

# 木構造最適化のための多段階探索交叉における 確率モデルに基づく近傍生成法の提案

松村 康平<sup>1</sup> 花田 良子<sup>2</sup> 小野 景子<sup>3</sup> 村中 徳明<sup>2</sup>

**概要:** 多段階探索交叉は局所探索に基づく交叉法の一つで、木構造の最適化問題において優れた探索性能を示している。ただし、ランダムに近傍解候補を生成するため、良好な解を効率的に生成することが困難である。ここでは、探索過程で生成される個体群の情報からノードの出現に関する確率モデルを構築し、それに基づき近傍解を生成することで、よい形質の保存および形成を促進させる。本研究では、単純な意思決定問題の一種である Santa Fe trail 問題を対象として、確率モデルによる性能を評価する。

## 1. はじめに

解が木構造で表現される遺伝的プログラミング (Genetic Programming: GP) においては、多くの組合せ最適化問題と同様に、両親の形質を受け継ぐような問題固有の構造、性質を利用した交叉の設計が重要である。多段階探索交叉 deterministic-Multi-Step Crossover Fusion(dMSXF)[1] は局所探索のメカニズムを利用した形質遺伝に優れた交叉である。これまでに我々は GP において、両親間に共通するグラフパターンを保存すべき形質ととらえ、それに基づく近傍と距離を導入した dMSXF を提案し、高い探索性能を有することを示した [2]。dMSXF の局所探索において近傍解を生成する際には、問題特有の制約条件を満たすために、しばしばランダムに発生させたノードを解に置換、挿入している。そのため、効率的に良好な近傍解を生成することが困難であると考えられる。本研究では、探索過程で生成される個体群の情報から、ノードの出現に関する確率モデルを構築する。それに基づきノードを発生させることで、近傍探索においてよい形質の保存および形成を促進させる。

## 2. 木構造における多段階探索交叉

dMSXF[1] は、親  $p_1$  から親  $p_2$  に向けて局所探索を行うことで、両親の形質の受け継ぎ方が多様な子個体群を生成

する。dMSXF のアルゴリズムを以下に示す。親  $p_1, p_2$  から生成される子個体群を  $C(p_1, p_2)$  と表す。

### 【dMSXF のアルゴリズム】

1.  $p_1, p_2$  を両親、その子個体群  $C(p_1, p_2) = \phi$  とする。
2. 探索初期点  $x_1 = p_1$ ,  $k=1$  とし、 $x_1$  を  $C(p_1, p_2)$  の要素として加える。
3. ステップ  $k$  における探索点  $x_k$  の近傍解を  $\mu$  個生成し、その集合を  $N(x_k)$  とする。ただし、 $N(x_k)$  のすべての近傍解  $y_i (0 < i < \mu)$  はかならず  $d(y_i, p_2) < d(x_k, p_2)$  を満たさなければならない。
4.  $N(x_k)$  の中で最も良い解  $y$  を選択する。 $x_{k+1} = y$  とし、 $x_{k+1}$  を  $C(p_1, p_2)$  の要素として加える。
5.  $k = k + 1$  とし、 $k = k_{max}$  あるいは  $x_k$  が  $p_2$  に等しくなれば終了。そうでなければ、3 にもどる。

## 3. 木構造における距離と近傍

### 3.1 距離の定義

木  $T_a, T_b$  それぞれの最大共通部分木について、対応する遺伝子座のノードと記号が異なるノードの集合を  $U_a, U_b$  とする。また、それぞれの木において、最大共通部分木に含まれないノードの集合をそれぞれ  $D_a, D_b$  とする。なお、最大共通部分木は、複数のグラフにおける共通の連結部分グラフのうちノード数が最大のものをいう。

木  $T_a, T_b$  の距離  $d(T_a, T_b)$  を次のように定義する。 $|\cdot|$  は要素数 (ノードの個数) を示す。ここで  $|U_a|=|U_b|$  である。

$$d(T_a, T_b) = |U_a| + |D_a| + |U_b| \quad (1)$$

### 3.2 近傍の生成

各ステップにおける暫定解  $x_k$  について、 $x_k$  と  $p_2$  に関する最大共通部分木を抽出する。それに基づき、以下を  $x_k$

<sup>1</sup> 関西大学大学院理工学研究科  
Graduate School of Science and Engineering, Kansai University  
<sup>2</sup> 関西大学システム理工学部  
Faculty of Engineering Science, Kansai University, Japan  
<sup>3</sup> 龍谷大学理工学部  
Department of Electronics and Informatics, Ryukoku University, Japan

に確率的に複数回適用し近傍解  $y_i (0 < i < \mu)$  を生成する。ここでは、 $[1, 2 \times d(p_1, p_2) / k_{max} - 1]$  の範囲のランダムな整数  $s$  とすると、 $d(x_k, y_i) \geq s$  となるまで適用する。

- 置換:  $U_{x_k}$  の内部ノード\*1を1つ選び、それと対応する位置にある  $U_{p_2}$  のノードに置き換える
- 削除:  $D_{x_k}$  からノードを1つ選び、 $x_k$  から削除する
- 挿入:  $D_{p_2}$  からノードを1つ選び、 $x_k$  に挿入する

