

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

谷村 勇輔 (知的システムデザイン研究室)

Yusuke TANIMURA (Intelligent Systems Design Laboratory)

Abstract In this paper, the characteristics of the typical two models of parallel genetic algorithms are compared. Those models are the coarse grained model and the micro grained model. Especially, the parallel efficiency on PC clusters that have not more than 10 nodes are discussed. The cluster used in this study have two kinds of network; FastEthernet and Myrinet. The characteristics are examined through the numerical examples; the Rastrigin function. In result, the micro grained model shows bad performance because of having not so high performance network on a PC-Cluster. The micro grained models shows good performance if the problems are properly big.

1 はじめに

近年のコンピュータ技術の急激な進歩により、高性能のハードウェアを低価格で入手できるようになった。ソフトウェアの面では、オープンソースなソフトウェアの普及によりオリジナルのコンピュータ環境を整えることも手軽に行えるようになった。それに加えて、いくつかの並列処理技術も徐々に一般化する傾向にあり、従来の高性能であるけれども高価な専用並列計算機ではなくとも、並列計算を行える土台が整ってきたといえる。PC-Cluster型の並列計算機は、こうした状況の先駆けとして考え出され、欧米を中心に盛んに研究 [1] がなされている。PC-Clusterは、LANに接続されたPCを1つの並列計算機として利用する手製の並列計算機のモデルであり、高いコストパフォーマンスを示すのである。

遺伝的アルゴリズム (GA: Genetic Algorithms) は、最適化問題を解くための強力な手法の1つである。GAは生物学の遺伝と進化を模倣したアルゴリズムであり、確率的多点探索を行う。GAの長所は、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ことであるが、解の探索に膨大な反復計算を必要とするために時間的コストの面から実用性に乏しいという欠点もある。一方、GAは多くの並列性を有したアルゴリズムである [2] と言われており、これまでも並列計算の分野において研究が行われている。並列遺伝的アルゴリズム (PGA: Pararell Genetic Algorithms) には、粗粒度モデルと細粒度モデルの2つの代表的なモデルが存在する。本論文では、以降の章において2つの並列モデルの比較を行い、計算時間や並列化効率を検証する。

2 並列遺伝的アルゴリズム

2.1 遺伝的アルゴリズム

GAは自然界の進化の仕組みから考え出された工学的なアルゴリズムである。図1に一般的なGAを示す。図中における各操作は遺伝的操作と呼ばれ、各世代において適度なパラメータに基づいて行われる。本論文で用いている遺伝的操作の特徴は、交叉には一点交叉、選択にはルーレット選択を用いていることである。

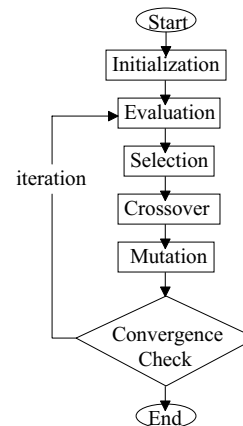


図 1: 一般的な GA

実用的な GA の並列モデルを考えると、通信によるオーバーヘッドやモデルの単純さの点から、粗粒度モデルと細粒度モデルを考えることができる。次節では、これらのモデルについて詳細を述べる。

2.2 粗粒度並列モデル

粗粒度モデルは一般に島モデルとも呼ばれる。このモデルでは、母集団を複数のサブ母集団に分割し、各母集団ごとに図1に示すような遺伝的操作を行う。そして数世代の反復計算ごとに、あるサブ母集団のいくつかの個

体を他のサブ母集団に移動させる．この遺伝的操作は移住と呼ばれる．本論文での移住は，全てのサブ母集団において同一の世代で適用し，母集団全体でみるとある世代における移住は，図 2 に示すようにリング型の個体交換モデルとなる．移住個体数は初期パラメータで与えるが，移住個体や移住先などは，移住世代ごとにランダムに選ばれることになる．粗粒度モデル GA の処理の流れを図 3 に示す．

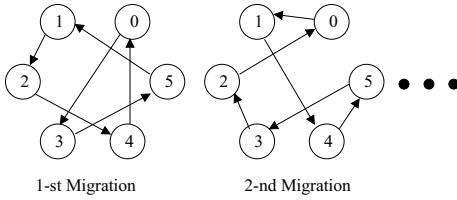


図 2: ある世代における移住の様子

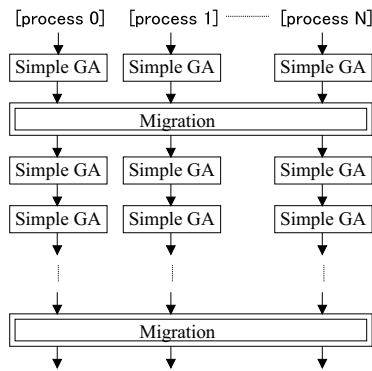


図 3: 粗粒度モデル GA の処理の流れ

ランダムに作成するリング状の移住形式では，プロセス間の通信にデッドロックが生じる可能性がある．例えば，あるサブ母集団が移住先に移住個体を送信しようとした時に，移住先の母集団が別のサブ母集団に個体を送信しようとしていて，受信できる態勢になっていないといった状況が考えられる．これを回避するために，ここでは `MPI_Sendrecv()` 関数を用いている．この関数は，通信サブシステムによって送受信の順序を調節し，周期的な依存関係を避けるように働くのである．

