岡山理科大学紀要 第36号 A pp. 183-190(2000)

集合被覆問題に対する遺伝的局所探索法のパラメータの影響分析

石田　洋一　　片山　謙吾*　　成久　洋之*
岡山理科大学大学院工学研究科修士過程情報工学専攻
*岡山理科大学工学部情報工学科

(2000年11月1日　受理)

1. まえがき

近年，さまざまな組合せ最適化問題に対して遺伝的アルゴリズム(Genetic Algorithm, GA)の適用がなされている．この手法に局所探索法(Local Search, LS)を加えることによって，大域的かつ局所的にバランスの良い探索法が実現される．これは一般に，遺伝的局所探索法(Genetic Local Search, GLS)と呼ばれる．また，GLS をある組合せ最適化問題に適用した際のパラメータ設定は，良好な結果を得るために重要な役目を担っており，その適切な設定は容易でないとされている．

本論文では，組合せ最適化問題の一つである，集合被覆問題(Set Covering Problem, SCP)を取り上げ，この問題に対して GLS を適用し，そのパラメータの影響分析を行う．SCP とは，与えられた問題のすべての行をコストが最小となるように，解の中の列の部分集合によって少なくとも一つは被覆するという問題である．つまり，与えられた問題のすべての行で，被覆されていない行がある場合の解の実行可能性は保証されない．そこで，本研究では，GLS で使用する LS において，被覆されていない行を見つけ，コストができるだけ低くなるようにすべての行を被覆し，解の実行可能性を保証する．また実行可能解でも，GA の操作である交叉や突然変異後は，解のビットが解を操作する以前と比べ変化しているので，実行可能解となる可能性は低い．そこで，遺伝的操作の後にも LS を施す．GLS 適用の際のパラメータには，生成される個体の数を決定する PS，交叉の確率を決定する交叉率 Pc，突然変異の確率を決定する突然変異率 Pm があり，それらの組合せによって構成されるパラメータ値設定により，よく知られた SCP のベンチマーク問題上で，GLS の探索傾向を調べた．その結果，Pc 及び Pm の設定値よりも PS の設定値の方が，GLS の適用により算出されるコストに，大きな影響を及ぼすことを示す．

2. 集合被覆問題

集合被覆問題(SCP)とは，よく知られた NP 困難な組合せ最適化問題の一つで，最小コストで列の部分集合によって m 行 n 列の 0-1 行列$[a_{ij}]$の行を被覆する問題である[2].

$x_j=1$：列 j（コスト $c_j>0$）が解の中にあるとき，
$x_j=0$：それ以外
と定義すると，SCP は次のように表される．

最小化　$\sum_{j=1}^{n} c_j x_j$

制約条件　$\sum_{j=1}^{n} a_{ij} x_j \geq 1$，$i=1,\ldots,m$

$x_j \in \{0,1\},\ j=1,\ldots,n$

この問題の一番目の制約式は，各行が少なくとも1つの列で被覆されることを示しており，2番目の制約式は整数条件を示している．この問題の応用例としては，鉄道，航空，バス等の運転士や車掌，パイロットなどの乗務員の乗務スケジュールなどがよく知られている．

3. 遺伝的アルゴリズム

遺伝的アルゴリズム(Genetic Algorithm:GA)は，生命が環境に適応していくメカニズムを進化としてとらえる考え方で，Lamarck や Wallace らの議論を経て Darwin の自然選択説により生物学の大きなテーマとして位置づけられた．一般に GA は，選択・交叉・突然変異の3つの遺伝的操作からなる．選択(selection)は，ある世代の個体群から適応度に応じて，次世代に残す子を選び出す操作である．交叉(crossover)は個体群の個体をランダムにペアリングし，両親の優れた形質を子に継承させる操作である．突然変異(mutation)は，各個体についてある確率でビットを選び出し，そのビットを反転させる操作である．これら3つの遺伝的操作を数世代にわたり繰り返すことによって徐々に優れた個体群を形成するという大域的な探索手法である．この手法に局所探索法(Local Search, LS)の処理を加えることにより，大域的かつ局所的な探索法が実現される．これは遺伝的局所探索法とよばれ，多くの場合大域的処理である GA のみの性能を大幅に改善できることが知られている．以下では，SCP に対する LS と GLS について記述する．

4. 局所探索法

実行可能解の集合 F を与えたとき，近傍 N は以下の写像と定義される[3].

