

細胞画像領域分割のための遺伝的プログラミングにおける ブロード抑制モデルの比較

山口 浩明

Hiroaki YAMAGUCHI

1 はじめに

近年, 失明の原因である角膜損傷や角膜減少障害を解決するために角膜再生医療が注目されている. 失明は, 角膜の最も内側に存在する角膜内皮細胞が, 一度損傷すると再生しないことが主な原因の一つである. そこで, この角膜内皮細胞を新たに培養し, 移植することで失明を防ぐ角膜再生医療の研究が進められている. 角膜再生医療において, 培養した角膜内皮細胞が良好な状態であるかの判断は, 細胞密度, 形状, 大きさなどの特徴量が用いられ, 現在, 特徴量を計測するための多くのソフトウェアの開発や研究が行われている^{?)}. しかし, これらのソフトでは, 特徴量を計測するための細胞領域分割のために, ソフトに組み込まれている画像処理フィルタを複数適用させる必要がある. そのため, 使用者には画像処理の知識が必要とされる. また, 対象の画像が変わるごとに組み合わせ方が変わるので, 使用者の負担が大きい. そこで, 自動で目的の画像処理を構築する Automatic Construction of Tree-structure Image Transformation や Genetic Programming based on Image Segmentation 手法などが考案されている^{?)}.

これらの手法は進化計算の一つである遺伝的プログラミング (Genetic Programming:GP) を用いた画像処理の最適な組み合わせを求めている. 一方で, 先行研究に用いられている GP では, GP の解探索において重要な問題とされているブロードの抑制に対して, 木の深さの制限や, サイズに応じたペナルティを与えるといった単純な方法しか行っていない. そこで, 本研究では, 現在提案されているブロード抑制のための GP モデルを適用し, 細胞画像の領域分割に適した GP モデルの比較を行った. 適用した GP のブロード抑制モデルは, Double Tournament^{?)}, Tarpeian^{?)}, Non-destructive Crossover^{?)}, Recombinative Hill-climbing^{?)}, Spatially Structure + Local elitism^{?)} モデルである.

2 GP を用いた画像処理フィルタ構築

一般的に画像処理は既知の単純な画像処理フィルタの組み合わせとして表現可能とされている. そこで, 画像処理はフィルタの組み合わせ最適化問題と捉え, その最適化問題の解を求めることで, Fig. 1 に示すような木構造状フィルタを構築する. これが GP を用いた画像処理フィルタの構築手法であり, 先行研究として ACTIT や GPIS 手法などが挙げられる.

Fig. 1 に示すような木構造フィルタを構築するために,

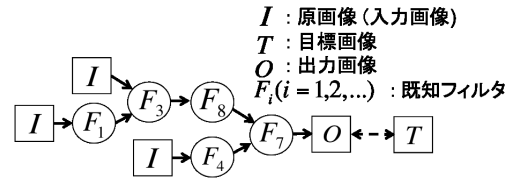


Fig.1 木構造フィルタ構築の原理

学習データである原画像 I と原画像に対して理想的な処理画像である目標画像 T のセットと, 複数の既知の画像処理フィルタが用意される. そして原画像から目標画像からへ近似するように, 既知の画像処理フィルタに対して組み合わせ最適化を行い, 木構造フィルタが構築される. この木構造フィルタでは, 終端ノード I から画像が入力され, 各非終端ノードに格納された既知フィルタの処理を順に行うことで, 1 枚の出力画像を作成する.

構築された木構造フィルタの適合度の計算には, 評価関数 (1) が用いられる.

$$fitness = \frac{1}{k} \sum_{k=1}^K \left\{ 1 - \frac{\sum_{x=1}^{W_x} \sum_{y=1}^{W_y} |O(x, y) - T(x, y)|}{W_x W_y V_{max}} \right\} \quad (1)$$

適合度は, 出力画像 $O(x, y)$ と目標画像 $T(x, y)$ の各ピクセル値の差分によって求められ, この差分が小さいものほど優れていると評価する. 適合度は 1.0 を最大値とする. K は学習画像セット (原画像と目標画像) の数, W_x, W_y は画像の横, 縦のサイズ, V_{max} は最大階調値を示す.

木構造フィルタの GP 個体は, 終端ノードに入力する画像, 非終端ノードに既知の画像処理フィルタが格納される. また非終端ノードに格納する既知の画像処理フィルタは, 入力系統数が 1 および 2 のものである, 使用する既知の画像処理フィルタ (非終端ノード) を Table 1, 2 に示す.

