F-003

最大多様性問題に対する個体間距離に基づく適応的交叉確率を用いた遺伝的局所探索法

Genetic Local Search with Crossover Probability Based on the Distance between Individuals and its Application to Maximum Diversity Problem

浅田 宏樹 † 外山 史 † 東海林 健二 † 宮道 壽一 †
Hiroki Asada Fubito Toyama Kenji Shoji Juichi Miyamichi

1 はじめに

最大多様性問題 (MDP) とは、与えられた n 個の要素から m 個の要素を選ぶとき、できるだけ多様性を有するように要素を選定する問題であり、VLSI 設計、マーケット計画やポートフォリオ選択など数多くの応用例が存在する。この問題に対し、遺伝的局所探索法 (GLS) を用いた手法 [1] が提案されている。GLS とは、遺伝的操作によって大域的な探索を行う遺伝的アルゴリズム (GA)[2] に局所探索 (LS) の処理を取り入れた探索手法である。文献 [1] の GLS では、親個体のペアリングはランダムに行われるため、個体間距離の小さい親個体が選択された場合、生成された子個体に LS を適用すると親と同じ個体に収束する可能性が高く、探索効率が低下してしまう。また、問題によっては交叉に適した個体間の距離が存在する場合があり、このような距離を持つ個体を交叉させることで効率的に解を探索できる場合がある。

そこで本研究では、探索途中の交叉や突然変異によって新たに得られた解とその親の解を元に、交叉時のペアリングに最適な個体間距離を動的に推測する GLS を提案し、MDP に適用する。実験では、従来の GLS[1] と提案する GLS の比較を行い、提案手法の有効性を確かめる。

2 最大多様性問題 (MDP)

MDP とは、n 個の要素および各要素間の距離を表す対称距離行列 d が与えられたとき、できるだけ距離に基づく多様性を有するように m 個の要素を選定する問題である。ここで、d_{ij} を対称距離行列 i 行 j 列の値、すなわち、要素 i と要素 j の距離値を表し、$d_{ij} = d_{ji}$、$d_{ii} = 0$ とする。MDP は、以下に示す目的関数が最大となるような解 x を求める問題である。

$$\text{maximize} \qquad f(x) = \sum_{i=1}^{n}\sum_{j=1}^{n} d_{ij}x_ix_j, \qquad (1)$$

$$\text{subject to} \quad \sum_{i=1}^{n} x_i = m, \quad x_i \in \{0,1\}, \forall i = 1, ..., n.$$

3 MDP に対する GLS

MDP に対して、GLS を用いた手法 [1] が提案されている。この手法で用いられている GLS には高性能な k-flip 局所探索法と呼ばれる局所探索法が使われており、MDP に対して有効に働くことが示されている。GLS の流れは、(1)PS 個の初期集団を生成、(2) 初期集団の各個体に LS を適用、(3) ランダムに選択した 2 つの親個体に対する一様交叉、(4) 生成された子個体に対して LS を適用、(5) 適応度の高い順に PS 個の個体を次世代の集団として選択、(6) 再スタート戦略 (30 世代最良解の更新が行われなかったときのみ実施)、(7) 終了条件を満たさない場合は (3) に戻る、という処理を繰り返す。ここで交叉などの遺伝的操作によって生成された個体は、実行可能性が保証されていないため、実行可能解の判別・変換を行う操作 (リペア法) を適用している。GLS を用いた手法の詳細は文献 [1] を参照されたい。

4 提案手法

本研究で提案する GLS は、従来の GLS に個体間距離を考慮したペアリングを行うという操作を加えたものである。これにより、個体間距離の小さい交叉によって子孫個体が親と同じ解に収束するのを防ぐだけでなく、その問題に対して探索に最適な個体間距離を動的に推測することができる。提案手法と従来手法とのアルゴリズムの違いは、交叉処理において提案手法では交叉確率テーブルに基づく交叉・突然変異を行う部分と、交叉後の子孫に LS を適用した後に、交叉確率テーブルの更新処理が追加されている点である。以下、具体的な交叉確率テーブルを用いた交叉・突然変異処理について説明する。