$$N : F \to 2^F$$

実行可能解 $x \in F$ で，$f(x) \le f(x')$，$\forall x' \in N(x)$ を満たすものを局所解と呼ぶ．現時点での解 x の近傍 $N(x)$ から選ばれる近傍解 $x' \in N(x)$ を生成して，その近傍が現時点より，良好な評価値の解であれば x' に現時点の解を移動させる操作を繰り返すものである．この LS によって最終的に得られる解 x は，$N(x)$ の中に，改善解が存在しなくなった時とされ，$N(x)$ のもとで局所的に最適な解（局所最適解）となる．

4.1 SCP に対する LS の適用

本研究での SCP に対する LS のアルゴリズム[1]を以下に示す．なお，解 x や問題に対して，I はすべての行の集合，J はすべての列の集合，α_i は行 i を被覆した列の集合，β_j は列 j によって被覆された行の集合，S は解の $x_j = 1$ となる列の集合，U は被覆されてない行の集合，w_i は行 i を被覆した列の数を表す．

① すべての行 i で S と α_i の共通部分の数を数える．

② すべての行 i で，w_i が 0 となる行 i を数える．

③ U の中のそれぞれの行 i（行 i が増加していく順に）に対して，
(a) U と S の共通部分でコストが最小な最初の列を見つける．
(b) S に(a)で見つけた列の部分を加えて，β_j のすべての行 i で，
$w_i := w_i + 1$，$U := U - \beta_j$ とする．

④ S の中のそれぞれの列 j（列 j が減少していく順に）で
もし，β_j の中のすべての行 i で，$w_i \ge 2$ なら，$S := S - j$，$w_i := w_i - 1$ とする．

①では，すべての行 i で解 x と α_i が共に 1 であるビットの数をそれぞれ数える．②では，w_i=0 となる行を数える．③では，被覆されていない行 i に対して(列 j の順番が増えていく順に)(a) 列のコストが最小な α_i の中で一番始めの $x_j=0$ となっているビットを見つける．(b) $x_j=0$ のビットを $x_j=1$ とすることによって β_j が所属する行の w_i を 1 増加させ，被覆されていない行は β_j 個減少することになる．この操作によってすべての行 i で $w_i \geq 1$ となり実行可能性が保証される． ④では，列 j(j が増加する順に)に対して，もし，β_j が所属する行で w_i が 2 個以上あるならば，$x_j=1$ となっているビットを $x_j=0$ として，β_j が所属する行の w_i を 1 つ減少させる．この操作によって，重なって被覆されている行が減少するので，よりコストの低い解を算出し易くなる．

5. 遺伝的局所探索法

一般に GA は，交叉，突然変異，選択からなる遺伝的操作を個体群に対して施し，この操作を数世代にわたり繰り返すことで，優れた集団を形成する大域的な探索手法である．この手法に局所探索法(LS)の処理を加えることにより，大域的かつ局所的な探索法が実現される．これは，遺伝的局所探索法(Genetic Local Search,GLS)とよばれ，多くの場合，大域的処理である．GA のみの性能を大幅に改善できることが知られている．一般に GLS では，交叉，または突然変異後に生成される個体に対して LS の処理を加えるので，集団内の全個体は局所的に最適化された解（局所最適解）となる．従って，GA では，大きな範囲に及ぶ空間を探索対象するのに対し，GLS では，局所最適解で構成される個体群に遺伝的操作が行われるので，探索対象とする空間領域は格段に減少する．以下では，SCP に対する GLS 適用の流れとそれぞれの操作について示す．

5.1 SCP に対する GLS の適用の流れ

① SCP の解 x の表現方法は，0-1 のビット表現とし，長さは問題のサイズ n とする．
② PS 個の個体をランダムに生成.
③ 交叉率 Pc に従い，一様交叉を行う.
④ ③で生成された子に対して，突然変異率 Pm に従い，突然変異を実施する.
⑤ 交叉，突然変異後の個体は実行不可能解である可能性が高いので，実行可能性を保証するためにここでも，LS を実施する.
⑥ 選択では次世代に残す個体群を親と子からコストの低い順に PS 個選ぶ.
⑦ ③から⑥までの遺伝的操作を終了条件を満たすまで繰り返す.