Table1 1 入力フィルタ

Filter	Effect
f1	Edge emphasis(sobel)
f2	Edge emphasis(laplacian)
f3	Dilation
f4	Erosion
f5	Median
f6	Light Pixel
f7	Dark Pixel
f8	Large Area
f9	Small Area
f10	Binartization
f11	Inversion
f12	Gradient
f13	Watershed

Table2 2 入力フィルタ

Filter	Effect
F1	Bounded Sum
F2	Bounded Prod
F3	Logical Sum
F4	Logical Prod
F5	Algebraic Sum
F6	Algebraic Prod
F7	Sub

3 GPにおけるブロート抑制モデル

遺伝的プログラミング (Genetic Programming:GP) は, John Koza らによって提案された遺伝的アルゴリズムを構造型 (木構造, グラフ構造) が表現できるように拡張した進化計算手法である^{?)}. この GP の探索過程では, プログラムサイズが増大するブロートと呼ばれる現象が問題とされており, ブロートが発生することによって, 探索時間の増加や解の多様性の喪失による局所解への早期収束, 解の過学習などの問題が生じる. このブロートを抑制するための代表的な手法は木の深さやサイズ (ノード数) の上限を制限する方法や, 個体評価の際に, サイズが大きい個体ほどペナルティーを与える方法である. またそれら以外にも様々なブロートを抑制するモデルが提案されている.

本稿では細胞領域分割に適した GP のブロート抑制モデルの調査を行う. 比較したモデルは^{?)}を参考にし, 標準的な GP の他, Double Tournament^{?)}, Tarpeian^{?)}, Non-destructive crossover(NDC)^{?)}, Recombinative hill-climbing(RHC)^{?)}, Spatial Structure + Elitism (SS+E)^{?)}モデルである. 以下にこれらモデルについての説明を行う.

3.1 Double Tournament

Double Tournament では, 適合度とサイズに基づく2つのトーナメント層が使用される. 一般的に, はじめに適合度に基づくトーナメントが行われ, その勝者に対してサイズに基づくトーナメントが行われ, 1つの個体が選択される. Double Tournament 手法で設定するパラメータは, 適合度トーナメントサイズ F , parsimony トーナメントサイズ D , そして各トーナメントを行う順序である.

3.2 Tarpeian

Tarpeian 手法では, 母集団の平均個体サイズ以上を持つ一部の個体を競争に参加させないようにする. 評価の際, 平均個体サイズ以上の個体には, ある一定の確率 W で非常に悪い適合度を設定する. これによりサイズが大きい個体は次世代に残る確率が低くなるため, ブロートを抑制することができる.

3.3 Non-Destructive Crossover

GP において通常用いられる交叉は個体に対して破壊的な作用を持っていると考えられており, イントロンの指数成長であるブロートの発生を引き起こすとされている. このブロートの発生を抑制する方法として, 非破壊的な性質を持つ交叉手法 (Non-destructive crossover: NDC) についての改良が多く提案されている. NDC は交叉により生成された子個体とその親個体の適合度よりも優れている場合のみ, 子個体は生存する操作である. NDC を用いることにより, 適合度の向上のためにのみ必要なプログラムの成長が行われるため, ブロートを抑制することが可能となる.

3.4 Recombinative Hill-Climbing

Recombinative hill-climbing(RHC) は, Hooper らによって提案された GP と山登り法を組み合わせたモデルである. RHC では, 移住母集団と呼ばれる元の母集団は次世代の母集団を生成するために訪問母集団にコピーされる. 移住母集団の各個体は訪問母集団のランダムに選ばれた個体とペアを組み, そのペアに対して子個体を生成するための遺伝的操作が適用される. RHC においても, 生成された子個体は親個体よりも適合度において優れている場合のみ, 移住母集団の親個体と交換される.

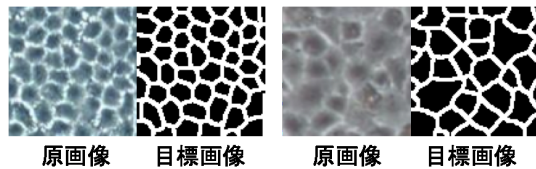
3.5 Spatial Structure + Elitism

Spatial Structure(SS) モデルでは, N 個の個体は N 個の要素をもつ行列 W にそれぞれ配置される. また, $W[i, j]$ に配置された個体とそのムーア近傍の8個体を部分母集団と定義する. 母集団に対して遺伝的操作を適用する際, $W[i, j]$ に配置された各個体は訪問され, それぞれの部分母集団内から親個体を選出し, 子個体が生成される. これが SS モデルのアルゴリズムである. また, 本稿では SS モデルに対して, 適合度の向上とブロートの抑制において良好なトレードオフ関係があるとされるローカルエリートを適用する. ローカルエリートでは, 生成された子個体が訪問されている $W[i, j]$ の個体よりも優れている場合のみ, その個体と子個体を入替える.

