

分散環境型並列分散遺伝的アルゴリズム

金子 美華^{*}, 三木 光範^{**}, 廣安 知之^{**}

^{*} 同志社大学大学院 ^{**} 同志社大学工学部

本研究では、遺伝的アルゴリズム (GA) における最適な突然変異率と交叉率を選択する作業を軽減する一つのアプローチを提案する。最適な突然変異率と交叉率は母集団サイズと問題に依存し、また、それらは単一の母集団の場合と複数の母集団の場合では異なる。提案した分散環境並列分散 GA、すなわち分散環境を持つ並列 GA は並列分散 GA におけるサブ母集団においてこれらのパラメータの値の種々の組合せを用いる。この新しいスキームの優秀な性能を実験的に検証するために4つの標準テスト関数を用いた。その結果、分散環境並列分散 GA は与えられた母集団サイズの下で、適切な交叉率と突然変異率が不明の場合には最良の解を最も早く与えてくれる方法であることが分かった。

A Parallel Distributed Genetic Algorithm with Distributed Environment Scheme

Mika KANEKO^{*}, Mitsunori MIKI^{**}, and Tomoyuki HIROYASU^{**}

^{*} Graduate School of Doshisha University ^{**} Doshisha University

This paper introduces an alternative approach to make the task of choosing optimal mutation and crossover rates easier by using a parallel and distributed GA with distributed environments. It is shown that the best mutation and crossover rates depend on the population sizes and the problems, and those are different between a single and multiple populations. The proposed distributed environment GA uses various combinations of the parameters as the fixed values in the subpopulations. The excellent performance of the new scheme is experimentally demonstrated for four standard test functions. It is concluded that the distributed environment GA is a useful method to find the best solution under a given population size and uncertainty for the appropriate crossover and mutation rates.

1 はじめに

遺伝的アルゴリズム (GA: Genetic Algorithms) は自然界の選択や遺伝学のメカニズムに基いた確率的探索アルゴリズムである [1]。GA を実問題に適用する場合、多くの世代数と個体が必要であり、良好な解を得るためには莫大な計算資源が必要である。そこで、並列コンピュータ上で GA を実行するための並列分散 GA について多くの研究がなされている [2]。

他方で、GA の性能は交叉率と突然変異率の選択に依存する。しかしながら、これらの適切な組み合わせを選択することは難しい。Tuson と Ross は最適な交叉率は解くべき問題に依存すると示した [3]。それらの結果は単一母集団の GA で得られたものであり、並列分散 GA において適切な交叉率と突然変異率を見つけるための研究はほとんどない。本論文は並列分散 GA においてそれらの遺伝的パラメータの影響を示し、いくつかのパラメー

タ調整の必要がない新しいアプローチを提案する。

2 並列分散 GA

GA の並列化に関する研究は (1) 個体の適合度評価のみを並列化する細粒度モデルと (2) 母集団を分割する粗粒度モデルの2つに大別される [4]。本論文では代表的な粗粒度モデルである島モデル [6] を並列分散 GA (PDGA: Parallel Distributed GA) と呼ぶ。

PDGA では、母集団を複数のサブ母集団 (subpopulation) に分割し、各サブ母集団毎に遺伝的操作を行う。このサブ母集団をプロセッサに割当てることによって並列処理を行う。また、一定世代毎に異なるサブ母集団間に移住 (migration) と呼ばれる個体の交換を行う。移住に関して、移住間隔と移住率というパラメータが必要となる。移住間隔は移住を行う世代間隔であり、移住率はサブ母集団の個体数に対する移住個体の割合であ

る。本論文で用いた移住の概念を図1に示す。図のように、移住元と移住先のサブ母集団は1対1の関係である。また、移住先のサブ母集団および移住個体は移住のたびにランダムに選ぶ。