2.3 細粒度並列モデル

GA においては評価の操作に要する時間が，全体の計算時間の大部分を占めることが多く，対象とする問題が複雑になるほどその傾向が強くなる．そこで非常に単純な並列モデルとして評価計算の部分を分散し，並列に計算を行う細粒度並列モデルが考えられる．これはマスター・スレーブ型をとり，評価を除く全ての遺伝的操作はマスターとなる 1 つのプロセスにおいて行う．評価の操作は，マスターから複数のスレーブに評価すべき個体のデータを送信し，スレーブにおいて実際の計算を行い，

結果をマスターに返すという手順となる．ロードバランスを考慮して，マスターからスレーブへの個体データの送信は 1 個体ずつとするため，実際には図 4 に示すようなモデルとなる．

細粒度モデルに関する研究においては PGAPack[3] が有名である．PGAPack は評価関数を記述するだけで，簡単に細粒度モデルを実現できる PGA のライブラリである．しかし PGAPack もそうであるが，細粒度モデルは実行環境として専用の並列計算機を想定している．そのために，ノード数の少ない PC-Cluster ではマスタープロセスを走らせるノードが 1 つ必要となるために，8 ノードを用いてもその並列化効率の上限は $7/8$ となる．さらにこのモデルでは通信時間に大きく影響される．個体の送受信に要する時間が個体の評価計算に要する時間よりも小さい時には，全てのスレーブノードを使って計算することはできず，計算資源に無駄が生じてしまうのである．

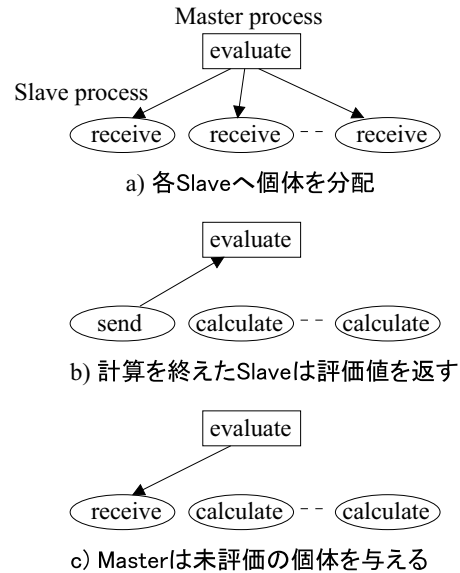


図 4: 個体の分散計算

3 粗粒度モデルと細粒度モデルの比較

前章では，2 つの PGA のモデルについて紹介した．本章では，それぞれのモデルにおいて計算時間を測定し，結果からわかる特徴について述べる．

3.1 使用する計算機環境

実験で用いた PC-Cluster 型の並列計算機は 8 台の PC から構成される．各ノードの接続には，FastEthernet と Myrinet を用いている．FastEthernet はスイッチング・ハブによって接続されている．Myrinet は Myricom 社により開発されたギガビットネットワークである [4] ．Myrinet のホストインタフェースには LANai と呼ばれる RISC プロセッサが搭載され，Myrinet 上の通信プロ

トコルを制御し、高速な通信を実現している。表 1 に用いた PC-Cluster のスペックを示す。また並列プログラムの記述には MPICH[5] と呼ばれる通信ライブラリを用いて、メッセージ・パッシング型の並列プログラムで PGA を記述している。

表 1: クラスタのスペック

CPU	Pentium2(Deschutes)	
Memory	128MB	
OS	Linux2.2.10	
Network	Ethernet	Myrinet
	TCP/IP	GM1.02
通信ライブラリ	MPICH1.1.2	

3.2 実験方法

計算速度の測定を行う実験において、テスト関数として Rastrigin 関数を用いる。Rastrigin 関数は式 (1) のように表される。ここで f は適合度値であり f を最大にする x_i を求める。実際には f の最大値は 0.0 であることがわかっている。

$$f = -10n - \sum_{i=1}^n (x_i^2 - 10\cos(2\pi x)) \quad (1)$$

Rastrigin 関数は比較的大きな局所解をもつが、設計変数間に依存関係がないために GA では容易に最適解を見つけることができる。本実験においては、5次元の Rastrigin 関数と 15次元の Rastrigin 関数を用いる。各設計変数は 10ビットで表現されるため、それぞれの問題において遺伝子長は 50ビットと 150ビットになる。それぞれの問題において、交叉率は 0.6、突然変異率は遺伝子長分の 1、移住間隔は 5、移住率はサブ母集団の個体数の 0.15 とした。個体数は総個体で 5次元の場合に 960個体、15次元の場合に 1920個体とした。つまり、粗粒度モデルにおいては各サブ母集団の個体数は総個体数/サブ母集団数となる。