4 細胞領域分割におけるブロート抑制モデルの比較実験

4.1 実験概要

本章では, 細胞領域分割問題に対して複数の GP モデルを比較し, 適したモデルの調査を行う. 本実験で使用した学習画像セットを Fig. 2 に示す. Fig. 2 に示す画像は光学顕微鏡を用いて50倍で撮像された角膜内皮細胞画像であり, Fig. 2(a) は細胞状態が良好, Fig. 2(b) は不良であると専門家によって判断されたものである. 目標画像は原画像における細胞壁領域を白, 細胞内を黒でマークした画像であり, 画像スケールは 100×100 である. また, 既知フィルタは Table 1, 2 に示すものを用いる.



(a) セット A (細胞状態: 良好) (b) セット B (細胞状態: 不良)

Fig.2 学習画像セット

比較する GP モデルは3章で示した, 標準モデル (SGP) に加え, Double Tournament, Tarpeian, NDC, RHC, SS+E モデルである. 本実験で使用した GP のパラメータを Table 3 に示す. 本実験では, 各モデルにおいて30試行を行った結果について検討する.

Table3 GP のパラメータ

パラメータ	値
世代数	300
母集団サイズ	484
選択	トーナメント (サイズ:7), バイナリトーナメント (SS+E)
交叉率	0.9 (SS+E:1.0)

4.2 実験結果

各モデルにおいて得られた適合度とその解のサイズの箱髷図を Fig. 3, 4 に示す。Fig. 3(a), 4(a) は、各試行で得られた最良解の適合度の平均値を示し、Fig. 3(b), 4(b) は最良解の個体サイズの平均値を示している。また、各モデルにおいて得られた最も低い適合度の出力画像を Fig. 5, 6 に示す。

得られた適合度において、Fig. 3(a), 4(a) から RHC と SSE モデルがバラツキの少ない高い適合度を示した。Fig. 5, 6 の出力画像においても両モデルは良好な細胞分割が行われていることが確認でき、他のモデルに比べ高い探索性能を示した。

解の個体サイズでは、Fig. 3(b), 4(b) の結果から NDC モデルが最も低い値を示している。また、RHC モデルでは、NDC モデルの次に個体サイズが小さく、かつバラツキの少ない結果となった。

適用対象である学習画像セットでは、細胞状態が不良である学習画像セット 2 の方が学習画像セット 1 に比べ、全体的に適合度が低く、個体サイズが増加した傾向となった。一方で、各モデルの細胞画像分割問題に対する解の探索結果の傾向は学習セット間で類似していた。以上の結果から、GP を用いた細胞画像分割問題では、RHC モデルが高いブロー抑制機能を持ち、かつ高い解の探索性能を有していることが確認できた。

5 考察

式 (1) の評価関数は、学習セットの目標画像と GP によって生成された出力画像とのピクセルごとの一致度を表している。この評価関数による最適解は、目標画像と出力画像の全ピクセルが一致した場合であるが、手作業で作成した目標画像と全ピクセルが一致する解が生成されることはほぼ不可能であると考えられる。また、Fig. 5, 6 の出力画像の結果からも、適合度が 1.0 である最適解でなくとも、十分に良好な細胞画像領域分割が可能であることが確認できる。そのため、本問題では準最適解をいかに効率良く探索できるかが重要な要素となる。

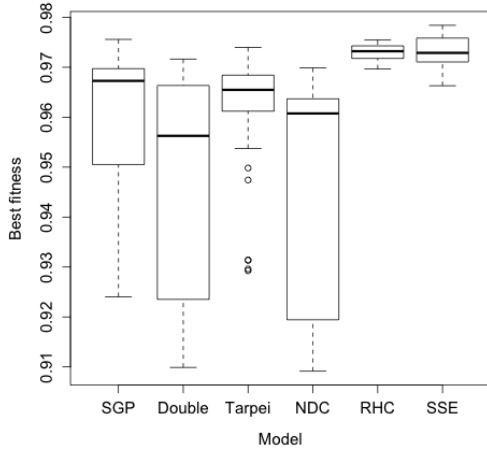
RHC は、局所探索法である山登り法を GP に加えたモデルであるため、今回比較したモデルの中で最も効率良く局所解の探索を行えたことが考えられる。SSE モデルでは、高い適合度をもつ解を探索することができたが、一方で個体サイズが増大する傾向がみられた。SSE は高い探索性能を有していることが先行研究²⁾で報告されている。そのため、SSE は局所解へ収束せず、最適解を探索し続けたゆえに個体サイズが増加したことが考えられる。