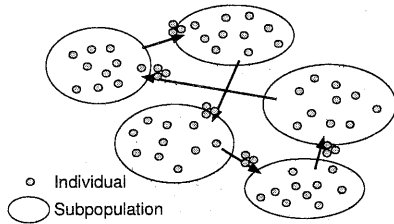


図1: 移住

3 交叉率および突然変異率の影響

3.1 最適な交叉率と突然変異率

交叉と突然変異の役割はGAでは極めて重要である[1]。交叉は母集団の中の個体間での情報交換を行うのに用いられる。一方、突然変異は進化が停滞するのを防ぐために用いられる。GAの性能は交叉率と突然変異率の最適な選択に依存している。従来の研究によれば、交叉率の最適な設定はアルゴリズム全体の他の設定すなわち、母集団サイズ、突然変異率および選択の方法などに依存している。一般的に用いられる交叉率は0.45から0.95である[5]。Tuson & Ross [3]は交叉率を0.05から0.95まで0.05刻みで変化させる大規模な研究を行った。その結果、最適な交叉率は解くべき問題によって変化した。一方、突然変異率の最適な設定も解くべき問題に依存していることが明らかとなった[7]。そして、最適な突然変異率の下限が $1/L$ ともいわれている[8]。ここで、 L は染色体の長さである。

3.2 対象問題と各種設定

並列分散GAの性能における交叉率と突然変異率の影響を調べるため、対象問題として次に示す4つの数学的関数、Rastrigin関数(F_{Ra})、Schwefel関数(F_{Sc})、Griewank関数(F_{Gr})およびRosenbrock関数(F_{Ro})[9]を10次元で用いた。

$$F_{Ra} = 10n + \sum_{i=1}^n (x_i^2 - 10 \cos(2\pi x_i)) \quad (1)$$

$$-5.12 < x_i \leq 5.12, n = 10$$

$$F_{Sc} = \sum_{i=1}^n -x_i^2 \sin(\sqrt{|x_i|}) \quad (2)$$

$$-512 < x_i \leq 512, n = 10$$

$$F_{Gr} = 1 + \sum_{i=1}^n \frac{x_i^2}{4000} - \prod_{i=1}^n \left(\cos\left(\frac{x_i}{\sqrt{i}}\right) \right) \quad (3)$$

$$-512 < x_i \leq 512, n = 10$$

$$F_{Ro} = \sum_{i=2}^n [100(x_1^2 - x_i)^2 + (x_i - 1)^2] \quad (4)$$

$$-2.048 \leq x_i \leq 2.048, n = 10$$

式(1)のRastrigin関数は $x_i = 0$ で最小値0をとり、その周辺に格子状に複数の準最適解をもつ多峰性の関数である。また、変数間に依存関係はない。式(2)のSchwefel関数も多峰性の関数であり最小値は0である。また、変数間の依存関係はない。式(3)のGriewank関数は $x_i = 0$ で最小値0をとる。大域的には単峰性であるが最小値付近では極めて多くの準最適解をもち、変数間に中程度の依存関係がある。式(4)のRosenbrock関数は $x_i = 0$ で最小値0をとる単峰性の関数であり、変数間に強い依存関係をもつ。Rastrigin関数、Schwefel関数およびGriewank関数では一つの変数を10ビットで表現し、10変数で染色体の長さは100ビットとなる。また、Rosenbrock関数では一つの変数を12ビットで表現し、10変数で染色体の長さは120ビットとなる。また、コード化にはグレイコードを用いた。

交叉率および突然変異率が解に与える影響を調べるため、以下の実験を行った。3種類の交叉率(0.3, 0.6, および1.0)と3種類の突然変異率(0.1/L, 1/L および10/L)を組み合わせた9種類のパラメータ設定が解に与える影響を調べた。ここでLは染色体長である。サブ母集団の数は9とした。サブ母集団サイズは予備実験の結果を基に対象問題に関して、小さいもの、大きいもの、そして中間のものとして、20, 50 および180を用いた。また、移住率を0.3そして移住間隔を20世代とした。選択はルーレット選択、交叉法は一点交叉を用い、エリート保存戦略を用いて1000世代まで計算を行った。結果は12試行のうちの最高と最低のものを除いた10試行の最良個体の適合度関数の値の平均で考える。このような取り扱いをすることにより、確率的事象においてごく稀に生じる異常値の影響を排除することが出来る。用いた並列計算機は64プロセッサを持つnCUBE2Eであり一つのプロセッサを並列分散GAにおける一つのサブ母集団に割り当てた。

