした。ペナルティーとして $h$ と$\alpha$の乖離度に応じて適応度を減少さすこともできるが、今回はＧＡのみの探索でどこまで制約付き MDP に適用できるかを検証するために今回は適応度を0とした。

$$i)\quad h^s \geq \alpha \quad \text{のとき} F^s = g^s$$
$$ii)\quad h^s < \alpha \quad \text{のとき} F^s = 0$$

**＜CASE2＞政策改良法+適応度(目的関数重視)**

ＣＡＳＥ２はＧＡと政策改良法とのハイブリット型である。グローバルな探索はＧＡ、ローカルな探索は政策改良法により行なう。ただし、どちらの手法に重点をおくかによって、探索効率は大きく変わることが予想できる。本研究では、制約ｈの値が$\alpha$を満たさない個体に対してのみ、政策改良法を用いて、ｈを増加させるような新しい政策を探索する。しかし、ローカル探索に政策改良法を用いた場合、制約ｈの値は増加されても目的関数ｇの値は減少することがある。このような個体を淘汰せずに、次世代に継承すると、制約は満たしているが、目的関数ｇの値は小さい個体が増加し、個体群の多様性が失われ、局所解に陥る可能性が高くなる。そこで、探索した新しい政策の評価指標をhg平面の傾き（ｇの増分／ｈの増分）としている。これにより、目的関数ｇの値が減少するような個体は淘汰される可能性が高くなり、個体群中の多様性を維持できる。

$i)$ $h^s \geq \alpha$ のとき
$F^s = g^s$
$ii)$ $g^s < \alpha$ のとき
$h^s + w_i^s = b_i^s + \sum_j p_{ij}^s w_j^s$ を満たす$h^s$、$w_i^s$を求める。次に
$b_i^k + \sum_j p_{ij}^k w_j^s$ を最大にする$k^*$を求め、
$F^s = (g^{k^*} - g^s)/(h^{k^*} - h^s)$

**＜CASE3＞政策改良法+適応度(制約条件重視)**

ＣＡＳＥ３では、政策改良法により目的関数ｇの値が増加されたときだけ新しい政策を次世代の個体として採用し、政策改良法によっても制約値$\alpha$を満たしていない個体に対してはペナルティーを課し、次世代では淘汰されやすいように設定した。

したがって、ＣＡＳＥ２よりも政策改良法により探索された個体が次世代に継承される条件は厳しく、ローカルな探索に依存する割合はＣＡＳＥ２よりも小さい。また、ＣＡＳＥ２の方法はhg平面の傾きを適応度としたため、目的関数値の増加を重視した適応度の設定となるが、hg平面の傾きが似通った個体が増加することにより多様性が失われ、局所解に陥る可能性も高くなる。

一方、ＣＡＳＥ３は制約を満たさない個体にはペナルティーを課し、制約条件を重視した適応度の設定となるため、制約を満たした個体の生存率は高く、解空間が狭い問題でも効率の悪い探索空間を避けて通ることが可能となる。その結果、早い世代で制約を満たした解を探索することができる。

ＣＡＳＥ３はＣＡＳＥ１とＣＡＳＥ２の性質を合わせ持つような構成となっている。

$i)$ $h^s \geq \alpha$ のとき
$F^s = g^s$
$ii)$ $h^s < \alpha$ のとき
$h^s + w_i^s = b_i^s + \sum_j p_{ij}^s w_j^s$ を満たす$h^s$、$w_i^s$を求める。次に
$b_i^k + \sum_j p_{ij}^k w_j^s$ を最大にする$k^*$を求め、
$iia)$ $h^{k^*} \geq \alpha$かつ $g^{k^*} \geq g^s$ であれば
$s$を$k^*$、$(g^s, h^s)$を$(g^{k^*}, h^{k^*})$に置換え
$F^s = g^{k^*}$
$iib)$ $h^{k^*} \geq \alpha$かつ $g^{k^*} < g^s$ であれば

$$F^s = g^s/\beta \qquad \text{ただし、}\beta > 1.0$$

*iii)* $h^{k^*} < \alpha$ であれば

$$F^s = 0$$

## 4 数値計算例

図2～図12は，個体数;20, 状態数;10, 決定数;5, 制約値 $\alpha = 20, 25, 30, 40$, $\beta = 2.0$ 直接利得a，直接利得b，推移確率を以下の表として，２種類の時間平均利得(h,g)をxy平面上に示したものである。

