

平成 30 年 6 月 25 日現在

機関番号：34416

研究種目：基盤研究(C) (一般)

研究期間：2014～2017

課題番号：26330290

研究課題名(和文) 形質の保全と戦略的獲得に基づく大規模木構造のための遺伝的プログラミング

研究課題名(英文) Genetic Programming for large tree optimization with considering inheritance and acquisition of traits

研究代表者

花田 良子 (Hanada, Yoshiko)

関西大学・システム理工学部・准教授

研究者番号：30511711

交付決定額(研究期間全体)：(直接経費) 3,500,000円

研究成果の概要(和文)：式や自律制御のためのプログラム生成など、木構造で表現される変数からなる設計問題において、特に、大規模な木構造に焦点をあて、最適な構造を効率的に探索するための遺伝的プログラミングの汎用性の高い探索手法を開発することを目的に、(1) これまで開発してきた高い形質保全性を有する多段階探索交叉の改良、(2) 過去の探索情報による良好な形質推定に基づく効果的な初期個体生成の開発、(3) 木を指向的に成長させる探索スキームの導入の3点を、研究期間を通して行った。主に、最適解が既知の問題である関数同定問題を用いた性能解析を通して手法の有効性および適用指針を示した。

研究成果の概要(英文)：The purpose of this study is to develop a new genetic programming for large tree optimization in order to solve the design problem consisting of variables expressed in tree structure such as expressions and program generation for autonomous control. The achieved works of this study are: (1) Improvement of a multi step crossover that has high trait integrity, (2) Development of an initialization scheme based on preferable trait estimation by past search information, and (3) Development of a search scheme for obtaining new traits. We showed the effectiveness and application guidelines of the proposed method mainly through the performance analysis using the function regression problem.

研究分野：最適化

キーワード：遺伝的プログラミング 木構造 局所探索 初期化

1. 研究開始当初の背景

遺伝的プログラミング (Genetic Programming: GP) は、自律制御のプログラム生成や関係式推定などの設計問題を対象とした進化計算の一種である。進化計算とは選択、交叉、突然変異に代表される遺伝的操作を繰り返し解 (個体) 群に適用することで近似解や満足解を得る最適化手法の総称である。進化計算では各個体が持つ情報を交換する交叉が主探索オペレータであり、その設計が解探索性能の向上の鍵となる。個体を木構造で表現する GP においては、一般に、個体を表現する染色体の長さは固定されない。そのため、個体の部分を交換する交叉の操作によって、探索の過程で木が急激に膨張するブロートとよばれる現象をしばしば引き起こすといった問題がある。これにより、木構造最適化における交叉設計では個体の成長を抑制することが最重要課題となっている。ただし、逐次処理や条件分岐の要素で記述される制御モジュールの設計など、大規模な木構造最適化問題に応用することを考えた場合には、個体の評価値に影響を与えない冗長なノードあるいは部分木 (イントロン) や部分木どうしの機能相殺の発生を抑制しつつ、個体の良好な成長を促進させる操作を併用するなど、木のサイズをコントロールしたもとで効率的に満足解を求める GP の手法の開発が求められる。

2. 研究の目的

木構造で表現される変数からなる設計問題において、とくに、大規模な木構造設計に焦点をあて、最適な構造を効率的に探索するための GP の汎用性の高い探索手法を開発することが本研究の目的である。具体的には、(1) これまで開発してきた高い部分分解 (形質) 保全性を有する多段階探索交叉 (dMSXF) の改良、(2) 効果的な初期個体生成の開発、(3) 木を指向的に成長させる探索スキームの導入の3点を、研究期間を通して行う。(1) と (2) については、過去の探索情報による良好な形質推定に基づく手法を考案する。

3. 研究の方法

これまでに開発してきた多段階探索交叉は、解の類似性を反映した距離の定義のもと、一方の親から他方の親へと距離を縮める方向に近傍探索を繰り返すことで探索を進め、その過程で得られた優良な子個体を次世代へと残す。手法の流れは次のとおりである。

[多段階探索交叉のアルゴリズム]

Step 1 p_1, p_2 を両親, その子個体群 $\mathcal{A}(p_1, p_2) = \phi$ とする。

Step 2 探索初期点 $x_1 = p_1, k=1$ とし, x_1 を $\mathcal{A}(p_1, p_2)$ の要素として加える。

Step 3 ステップ k における探索点 x_k の近傍個体を μ 個生成し, その集合を $M(x_k)$ とす

る。ただし, $M(x_k)$ のすべての近傍解 y_i ($0 \leq i \leq \mu$) 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。