3.3 実験結果

サブ母集団サイズ 20 での最適な交叉率 P_c と突然変異率 P_m の組み合わせ (P_c, P_m) は, Rastrigin 関数では (0.1, 1/L), Schwefel 関数では (1.0, 0.1/L), (1.0, 1/L) あるいは (0.6, 1/L), Griewank 関数では (1.0, 0.1/L), そして Rosenbrock 関数では (1.0, 0.1/L) であった. このように, 最適な交叉率と突然変異率は対象問題によって異なる. また, 交叉率と突然変異率の違いによって得られる解は大きく異なっていた. さらに, 母集団サイズ 180 での最適な組み合わせは母集団サイズ 20 の結果とは異なっていた. 実験結果より, 交叉率と突然変異率の値は, 対象問題および個体数に依存し, 解に大きな影響を与えることが分かった. したがって, 最良な解を得るためには最適な交叉率および突然変異率を設定する必要がある, そのためには多くの予備実験が必要となる

4 分散環境型並列分散 GA

4.1 分散環境スキーム

以上の結果より, 最良の解を得るためには, 交叉率および突然変異率を適切に調節しなければならない. しかしながら, 最良の突然変異率と交叉率を求めることは非常にコストのかかる仕事となる.

この問題を克服するために, 本論文では分散環境を持つ新しい並列分散 GA を提案する. このスキームでは並列分散 GA において各サブ母集団における突然変異率や交叉率という GA パラメータを互いに異なる値で設定する. 移住操作は慣用的な並列分散 GA と同様に行う. このスキームを分散環境並列分散遺伝的アルゴリズム (PDGA/DE: Distributed Environment) と呼ぶ. PDGA/DE の概念を図 2 に示す. ここでは温度計が突然変異率を表し, ハートのマークが交叉率を表している. 高い温度は高い突然変異率を示し, 大きなハートは高い交叉率を示す. このようにサブ母集団毎に異なる交叉率と突然変異率を用いることによって, サブ母集団毎に異なる環境で GA を行うため, ビルディングブロックが種々のサブ母集団で進化し, 移住操作が大域的最適解を生み出すことが期待される. また, PDGA/DE では多くのサブ母集団において交叉率と突然変異率の複数の組合せが用いられているため, これらのパラメータ調節が不要である. したがって, PDGA/DE を用いることにより, 予備的な実験なしで容易に大域的最適解を求

めることが出来る.

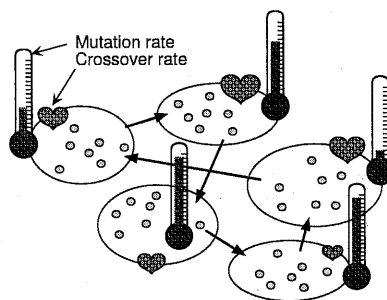


図 2: 分散環境並列分散 GA の概念図

4.2 PDGA/DE の有効性

サブ母集団数が 9 である PDGA/DE について, 前述の 4 つの数学的関数を対象問題として実験を行った. 9 個のサブ母集団における交叉率と突然変異率は, 3 種類の交叉率 (0.3, 0.6, および 1.0) と 3 種類の突然変異率 (0.1/L, 1/L および 10/L) の 9 種類の組み合わせを設定した. PDGA/DE の有効性を示すために, PDGA/DE の 9 種類の交叉率と突然変異率の一つの組み合わせを全てのサブ母集団で用いた 9 種類の並列分散 GA (これをここでは一定環境並列分散 GA ; PDGA/CE: Constant Environment と呼ぶ) と比較した. 用いた計算機は 64 プロセッサの nCUBE2E であり, 一つのプロセッサを一つのサブ母集団に割り当てた. パラメータは前述の実験と同様である.

サブ母集団サイズ 50 個体での PDGA/DE と 9 種類の一定環境の並列分散 GA (PDGA/CE) の比較を図 3 に示す. この図では, 1000 世代目に行われた解の適合度を示している. なお, ここでの結果は, 12 試行から最高値と最低値を除いた 10 試行の最良個体の適合度の値の平均を示している.

図 3 より, 先ず交叉率と突然変異率の違いによって PDGA/CE で得られる解は大きく異なることが確認できる. また, 対象問題によって, 交叉率と突然変異率が解に与える影響が異なることも確認できる.

次に, PDGA/CE と PDGA/DE の性能を比較する. Rastrigin 関数および Rosenbrock 関数では, PDGA/DE は最も良い値であった. また, Schwefel 関数および Griewank 関数では最も良い値ではなかったものの 9 種類の PDGA/CE と比較して良好な値であるといえる.