5.2 操作の概要

①では，SCP に対する解 x のコーディング(0-1 ビット表現)を決定する．②では，乱数を用いて，個体を PS 個生成する．③で用いている一様交叉について以下に説明する．

・一様交叉

一様交叉とは，ペアリングする親の持つ共通のビットはそのまま次世代の子に残し，それ以外のビットにランダムに 0,1 を入れる．以下に，一様交叉の一例を示す．

親 1　1001010010111001

子　1100010010100001

親 2　1110001110101100

④で用いている突然変異について以下に説明する.

・突然変異

両親の形質を受け継いだビットから，ランダムに 1 つ選び出しそのビットの値を反転させる．以下に，突然変異の一例を示す.

子　1100010010100001 ──▶ 突然変異後　1100010010000001

⑥の選択では，コストの高い個体は子の個数だけ淘汰される.

6. 実験結果

本実験では, SCP に対して GLS を適用した際に算出された解がパラメータの設定によってどのような影響を及ぼすかに関して分析する．上述した GLS のパラメータ（集団数 PS，交叉率 Pc，突然変異率 Pm）の影響を調べるために，いくつかの良く知られたベンチマーク問題を利用し，実験を実施する．各パラメータ値は，PS={10，50，200，500}，Pc={0.2，0.4，0.6，0.8，1.0}，Pm={0.2，0.4，0.6，0.8，1.0}とし，これらのパラメータ値の組合わせ（100 通りの実施が可能）ごとに，5 回の試行を実施する．GLS の終了条件は計算打ち切り時間 30 秒とする．ここでは，OR-Library から scp4.1(200 × 1000, Density 2%)(m のサイズ × n のサイズ)及び scp4.2(200 × 1000, Density 2%), scp5.1(200 × 2000, Density 2%), scp5.2(200 × 2000, Density 2%), scp6.1(200 × 1000, Density 5%), scp6.2(200 × 1000, Density 5%)の 6 つの問題例に対してその結果を示す.

図 1~図 6 は，各パラメータの組合せにおける 5 回の試行で算出されたコストの平均値の推移を表している．図中の x 軸は上で述べた１００通りの組合せを，PS=10 の区間，PS=50 の区間，PS=200 の区間，PS=500 の区間の 4 区間で表している．それぞれの区間の組合せに対する対応番号を 1~25, 26~50, 51~75，76~100 とする．また，それぞれの区間では，Pc と Pm の組合せを 25 通り表しており，表記の数字はその組合せのパターンを表している．以下にそのパターンがどの Pc と Pm の組合せに対応しているかを表 1 に示す（パターンは 25 周期で PS=10 から PS=50,200,500 と移っていく）．図中の y 軸はコストを表している.

表 1　パラメータ値の組合せとその対応番号

Pc \ Pm	0.2	0.4	0.6	0.8	1.0
0.2	1,26,51,76	2,27,52,77	3,28,53,78	4,29,54,79	5,30,55,80
0.4	6,31,56,81	7,32,57,82	8,33,58,83	9,34,59,84	10,35,60,85
0.6	11,36,61,86	12,37,62,87	13,38,63,88	14,39,64,89	15,40,65,90
0.8	16,41,66,91	17,42,67,92	18,43,68,93	19,44,69,94	20,45,70,95
1.0	21,46,71,96	22,47,72,97	23,48,73,98	24,49,74,99	25,50,75,100

図 1 は，scp4.1(200 × 1000, Density 2%)で，PS=50 の区間が比較的良好な解を算出している事が観測された．中でも，組合せが 58,61 の時に良好な解が算出されていることが観測され，この時の(Pc, Pm)は，それぞれ(0.6,0.2)，(0.4,0.6)であった．図 2 は，scp4.2(200 × 1000, Density 2%)で，PS が大きくなるにつれ，良好な解が算出できているのが観測でき、特に組合せが 97 の時に最も良好な解が算出されており，その時の(Pc, Pm)は(1.0,0.4)であった．図 3 は，scp5.1(200 × 2000, Density 2%)で，PS=50,200 のとき似たような値のコストを同じ回数算出しているが，PS=500 の時は，比較的良好な