前節で提案した３つのＣＡＳＥについて、時間平均利得bの制約値$\alpha$を変化させて、数値計算を行ったのでその結果を示す。

これらの数値計算は全て同じ初期解でいずれも３００世代まで計算した結果である。図2～図4は$\alpha$が２０、２５、３０、４０のときのＣＡＳＥ１での世代推移における(h,g)の値の変化を示したものである。図中の数字は世代数を示している。

図5～図8はＣＡＳＥ２での世代推移における(h,g)の値の変化を示したものである。

図9～図12はＣＡＳＥ３での世代推移における(h,g)の値の変化を示したものである。

## 5 考察

ＣＡＳＥ１のＧＡだけの探索では予想以上の効果があった。しかし、制約値$\alpha$の値が大きくなるにつれ、制約を満たした解を見つけるまでに時間がかかっている。また$\alpha = 40$の時には３００世代でも制約を満たした解を探索することができなかった。

ＣＡＳＥ２のＧＡと政策改良法のハイブリッド型では、ＣＡＳＥ１よりも早期に制約を満たした解を探索していることが判る。また、ＣＡＳＥ１では探索不可能であった$\alpha = 40$の時でもわずか１８世代で最適解に到達している。

ＣＡＳＥ３のハイブリット型＋適応度にペナルティーを与えるＧＡでは、ＣＡＳＥ１の約半分の世代でＣＡＳＥ１同等もしくはそれ以上の探索能力を発揮している。また、$\alpha = 20$の時にはＣＡＳＥ２の方が早く最適解に到達しているように見えるが実はＣＡＳＥ２は最適解には到達してはおらずgの値は16.7であった。しかし、ＣＡＳＥ３では最適解g=16.8（$\alpha = 25$の時と同じ）に達していた。また、３００世代以内でＣＡＳＥ３は$\alpha$の値がいずれの時も最適解に達していた。

これらのことから、ＧＡだけのランダム探索よりもＧＡと政策改良法のハイブリット型で構成した探索法の方が効率的な探索が実現できていることが判る。

時間平均利得hの制約値$\alpha$は大きくなるほど、探索空間は小さくなり、実行可能解でさえ探索は困難になる。逆に、制約値$\alpha$が小さくなるほど、探索空間は広がり、実行可能解の中から最適解を探索することが困難となる。今回の数値実験ではどちらの場合でも政策改良法とＧＡのハイブリット型がＧＡだけの探索よりも有効であることが確認された。

また、前者の場合ＣＡＳＥ２の設定の方が有効であり、後者の場合ＣＡＳＥ３の設定の方が有効であろう。いづれにせよ、制約を満たしていない個体（政策）を如何に淘汰し、制約を満たした個体から如何に優性な決定を次世代に継承するかがポイントとなり、個体群内に無駄な探索となる個体を留めないことが重要であると考えられる。

## 6 おわりに

本研究では制約付き MDP 問題について、ＧＡと政策改良法のハイブリット型を提案したが、非常に良い結果が得られた考える。今回は政策改良法は各個体（政策）に１回しか行っていないが、繰り返し行えば必ず制約を

満たす個体を生成することも可能である。これは次回の課題としたい。

適応度の設定方法の違いによって、同じハイブリット型でもＣＡＳＥ２，ＣＡＳＥ３のように探索過程が異なってくることは興味深い。また、適応度の設定方法は今回の数値実験を行った方法以外にも、様々な方法が考えられる。

制約についても、今回は１つであったがＧＡでは複数の制約も取扱うことが可能である。しかし、その際には適応度の設定方法をよく考慮しておかないと効率的な探索は行えないであろう。

今後、これらの課題についても研究を継続していきたい。

## 参考文献

[1] Kenji Muro,masamitsu Ohnishi, and Toshihide Ibaraki:Markov Decision Processes with Multiple Constraints. Proceedings of the Seminar on Queueing Theoty and Its Applications,May 11-13,1987

[2] Beutler,F.J. and Ross,K.W.:Optimal Policies for Controlled Malkov Chains with a Constraint, J.Math.Anal.Appl., Vol.112, pp.236–252,1985.

