

福岡工業大学 学術機関リポジトリ

The Effects of Mutations on Solutions in Genetic Algorithm

メタデータ	言語: jpn 出版者: 公開日: 2021-02-26 キーワード (Ja): キーワード (En): 作成者: 百原, 武敏, 荒屋, 真二 メールアドレス: 所属:
URL	http://hdl.handle.net/11478/00001667

遺伝的アルゴリズムにおける突然変異の 解におよぼす影響について

百 原 武 敏 (情報工学科)
荒 屋 真 二 (情報工学科)

The Effects of Mutations on Solutions in Genetic Algorithm

Taketoshi MOMOHARA (Department of Computer and Communication Engineering)
Shinji ARAYA (Department of Computer and Communication Engineering)

Abstract

This paper discusses the characteristics of genetic algorithms from the view point of mutation. Two kinds of mutation with fixed and variable probabilities are used for the Knapsack problem in our experiments. The following results are obtained :

- (1) The quality of solution is better when using mutation of a suitable probability value than not.
- (2) There exists an optimal value of mutation probability.
- (3) The variable mutation method where the probability decreases together with generations could converge to the best solution.

Key Words : *Genetic algorithm, Mutation, Knapsack Problem*

1. 緒 言

遺伝的アルゴリズム (GA : Genetic Algorithm) は、生物の遺伝と進化過程をモデルとした探索アルゴリズムであり、最適化問題の求解法としても有望視されている。GAにおいて解に影響をおよぼす基本操作に、交叉と突然変異がある。交叉については多くの論文で取り上げられているが、突然変異についてはその特性を論じたものはきわめて少ない。本研究では、最適化問題であるナップザック問題を取り上げ、突然変異が解におよぼす影響について実験的に特性解析を行

う。2章ではGAによる最適化問題解法の概要を述べる。3章では、問題と実験方法について述べ、実験結果を示すと同時に、突然変異と解の質との関連性について解析・検討する。

2. 遺伝的アルゴリズムについての概要

2.1 遺伝的アルゴリズムの概要

生物は個体の集合体として、ある規模の集団を構成している。各個体は遺伝子を持ち、それがその個体の特徴づけている。各個体は集団の中で生存競争を行い、勝ち抜いた個体は生き残り、次世代へ子孫を残す。新しい世代は、親の遺伝子を組み合わせた新しい遺伝子とそれに基づく能力を持つ。そして新しい世代で再度

生存競争を経て、さらに次世代が生成される。この過程における変化の一つとして、突然変異がある。突然変異においては、親の遺伝子の組み合わせから生ずる能力以上あるいは以下の能力を持つ個体が生成される。

染色体とは遺伝情報を伝える実体であり、コンピュータ上では配列として定義されるデータ領域である。遺伝子座は染色体の各位置であり、この各位置にどのような遺伝情報が格納されるかが決まっている。コンピュータ上では配列の各要素に対応する。遺伝子座には各個体の遺伝的特徴を決定する情報、すなわち遺伝子が格納されており、遺伝子の組み合わせのパターンを遺伝子型と呼ぶ。遺伝子型に基づいて形成された個体を表現型と呼ぶ。

2. 2 遺伝的アルゴリズムのシミュレーション

このような進化の仕組みをコンピュータ上でシミュレーションし、最適化問題を解く方法を以下に概説する。ただし、GAは緩やかな枠組みであり、個々の問題解決において、以下のすべてのステップを必要とするものではない。

STEP 1 : 問題事象のモデル化

GAでは個体の遺伝子を、コンピュータ内部の記号列として表現する。すなわち問題の事象を、STEP 2以降の処理に都合のいい形式でコーディングする。個体としての記号列表現は、対象とする問題の探索空間を符号化したものであり、それは問題に依存し設計者の決定に任される。

STEP 2 : 初期個体集団の生成

STEP 1で決定したコーディング法に基づき、個体の集団を生成する。各個体は探索空間における探索点に担当し、個体集団の生成は複数の探索点を定めることである。各個体は遺伝子型の違いによりそれぞれ個性、すなわち問題環境に対する適応度を持つ。