#### 4. 確率モデルに基づく近傍生成

本研究では、探索過程で生成される個体群の情報により得られるノードの出現に関する確率モデルに基づき、終端ノードを生成することを考える。確率モデルの利用として、変数間の依存関係を考慮した優れた手法が提案されているが、ここでは初歩的な検証として、GP における最も単純な手法の一つである Probabilistic Incremental Program Evolution(PIPE)[3] で用いられる確率木 (Probabilistic Prototype Tree: PPT) を導入する。確率木の各ノードにおいては記号それぞれの出現頻度が保存されている。近傍探索における終端ノードの生成では、それと対応する確率木のノードにおける度数分布に従い、記号を選択する。また、確率木において、探索の過程でノードが出現する頻度も保持し、削除対象のノードは、出現頻度が低いものほど高い確率で選択する。

#### 5. 数値実験

dMSXF の近傍探索に確率モデルを導入することの有効性を示す。ここでは確率木を導入した手法を dMSXF+PPT と呼ぶ。検証にあたり Santa Fe Trail 問題を用いる。Santa Fe Trail 問題はエージェントの意思決定問題の最も単純なテスト問題の一つである。32 × 32 の盤に 89 個の餌が配置され、人工蟻が所与の体力のもとですべての餌を回収することを目的とした行動則を決定する。人口蟻の行動則は MOVE (1マス前進)、LEFT (向きを 90° 左に回転)、RIGHT (向きを 90° 右に回転) の動作を示す 3 種の終端記号、および条件分岐を表す IF\_FOOD\_AHEAD と逐次処理を表す PROG2, PROG3 の 3 種の非終端記号で記述される。目的関数は蟻が体力 0 となった時点で回収した餌の個数で定義され、最適解の目的関数値が 89 の最大化問題である。蟻の初期体力を 400 とし、終端記号が実行されるごとに 1 ずつ減少させる。

本問題を解くにあたり、dMSXF および dMSXF+PPT における母集団サイズを 50、計算終了世代を 200 とする。世代交代モデルは dMSXF の原論文 [1] のものを用いる。初期解はあらかじめ定めたサイズの範囲内でランダムに生成する。ここでは [10,20],[25,35] の 2 パターンで検証する。dMSXF, dMSXF+PPT における局所探索のステップ数、近傍数は  $(k_{max}, \mu) = (3,4), (4,5)$  の 2 パターンとした。

表 1 2 手法の比較

size	dMSXF				dMSXF+PPT			
	(3,4)		(4,5)		(3,4)		(4,5)	
[10, 20]	#opt	#eval	#opt	#eval	#opt	#eval	#opt	#eval
	9	11x10 <sup>4</sup>	19	16x10 <sup>4</sup>	6	9.7x10 <sup>4</sup>	8	14x10 <sup>4</sup>
[25, 35]	5	11x10 <sup>4</sup>	14	18x10 <sup>4</sup>	13	9.8x10 <sup>4</sup>	30	14x10 <sup>4</sup>

( $k_{max}, \mu$ )

dMSXF+PPT において、確率木はすべての葉の深さが 8 の完全 3 分木とし、いずれのノードも初期の記号の度数分布を頻度 1 の一様分布とする。確率木は交叉の適用ごとに生成された近傍解集合において目的関数値が上位半分で更新される。確率木の参照時、更新時には解と確率木でマッチングを行い、ノードの位置を特定する。記号の発生確率はそのノードにおける度数分布に従うものとする。一方、生成された解において深さが 8 よりも深い位置にあるノードは確率木でマッチングするものがない。更新の場合にはそのノードを無視、参照の場合にはすべての記号の出現確率は一様分布に従うものとする。確率木は 1 回目の試行の開始と同時に初期化し、2 回目以降の試行は前の試行で更新された確率木を引き継ぐ。

表 1 に dMSXF と dMSXF+PPT の比較結果を示す。これらは 50 試行の結果であり、最適解を得た回数 (表中, #opt) と最適解を得たときの総評価計算回数の平均 (#eval) である。表 1 から初期の木をサイズを [25,35] にしたとき、確率木を用いることによって最適解に到達する確率が上がっていることがわかる。また、いずれの設定においても最適解を得るのに要した評価計算回数が少ないことが確認できる。一方、初期の木をサイズを [10,20] にしたときは、性能が低下している。

#### 6. おわりに

木構造最適化のための多段階探索交叉 dMSXF において、良い形質の保存および形成を促進させることを目的とし、ノードの出現に関する確率モデルを近傍解生成に導入した。数値実験より、Santa Fe Trail 問題において、探索過程で得られた確率木を用いることで、初期の木が大きい場合にすくない評価計算回数、かつ高い確率で最適解を得られることがわかった。

謝辞 本研究は科研費 26330290, 26730133 の助成を受けたものです。

#### 参考文献

- [1] Ikeda, K., and Kobayashi, S.: deterministic Multi-step Crossover Fusion: A Handy Crossover for GAs, Proc. Parallel Problem Solving from Nature VII, pp.162-171 (2002).
- [2] Hanada, Y., Hosokawa, N., Ono, K. and Muneyasu, M.: Effectiveness of Multi-step Crossover Fusions in Genetic Programming, Proc. IEEE Congress on Evolutionary Computation (CEC 2012), pp. 1743-1750 (2012).
- [3] Salustowicz, R. P. and Jurgen Schmidhuber, J: Probabilistic Incremental Program Evolution, Evolutionary Computation, Vol. 5, No. 2, pp. 123-141, 1997

\*1 最大共通部分木における内部ノード