[3] 北川敏夫：マルコフ過程、共立出版

[4] 茨木俊秀：組合せ最適化法をめぐる最近の話題、モダンヒューリスティックスの新展開－Genetic Algorithm,Simulated Annealing,Tabu Search,Neural Net法は本当に有効か？－、日本オペレーションズ･リサーチ学会第３０回シンポジウム、pp. 1–10(1993).

[5] 北野宏明：遺伝アルゴリズム、産業図書、(1993) p.871～p.883

[6] 三宮信夫：遺伝アルゴリズムによる最適化問題の解法、第３６回システム制御情報学会研究発表公演会p.9～p.18

[7] Branko,Soucek,and The IRIS Group：DYNAMIC, GENETIC,AND CHAOTIC,PROGRAMING,WILEY INTER SCIENCE.

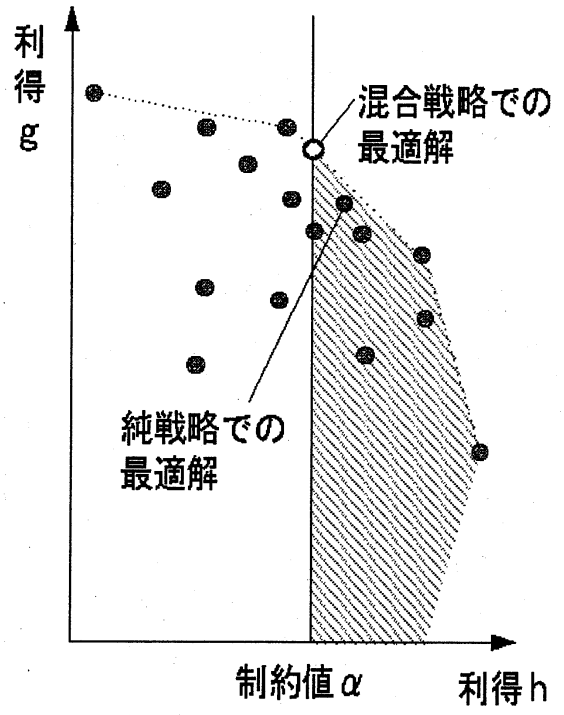