STEP 3 : 各個体の評価

STEP 2で生成された個体集団の各個体は、環境への適応度が高いものから低いものまで様々である。適応の度合いを表すために、評価関数を用いて各個体の評価値(適応度)を求める。評価値の高い個体ほど環境に対する適応度が高く、優れた個体であるとみなす。評価関数は問題に応じて設計者が定義する。

STEP 4 : 淘汰

STEP 3の評価結果に基づき、環境に対する適応度の低い個体のある一定数削除する。

STEP 5 : 増殖

STEP 4の淘汰の結果減少した個体数分を、現在の集団中の個体を選択して増殖させる。

STEP 6 : 交叉

個体集団中からある方法で個体の対を選び、それらの特定の部位を入れ換える。その結果、これまで残っている集団中の個体よりも、評価値の高い個体が生成される可能性がある。

STEP 7 : 突然変異

突然変異は、個体中のある部位をある確率で変化させる。この操作により、突発的に評価値の高い個体が生成される可能性がある。

STEP 8 : 収束判定

生成された次世代の個体集団が、シミュレーションを終了するための収束条件を満たしているかどうかを調べ、満足していればシミュレーションを終了し、満足していなければSTEP 3に戻る。

3. 実験と考察

3. 1 ナップザック問題

重さ w_i 、コスト c_i が既知の n 個の物体があるとき、制限重量 W の範囲内でコストの総和 C が最大になる物体の組み合わせを求め問題であり、次のように定式化される。

$$\max C = \sum_{i=1}^{i=n} c_i g_i \quad (1)$$

$$\sum_{i=1}^{i=n} w_i g_i \leq W \quad (2)$$

ただし g_i は物体 i を選択する場合は1、選択しない場合は0である。

3. 2 実験の条件

本研究ではSGA⁽¹⁾を用いて実験を行った。SGAは個体の生成、交叉、突然変異の3つの基本操作に基づいたアルゴリズムである。ただし、以下のような変更を加えた。すなわち個体の生成、交叉、突然変異な

どの操作に関与する乱数系列が、実行時に与える値に基づいて決定されるのを、時刻を元に実行ごとに異なる乱数系列を生成するようにした。以下このアルゴリズムを単にSGAと記す。さらにエリート保存戦略(個体集団中で最大の評価値を持つ個体を次世代に残す方法)を組み込んだアルゴリズムも用いた。これをE-SGAと記す。

さらに、突然変異として2種の方式を用いた。
(突然変異方式1)

各個体のすべての遺伝子を対象にして、突然変異の生起確率(Pm)で、0を1、1を0に変更する。

(突然変異方式2)

染色体のある部分からある長さの範囲内で、上位と下位の遺伝子を入れ替える方式で、逆位反転法という。入れ換える遺伝子の長さを、逆位反転のビット長と呼ぶ。

(実験条件)

- 問題データ 表1に示す35個のデータ
- 打ち切り世代数 1,000世代
- 集団個体数 200
- 交叉の生起確率 0.03
- 交叉法 一点交叉

(以上の条件は固定)

- 突然変異の生起確率(Pm) 次節に述べる実験方法ごとに表2, 3, 4に示す。実験によっては必ずしもすべてのパラメータについて測定を行ってはいない。
- 逆位反転のビット長 表5に示す。

3.3 実験方法

突然変異が解におよぼす影響を調べるために、アルゴリズムをSGAとE-SGAの2種、突然変異の方式を2種設定した。また突然変異を①Pm一定、②Pmの値を次第に減少させる、③Pmの値を次第に増加させる、という3種の方法で作用させた。①は普通の方法である。②は次のような状況を想定している。すなわちPmの値が大きな初期の世代で突然変異が頻繁に起こり、評価値の高い個体が生成され、Pmの値を漸減する事によりその個体が生き残り、解が評価値の高い点に収束するのではないかとということである。③は

