

疎結合分散計算環境を前提とした分散 GA の提案*

伊藤 強†

株式会社シーエーシー‡

1 はじめに

GA の組合せ最適化問題において、通常の単一の染色体による手法では遺伝子型へのコーディングが複雑になる場合がある。このような場合にも簡潔なコーディングを実現するために多種染色体による遺伝子コーディングを試行している。

本研究では、この多種染色体 GA を分散並列化することでより効率よく計算を行うことを目指した。特にここでは、近年注目されている Grid 計算環境のように、結びつきが弱く、動的に計算資源が変化し、計算資源が非均等であるような分散環境を前提とし、これを有効に利用できよう分散手法を提案する。また、この手法と非分散 GA / 通常の分散 GA とを比較検証する。

2 スキル要件による人員配置

2.1 多種染色体 GA の適用

業務アプリケーション開発プロジェクトにおける人員配置について、GA の適応を試行してきた [3]。この問題において、下記の条件に考慮し GA によって最適な組合せを求める。

- 必要なスキル要件 (技術 + 習熟度) を持ったメンバーが必要な人数いなくてはならない
- 人員コスト、すなわち要員のそれぞれのコストの合計をなるべく小さくする
- 一人の要員は一つまたは複数のプロジェクトで複数の役割をもつことができる

このとき組合せは「スキル」と「要員」という二つの軸をもつ平面上に「時間」を割り当てる問題に帰着される。このような二次元以上の情報を GA で扱う場合、従来の単一染色体を用いた計算よりも、複数の染色体を用いた計算を行った方が遺伝子コーディングが簡易になり、特別な交叉法を使用するといった必要もなくなる。また、

解の探索においてもより良い結果が得られることがある。この問題においても多種染色体 GA を用いた場合の方が良い結果を得ることができた。

2.2 分散 GA による拡張

分散 GA は GA の母集団を複数の副母集団に分割して、各母集団で独立して計算を行う手法である。この手法は各母集団の計算を並列で行うことで、計算時間の短縮が可能となるが、それだけでなく、単一母集団でおこなう GA よりも高品質な解を得られることが報告されている [2]。

この高品質な解が求められる点に着目して、複数の母集団を順次計算していく疑似的な分散 GA を試し、この問題における有効性を検証した。

3 疎結合分散計算環境を前提とした GA

今回の実験では実際に並列計算を行うことで計算時間の短縮を図った。特にここでは、近年注目されている Grid 計算環境を視野に入れ、疎結合の分散計算環境を想定して実装を行った。今回想定したのは以下のような環境である。

- 計算は複数の資源にまたがって行われる
- あるプロセスが計算資源中で使用できる割合は動的に変化する
- 計算資源を繋ぐネットワークには遅延/切断が起こりえる
- 計算資源自体が停止する可能性がある
- 非均等な計算能力の資源が混在する

このような環境下では従来の分散 GA の手法で計算を行なった場合、その効果が得られない可能性がある。そこで、ここでは分散 GA に以下のような世代交代や移住操作に関する手順の変更を加えた。

- (1) 初期集団の大きさは与えられたパラメータの値より小さい
- (2) ある世代での集団の大きさは、前世代の適応度の平均成長率をもとに、パラメータで与えられた値を最

* Distributed GA for Loosely Coupled Environment

† Tsuyoshi Itoh

‡ CAC Corporation

適値として確率的に決定される

- (3) 移住は集団の大きさによって確率的に発生する
- (4) 移住先の決定はそのつど存在する集団をさがし、その中から決定する

4 実験結果と考察

テストデータとして表1のような要件を必要とする開発プロジェクトを想定し、要員のコストや保有スキルのデータとして当社技術者のデータを使用した。

表1 要件データ

	中レベル	低レベル
顧客業務知識	2人	
サーバ(HW)		1人
サーバ(OS)	1人	
Webサーバ	1人	2人
Javaプログラム	2人	5人

このデータを使って非分散の多種染色体GA(mGA)と逐次実行による疑似分散GA(dGA)、今回提案する手法(lGA)の結果を図1に示す。また図中のsGAは通常の(非分散の)単一染色体によるGAの結果である。

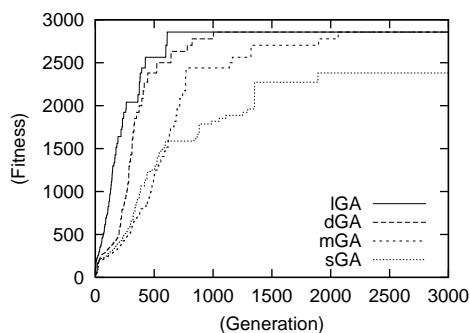


図1 実行結果

これからわかるように、提案手法は通常の分散GA(逐次計算による疑似分散GA)を用いた場合より探索効率が良くなっている。これは、計算資源を効率的に利用するために加えた手順が探索効率の向上につながったものと考えられる。

実行時間の比較が表2である。計算時間も大幅に短縮することができた。提案手法では準最適解が求まるまでの時間が、世代数に比例していないのは初期の集団の平均適応度の成長率が大きい場合それに比例して集団内の個体数が大きくなるためである。

表2 実行時間

	3000世代	準最適解まで
mGA	25分前後	15分前後
dGA	40分前後	15分前後
lGA	15分前後	8分前後

今回の実験において逐次計算による疑似分散GAでは、探索効率を上げるため各々の副集団で異なった遺伝パラメータを使用してチューニングしているため、設定すべきパラメータが多い。提案手法では、さらに集団サイズを決定する確率や移住を行う確率という新たなパラメータの設定が必要になる。しかしながら、このパラメータはeil51-巡回セールスマン問題で行ったチューニング結果の値が0/1ナップザック問題、人員配置問題とともに有効であったため、問題ごとに値を再設定する必要がないと考えている。また、提案手法では各々の副集団に異なった遺伝パラメータを設定して計算した場合も、すべての副集団に同じパラメータを設定して計算した場合と比べて探索効率の向上は見られなかった。このため設定すべきパラメータの数を大幅に減らすことができた。

5 まとめ

これらより提案手法には以下のような利点があると考えられる。

- (a) 並列計算を行うことで計算時間が短縮される
- (b) 分散GA特有のパラメータ設定は移住個体の数だけでよい
- (c) 通信の遅延/切断やある集団の計算を行う資源の停止があっても全体として計算を続行できる
- (d) 従来の分散GAより比較的探索効率にすぐれている

多種染色体GAを用いて簡便なインプリメントを可能にするという目標を考慮すると、(b)の利点は特に重要であると言える。

参考文献

- [1] John H. Holland: *Adaptation in Natural and Artificial Systems*, University of Michigan Press (1996)
- [2] 三木光範, 畠中一幸: 並列分散GAによる計算時間の短縮と解の高品質化, 日本機械学会第3回最適化シンポジウム pp.59-64,(1998)
- [3] 伊藤 強: 多種染色体GAの人員配置問題への応用, 第65回情報処理学会全国大会 pp.2-127,128,(2003)