

PC-Clusterにおける並列分散GAの実装

IMPLEMENTATION OF PARALLEL AND DISTRIBUTED GENETIC ALGORITHMS USING A PC-CLUSTER

三木光範¹⁾, 廣安知之¹⁾, 三村泰成²⁾, 谷村勇輔³⁾

Mitsunori MIKI, Tomoyuki HIROYASU Yasunari MIMURA and Yusuke TANIMURA

1)同志社大学工学部

2)大分大学工学部

3)同志社大学大学院

A PC-cluster is one of parallel computers. PC-clusters have high cost performance, but they still do not have good computing performance. This paper examines the characteristics of parallel genetic algorithms (PGAs) on a PC-cluster. PGAs can be classified into two approaches. In this study, these two approaches are compared from the view point of calculation time and calculation cost. It is made clear that parallel distributed genetic algorithms (PDGAs) using the island models can find optimum solutions in short time and PDGAs are suitable for PC-clusters.

Key Words : *Parallel Computing, Cluster Computing, Genetic Algorithms*

1. はじめに

近年のコンピュータ技術の著しい進歩により、PCの性能も飛躍的に向上してきている。PC-Clusterは、そのようなPCをネットワークで接続した並列計算システムである。PC-Clusterは、従来の専用並列計算機に比べて高いコスト・パフォーマンスを提供するため、近年注目されている並列処理が容易に実現できるようになり、一般にも高く評価されるようになってきた。そこで本研究では、時間的コストがかかる遺伝的アルゴリズム（以下GA）をPC-Cluster上に実装し、その性能を検証する。GAを並列計算システム上に実装するモデルとしては、単一母集団モデルと分割母集団モデルがある。そこで本研究では、母集団を分割してGAを行う時の解の探索性能を調べ、それを踏まえて単一母集団モデルと分割母集団モデルを並列化して実行した時に短縮される計算時間を比較する。本論文では、これらの結果からそれぞれのモデルでどのような並列・分散効率があるかを明確にし、PC-Cluster上で並列分散GAを行う際の指針を示す。

2. PC-Clusterシステム

通常、PC-Clusterシステムは一般的に手に入りやすいハードウェアと、MPIやPVMなどの汎用の並列通信ライブラリを用いて構築される。これにより、従来から用いられている専用設計の並列計算システムに比べて、コスト・パフォーマンスの高いシステムを構築することが可能である。本研究で用いたPC-Clusterシステムの仕様を表1に示す。

表1 用いたPC-Clusterの仕様

Processor	Pentium 300MHz × 8
Memory	64MB × 8
BCL	10Mbps Ethernet, TCP/IP
OS	Linux 2.0.35
通信ライブラリ	MPICH 1.0.13

3. GAとその並列化

GA（遺伝的アルゴリズム）は、生物の進化を模倣した確率的多点探索手法であり、探索領域が連続な問題にでも離散的な問題にでも対応できる最適化アルゴリズムの1つである。しかしその計算負荷は高く、大規模な問題を解くには実用性に乏しいという問題点もある。その問題点の解決方法の1つとして並列計算機を用いた並列GAが考えられる。

GAの並列化には、単一母集団型と分割母集団型の2つのモデルが考えられる。前者は、GA計算の最も時間を要する評価計算の部分をデータ並列化するモデルが考えられ、PGAPack[1]等が挙げられる。後者は並列化を行う前に、GAを行う母集団を分割して複数の島に分け、各島ごとにGAを行うモデルが考えられる。またこのモデルでは、ある島間で個体を交換する移住と呼ばれる操作が必要になる。本研究では、前者の実装モデルをa)評価計算分散アプローチ、後者をb)島モデル分散アプローチと呼ぶ。

4. 並列GAモデルの検討

4.1 分散効果の影響

まず基本実験として分割母集団型GAにおける解の探索性能を調べる．本研究では，式(1)に示すRastrigin関数の最小値探索問題において，GAにおける母集団の分割による影響を調べた．

$$f = (10 \cdot N) + \left[\sum_{i=1}^N (x_i - 10 \cos(2\pi x_i)) \right] \quad (1)$$

$$x \in [-5.12, 5.12]$$

分散効果の影響を調べるために，以下を定義する．まず用いるGAのモデルを，突然変異を行わずにスケールリングを用いた選択によってGAを収束させるものとした．これにより，解の探索空間が初期個体数に大きく依存することになる．そして，9割以上の確率で既知の最適値が求まる初期個体数を必要個体数と呼ぶ．本研究では，最初に必要個体数を調べ，次にその必要個体数でGAを行った時の評価計算回数を各島ごとに調べた．そして，「各島の評価計算回数×島数」が単一母集団の評価計算回数を下回っていた時に，分散効果が得られたと考えることにする．

