岡山理科大学紀要　第 37 号 A pp.137-143(2001)

集合被覆問題に対する効率的な遺伝的オペレータ

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課程情報工学専攻
*岡山理科大学工学部情報工学科

(2001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．

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，SCP の探索空間の地形解析にもとづいた遺伝的交叉オペレータを GLS に組込んだ場合と，標準的な遺伝的オペレータを GLS に組込んだものを比較検討する． 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．本研究で示す効率的な遺伝的交叉オペレータは，一様交叉に基づいており，地形解析によって得られる局所最適解と真との最適解のハミング距離を利用したものである．この観測された特徴にもとづいた遺伝的交叉オペレータと，標準的な遺伝的オペレータを GLS に組込んで，よく知られた SCP のベンチマーク上で，比較検討した．その結果，単純に両親の形質を継承させる一様交叉よりも，対象となる問題の地形解析にもとづいた交叉の方がより強力な性能をもつことを示す．

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].

$$N : F \rightarrow 2^F$$

実行可能解 $x \in F$ で，$f(x) \leq f(x')$，$\forall x' \in N(x)$ 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，$w_i \geq 2$ 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．

6．地形解析

本研究では，地形解析を行うことによって SCP の探索空間の構造を解析する．ランダム解を初期解とする LS[1][4]を 2 千回試行することによって，与えられた初期解は，局所最適解となる．得られた局所最適解と真の最適解との関係を調べる．図 1 〜3 は，地形解析によって得られた 2 千個の局所最適解で，横軸に真の最適解とのハミング距離，縦軸に真の最適解とのコスト差をとったものである．図 1 〜図 3 において LS 後の局所最適解は，真の最適解に向ってクラスタとなり分布していることが観測された．

7．効率的な遺伝的交叉オペレータ

本研究で示す効率的な遺伝的交叉オペレータは，一様交叉にもとづいている．一様交叉とは，両親の各ビットにおいて共通する遺伝子はそのまま継承させ，異なる場合は，ランダムに 0 か 1 を選び子の遺伝子とする操作である．まず，$e_j = c_j / \sum_{i=1}^{m} a_{ij}$ を計算する．この e_j の値は，多くの行を被覆することができ，比較的コストの良い解のビットの位置を模索する基準値となると考える．本研究での交叉は，地形解析によって得られた真の最適解と LS 後の局所最適解との間に観測されたハミング距離分，両親の各ビットの遺伝子が共通して 0 である場合，対応する子のビットを 1 にする．この時，親と子のハミング距離が観測されたハミング距離を満たすまで，e_j の値が小さい順に，それに対応する子のビットを 1 とし，子を生成する．その他のビットに対する操作は，標準の一様交叉と同様である．いかに効率的な交叉を用いた GLS の流れについて示す．

① PS 個の個体をランダムに生成しその個々に LS を施す．
② LS 後の局所最適解に対して地形解析や e_j にもとづいた交叉を行う．
③ 交叉によって生成された子に対して LS を施す．

④ 次世代に残す個体群を親と子からコストの良い順に PS 個選び出す.
⑤ ②から④までの操作を終了条件を満たすまで繰り返す.

8. 実験結果

本実験では，SCP に対して効率的な交叉法を用いた GLS と，一様交叉を用いた GLS を比較検討するために，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 scp5.1(200 × 2000, Density 2%)，scp6.1(200 × 1000, Density 5%)，scpa.1(300 × 3000, Density 2%)，scpb.1(300 × 3000, Density 5%)，scpc.1(400 × 4000, Density 2%)，scpd.1(400 × 4000, Density 5%)，scpnre.1(500 × 5000, Density 10%)，scpnrf.1(500 × 5000, Density 20%)，scpnrg.1(1000 × 10000, Density 2%)，scpnrh.1(1000 × 10000, Density 5%)，の 11 個の問題例に対して GLS を施した結果を表 1 に示す．各パラメータ値は，集団数 PS=50，交叉率 Pc=1.0，突然変異率 Pm=0 とした．また，各問題例における GLS の終了条件は世代数 500 とし，1 回の試行における LS の実施回数は 25000 回とした．