その対比である。

Pmの設定値については予備実験の結果、同じ桁の数値を変えたぐらいではあまり顕著な差異は見られず、広範囲の値に設定する必要性があった。また突然変異方式2では、方式1の場合よりもPmの値が大きな領域で評価値が高くなる傾向があった。

実験1～8は、突然変異方式1による実験である。Pmをパラメータとし、以下の実験方法において各10回ずつ測定する。

(SGAでの実験)

- ・実験1 Pmの値を一定。
- ・実験2 Pmの値を100世代ごとに2分の1にする。
- ・実験3 Pmの値を1000世代で元の値の100分の1になるよう100世代ごとに変化させる。
- ・実験4 Pmの値を1000世代で元の値の100倍になるよう100世代ごとに変化させる。

(E-SGAでの実験)

- ・実験5 Pmの値を一定。
- ・実験6 Pmの値を100世代ごとに2分の1にする。
- ・実験7 Pmの値を1000世代で元の値の100分の1になるよう100世代ごとに変化させる。
- ・実験8 Pmの値を1000世代で元の値の100倍になるよう100世代ごとに変化させる。

さらに突然変異方式2において、Pmと逆位反転のビット長をパラメータとして、以下の実験方法においてそれぞれ5回ずつ、測定を行った。

(SGAでの実験)

- ・実験9 Pmの値を一定。
- ・実験10 Pmの値を1000世代で元の値の100分の1になるよう100世代ごとに変化させる。

(E-SGAでの実験)

- ・実験11 Pmの値を一定。
- ・実験12 Pmの値を1000世代で元の値の100分の1になるよう100世代ごとに変化させる。

3.4 実験結果

実験結果のデータ量は膨大である。結果の一部を図1～7に示す。これらの図は、世代を横軸に、個体集

団の評価値の平均値(評価平均値)を縦軸にとっている。図1～3は実験1の結果である。SGAにおいて、Pmの値が0, 0.0005および0.05一定の場合である。Pmの値が増加するにつれ、評価平均値の変動が大きくなり、またその値が小さくなっている。突然変異の発生が多量になると、評価値の高い個体の生成が見られる一方、致死遺伝子(遺伝子の組み合わせが制限重量を越え、問題の解としての条件を満たさないもの)も発生し、評価平均値が低下している。図4に実験2、

図5に実験3の結果を示す。図4は、SGAでPmの初期値を0.05, 100世代ごとに1/2に減少させた場合である。図3(Pm一定)と比較してみると、評価平均値がスムーズに収束していることがわかる。図5は、SGAでPmの初期値を0.01, 1000世代で初期値の1/100になるよう100世代ごとに均等に減少させた場合である。図4と比較すると、Pmの値の変化のさせ方による評価平均値の変化に差異があることがわかる。さらに図6に実験4の結果を示す。SGAでPmの初

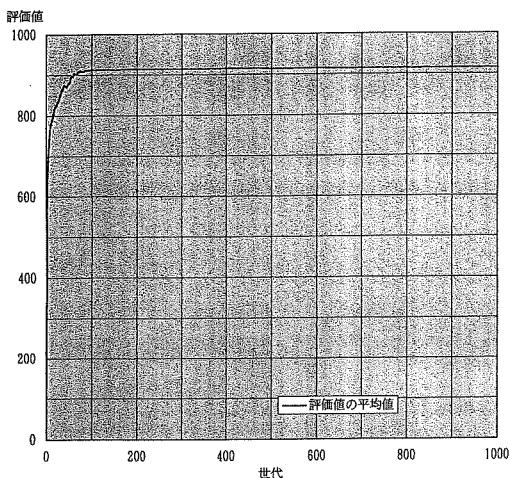


図1 実験1 世代に対する評価平均値の変化 (SGA, Pm=0一定)

