

グローバルコンピューティング環境を利用した最適化手法に関する研究

Optimization Methods using a Global Computing Environment

谷村 勇輔

Yusuke TANIMURA

Abstract: In this study, a new model of Genetic Algorithms (GA) that is suited for the global computing environment is proposed. We assumed that the proposed model is a client server system. GA is executed in the server when the request is occurred from the client. Moreover, the client has a checkpointing function with queuing mechanism. As the numerical experimental results, we confirmed that the proposed model is an adequate model on the global computing environment.

1 はじめに

インターネットの普及と共に、遠隔地のコンピュータを結びつけるための高速なネットワーク環境が整備されてきた。これを利用して、広域ネットワーク上に分散した計算資源や情報資源を積極的に活用し、科学技術計算に代表される大規模計算を処理することが期待されている。これらは一般にグローバルコンピューティングと呼ばれる。グローバルコンピューティングに関する研究は、現在テストベッドを構築・維持するという形で欧米を中心に推し進められている。しかし、グローバルコンピューティング環境を実際の計算へどのように応用するのか、あるいはそこで何ができるのかといった応用研究は、H.Casanova らの NetSolve の研究¹⁾などが挙げられるが、ほとんど行われていない。

そこで、本研究では遺伝的アルゴリズム (GA) に着目し、それをグローバルコンピューティング上で実行するためのモデルを提案する。提案モデルはクライアント・サーバの形をとり、GA における解探索能力と障害に対するロバスト性を考慮したモデルとなっている。本研究では提案モデルをシミュレーションとして実装し、その有効性を検証した。

2 グローバルコンピューティングの特徴

グローバルコンピューティング環境で動作するアプリケーションは、主に次に述べるようなことを考慮する必要がある。本研究では特に後半の 2 者に着目し、それを考慮したシミュレーション実験を行う。

- 各計算拠点間を結ぶネットワークのスループットは非均質であり、動的に変化する。
- 各計算拠点は等しい計算資源をもっているわけではない。また、各計算機毎の計算負荷は動的に変化する。
- 様々な障害により、ネットワークに大きな遅延が発生する可能性がある。
- 様々な障害により、ある計算拠点の計算資源が利用できなくなる可能性がある。

3 2 個体分散遺伝的アルゴリズム

GA は生物の進化と淘汰を模倣した確率的な多点探索手法である。GA は強力な最適化手法の 1 つであるが、その欠点の 1 つとして高い計算負荷が挙げられる。これを克服するための 1 つの方法が並列処理である。GA は潜在的に高い並列性を有しているため、これまでに数多くの並列 GA の研究が行われている。²⁾ 代表的なモデルとしては単純 GA の並列モデル、分散 GA の並列モデル、セルラ GA の並列モデルが挙げられる。

2 個体分散遺伝的アルゴリズム (Dual DGA) は分散 GA を拡張したモデルである。³⁾ Dual DGA では 1 つの島内の個体数を 2 と設定し、アルゴリズムの簡易化を図ることで、ユーザが考慮すべきパラメータの数を減らしている。Dual DGA は高い多様性を維持することができるため、単純 GA や分散 GA に比べて優れた解探索能力をもつ。しかし、高い計算負荷が大幅に改善されるわけではないため、その並列化が望まれる。

Dual DGA の並列モデル (Dual PDGA) では、逐次モデルと異なり 2 段階の移住操作を適用する。これは、逐次モデルの時と同様にプロセッサ内で行う移住と、並列モデルにする際に新しく加えるプロセッサ間で行う移住である。前者の移住では島間で個体の交換を行い、後者の移住ではプロセッサ間で島の交換を行う。さらに後者の移住は十分な世代間隔を保って行うため、プロセッサ間の通信量を小さくすることができる。これにより、Dual PDGA は並列計算機を用いて高速に処理することが可能となる。

以下に Dual PDGA の特長を示す。これらの特長はグローバルコンピューティング環境においても十分に発揮されると思われるため、本研究では Dual PDGA を基本とした GA のグローバルコンピューティングモデルを考えることにする。

[Dual DGA]

- 分散 GA の仕組みを受け継いでいるが、分散 GA に比べてアルゴリズムが簡易化されている。
- これまでの単純 GA や分散 GA に比べて、解探索能力が高い。

[Dual PDGA]

