

グリッド環境を想定した多目的遺伝的アルゴリズムの 並列モデルの提案およびその検討

吉井 健吾[†], 廣安 知之^{††}, 三木 光範^{††},

[†] 同志社大学大学院 ^{††} 同志社大学工学部

本研究ではグリッド環境に対応した多目的 GA の並列モデルを提案し、ヘテロな計算環境を使用した数値実験によりその有効性を検討する。提案モデルは通常の多目的 GA に近傍交叉、計算資源の性能に適応した生成子個体数の変化の 2 つの機能を拡張させたものであり、マスタースレーブモデルを利用して並列化を行う。計算資源の性能に適応して生成子個体数を動的に変化させることにより、全ての計算資源を最大限に利用することができると考えられる。本稿では提案モデルを NSGA-II に組み込み、生成子個体数の変化による解探索能力を調査することにより提案手法のアルゴリズムとしての有効性を検証するとともに、ヘテロな計算環境における数値実験により提案モデルの有効性を検証した。数値実験の結果、提案モデルでは全ての計算資源を最大限に利用し、かつオーバーヘッドの影響を軽減することができ、多様性の優れた幅広い非劣解集合を維持しながら探索を行うことができることを確認した。

Discussion of Parallel Model of Multi-Objective Genetic Algorithms on Grid

Kengo YOSHII[†], Tomoyuki HIROYASU^{††}, Mitsunori MIKI^{††}

[†] Graduate Student of Engineering, Doshisha University

^{††} Knowledge Engineering Dept., Doshisha University

[‡] Undergraduate student, Doshisha University

In this paper, a parallel model of Multi-objective Genetic Algorithm supposing a hetero calculation environment is discussed. In this parallel model, neighborhood crossover is applied to the conventional Multi-objective Genetic Algorithm. The neighborhood crossover performs the crossover operation between individuals that are close to each other in the objective space. In the proposed model, the number of offsprings generated by neighborhood crossover is changed dynamically adapting to the performance of the calculation resources. Through the numerical experiments on heterogeneous computational resources, we found that our proposed parallel model enabled to utilize the maximum performance of all calculation resources and obtain a wider variety of individuals in the objective.

1 はじめに

実最適化問題の多くは複数の評価基準が存在し、評価基準が互いにトレードオフの関係にあることが多い。このような問題を多目的最適化問題として捉え、どの解にも劣らない解の集合であるパレート最適解集合を一度に求める進化的多目的最適化 (Evolutionary Multiobjective Optimization: EMO) に関する研究が盛んに行われている。一方、EMO の問題点として高い計算負荷が挙げられる。実問題の多くは膨大な計算時間を必要とすることから、並列処理により計算時間を短縮させることは重要な課題であると言える。EMO の並列に関する研究も数多くなされているが、その並列度は小さい場合が多い^{1, 2)}。一方、世界中の計算資源を仮想的に

統合したグリッドコンピューティングの技術の進歩により、数多くの資源を容易に使用可能になりつつある。そのためこのグリッド環境を利用した並列モデルについても考慮する必要がある。

そこで本研究では、グリッド環境を想定した多目的遺伝的アルゴリズムの並列モデルを提案する。提案モデルには多目的遺伝的アルゴリズムに近接した個体同士を交叉させる近傍交叉を組み込み、また計算資源の性能に適応させて生成子個体数を変化させることを考える。本稿ではまずアルゴリズムとしての性能評価を、テスト関数を使用した数値実験により行い、その後ヘテロな計算資源を使用した数値実験により提案モデルの有効性を検討する。

2 グリッド環境を想定した多目的遺伝的アルゴリズムの並列モデル

本章ではヘテロなグリッド環境を想定した並列多目的遺伝的アルゴリズムを提案する。まずグリッド環境における並列モデルとして考慮すべきことを以下に示す。

- グリッド環境に存在する計算資源はそれぞれ性能が異なる。そのため、ヘテロな環境に対応したアルゴリズムを考慮する必要がある。
- 通信時間などのオーバーヘッドは、評価計算時間と比較して十分小さいものである必要がある。