図 1: 混合政策と純正策での最適化

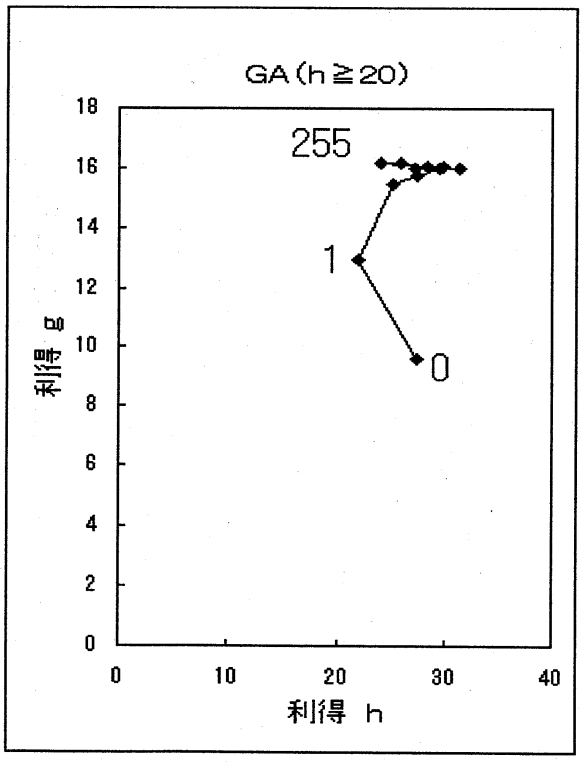


図 2: CASE1($h \geq 20$) での (h,g) の変化

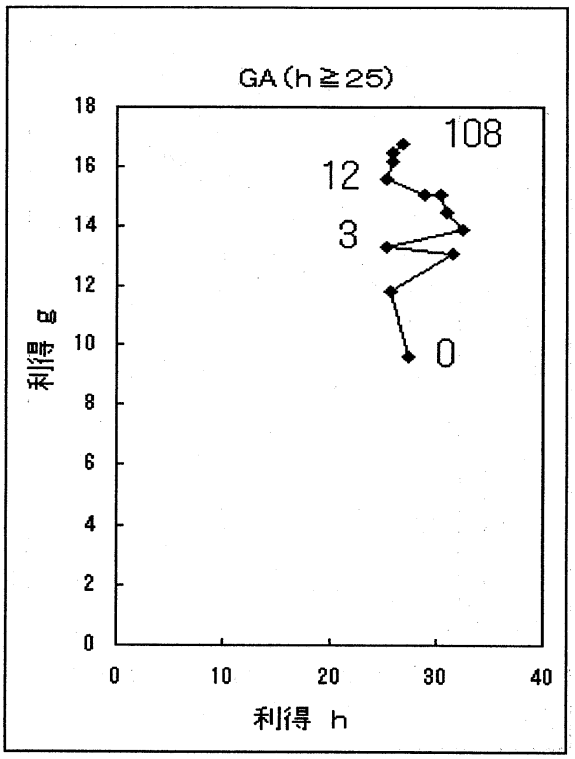


図 3: CASE1($h \geq 25$) での (h,g) の変化

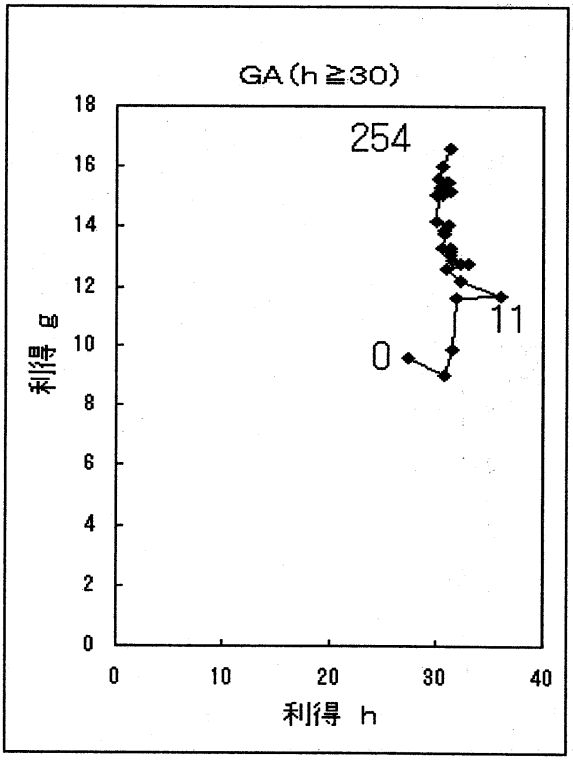


図 4: CASE1($h \geq 30$) での (h,g) の変化

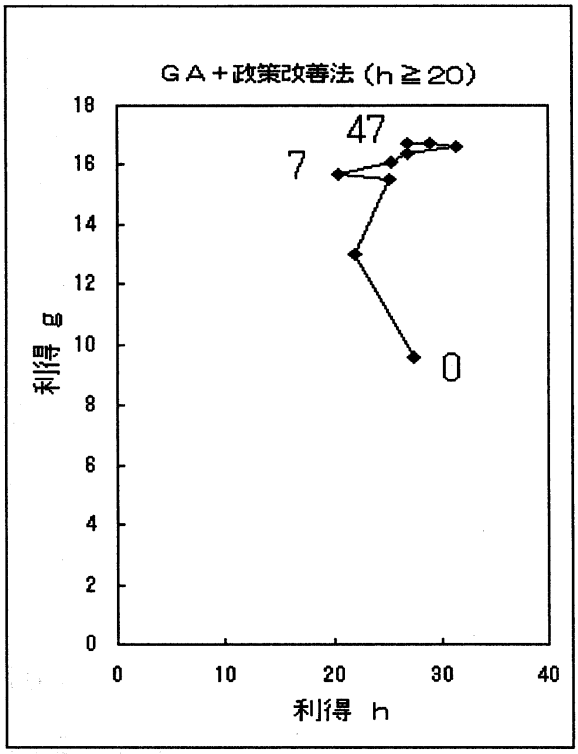


図 5: CASE2($h \geq 20$) での (h,g) の変化

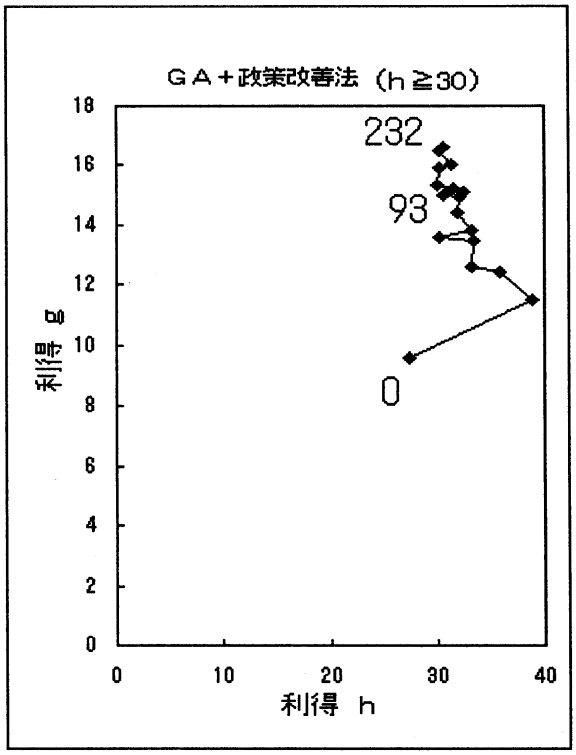


図 7: CASE2($h \geq 30$) での (h,g) の変化