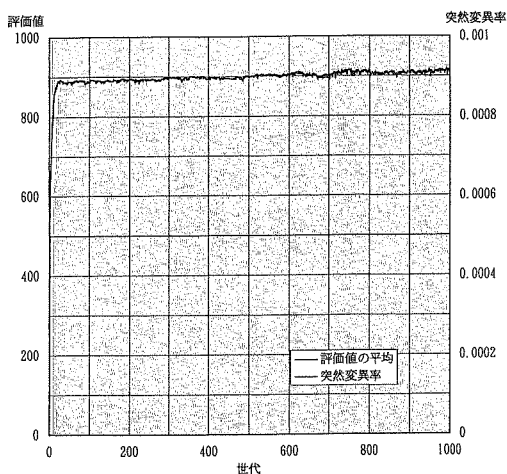


図2 実験1 世代に対する評価平均値の変化 (SGA, Pm=0.0005一定) (突然変異方式1)

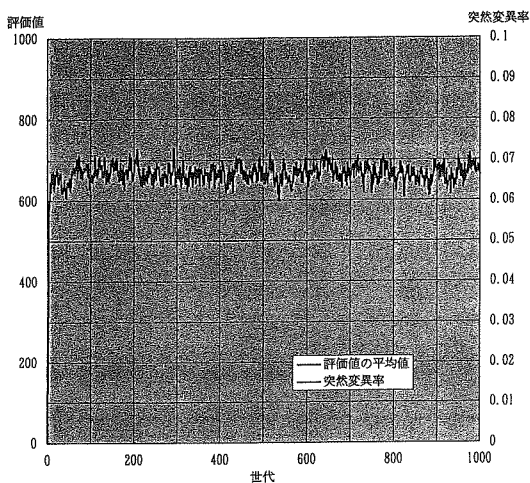


図3 実験1 世代に対する評価平均値の変化 (SGA, Pm=0.05一定) (突然変異方式1)

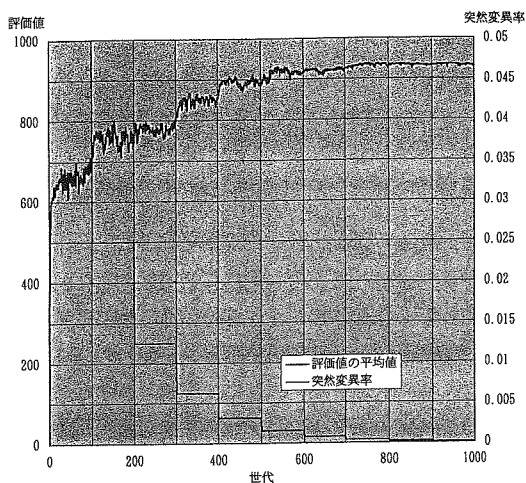


図4 実験2 世代に対する評価平均値の変化 (SGA, Pm=0.05, 100世代ごとに1/2) (突然変異方式1)

期値を0.001とし、1000世代で初期値の100倍になるよう100世代ごとに増加させた場合である。この場合は、評価平均値が世代とともに低下し、全体として発散していることがわかる。これらの傾向は、実験9～12すなわち突然変異方式2（逆反転法）においても同様である。図5の実験3の結果に対応するものとして、図7に実験10の結果を示す。SGAでPmの初期値を0.1とし、1000世代で初期値の1/100になるように100世代ごとに減少させた場合である。

実験結果を別の観点から見てみる。図8は、実験1（SGA, Pm一定）の結果を横軸にPmの値、縦軸に評価平均値をとったグラフである。ここでavg-maxは、各Pmの値ごとに10回ずつ測定したデータの評価平均値の平均値、同じくavg-minは評価平均値の最小値、avg-maxは評価平均値の最大値である。これを見ると、評価平均値はPm=0.001を境に減少に転じている。図9に、実験2（SGA, Pmが100世代ごとに1/2）の結果を同様に表したものを示す。こ