解を算出しており，中でも 87,91,100 の時に良好な解が算出されていることが観測され，その時の(Pc, Pm)はそれぞれ，(0.6,0.4)，(0.8,0.2)，(1.0,1.0)であった．図 4 は，scp5.2(200×2000, Density 2%)で，PS=50,200 のときは似たような値のコストを算出しているが，その算出回数は PS=200 のときの方が比較的多く算出していることが観測された．最も良好なコストを算出したのは 92 であり，その時の(Pc, Pm)は(0.8,0.4)であった．図 5 は，scp6.1(200×1000, Density 5%)で，PS=200 から PS=500 へ移った後に，他の場合と比較して，より良好なコストを算出していることが観測された．図 6 は，scp6.2(200×1000, Density 5%)で，この場合も scp6.2 と同じで，PS=200 から PS=500 に移った後，より良好なコストを算出している．また，PS=500 以外の組合せで算出されたすべての解のコストよりも，良好なコストが，PS=500 のパラメータの組合せすべてで算出されていることが観測された．

図 1~図 6 に示したように，SCP に対する GLS 適用の際のパラメータの組合せによる結果から，Pc 及び Pm の設定値よりも，PS の設定値が大きく影響することが観測された．これは，PS が 500 の時と，PS が小さい場合と比較して，良好な解を多く算出可能であった観測によるものである．また，Pc が小さい場合，Pm を大きくした方が，比較的良好な解を算出し易い傾向が多くあることが観測された．

7. むすび

本論文では，集合被覆問題(SCP)に対して遺伝的局所探索法(GLS)のパラメータの影響について検討した．GLS を組合せ最適化問題に適用した際のパラメータ値設定は容易でないとされている．そこで，GLS のパラメータ（集団数 PS，交叉率 Pc，突然変異率 Pm）がどのような影響を及ぼすのかを調べるために，いくつかの良く知られたベンチマーク問題を利用し，実験を実施した．各パラメータ値は，PS={10，50，200，500}，Pc={0.2，0.4，0.6，0.8，1.0}，Pm={0.2，0.4，0.6，0.8，1.0}とし，各パラメータ値の組合わせ（100 通りの実施が可能）にもとづき，その各組合せに対して，5 回の試行を実施し，コストの平均値の推移を観測した．その結果，Pc 及び Pm の設定値よりも，PS の設定値が大きく影響することが観測された．また，PS，Pc が小さい場合（生成される個体が少ない）は，良好な解が算出されにくい傾向が多く観測されたことから，探索領域がせばまっていると推測される．良好なコストを有する解を算出するためには，探索領域を広げることが重要であることが観測された．最後に，本研究でのパラメータ値設定の結果を今後の SCP に対する GLS の研究に活かす予定である．

8. 参考文献

[1] J.E.Beasley and P.C.Chu, “A genetic algorithm for the set covering problem,” European Journal of Operational Research 94 (1996)392-404.

[2] C. R. Reeves, モダンヒューリスティックス，日刊工業新聞社，(1997).

[3] 久保幹雄，“メタヒューリスティックス，” 離散構造とアルゴリズムⅣ，近代科学社，pp.171-230，(1995).

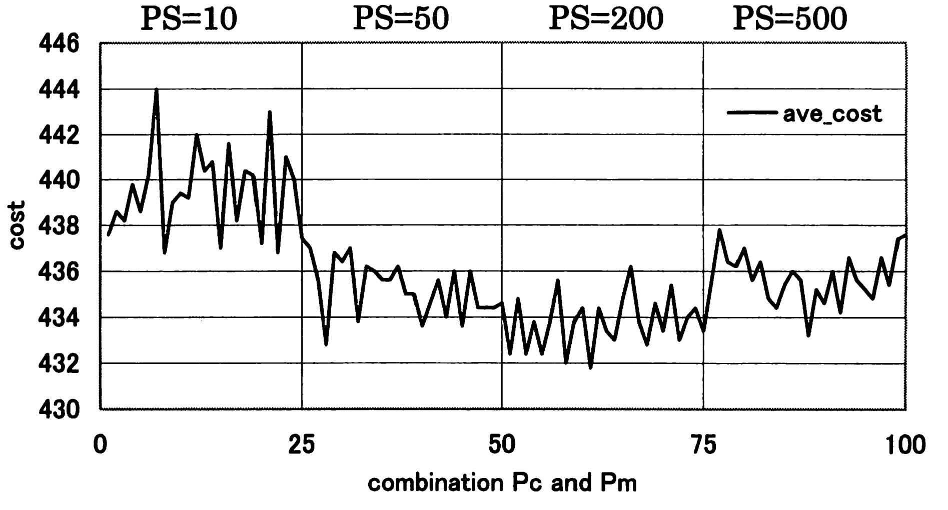

図 1 scp4.1(200 × 1000, Density 2%)