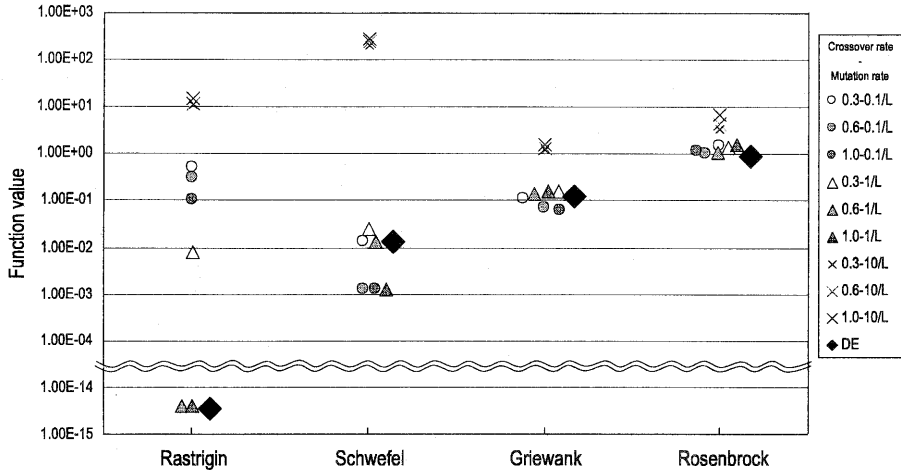


図 3: PDGA/DE と PDGA の比較

PDGA/DEにおいて良好な解が得られるメカニズムは明らかではない。原因の1つとして少なくとも1つの最適な設定のサブ母集団が存在するためであると考えられる。また別の原因として、環境の異なるサブ母集団毎に異なる性質の解探索を行うことによって、局所的かつ大域的に解探索が行われ、移住によってそれらの解が組み合わせられるためであると考えられる。

5 結論

単一母集団のGAにおいても並列分散GAにおいても交叉率と突然変異率はパフォーマンスに影響を与える。また、それらの最適な組み合わせは母集団サイズや対象問題によって異なる。並列分散GAでの最適な交叉率と突然変異率の値は単一母集団GAでの値と異なる。したがって、並列分散GAにおいても交叉率と突然変異率の最適な組み合わせを知るのは難しい。分散環境型並列分散GA (PDGA/DE) を提案し、実験的にその有効性を示した。PDGA/DEは与えられた母集団サイズの下で、適切な交叉率と突然変異率が不明の場合には最良の解を最も早く与えてくれる方法であるといえる。

参考文献

[1] J.H.Holland. *Adaptation in Natural and Artificial Systems*. University of Michigan Press, 1975.

MIT Press, 1992.

- [2] Marco Dorigo & Vittorio Maniezzo. *Parallel genetic algorithms: Introduction and overview of current research*. *Parallel Genetic Algorithms: Theory and Applications*, pp. 5-42, IOS Press, 1993.
- [3] Peter Ross Andrew Tuson.: *Cost based operator rate adaptation: An investigation*. *Proc. 4th Conference of Parallel Problem Solving form Nature*, 1996.
- [4] Cantu-Paz, E. *A survey of parallel genetic algorithms*. *Calculateurs Paralleles*. Vol. 10, No. 2. Paris: Hermes, 1998
- [5] Lashon Booker.: *Handbook of Evolutionary Computation*, chapter 3.3.1. Oxford University Press, 1997.
- [6] V. S. Gordon & D. Whitley.: *Serial and Parallel Genetic Algorithm an Function Optimizers*. *Proc. 5th International Conference on Genetic Algorithms*, pp.177-183, 1993.
- [7] Hinterding, R., Gielewski, H. & Peachey, T.C. : *The Nature of Mutation in Genetic Algorithms*. *Proc. 6th Int. Conf. on Genetic Algorithms*, pp. 65-72.
- [8] Thomas bäck.: *Optimal Mutation Rates in Genetic Search*. *Proc. 5th Int. Conf. on Genetic Algorithms, Morgan Kaufmann*, pp. 2-8.
- [9] Whitley, D., Mathias, K., Rana, S. and Dzubera, J.: *Building Better Test Functions*. *Proc. 6th Int. Conf. on Genetic Algorithms, Morgan Kaufmann*, pp. 239-246.