問題	Dens %	opt	一様交叉		効率一様交叉		ハミング距離
			min	ave	min	ave	hd
Scp4.1	2	429	431	432.8	430	432.8	50
Scp5.1	2	253	258	263.4	253	258.0	50
Scp6.1	5	138	140	143.0	138	141.7	30
Scpa.1	2	253	255	256.7	256	256.7	50
Scpb.1	5	69	69	70.1	69	69.9	30
Scpc.1	2	227	231	232.5	229	230.4	50
Scpd.1	5	60	60	60.2	60	60.0	20
Scpnre.1	10	29	29	29.0	29	29.0	10
Scpnrf.1	20	14	14	14.4	14	14.2	10
Scpnrg.1	2	176	181	181.7	180	181.7	30
Scpnrh.1	5	63	65	66.6	64	65.8	20

表 1 で，Dens は，m行 × n列の行列 a_{ij} に 1 が存在する割合を示す．Opt は既知の最良解を表している．hd は，地形解析で観測されたハミング距離を参考に定めた．Min，ave は，各問題例に対して効率的な交叉オペレータを用いた GLS，一様交叉を用いた GLS をそれぞれ 10 回試行して算出された最小値，平均である．結果から，一様交叉よりも地形解析にもとづいた交叉の方が，強力であることが観測された．これは，LS 後の局所最適解と真の最適解とのハミング距離や e_j を考慮にいれた交叉が標準的な一様交叉に比べ，真の最適解に近づくように，より多くの行を被覆でき，比較的コストの良い解のビットを 1 にした結果だと推測される．

9. むすび

本論文では，SCP の探索空間の地形解析を行うことによって観測された特徴にもとづいた遺伝的交叉オペレータと，標準的な遺伝的オペレータの比較検討をした．単純に両親の形質を継承させる一様交叉よりも，対象となる問題の地形解析にもとづいた交叉の方が，強力な性能を示した．

10. 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, pp.392-404, (1996).

[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).

[4] 石田洋一, 片山謙吾, 成久洋之, "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," 電気・情報関連学会中国支部第51回講演論文集, p.74, (2000).

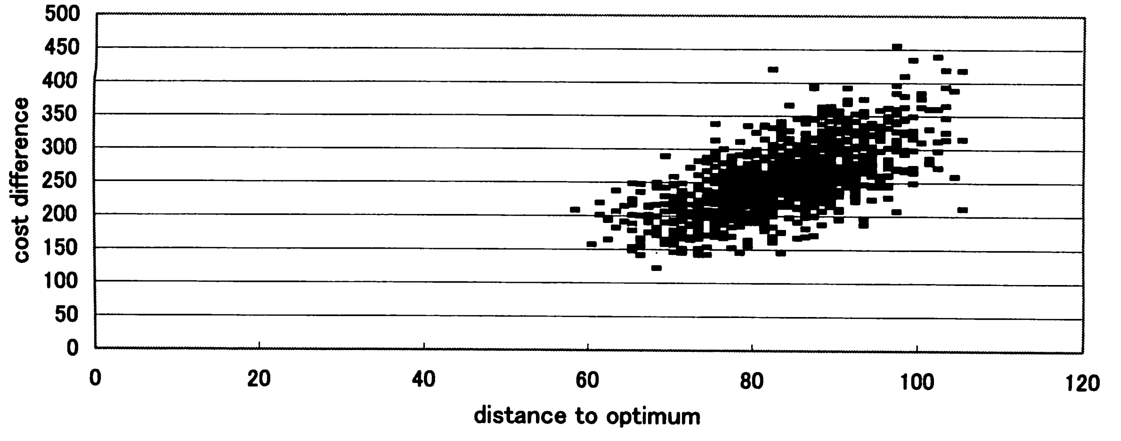

図 1 scp5.1(200×2000,Density 2%)