実験では、さらにこの評価関数に一定の待ち時間を埋め込んで、評価計算に要する時間と通信時間の関係から、全体の計算時間がどのように変化するかを調査する。

3.3 粗粒度並列モデル GA の実験結果

粗粒度並列モデル GA を Ethernet-Cluster で行った時の計算時間の測定結果を図 5 に、Myrinet-Cluster での測定結果を図 6 に示す。これより Ethernet では、評価計算時間が小さい時にはプロセッサ数を増やしても計算時間が短縮されていない。それに対して、Myrinet では

時間短縮が行われている。また遺伝子長が 150bit になると、通信量が 3 倍となるため、全体の計算時間に影響が現れているのがわかる。

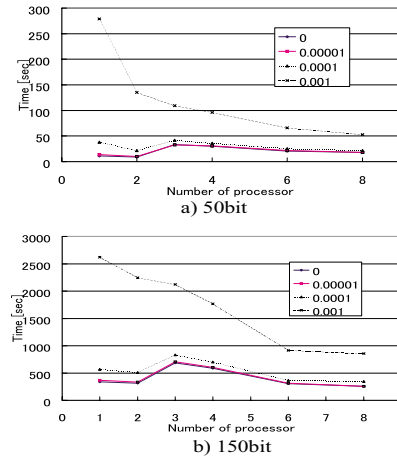


図 5: 粗粒度並列モデル GA の実行時間 (Ethernet)

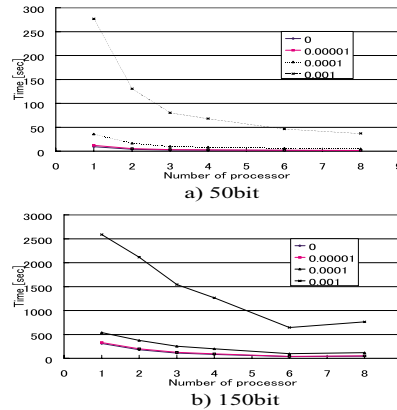


図 6: 粗粒度並列モデル GA の実行時間 (Myrinet)

3.4 細粒度並列モデルの実験結果

細粒度並列モデル GA を Ethernet-Cluster で行った時の計算時間の測定結果を図 7 に、Myrinet-Cluster での測定結果を図 8 に示す。細粒度並列モデルでは Ethernet でも Myrinet でも、評価計算部分に入れる待ち時間を非常に大きい場合でないと、計算時間の短縮を行うことができなかつた。これは粗粒度並列モデル GA に比べて、通信量および通信回数が大きくなるために、その部分のオーバーヘッドがかなり生じてしまうからであると考えることができる。さらに図中では、ある数以上のプロセッサを用いても実行時間が変化していないことがわかる。これは個体の分配を行う通信に時間がかかり、ほとんど未評価の個体を与えられないプロセッサが存在するためである。

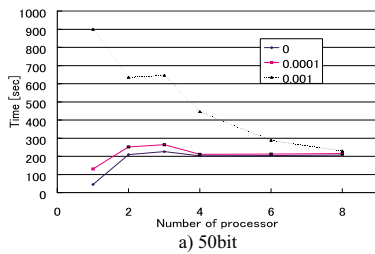


図 7: 細粒度並列モデル GA の実行時間 (Ethernet)

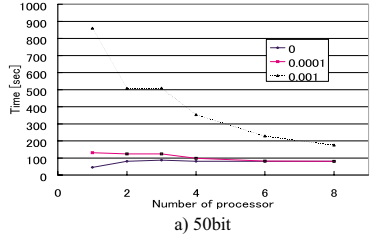


図 8: 細粒度並列モデル GA の実行時間 (Myrinet)

4 結論

本論文では、粗粒度並列モデル GA と細粒度並列モデル GA の計算時間の比較を行った。これは対象問題の複雑さから生じる個体評価の計算時間と、並列に処理を行う時に必要な通信のオーバーヘッド時間との関係によって決まる。実験結果より、今回用いたような小規模な PC-Cluster では特に通信に気を配る必要があり、細粒度並列モデル GA は通信量が過多であるためにあまり PC-Cluster に適したモデルではないといえる。それに対して、粗粒度並列モデル GA では通信量がそれほど多くないので、評価計算に要する時間がある程度大きい場合や高速なネットワークを使うことで、十分に速度向上が望めるといえる。今後は、高速に動作する移住モデルの構築やノード数の多い PC-Cluster での研究が必要であると思われる。

参考文献

- [1] <http://beowulf.gsfc.nasa.gov/>
- [2] Goldberg, D.E., Genetic Algorithms in Search, Optimization and Machine Learning, Addison-Wesley, Reading, MA, 1989
- [3] D., Levine, A Parallel Genetic Algorithm for the Set Partitioning Problem, MCS-P458-0894, 1994
- [4] <http://www.myri.com/>
- [5] <http://www.mcs.anl.gov/mpi/index.html>