図1に分散効果の測定結果を示す．この図は単一母集団でGAを行った時の分散効果を100として，それを基準とした相対値を示している．この結果，Rastrigin関数の最小値問題に対しては非常に良い分散効果が得られた．そして16島が最適島数であることがわかった．

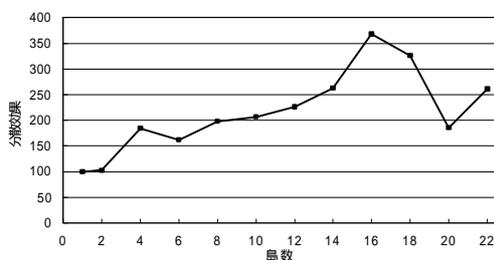


図1 分散効果の測定

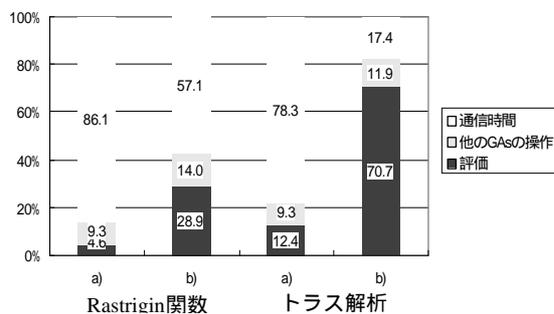


図2 計算時間の内分け

4.2 2つの並列モデルの検討

第3章におけるa)とb)の並列GAのモデルに関して，計算速度の検討を行うために計算時間の内分けを測定した．計算時間の内分けは，評価の操作時間とそれ以外のGA

の操作時間，そして並列化によって生じた通信によるオーバーヘッド時間の3つに分けられる．しかし評価計算時間は対象問題に依存するため，簡単な評価として4.1節で扱ったRastrigin関数の評価，複雑な評価として6節点10部材トラスの構造解析の評価の2つの場合を別々に測定した．図2にその結果を示す．ただしこれは，分散効果の影響が現れないようにするために，4.1節で説明した手法を用いて収束させた結果ではなく，一定の世代数でGAを終了させた結果である．

4.3 島モデル分散アプローチの島数について

島モデル分散アプローチでは図1に示したように最適島数が存在する．しかし，実際には用いることのできるPE数と最適な島数が一致しないことが多いと考えられる．図3はRastrigin関数の最小値探索問題における島数とPE数の関係について実験を行った結果である．

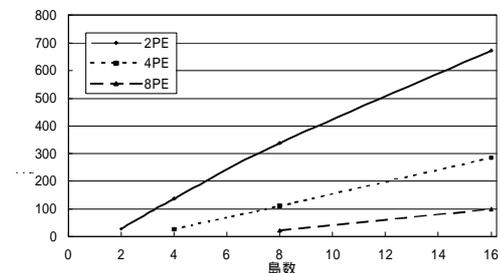


図3 島数とPE数の関係

4.4 考察

図2より，a)のアプローチはb)のアプローチに比べて通信のオーバーヘッドが大きい．さらに評価計算に時間がかかる問題においても大きなオーバーヘッドが生じるため，並列化により時間短縮を図ることはできなかった．

それに対してb)は通信量が少ないために，評価計算に時間がかかる問題に対しては，並列化により大きく時間短縮を図ることができた．そして，GAの分散効果と並列化による効果の両方が期待できることから，実際にはさらに短い時間でGAを行うことができると考えられる．また，実際に利用可能なPE数が最適島数より少ない場合には，図3よりPE数を島数としてGAを行うのが良いと考えられる．

6. 結論

以上より，PC-Clusterシステムにおいて並列GAを実装するには，島モデル分散アプローチを用いるのが良いと考えられる．今後の課題としては，島モデル分散アプローチにおける必要個体数や最適島数を知る方法，大規模な並列計算システムへの対応の問題等が挙げられる．

参考文献

- 1) D., Levine, A Parallel Genetic Algorithm for the Set Partitioning Problem, MCS-P458-0894, 1994