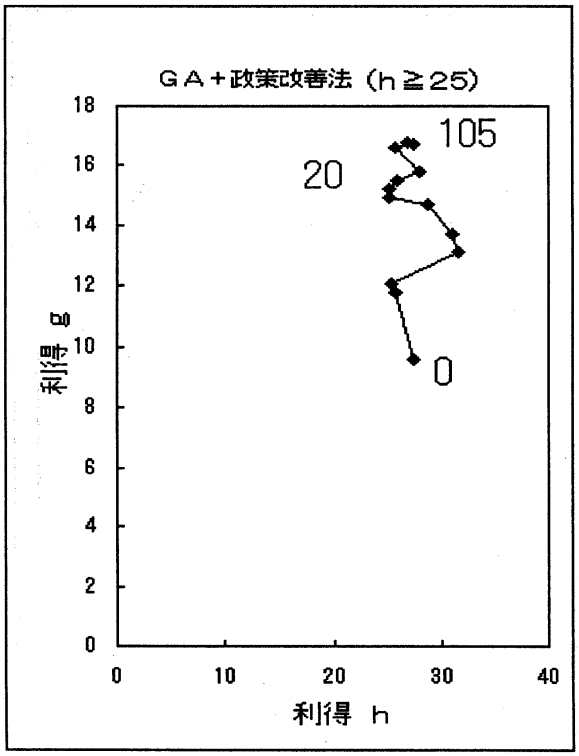


図 6: CASE2($h \geq 25$) での (h,g) の変化

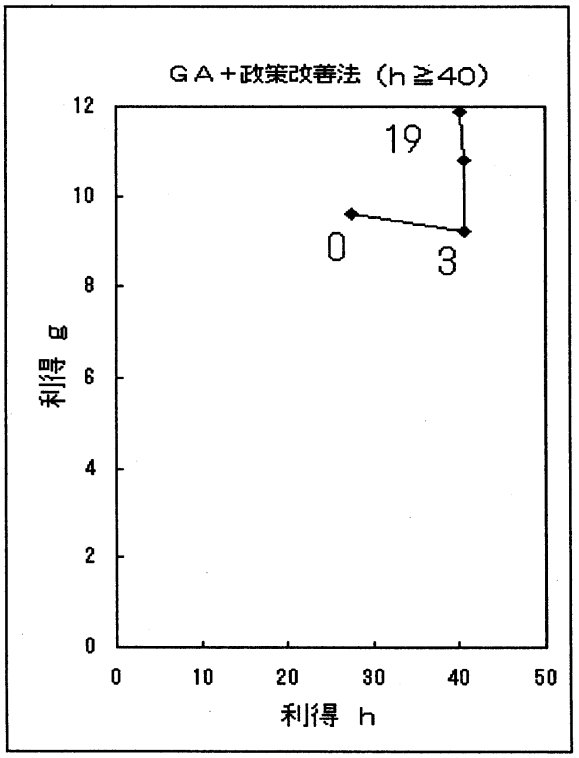


図 8: CASE2($h \geq 40$) での (h,g) の変化

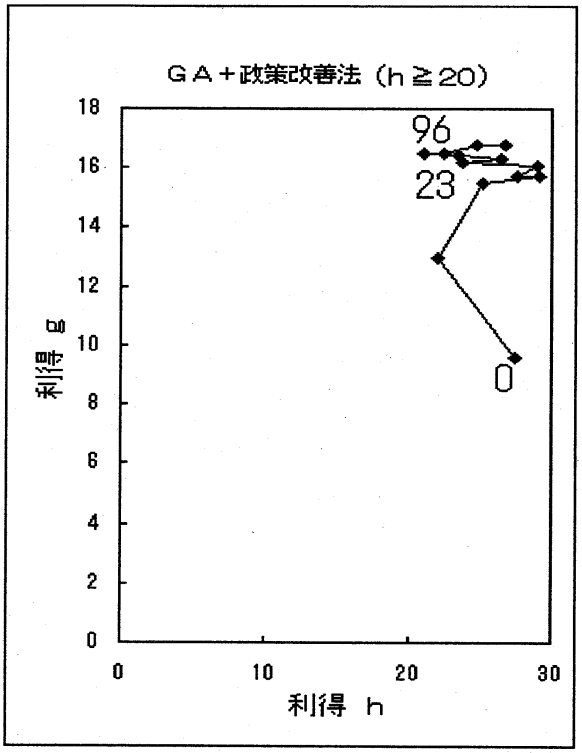


図 9: CASE3($h \geq 20$) での (h,g) の変化

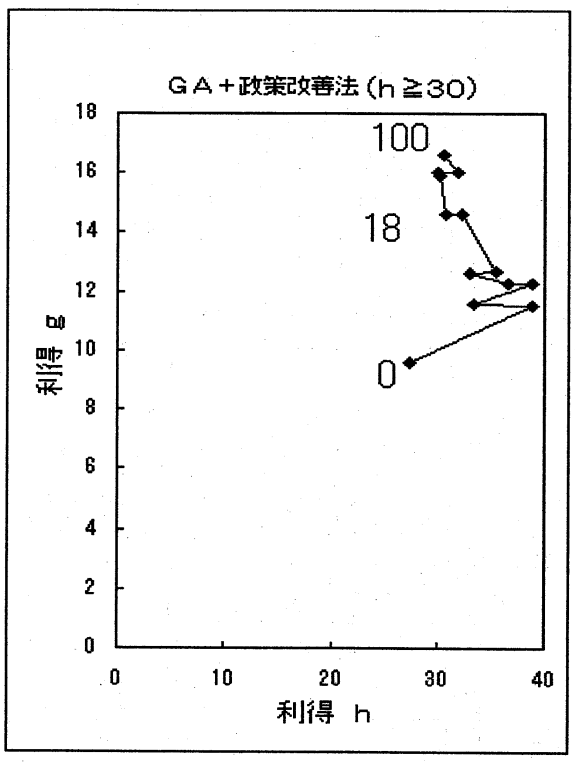


図 11: CASE3($h \geq 30$) での (h,g) の変化

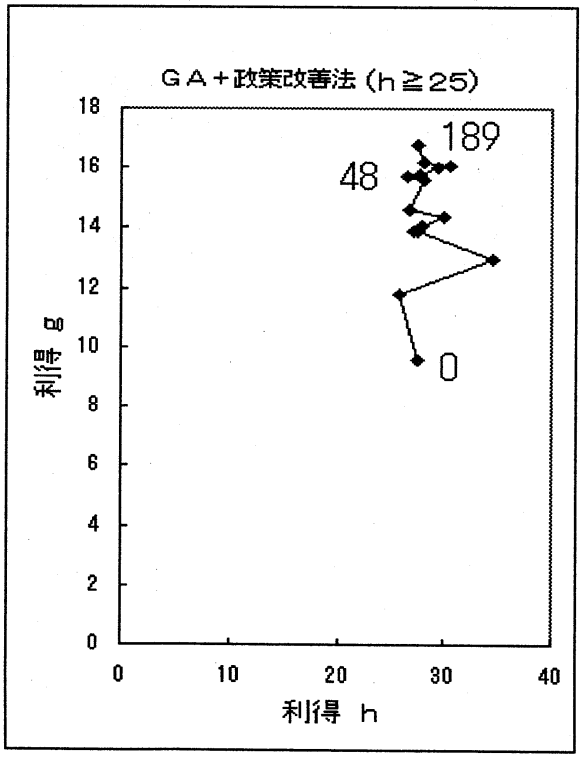


図 10: CASE3($h \geq 25$) での (h,g) の変化

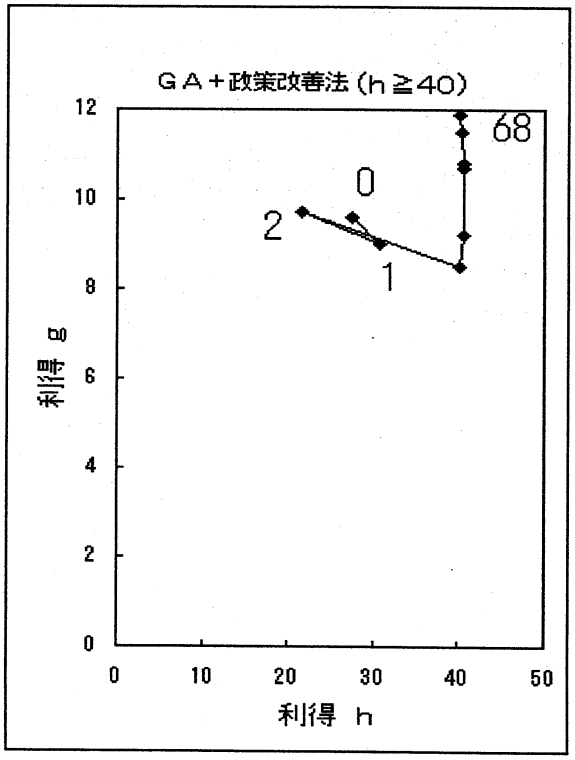


図 12: CASE3($h \geq 40$) での (h,g) の変化

政策1での直接利得a, 直接利得b, 推移確率(1/1000)

| j | a | b | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| i=1 | 8 | 8 | 165 | 835 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| i=2 | 14 | 31 | 276 | 503 | 221 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| i=3 | 13 | 20 | 0 | 404 | 28 | 569 | 0 | 0 | 0 | 0 | 0 | 0 |
| i=4 | 2 | 16 | 0 | 0 | 179 | 579 | 243 | 0 | 0 | 0 | 0 | 0 |
| i=5 | 2 | 21 | 0 | 0 | 0 | 256 | 128 | 617 | 0 | 0 | 0 | 0 |
| i=6 | 10 | 41 | 0 | 0 | 0 | 0 | 762 | 200 | 38 | 0 | 0 | 0 |
| i=7 | 19 | 37 | 0 | 0 | 0 | 0 | 0 | 168 | 358 | 474 | 0 | 0 |
| i=8 | 17 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 336 | 347 | 317 | 0 |
| i=9 | 3 | 20 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 354 | 146 | 500 |
| i=10 | 11 | 38 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 504 | 496 |