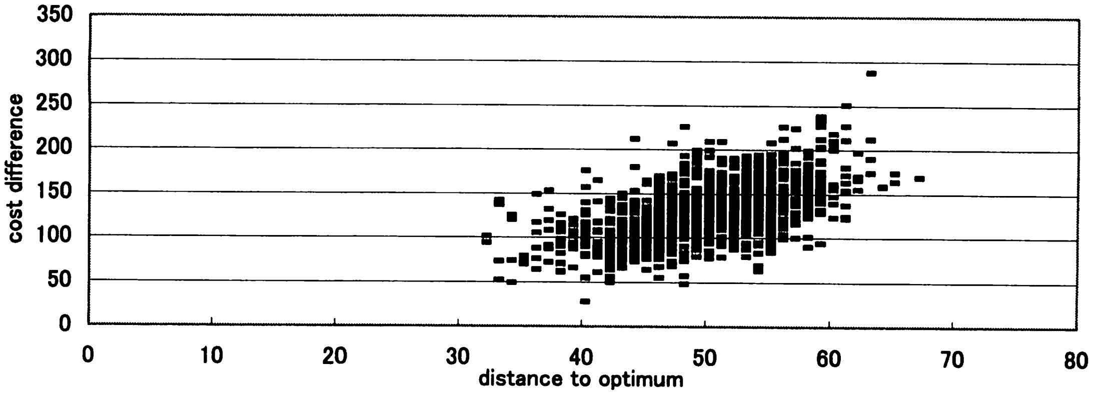

図 2 scp6.1(200×1000, Density 5%)

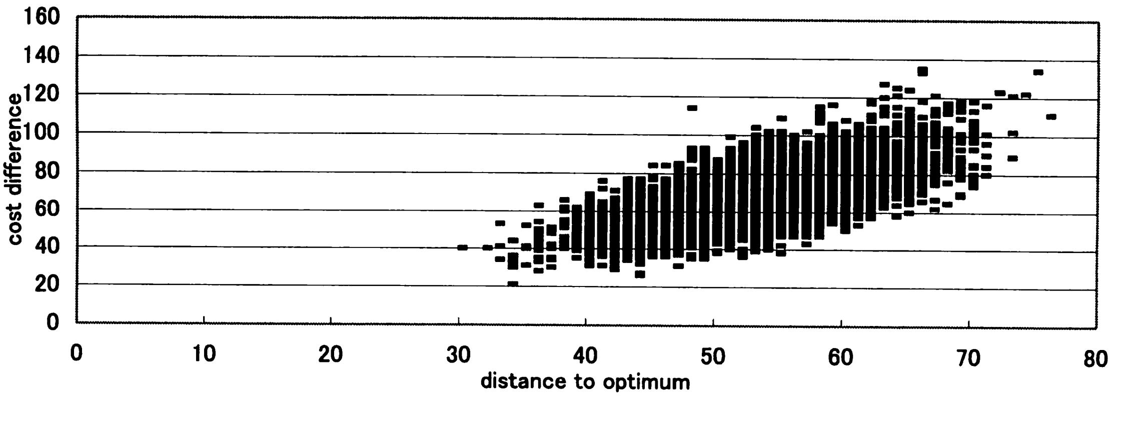

図 3 scpb.1(300×3000, Density 5%)

Efficient Genetic Operator 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, 2001)

A genetic local search (GLS) is known to be one of the most powerful heuristic algorithms. In this paper, we consider the set-covering problem which is one of the combinatorial optimization problems. We compare genetic crossover operator based on the geographical feature analysis of the search space of SCP with standard genetic crossover operator on the benchmark of SCP known well by applying to GLS. Efficient genetic crossover operator shown by this research is based on uniform crossover, and uses the humming distance of the local optimum solution and the true optimum solution which are obtained by geographical feature analysis. we show that the way of crossover based on geographical feature analysis of the target problem has a more powerful performance rather than uniform crossover which makes parents' character inherit simply.