6 まとめ

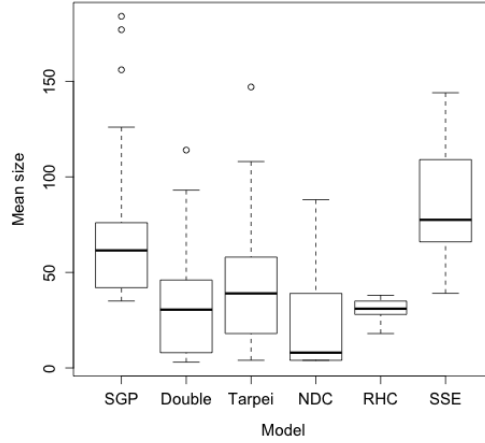
本稿では、細胞画像領域分割のための GP におけるブロー抑制モデルの比較を行った。実験の結果、RHC と SSE モデルは適合度の高い解、RHC と NDC モデルでサイズの小さい解を生成することが確認できた。そのため、小さいサイズかつ適合度の高い解を生成することができる RHC が、細胞画像領域分割のための GP モデルに適していることが結論づけられる。

参考文献

- 1) Abramoff MD, Magalhaes PJ, and Ram SJ. Image processing with imagej. *Biophotonics International*, Vol. 11, pp. 36–42, 2004.
- 2) T. Nagao and S. Masunaga. Automatic construction of image transformation processes using genetic algorithm. *Processing of ICIP'96*, Vol. 3, pp. 731–734, 1996.
- 3) S.Luke and L.Panait. A comparison of bloat control methods for genetic programming. *Evol Comput*, Vol. 14, No. 3, pp. 309–344, 2006.
- 4) Whigham.P.A and Dick.G. Implicitly controlling bloat in genetic programming. *Evolutionary Computation, IEEE Transactions on*, Vol. 14, pp. 173–190, 2010.
- 5) J.Koza. *Genetic programming, on the programming of computers by means of natural selection*. MIT Press, 1992.

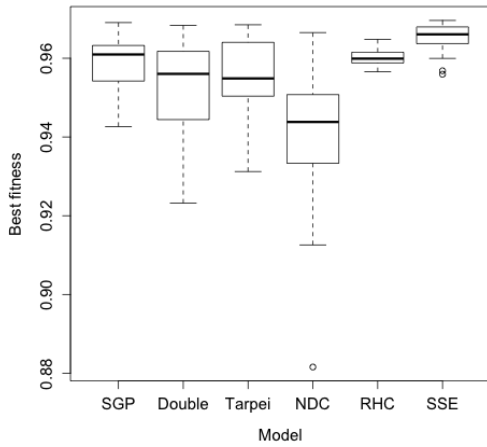


(a) 適合度

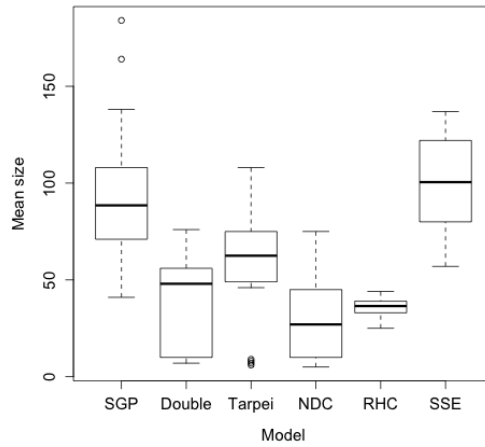


(b) 個体サイズ

Fig.3 学習セット 1 における各モデルの結果 (30 試行平均)



(a) 適合度



(b) 個体サイズ

Fig.4 学習セット 2 における各モデルの結果 (30 試行平均)

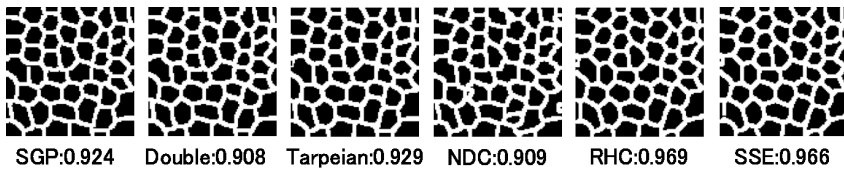


Fig.5 各モデルにおいて得られた最も低い適合度の出力画像 (学習セット 1)

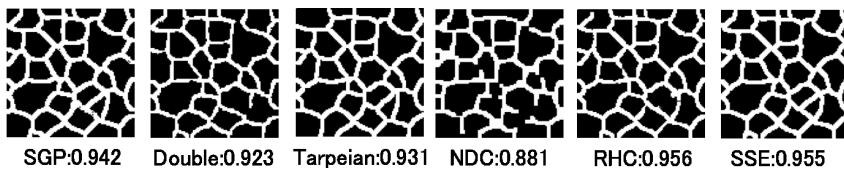


Fig.6 各モデルにおいて得られた最も低い適合度の出力画像 (学習セット 2)