4.1 個体間の距離による交叉確率テーブルの定義

$P_{cross}[h](h = 1, 2, \cdots, 2m)$ を個体間の距離による交叉確率テーブルとして定義する。$P_{cross}[h]$ には個体間の距離 h に対する交叉確率が格納されている。ここで、h は個体間のハミング距離を表している。交叉において 2 つの親個体が選択された際、親個体間の距離 (ハミング距離) を計算し、その距離 h に対する交叉確率を交叉確率テーブル $P_{cross}[h]$ から参照して交叉を行うかどうかを決定する。この交叉確率テーブルは、探索途中に得られた交叉・突然変異の結果によって動的に変動する。なお、確率の初期値はすべて等しい値 P_0 が与えられる。

4.2 交叉確率テーブルの更新

交叉確率テーブル P_{cross} の更新は交叉・突然変異を行い局所探索法を適用した後に実施される。更新方法は親個体と子個体の関係により 4 通りに分類される。p_1, p_2 を親個体、c をその 2 つの個体から生成された子個体に局所探索を施した解とし、$f(p_1)$, $f(p_2)$, $f(c)$ をそれぞれの評価値とする。また α, β, γ を確率を変動させるパラメータとする。以下に交叉確率テーブルの更新処理を示す。ここで pd は、p_1 と p_2 とのハミング距離である。

Case1 : $p_1 = c$ または $p_2 = c$ の場合

† 宇都宮大学 工学研究科,
Graduate School of Engineering,Utsunomiya University

表 1: 従来手法による GLS の結果

MDP instance	n	m	best-known	avg.	(quality %)	b/30	t1/s	(gns)	t2/s	(gns)
mdp1000	1000	100	299730	299730.0	(0)	30/30	116.7	(190)	-	(-)
	1000	200	1125264	1125264.0	(0)	30/30	112.3	(74)	-	(-)
	1000	300	2458316	2458316.0	(0)	30/30	93.6	(70)	-	(-)
	1000	400	4292438	2492438.0	(0)	30/30	20.2	(8)	-	(-)
mdp2500	2500	250	1775366	1774634.4	(0.000412)	0/30	-	(-)	3000	(515)
	2500	500	6794586	6793997.8	(0.000087)	1/30	629.0	(56)	3000	(273)
	2500	750	14988436	14987905.0	(0.000035)	3/30	1095.3	(62)	3000	(193)
	2500	1000	26332276	26332153.8	(0.000005)	20/30	1159.0	(85)	3000	(244)

表 2: 提案手法による GLS の結果

MDP instance	n	m	best-known	avg.	(quality %)	b/30	t1/s	(gns)	t2/s	(gns)
mdp1000	1000	100	299730	299730.0	(0)	30/30	106.5	(146)	-	(-)
	1000	200	1125264	1125264.0	(0)	30/30	108.4	(74)	-	(-)
	1000	300	2458316	2458316.0	(0)	30/30	108.0	(51)	-	(-)
	1000	400	4292438	2492438.0	(0)	30/30	28.3	(7)	-	(-)
mdp2500	2500	250	1775366	1774711.9	(0.000368)	8/30	1527.8	(223)	3000	(467)
	2500	500	6794586	6794000.9	(0.000086)	4/30	1052.5	(106)	3000	(398)
	2500	750	14988436	14988076.7	(0.000024)	10/30	991.7	(84)	3000	(291)
	2500	1000	26332276	26332186.5	(0.000003)	22/30	1324.4	(96)	3000	(227)

$h_i \leq pd$ に対して以下の式を適用。

$$P_{cross}[h_i] \leftarrow P_{cross}[h_i] \times (1-\alpha)$$

Case2：$f(p_1) > f(c)$ かつ $f(p_2) > f(c)$ の場合
全ての h_i に対して以下の式を適用。

$$P_{cross}[h_i] \leftarrow P_{cross}[h_i] \times (1-\beta e^{-\frac{(h_i-pd)^2}{2\sigma^2}})$$

Case3：$f(p_1) < f(c)$ かつ $f(p_2) < f(c)$ の場合
全ての h_i に対して以下の式を適用。

$$P_{cross}[h_i] \leftarrow P_{cross}[h_i] \times (1+\gamma e^{-\frac{(h_i-pd)^2}{2\sigma^2}})$$