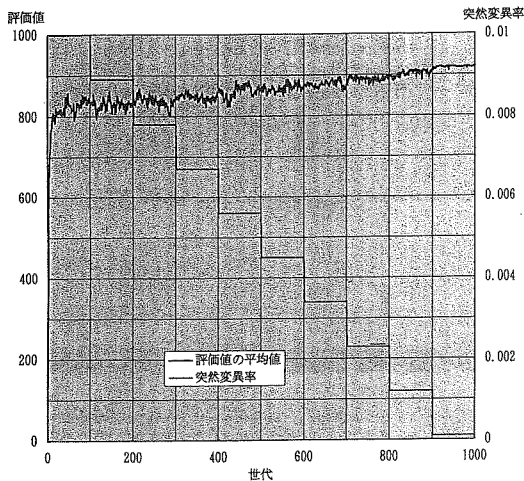


図5 実験3 世代に対する評価平均値の変化 (SGA, Pm=0.01, 1000世代で1/100) (突然変異方式1)

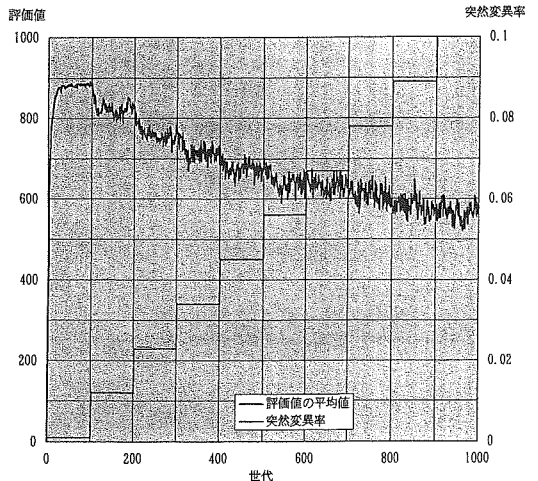


図6 実験4 世代に対する評価平均値の変化 (SGA, Pm=0.001, 1000世代で100倍) (突然変異方式1)

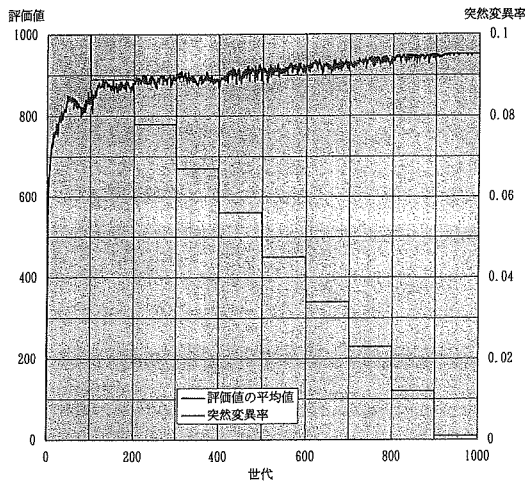


図7 実験10 世代に対する評価平均値の変化 (SGA, Pm=0.1, 1000世代で1/100) (突然変異方式2, 反転幅1=8)

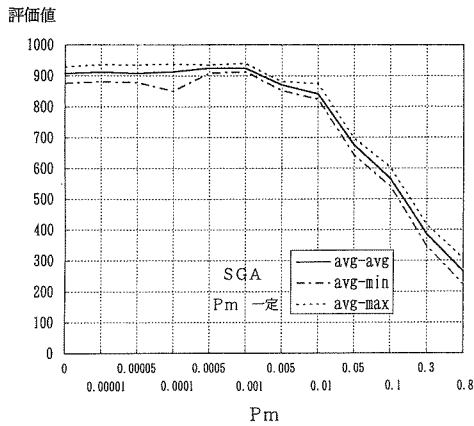


図8 実験1 Pmと評価平均値の関係 (SGA, Pm=一定) (突然変異方式1)

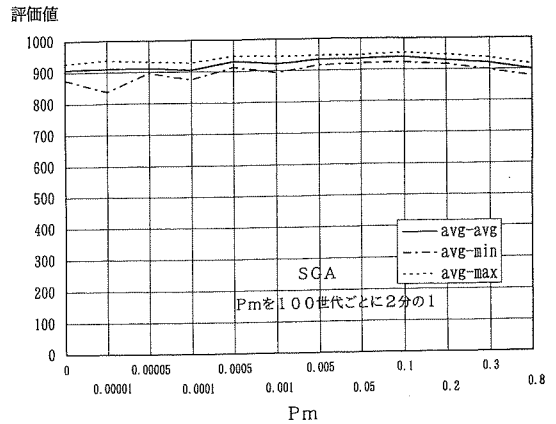


図9 実験2 Pmと評価平均値の関係 (SGA, Pm=100世代ごとに1/2) (突然変異方式1)