Step 4 $M(x_k)$ の最良解を x_{k+1} とし, x_{k+1} を $\mathcal{A}(p_1, p_2)$ の要素として加える。

Step 6 $k = k+1$ 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, Step 3 にもどる。

ここで d は 2 個体間の距離であり, 異なる染色体長の個体間で遺伝子座を定義したもとで, それぞれの親特有の遺伝子 (ノード) の個数の総和で求められる。局所探索の各ステップにおいて p_2 の特有のノードを x_k に置換, 挿入, および x_k 特有のノードを x_k から削除するといった 3 種類の操作を適用することで, p_1 から p_2 に徐々に近づく方向に近傍個体が生成される。多段階探索交叉は木の変化を限定的にすることで, 親の形質を保持しつつ, 不要なノードの発生を抑制している。

木構造を有する問題においては, 例えば, 算術演算の場合には演算子は 2 項の非演算子をもつといったように, 内部ノード (非終端ノード) には子ノードの数がその記号によってあらかじめ定められているものが多い。これまでの多段階探索交叉での近傍解生成では, 子ノード数に関する制約条件を満たすために, 適宜, ランダムにノードを挿入, 削除を行ってきた。この操作に, 過去の探索情報から構築したノードの分布を利用することで, 効率的な近傍解生成を実現する。また, 多段階探索交叉は木の成長を抑制するために, 大きな木構造を構築しなければならない問題については, あらかじめ十分な木のサイズ (ノード数) をもつ初期個体を準備する必要がある。その場合, 木のサイズが大きいほど, 個体の評価値に致命的な影響を与える形質が混入する確率が高くなり, 効率な探索が困難となる。そのため, 近傍解の生成と同様にノードの分布を利用する。さらに, 親の持たない形質を獲得しつつ, 木を成長させるため, 多段階探索交叉を, 形質が大きく異なる, かつ, サイズが大きい個体に対して適用することで, 木を指向的に成長させる探索を実現する。手法の性能検証には主に, 最適解が既知の問題である関数同定問題を用いる。関数同定問題は所与の入力と出力のセットから, その関係を演算子 (関数) および被演算子 (変数・定数) の種々の結合で推定する問題である。従来手法と比較することで開発手法の有効性を示す。

4. 研究成果

GP が対象とする問題では, 個体である木の階層的な性質により, 個体の評価値に致命的な影響を与える形質 (致死形質) や, その組合せで形成される構造 (致死構造) がしばしば発生する。これにより解の探索に遅滞や, 十分な探索が進む前に, 良質な構造を生み出す形質が損なわれるといった問題が生じる。

また、このような致死形質や致死構造は木のサイズが大きくなるほど、発生しやすくなる。

本研究では多段階探索の近傍生成および GP の初期個体生成において、過去の探索で得られたノードの出現に関する頻度分布をもとにノードの発生確率に偏りをもたせることで、致死形質・構造の発生の抑制、および良質な形質・構造の形成を実現した。ノードの頻度分布を扱う方法として、Salustowicz らによって提案された、木構造の推定を行うためのデータ構造である確率木を採用した。確率木は N 分木 (N は出現する非終端ノードの中で最大の引数の数) で表され、確率木の各ノードはその位置の記号の出現率に関する表を保持している。評価値の高い個体には良好な形質が含まれるということを前提に、過去の複数の探索の過程で得られた評価値の高い個体により確率木を構築した。また、各ノードの確率表には、(a) 自身のノードのみの記号発生率、(b) 親と自身のノードの記号のペアに関する発生率、の 2 種を検証した。

図 1 は多段階探索交叉 (dMSXF) の近傍生成について、ランダム、(a) に基づく近傍生成、(b) に基づく近傍生成、としたときの、それぞれの探索の成功率を示したものである。例題には次の 2 つを用いた。式と同等の関数が得られたときに探索成功としている、図中 [*,*] は生成する初期木のサイズを示しており、この範囲でランダムに個体を生成している。また、図 2 に致死形質の発生率を示す。これらはいずれも 50 試行の結果である。