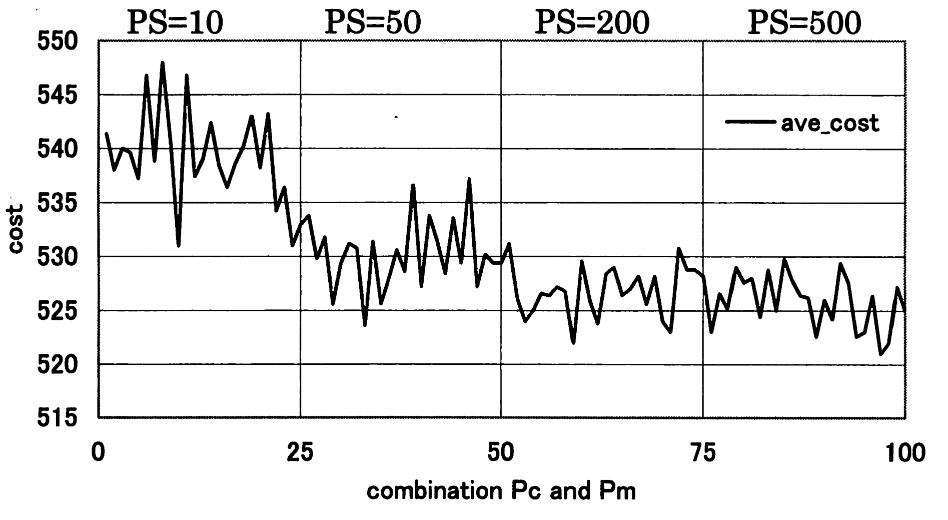

図 2 scp4.2(200 × 1000, Density 2%)

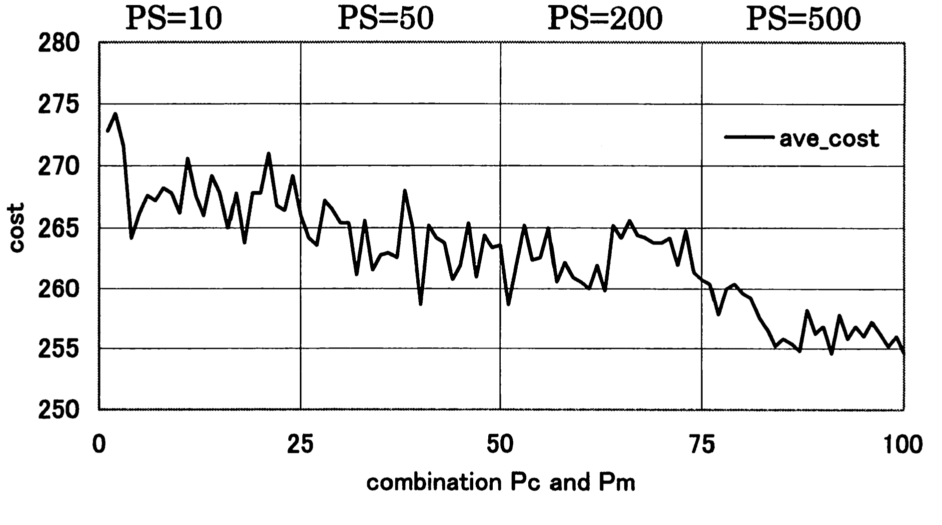

図 3 scp5.1(200 × 2000, Density 2%)

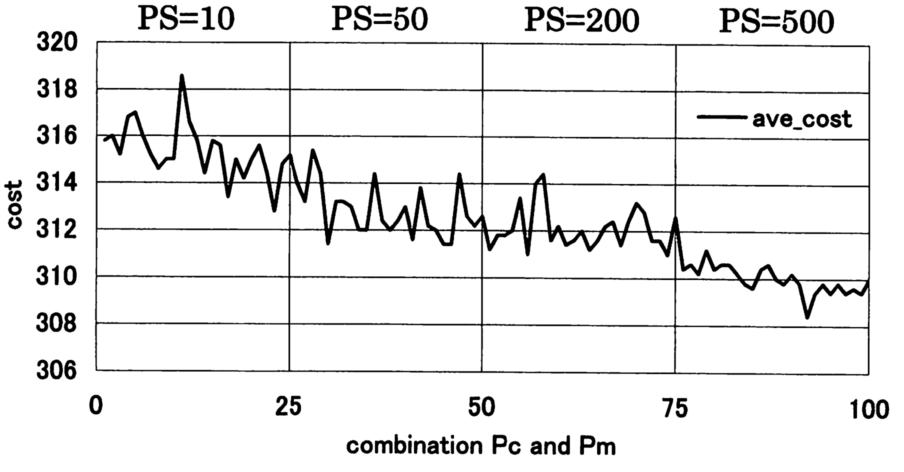

図 4　scp5.2(200 × 2000, Density 2%)