グリッド環境では性能の異なる計算資源が存在するため、全ての計算資源に同じ個体数の評価を行うように並列化する場合、性能の劣った計算資源が評価に時間がかかり、世代交代に悪影響を及ぼすことになる。そのため、計算資源に適応した仕事量の決定を行う必要がある。我々はこれらの問題を考慮し、マスタースレーブモデルを拡張した多目的遺伝的アルゴリズムの並列手法を提案する。

2.1 基本モデル

本節では、提案手法であるグリッド環境に対応した並列多目的遺伝的アルゴリズムの説明を行う。提案手法では、NSGA-IIやSPEA2などの多目的GAに2つの機能を拡張させたものである。一つ目は近傍交叉であり、目的関数空間で近接している個体同士で交叉を行う。二つ目は、計算資源の性能に適応した生成子個体数の変化である。提案手法では目的関数空間で近接している2個体が交叉ペアとしてスレーブプロセスに送信され、スレーブプロセスは性能に適応して交叉回数を変化させ、交叉回数の数の子個体を生成する。そして一定時間後、マスタープロセスは全てのスレーブプロセスから最良の2個体を受信することにより、世代交代を円滑に行うことが可能となる。提案手法のモデルを図1に示す。

2.2 近傍交叉

一般に代表的な多目的遺伝的アルゴリズムの手法では1点交叉もしくは多点交叉を行う。しかし、交叉ペアとなる個体はランダムに選ばれ、個体間の設計変数空間および目的関数空間における距離が大きく離れ、効果的な探索ができないという問題点が存在する。そのため、我々は目的関数空間で近接している個体同士で交叉を行う近傍交叉を提案している。近傍交叉を用いることにより、多様性の優れた解集合を得ることができる。

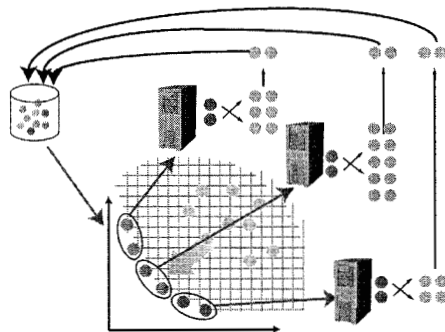


図 1: Basic model of the proposed parallel EMO

表 1: Calculation Resources

	CPU 数	CPU	メモリ
クラスタ A	10	Pentium4 2.8GHz	1GB
クラスタ B	15	Xeon 2.4GHz	1GB
クラスタ C	15	PentiumIII 1GHz	512MB
クラスタ D	10	PentiumIII 600MHz	256MB

2.3 生成子個体数の変化

一般的な多目的GAでは2個体の親個体から1回の交叉により2個体の子個体を生成するが、提案モデルでは計算資源の性能に適応して交叉回数が増加し、生成される子個体の数も変化する。つまり性能の優れた計算資源では多くの子個体を生成し、性能の劣る計算資源では少ない数の子個体を生成することになる。

このアルゴリズムでは、生成子個体数が増加するほど、1世代あたりの評価計算も多くなる。また生成子個体母集団のうち2個体以外は淘汰されるため、精度の高い個体群のみ生き残る仕組みになっている。

3 数値実験

本章では、ヘテロな計算環境を使用して数値実験を行い、提案並列モデルの有効性を検討する。

3.1 実験環境と実行手順

本実験では1台のマスタープロセスと4つの異なる性能を持つPCクラスタを用い、計50CPUのスレーブプロセスを利用して提案手法の有効性を検証する。使用した計算資源を表1に示す。クライアントから各PCクラスタへのジョブの投入にはGridRPCの一つであるNinf-G(version 2.4)を利用し、各PCクラスタ内でのスケジューリングにはOpenPBS(version 1.2)を用いた。

またジョブの実行の流れを図2に示す。クライア

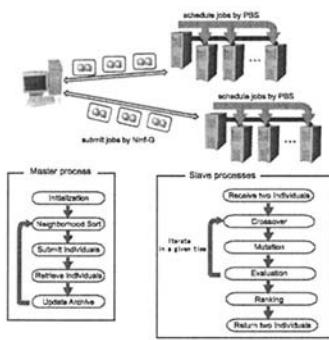


図 2: Execution Flow of Proposed Algorithm on heterogeneous computational resources