$$f_1 = x^4 - x^3 + x^2 - x \quad (1)$$

$$f_2 = x \sin x (\cos x - 1) \quad (2)$$

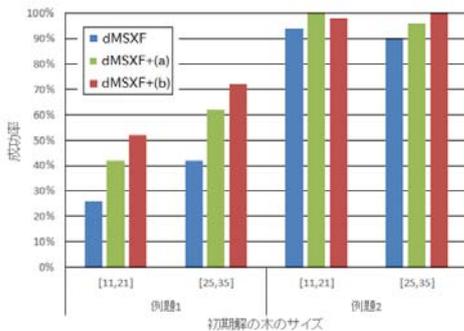


図 1 成功率の比較

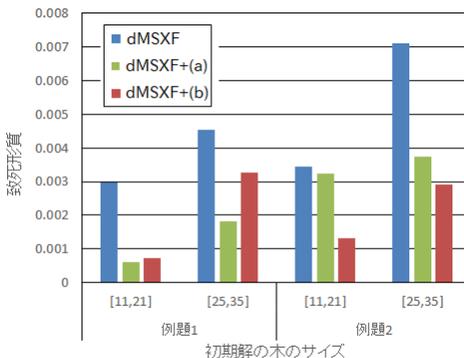


図 2 致死形質の発生率

図 1 より、過去の探索の情報を利用した dMSXF+(a)、dMSXF+(b) が従来の多段階探索交叉 (dMSXF) よりも高い割合で最適解が得られることを示した、中でも、親子の関係を考慮した確率木を用いた dMSXF+(b) がもっとも高い性能を示すことがわかった。また、図 2 より、木のサイズが大きいくほど致死形質の発生率が高くなること、確率木を採用することでその発生が抑制されることが確認された。

次に過去 50 試行で構築した確率木を初期化に利用した。確率木の品質は過去の探索の性能に大きく依存する。過去の試行で成功数が多い状況で確率木を生成した場合、局所解に陥った回数が多い状況で確率木が構築された場合の 2 通りを検証した。前者の結果を確率木 (良好)、後者の結果を確率木 (不良) とよぶ。図 3 に例題 (1) における初期個体の評価値分布を示す。目的関数は誤差の総和であり、小さいほどよい。

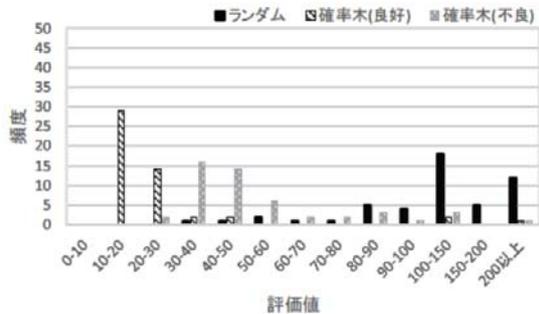


図 3 初期個体の分布

図 3 より、初期個体をランダムに生成する一般的な初期化より、確率木を用いて、ノードの発生率に偏りを持たせた初期化の方が良質な初期個体を多く生成していることが確認できた。とくに、用いる確率木の質が高い方が、評価の高い初期個体を多く生成していることがわかった。

これらの個体群を初期母集団として、交叉に一点交叉を用いて、通常の GP で探索したときの探索性能を表 1 に示す。これらは 50 試行の結果である。木のサイズが 100 を超えたときはプロットが起こったものとし、探索を止めている。なお、表中、“従来手法”は初期化をランダムにしたときの結果を示している。

表 1 性能比較

| | 成功数 | 深さ | サイズ | プロット数 |
|----------|-----|------|------|-------|
| 従来手法 | 0 | 12.7 | 55.8 | 43 |
| 確率木 (良好) | 22 | 9.70 | 42.8 | 17 |
| 確率木 (不良) | 12 | 11.1 | 48.1 | 34 |

表 1 より初期母集団生成に過去の探索における良質な解の情報を利用することで、高い割合で同定が成功していることがわかる。単純な部分木の交換を行う 1 点交叉はプロットを起こししやすい交叉であるが、その発生も抑制

できていることがわかる。ここでは、性能の差を強調するために単純な一点交叉を用いた結果を示したが、多段階探索交叉を用いることによってさらに高い探索性能が得られることを確認している。

確率木を用いることで、過去の探索により偏った形質が多く生成される。そのため、母集団に見られない形質を獲得するオペレータが必要となる。本研究では、多段階探索交叉を適用する親2個体について、それらと距離が離れた解の形質を近傍解に与えることで、探索の偏りを緩和させ、かつ個体である木を成長させる手法を現在、開発している。これらについては、今後も継続する予定である。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計 7 件)

[1] Kohei Matsumura, Yoshiko Hanada, Keiko Ono: Probabilistic Model-Based Multistep Crossover Considering Dependency Between Nodes in Tree Optimization, Studies in Computational Intelligence book series (SCI), Vol. 721, pp. 187--200, 2017