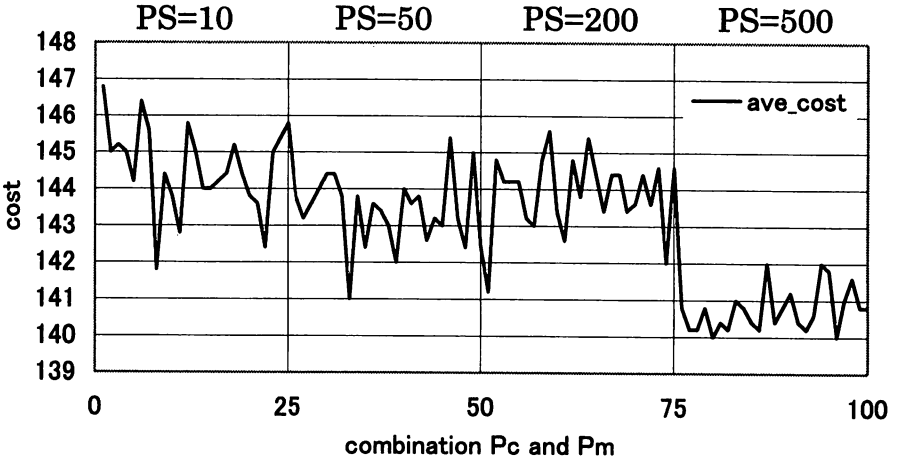

図 5　scp6.1(200 × 1000, Density 5%)

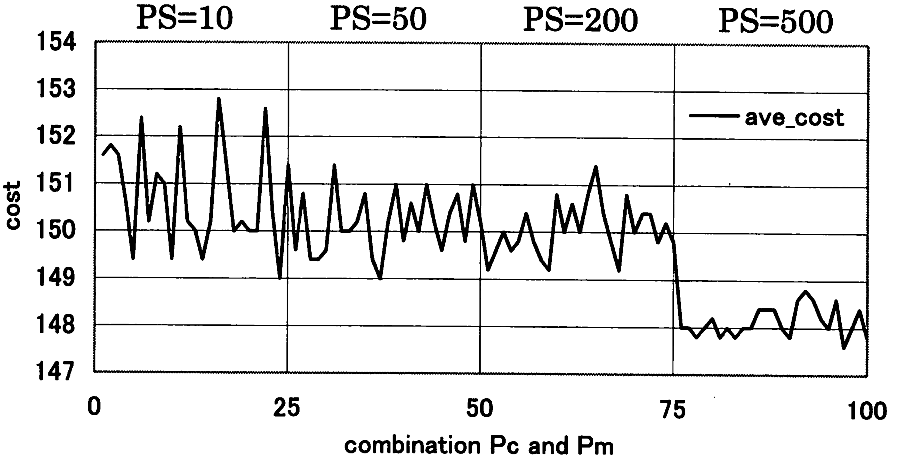

図 6　scp6.2(200 × 1000, Density 5%)

Analysis of Influences of Parameter Values for the Genetic Local Search in the Set-Covering Problem

Yoichi ISHIDA, Kengo KATAYAMA* and Hiroyuki NARIHISA*

Graduate school of Engineering

**Department of Information and Computer Engineering*

Faculty of Engineering

Okayama University of Science,

Ridai-cho 1-1, Okayama 700-0005, Japan

(Received November 1, 2000)

A genetic local search (GLS) is known to be one of the most powerful heuristic algorithms. However, it is considered that the parameter settings of the GLS such as population size, crossover rate, and mutation rate, are one of the major issues in finding good final solutions within reasonable computation times. In this paper, we consider the set-covering problem which is one of the combinatorial optimization problems. For the SCP, we analyze the influences of different parameter settings of the GLS, by conducting extensive experiments on several benchmark problem instances, for which optimal solutions are known. Results show that in obtaining good solutions, the population size is more important than the other parameters.