政策2での直接利得a, 直接利得b, 推移確率(1/1000)

| j | a | b | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| i=1 | 7 | 10 | 162 | 838 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| i=2 | 4 | 5 | 639 | 14 | 347 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| i=3 | 18 | 22 | 0 | 209 | 442 | 349 | 0 | 0 | 0 | 0 | 0 | 0 |
| i=4 | 3 | 22 | 0 | 0 | 256 | 279 | 465 | 0 | 0 | 0 | 0 | 0 |
| i=5 | 15 | 36 | 0 | 0 | 0 | 399 | 45 | 556 | 0 | 0 | 0 | 0 |
| i=6 | 5 | 32 | 0 | 0 | 0 | 0 | 161 | 0 | 839 | 0 | 0 | 0 |
| i=7 | 16 | 23 | 0 | 0 | 0 | 0 | 0 | 14 | 183 | 803 | 0 | 0 |
| i=8 | 2 | 44 | 0 | 0 | 0 | 0 | 0 | 0 | 159 | 483 | 358 | 0 |
| i=9 | 0 | 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 89 | 278 | 633 |
| i=10 | 12 | 6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 917 | 83 |

政策3での直接利得a, 直接利得b, 推移確率(1/1000)

| j | a | b | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| i=1 | 9 | 17 | 497 | 503 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| i=2 | 7 | 4 | 252 | 290 | 457 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| i=3 | 19 | 33 | 0 | 237 | 437 | 326 | 0 | 0 | 0 | 0 | 0 | 0 |
| i=4 | 19 | 21 | 0 | 0 | 10 | 367 | 622 | 0 | 0 | 0 | 0 | 0 |
| i=5 | 11 | 48 | 0 | 0 | 0 | 365 | 170 | 465 | 0 | 0 | 0 | 0 |
| i=6 | 5 | 20 | 0 | 0 | 0 | 0 | 650 | 317 | 33 | 0 | 0 | 0 |
| i=7 | 19 | 17 | 0 | 0 | 0 | 0 | 0 | 95 | 715 | 190 | 0 | 0 |
| i=8 | 19 | 41 | 0 | 0 | 0 | 0 | 0 | 0 | 332 | 328 | 341 | 0 |
| i=9 | 2 | 24 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 295 | 210 | 495 |
| i=10 | 4 | 39 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 709 | 291 |

政策4での直接利得a, 直接利得b, 推移確率(1/1000)

| j | a | b | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| i=1 | 3 | 39 | 446 | 554 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| i=2 | 8 | 38 | 436 | 221 | 344 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| i=3 | 19 | 48 | 0 | 620 | 315 | 65 | 0 | 0 | 0 | 0 | 0 | 0 |
| i=4 | 0 | 5 | 0 | 0 | 196 | 297 | 507 | 0 | 0 | 0 | 0 | 0 |
| i=5 | 10 | 48 | 0 | 0 | 0 | 127 | 253 | 620 | 0 | 0 | 0 | 0 |
| i=6 | 5 | 25 | 0 | 0 | 0 | 0 | 240 | 625 | 135 | 0 | 0 | 0 |
| i=7 | 11 | 46 | 0 | 0 | 0 | 0 | 0 | 347 | 189 | 463 | 0 | 0 |
| i=8 | 3 | 35 | 0 | 0 | 0 | 0 | 0 | 0 | 301 | 610 | 89 | 0 |
| i=9 | 4 | 44 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 109 | 341 | 551 |
| i=10 | 15 | 35 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 673 | 327 |

政策5での直接利得a, 直接利得b, 推移確率(1/1000)

| j | a | b | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| i=1 | 6 | 28 | 549 | 451 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| i=2 | 6 | 15 | 30 | 626 | 343 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| i=3 | 6 | 40 | 0 | 559 | 235 | 206 | 0 | 0 | 0 | 0 | 0 | 0 |
| i=4 | 13 | 14 | 0 | 0 | 264 | 595 | 142 | 0 | 0 | 0 | 0 | 0 |
| i=5 | 16 | 12 | 0 | 0 | 0 | 258 | 379 | 363 | 0 | 0 | 0 | 0 |
| i=6 | 11 | 26 | 0 | 0 | 0 | 0 | 409 | 181 | 409 | 0 | 0 | 0 |
| i=7 | 18 | 26 | 0 | 0 | 0 | 0 | 0 | 316 | 377 | 307 | 0 | 0 |
| i=8 | 11 | 11 | 0 | 0 | 0 | 0 | 0 | 0 | 156 | 269 | 575 | 0 |
| i=9 | 18 | 15 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 215 | 355 | 430 |
| i=10 | 15 | 29 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 603 | 397 |