Case4：Case1、Case2、Case3 以外の場合
確率テーブルの更新は行わない。

上記のように、生成された子個体が親と同じ解に収束した場合や、親個体よりも評価値が悪くなったときは、交叉を行った親個体間の距離が適切ではないと判断し、その距離周辺での交叉確率を下げる。一方、生成された子個体が親個体より評価値が良くなったときは、交叉を行った親個体間の距離が適切であると判断し、その距離周辺での交叉確率を上げる。以上のように交叉確率テーブルを変化させることで、親個体のペアリングにおける最適な個体間の距離を推測し、効率的な探索を行うことができると考えられる。

4.3 交叉確率テーブルを用いた交叉と突然変異

本 GLS において、交叉・突然変異は交叉確率テーブル P_{cross} に従って行われる。親個体を p_1, p_2、親の個体間の距離を pd とする。交叉を行う際、交叉確率 $P_{cross}[pd]$ を参照して交叉を行う。交叉が行われなかった場合、確率 P_m で突然変異を行う。突然変異は、P_{cross} の中から確率に応じたルーレットで選択したサイズ (距離) に基づいてビット値の反転を行う。突然変異が行われなかった場合は、親を再度選択し交叉処理をやり直す。交叉または突然変異により生成された子孫に LS を施した後、前述したように交叉確率テーブルが更新される。

5 実験

提案手法の有効性を示すために、従来の GLS と交叉時に個体間距離に基づいた交叉確率を導入した本 GLS をそれぞれ MDP の問題例に対して適用し、比較を行った。交叉確率テーブルの効果が現れるのは、ある程度世代が進み確率テーブルが更新されてからであるため、実験では多くの世代数が必要とされる問題サイズの大きい MDP を用いて比較を行う。MDP の問題例は、公開されているデータベース [3] の中から、問題サイズの大きい n が 1000, 2500 の問題群 (mdp1000, mdp2500) の 8 つの問題を使用した。実験は両手法とも同じ計算機上 (CPU：Intel Xeon 2.4GHz, メモリ：1GB) で行った。両手法において、個体数 PS は 40、終了条件は計算時間が 3000 秒に達するまでとした。提案手法のパラメータは $P_0 = 0.8$、$\alpha = 0.10$、$\beta = 0.03$、$\gamma = 0.50$、$P_m = 0.03$、$\sigma = 5.0$ とした。

表 1, 表 2 に、各問題に対し従来手法および提案手法をそれぞれ 30 回適用した結果を示す。表は、左から順に、問題群の名前、問題サイズ n、選択要素数 m、これまでに報告されている既知の最良解値 Best-known、1 回の試行で得られた最良解の平均の解値 Avg とその解質 (解質 (%)=100×(既知の最良解値/得られた解の値)-1.0)、30 回の試行において既知の最良解最を算出した回数 b/30、既知の最良解を得るまでに要した平均時間 t1/s(秒) と対応する平均世代数 gns、1 試行あたりの計算打ち切り時間 t2/s と対応する平均世代数 gns を示している。t1/s では、30 回の試行において既知の最良解を算出できなかった場合、t2/s では、打ち切り時間内に既知の最良解を全て算出した場合に"-" と記した。

表 1 および表 2 の結果より、1000 変数の問題に対しては、両手法共に試行 30 回中 30 回とも同じ最良解を算出しており、最良解を得るまでに要した時間にも大きな差が見られない。このことから、同等の結果が得られていると言える。一方、2500 変数の問題では、全ての問題において最良解の平均値 Avg およびその解質、既知の最良解を見つけた回数 b/30 の値が従来手法よりも提案手法の方が良い結果が得られている。以上より、サイズの大きい問題に対して提案手法が従来手法よりも有効に働いていることが確認できた。

6 おわりに

本研究では、MDP に対する個体間の距離に応じた交叉確率を用いた GLS を提案した。従来 GLS と比較した結果、提案する GLS の有効性を確かめることができた。

参考文献

[1] 片山謙吾, 成久洋之: "大規模な最大多様性問題に対する遺伝的局所探索", 情処学論 Vol.45, No.SIG2(TOM 10), pp.99-109, 2004.

[2] 北野宏明: "遺伝的アルゴリズム", 産業図書, 1993.

[3] Maximum Diversity Problem (MDP) instances, http://k2x.ice.ous.ac.jp/ katayama/bench/mdp /index.html