ントはまず初期母集団を生成し、目的関数空間において近接している順にソートを行う。次にソートされた母集団に対して、隣り合う 2 個体ずつ各 PC クラスタのマスに Ninf-G によりジョブの投入を行う。この際送信するデータは 2 個体の遺伝子情報である。そしてこれらの情報を受信した各 PC クラスタのゲートウェイノードは PBS によりスケジューリングを行いスレーブにジョブを配布する。各スレーブは受信した 2 個体に対し交叉、突然変異、評価を一定時間の間繰り返す、一定時間後ランキング処理を行って最も良い 2 個体を選択し、クライアントに送信する。この際クライアントに送信するデータは 2 個体の遺伝子情報、そして目的関数値になる。これらの処理を全てのスレーブに対して同期通信により行う。

本実験では各計算資源は一定時間の間、交叉、突然変異、評価を繰り返すことにより、計算資源の性能に適應した生成子個体数の変化を実現している。これにより、全ての計算資源は一定時間後ほぼ同時に処理を終了させることができ、性能の劣る計算資源による遅延を防ぐことができる。なお対象問題の計算負荷と利用する計算資源を考慮して設定する時間を考慮する必要がある。

対象問題は Kursawe の数値実験に使用された KUR を用い、無駄な計算を付加させることにより計算負荷を増大させている。計算負荷を増大させたときの各 PC クラスタの計算資源による 1 個体の評価時間はそれぞれ PC クラスタ A で 5.82 秒、B で 8.62 秒、C で 10.17 秒、D で 17.06 秒であった。このとき、最も性能の劣る PC Cluster D の計算資源が最低 2 個体を生成できるように、1 分間の間繰り返し交叉、突然変異、評価を行うようにした。

比較する対象として、オリジナル NSGA-II を同じ環境と同じ計算負荷の評価時間でマスタースレー

ブモデルのよる並列化を行った場合を用いた。終了条件は 2 時間経過した時点とした。なお、パラメータは、個体数 100、次元数 100、染色体長 20 × 次元数、交叉率 1.0、突然変異率 1/染色体長とした。

3.2 評価手法

また得られた非劣解集合を評価する手法は様々存在するが、本研究では I_{cover} , Spread, Hypervolume, Ratio of Non-dominated Individuals:RNI の 4 つの評価方法を使用する。

I_{cover} は目的関数空間におけるパレート最適解領域において、解集合が均一に分布しているかを評価する方法である。 I_{cover} は 0 から 1 の実数値で表現され、1.0 に近いほど解が全領域に求まっていると評価される。

Spread は、得られた解集合の各目的に対する最大値と最小値の差を足し合わせたものであり、値が大きいくほど幅広いパレート最適解を得ていることがわかる。

Hypervolume は得られた非劣解集合が支配している領域の大きさを計算したものであり、パレートフロントへの収束度、多様性、幅広さを総合的に評価することができる。

RNI は、2 つの非劣解集合を相対的に比較する手法であり、2 つの手法で得られた解集合の和集合をとり、その和集合から非劣解が存在する割合を各手法に対して導き出す。このため、この割合は最大値の 100 % に近いほど、もう一方の手法を優越していると判断することができる。

3.3 実験結果

図 3 に実験結果を示す。実験結果は 3 回試行における平均値である。図 3 から、提案並列モデルは標準のマスタースレーブモデルと比較して全ての評価手法において優れた結果を得ていることが確認できる。特に I_{cover} , Spread においては約 2 倍の性能を得ることが確認できる。図 4 に 3 回試行で得られた全てのパレート最適解をプロットした図を示す。図 4 から提案モデルは幅広い多様性の優れたパレート解集合が得られていることが視覚的にも確認できる。

3.4 考察

提案モデルでは、全ての計算資源は一定時間後ほぼ同時に処理を終了するため、計算資源のアイドル時間を最小限にすることができる。またスレーブでの処理時間を長くすることにより、通信時間やスケジューリングに要する時間などのオーバーヘッドも削減することが可能となる。これらのことを確認するため、各スレーブの数値実験中における

