Vol.2015-MPS-102 No.22
2015/3/4

GPにおける初期の木のサイズが多段階探索交叉の解探索性能に与える影響

松村 康平[1]　花田 良子[1]　小野 景子[2]　村中 徳明[1]

概要：解を木構造で表現した遺伝的プログラミングにおいては，しばしば，交叉などの遺伝的オペレータにより，木が爆発的に膨張するブロートの発生が見られる．これまでに我々は，木の成長を抑制しつつ，両親の形質を受け継がせることを目的とした，木構造のための多段階探索交叉を提案している．多段階探索交叉では初期母集団の木のサイズの設定が重要となる．本研究では，関数同定問題の複数の例題を対象として，初期の木のサイズが解探索性能に与える影響を検証する．

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，木が爆発的に膨張するブロートを抑制しつつ，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 を提案している [2]．dMSXF はブロートを抑制するため初期母集団の木のサイズの設定が非常に重要となる．本研究では関数同定問題の例題を対象として，初期の木のサイズが解探索性能に与える影響について検証する．

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 μ 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 において暫定解 x_k から生成する近傍の解 y_i を解 x_k よりも p_2 に近い個体に制限し，x_{k+1} が x_k よりも劣っていたとしても必ず探索を進めることで解遷移を決定的に行う．

3. 木構造における距離と近傍

3.1 距離の定義

木 A，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 A，B の距離 $d(A, B)$ を次のように定義する．$|\cdot|$ は要素数（ノードの個数）を示す．

$$d(A, B) = \frac{1}{2}(|U_A| + |D_A| + |U_B| + |D_B|) \qquad (1)$$

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 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．

[1] 関西大学システム理工学部
Faculty of Engineering Science, Kansai University, Japan
[2] 龍谷大学理工学部
Department of Electronics and Informatics, Ryukoku University, Japan

Vol.2015-MPS-102 No.22
2015/3/4

- **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
- **入替**：U_{x_k} の末端にあるノード*2を 1 つ選び，x_k 内でそのノードの兄弟のノードと位置を入れ替える

4. 数値実験

初期解の木のサイズ（ノード数）の設定が dMSXF の解探索性能に与える影響を検証する．また，比較として一点交叉（1X）と size fair 交叉（SFX）[3] を用いる．SFX は 1X を改良した交叉で，両親間で交換部分木のサイズを制限することで木の成長を抑制している．検証にあたり連続関数同定問題を用いる．検証に用いる例題を以下に示す．

$$f_{opt} = x^4 - x^3 + x^2 - x \tag{2}$$

変数の定義域をそれぞれ [-1, 1] とし，サンプル点を 21 点，等間隔に与える．推定するにあたり，非終端記号の集合 V^{NT}={+, -, *, /}，終端記号の集合 V^T={x, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9} を用いる．非終端記号はすべて 2 項演算子である．終端記号においては x のみが変数で，他は定数である．また，本例題において f_{opt} を表現する木の最小サイズは 13 である．

母集団サイズを 20 と 50，計算終了世代を 100 とした．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 [5, 15]，[11,21]，[15, 25] および [21, 31] の 4 パターンを検証する．dMSXF では k_{max} を 3，μ を 4 とした．他の 2 手法は 1 回の交叉で 12 個体生成する．

表 1，2 に 3 手法の成功回数（#opt）と最良解の木の平均サイズ（size）および平均の深さ (dep) の比較結果を示す．いずれも 20 試行の結果である．また，図 1 に母集団サイズを 50 としたときの初期解サイズ [5, 15] および [21, 31] における最良解の木のサイズの推移を示す．

表 1，2 の木のサイズに着目すると，1X でブロートを起こす試行が存在する一方で，SFX と dMSXF ではブロートが起こらず，1X と比較して小さい木が生成されることがわかる．また，SFX と dMSXF を比較すると dMSXF がより木の成長を抑制していることがわかる．dMSXF では初期解のサイズが最適解の発見率（同定成功率）に大きな影響を与えていることがわかる．母集団サイズおよび初期の木のサイズが小さい場合，関数の同定は成功していない．これは母集団に含まれる部分木の種類が少ないこと，木の成長が抑制されることに起因する．図 1 の dMSXF の [5, 15] の推移では探索の過程で木の成長は見られるものの，最適解のサイズに到達せず，最適解を得るに至っていない．

*1 最大共通部分木における内部ノード
*2 最大共通部分木で葉に位置するノード

表 1 3 手法の比較（p=20）

	1X			SFX			dMSXF		
	#opt	size	dep	#opt	size	dep	#opt	size	dep
[5, 15]	10	18.5	6.2	5	14.8	5.3	0	12.6	4.2
[11, 21]	11	100.6	11.4	12	19.2	5.5	10	15.5	4.6
[15, 25]	8	128.2	18.3	7	22.5	5.7	10	17.3	4.9
[21, 31]	11	29.7	7.6	7	28.4	7.0	16	20.7	5.9

表 2 3 手法の比較（p=50）

	1X			SFX			dMSXF		
	#opt	size	dep	#opt	size	dep	#opt	size	dep
[5, 15]	17	17.7	5.8	13	16.7	5.5	12	15	4.1
[11, 21]	18	22.1	7.4	16	19.1	6.0	18	15.3	4.6
[15, 25]	18	96.2	12.9	16	20.6	5.5	20	17.5	5.0
[21, 31]	14	41.8	8.1	13	24.9	6.1	20	19.1	5.5

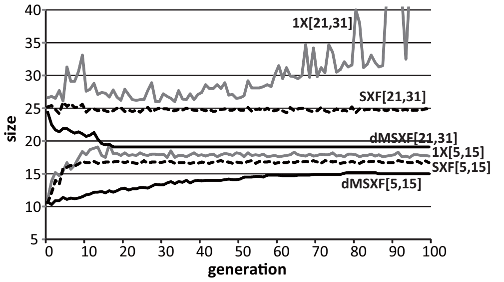

図 1 サイズの推移

一方で，母集団サイズあるいは木のサイズいずれかを大きくすることで最適解が得られるようになり，大きな母集団である程度の初期の木のサイズを確保することで他の手法と比較して高い最適解発見率を示すことがわかる．

5. おわりに

本研究では dMSXF において初期母集団サイズおよび初期解の木のサイズが解探索性能に与える影響について検証した．dMSXF は木の成長を強く抑制するため，小さい母集団かつ小さな初期解の場合には，木が十分成長せず，高い解探索性能が得られない．一方で，母集団サイズ，初期解のサイズを大きく設定した場合には，非常に高い最適解発見率を示すことがわかった．今後の課題として，小さな初期解のもとでも最適解が得られるよう，木の急激な成長を起こさずに，木を成長させるオペレータの導入が挙げられる．

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 fron 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] Langdon, W. B.: Size fair and homologous tree crossovers for tree genetic programming. genetic programming and evolvable machines, Proc. Genetic Programming and Evolvable Machines, pp. 95–119 (2000)