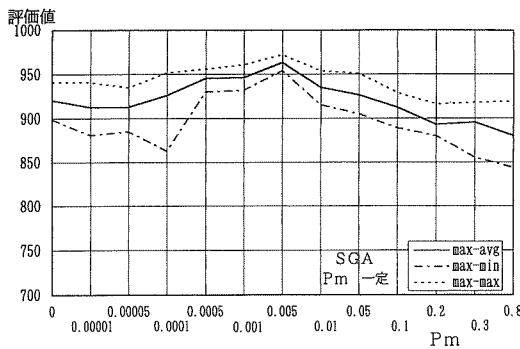


図10 実験1 Pmと評価値(最大値)の関係 (SGA, Pm=一定) (突然変異方式1)

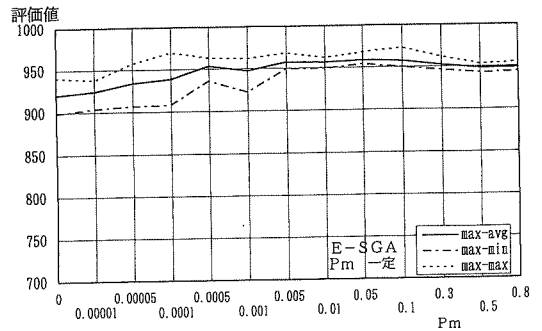


図11 実験5 Pmと評価値(最大値)の関係 (E-SGA, Pm=一定) (突然変異方式1)

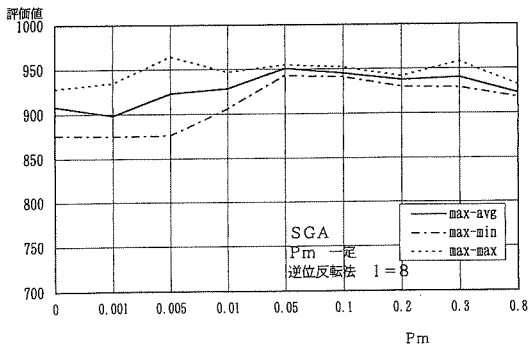


図12 実験9 Pmと評価値(最大値)の関係 (SGA, Pm=一定) (突然変異方式2, 反転幅 l=8)

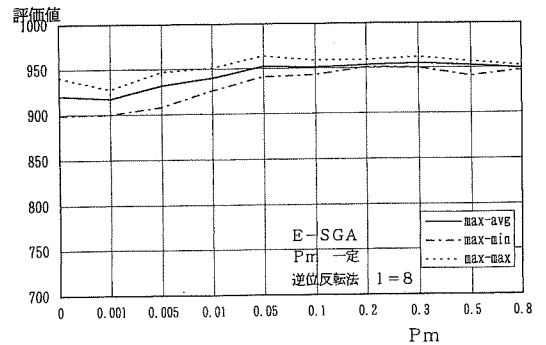


図13 実験11 Pmと評価値(最大値)の関係 (E-SGA, Pm=一定) (突然変異方式2, 反転幅 l=8)

表1 ナップザック問題のデータ

物体	重さ	コスト	物体	重さ	コスト
1	1	1	19	4	1
2	7	1	20	1	100
3	13	50	21	50	5
4	5	3	22	16	8
5	10	1	23	5	6
6	5	2	24	14	7
7	20	1	25	7	1
8	2	5	26	1	120
9	8	10	27	5	6
10	15	12	28	9	11
11	10	1	29	3	9
12	13	1	30	8	20
13	18	8	31	1	50
14	10	7	32	4	100
15	3	6	33	100	50
16	2	10	34	2	80
17	7	9	35	10	300
18	17	17			

表2 突然変異の確率 (Pm)
(実験1～3および5～7)