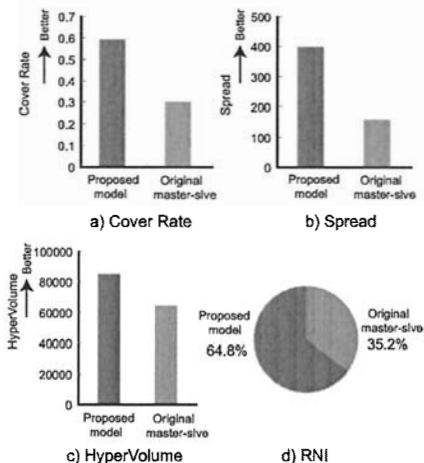


図 3: KUR: Results of Icover, Spread, and RNI

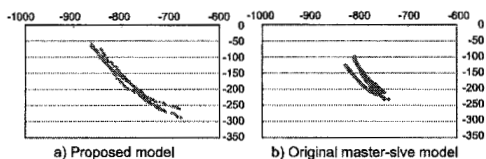


図 4: KUR: Solution set of proposed algorithm and original NSGAI

平均 CPU 使用率を表 2 に、実行時間中における総オーバーヘッドを表 3 にそれぞれ示す。2 から、標準のマスタースレーブモデルでは性能が優れた計算資源ほどアイドル時間が長くなっており、性能の劣る計算資源が足を引っ張っていることが確認できる。一方、提案モデルでは全ての計算資源が均一に負荷がかかっており、最大限に使用されていることが確認できる。また 2 からオーバーヘッドに関しても、提案モデルでは総実行時間に対して少い割合維持されていることがわかる。これらのことから提案モデルでは全ての計算資源を最大限に使用可能とし、オーバーヘッドの影響も抑えることができるため、グリッド環境に適した並列モデルであると考えられる。

最後に、図 4 に提案モデルと標準マスタースレーブモデルの探索履歴を示す。図 4 は実行時間が 30 分、60 分、120 分経過した時点での非劣解集合をそれぞれプロットしたものである。標準マスタースレーブモデルと比較して、提案モデルでは、探索の序盤から幅広い多様性の優れた非劣解集合が得られている。つまり、提案モデルでは多様性を維持しながら幅広く探索が可能である。

表 2: Average CPU usage rate of a process in each PC cluster

	A	B	C	D
提案モデル	88%	91%	95%	91%
標準マスター スレーブモデル	18%	44%	53%	89%

表 3: Overhead time and rate in total execution time

	Time	Rate
提案モデル	11m55s	17%
標準マスター スレーブモデル	55m9s	46%

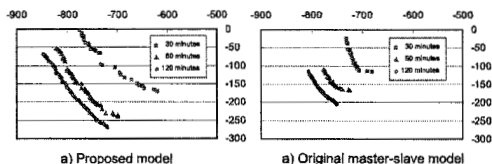


図 5: KUR: Search history of the proposed algorithm and the original NSGAI

4 まとめ

本稿はグリッド環境を想定した進化的多目的最適化の並列モデルを提案し、その有効性をヘテロな計算環境を使用した数値実験により検討を行った。提案モデルでは、NSGAIなどの多目的 GA に我々の近傍交叉を組み込み、計算資源に適応して生成子個体数を変化させることを考えた。数値実験の結果、提案モデルでは全ての計算資源を最大限に利用し、かつオーバーヘッドの影響を軽減することができ、多様性の優れた幅広い非劣解集合を維持しながら探索を行うことができることを確認した。

参考文献

- 1) Kalyanmoy Deb, Pawan Zope, and Abhishek Jain. Distributed computing of pareto-optimal solutions with evolutionary algorithms. In *EMO*, pages 534–549, 2003.
- 2) F. Streichert, H. Ulmer, and A. Zell. Parallelization of multi-objective evolutionary algorithms using clustering algorithms. In Carlos A. Coello Coello, Arturo Hernandez Aguirre, and Eckart Zitzler, editors, *EMO 2005*, volume 3410 of *LNCS*, pages 92–107, Guanajuato, Mexico, 9–11 March 2005.