- 高速に計算を行える。
- 通信量が少ない。
- 1つの島を粒度の単位とすることで、並列化の粒度を柔軟に変更できる。
 - 実行する並列計算機の（通信 / 計算）の割合に合わせた粒度を用いることができる。
 - ロードバランスがとりやすい。

4 グローバルコンピューティング環境における Dual DGA モデルの提案

提案モデルでは、複数の計算拠点にクライアント・サーバの関係を導入する。Fig. 1 に提案モデルの概要を示す。本研究では1つの計算拠点をサイトと呼ぶ。クライアントサイトは1つだけ存在し、それ以外はサーバサイトとなる。クライアントサイトはユーザの要求を受け、各サーバサイトに対して Dual PDGA を実行するように命令を下す。しかしグローバルコンピューティング環境では、障害により計算が途中で中断されたり、クライアント・サーバ間のネットワークに大きな遅延が発生したりすることが予想される。これらの対策として、提案モデルではクライアントサイトにチェックポイントの役割をもたせる。これは、ある一定時間毎にクライアントサイトがサーバサイトの稼働状況、および計算負荷などの情報を得る。また同時に、サーバサイト上で走っている Dual PDGA のその時点での探索点情報を取得する。取得した探索点情報はキューに格納する。これらは、他のサーバサイト上で走っている Dual PDGA や障害から復旧したサイトで実行される Dual PDGA に対して分配

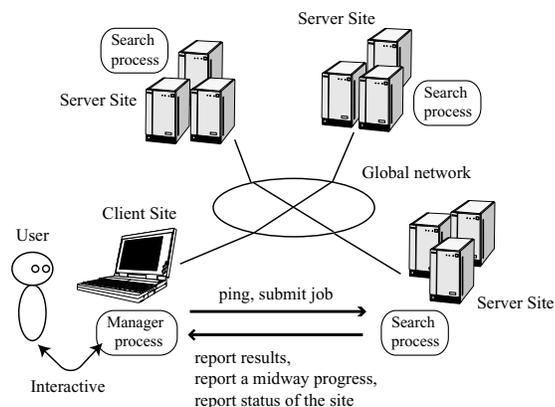


Fig. 1 Concept of proposed model

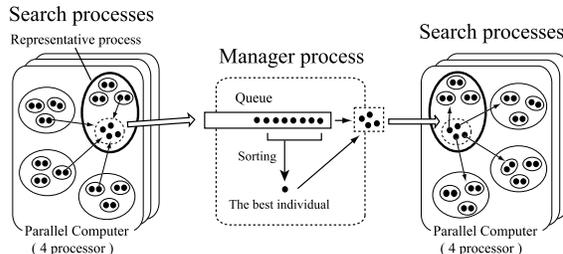


Fig. 2 Queuing operation

されることになる。キューに関する操作について Fig. 2 に示す。

このようなキュー操作を備えたチェックポイントの仕組みにより、途中の探索点情報が随時キューに保存されることになる。これらの探索点情報を利用することで、障害から復旧したサイトで実行される探索プロセスは計算を途中から始めることができる。また、他の探索プロセスと探索点情報を交換し合うことにより、探索が遅れているプロセスに対して探索を促進させる効果や、各サイトの探索プロセスが局所解へ陥るのを防ぐ効果を楽しむことができる。

5 まとめ

本研究のシミュレーションにおいては、グローバルコンピューティング環境を想定して計算途中で計算環境を動的に変化させた。どのように変化させるかを定めるシナリオとしては、「問題なく計算を実行できるシナリオ」「サーバサイトが緊急停止するシナリオ」「サイト間のネットワークに遅延が生じるシナリオ」の3つを用意した。シミュレーション実験の結果、提案モデルは複数の計算資源を有効に利用し、かつ障害に対してもロバストなモデルであることが確認できた。

参考文献

- 1) Henri Casanova, Jack Dongarra, "NetSolve: A Network Server for Solving Computational Science Problems", The International Journal of Supercomputer Applications and High Performance Computing, Volume 11, 1997
- 2) Enrique Alba, Jose M. Troya, "A Survey of Parallel Distributed Genetic Algorithms", Complexity Vol.4 No.4, John Wiley & Sons, Inc., 1999
- 3) Tomoyuki Hiroyasu, Mitsunori Miki, Masahiro Hamasaki and Yusuke Tanimura, "A New Model of Distributed Genetic Algorithm for Cluster Systems: Dual Individual DGA", Proceedings of CC-TEA, 2000