0
0.00001
0.00005
0.0001
0.0005
0.001
0.005
0.01
0.05
0.1
0.2
0.3
0.5
0.8

表3 突然変異の確率 (Pm)
(実験4, 8)

0
0.000001
0.000005
0.00001
0.00005
0.0001
0.0005
0.001
0.005

表4 突然変異の確率 (Pm)
(実験9～12)

0
0.001
0.005
0.01
0.05
0.1
0.2
0.3
0.8

表5 逆位反転のビット長
(実験9～12)

5
8
12
15

のグラフでは、 P_m の値が大きくなっても評価平均値の低下は認められない。これらの結果から、 P_m の大小とGAの評価値の変動との関連性、世代の進展とともに P_m の値を制御することにより、評価値の収束すなわちGAの結果が1つの解に収束していく具合を制御可能なことがわかる。しかし、突然変異が解の質にどのような影響を与えているのかを読みとることは困難である。

3. 5 実験結果の検討

前節で述べたようにこれまで表したような実験結果から、突然変異が解の質にどのような効果をもたらしているのかを、明確にすることは困難である。

そこで各実験ごとに評価値の最大値に着目し、 P_m を横軸、評価値を縦軸とし、各 P_m の値ごとに10回ずつ(実験9~12においては5回)測定したデータの、評価値の最大値の中での最大値(max-max)と最小値(max-min)および平均値(max-avg)をグラフ化したものが図10~13である。

図10~11は、突然変異方式1で、図10は実験1(SGA, P_m 一定)、図11は実験5(E-SGA, P_m 一定)の場合である。また、図12~13は、突然変異方式2(反転幅8)で、図12は実験9(SGA, P_m 一定)、図13は実験11(E-SGA, P_m 一定)の場合である。 $P_m=0$ では突然変異は全く起こらない。 P_m の値の上昇に伴って、評価値の最大値も上昇している。そして P_m の値をさらに上昇すると、評価値は逆に減少に転じている。そしてこのような変化は、突然変異方法1および2, SGAおよびE-SGAいずれの場合にも共通した結果である。

これらの結果から、 P_m の値の上昇につれて、評価値の最大値は緩やかな放物線を描いており、個体集団の評価値には、ある P_m の範囲においてピークになる領域が存在することがわかる。

4. 結 言

今回の研究で、GAにおいて突然変異が解におよぼ

す影響・効果について検討した。その結果、突然変異の生起確立 P_m の値によって、個体集団の評価平均値の変動幅が変化し、 P_m の値を世代とともに変化させることにより、解の収束を制御することができる、評価値が最大になるような P_m の領域が存在する、等のことがわかった。また $P_m=0$ 、すなわち突然変異が全く起こらず交叉のみによる場合よりも、突然変異が発生した場合に優れた個体が生成されることもわかった。これは現実の世界に置き換えても、納得できる事象であろう。しかし、 P_m のどの領域で評価値が最大値を示すか、すなわち解の最適値を得られるかということを確認するためには、さらに多くのデータを収集しなければならない。ナップザック問題でいうならば、個体数・交叉の確率・交叉法・データ個数等々変化する要素が多数である。今後さらに多くの場合のデータを収集し、検討を試みたい。

謝 辞

本研究を進めるにあたっては、1996年3月卒業の、本学情報工学科卒業研究生の協力によるところが大きい。下にメンバーを紹介し、心よりの謝意を表す。

神田 泰孝, 福田 忠晃, 山本 耕司
青木 保孝, 井上 健一, 今道 浩幸
中島 啓敬, 帆足勝次郎, 松本 美保
村上 正起, 横田 雅恵, 雷 興
以上 12名

参 考 文 献

- [1] R. E. Smith, D. E. Goldberg and J. A. Earicks on, SGA-C: A C-language Implementation of a Simple Genetic Algorithm, TCGA Report No. 91002, (1994).
- [2] 北野, 遺伝的アルゴリズム, 1993年7月, 産業図書
- [3] 安居院, 長尾: ジェネティックアルゴリズム, 1995年3月, 昭晃堂