[2] Kohei Matsumura, Yoshiko Hanada, Keiko Ono: Probabilistic Model-Based Multistep Crossover for Genetic Programming, 17th International Symposium on Advanced Intelligent Systems (ISIS), pp.154--159, 2016

[3] Keiko Ono, Yoshiko Hanada: Distributed Light Brightness Control based on cuSASGP, Proc. IEEE Congress on Evolutionary Computation (CEC 2016), pp.839--845, 2016

[4] 小野景子, 花田良子: GPにおける頻出構造に着目した複合的ブロード抑制戦略, 情報処理学会論文誌数理モデル化と応用 (TOM), No. 8, Vol. 2, pp. 77--87, 2015

[5] Keiko Ono, Yoshiko Hanada: A CUDA-based Self-adaptive Subpopulation Model in Genetic Programming: cuSASGP, Proc. the 2015 IEEE Congress on Evolutionary Computation (CEC), pp.1543--1550, 2015

[6] Keiko Ono, Yoshiko Hanada: Assembling Bloat Control Strategies in Genetic Programming for Image Noise Reduction, Proc. the 14th International Conference on Intelligent Systems Design

and Applications (ISDA), pp. 113--118, 2014

[7] Yoshiko Hanada, Kotaro Minami, Keiko Ono, Yukiko Orito, Noriaki Muranaka: A Study on Neighborhood and Temperature in Multi-step Crossover Fusion for Tree Structure, Proc. the 18th Asia Pacific Symposium on Intelligent and Evolutionary Systems (IES), Proceedings in Adaptation Learning and Optimization, Vol.2 pp.519--531, 2014

[学会発表] (計 9 件)

[1] 雲山大地, 松村康平, 花田良子, 小野景子: 遺伝的プログラミングにおける確率モデルに基づく初期解生成の提案, 第 61 回システム制御情報学会研究発表講演会, 143-5, pp.1--2, 2017 年 5 月 23 日, 京都府

[2] 松村康平, 花田良子, 小野景子: 親子の依存を考慮した確率モデルが多段階探索交叉の近傍生成に与える効果, 情報処理学会数理モデル化と問題解決 (MPS) 研究報告, 2016-MPS-112, pp.1--5, 2017 年 2 月 27 日, 岩手県

[3] 松村康平, 花田良子, 小野景子: 木構造最適化におけるノードの依存関係を考慮した近傍探索に基づく多段階探索交叉, F2B1, pp.1--4, 第 26 回インテリジェント・システム・シンポジウム, 2016 年 10 月 28 日, 大阪府

[4] 松村康平, 花田良子, 小野景子, 村中徳明: 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, 情報処理学会 数理モデル化と問題解決 (MPS) 研究報告, 2016-MPS-107, pp.1--2, 2016 年 3 月 8 日, 山口県

[5] 古謝望, 花田良子, 折登由希子, 村中徳明: ポートフォリオ最適化問題のための進化計算の効果的な初期集団: 縁付きヘッセ行列による多段階回アセット選択の有効性, 計測自動制御学会システム・情報部門学術講演会, NO.SS19-1, 2015 年 11 月 19 日, 北海道

[6] 松村康平, 花田良子, 小野景子, 村中徳明: 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有効性, 計測自動制御学会システム・情報部門学術講演会, NO.SS19-1, 2015 年 11 月 19 日, 北海道

[7] 松村康平, 花田良子, 小野景子, 村中徳明: 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, 情報処理学会 数理モデル化と問題解決 (MPS) 研

究報告, 2015-MPS-102, pp.1--2, 2015年3月4日, 長崎県

[8] 花田良子, 小野景子: 木構造で表現されたスタックフィルタ設計における多段階探索交叉の有効性の検証, 計測自動制御学会システム・情報部門学術講演会 2014, pp.762--767, 2014年11月21日, 岡山県

[9] 小野景子, 花田良子: GPにおける頻出構造に着目した複合的ブロート抑制戦略, 情報処理学会 数理モデル化と問題解決 (MPS) 研究報告, Vol.2014-MPS-100, pp.1--6, 2014年9月26日, 東京都

6. 研究組織

(1) 研究代表者 花田良子

花田 良子 (HANADA, YOSHIKO)

関西大学・システム理工学部・准教授

研究者番号: 30511711

(2) 研究分担者

小野 景子 (KEIKO ONO)

龍谷大学・理工学部・講師

研究者番号: 80550235