

PC クラスタシステムにおける並列分散遺伝的アルゴリズム

Parallel Distributed Genetic Algorithms on PC Cluster Systems

正 廣安 知之(同志社大工) 正 三木 光範(同志社大工)

Tomoyuki HIROYASU, Doshisha University, tomo@is.doshisha.ac.jp
Mitsunori MIKI, Doshisha University, mmiki@mail.doshisha.ac.jp

In this study, the models of parallel distributed genetic algorithms are summarized and a new model of distributed genetic algorithm is proposed. The models of parallel distributed genetic algorithms are classified into three categories; those are a master slave model, a distributed GA model, and a cellular GA model. The models of master slave and distributed GA are suited for PC cluster systems. The characteristics of these models are compared and discussed thorough the numerical examples. From this comparison and discussion, a distributed GA model has higher searching ability and it is much suitable for PC cluster systems. We also proposed a new model of distributed GA; that is called "Dual Individual Genetic Algorithm: DuDGA". In this model, there are only two individuals. Because it is needles to set some parameters of distributed GAs in this model, this model is very useful for users. More than that, DuDGAs have high searching ability. This characteristics are also discussed through the numerical examples.

Key word: Evolutionary Computation, Parallel Processing, Distributed Processing, Genetic Algorithms, PC Cluster Systems, Master Slave Model, Distributed GA Model, Cellular GA Model

1 はじめに

遺伝的アルゴリズム (Genetic Algorithms: GA) は生物の遺伝と進化を模擬した確率的探索手法の一つである。GA では、連続空間問題だけでなく離散空間問題の最適化も行える点、コーディングとデコーディングと呼ばれる探索空間変換のメカニズムにより、局所解を多く含むような問題が単純な問題に変換できるようになる場合がある点、真の最適解が求まるかは別にして複雑な問題も比較的容易に GA により解を求めることが可能な点、GA のアルゴリズムが生物的背景を持つ点など、さまざまな特徴により、近年、広く利用されている。^{1, 2, 3)}

しかしながら、遺伝的アルゴリズムは多点探索であるため、設計変数の数が小・中規模な問題においては、従来の傾斜法などと比較して繰り返し計算が多く必要となるという問題が生じる。特に実問題においては、目的関数値や制約条件値を求める部分が計算コストが高くなり、このような場合には GA での最適化は非常に計算時間が必要な手法となる。この問題を解決する手法の一つが GA を並列計算機上で処理し、高速に解を求めるという方法であろう。

上記のような背景以外に、GA の並列処理が着目されている理由として、近年の計算機環境の劇的な変化という側面もあげられる。すなわち、ハードウェアの性能の向上とそのコストの急落により、従来、単体として使用されてきたワークステーションや PC をネットワーク接続により複数台連結して使用することで、能力の高い並列計算機を構築することが可能となってきたのである。これがいわゆるクラスタ型並列計算機である。アルゴリズムを並列にて処理を効率良く行うためには、アルゴリズムの

多くの部分が並列にて処理可能である必要がある。これまでに、計算機環境が変化してきたのに対して、従来の傾斜法などの最適化手法では、そのアルゴリズムを効率良く並列にて処理することが非常に難しいという問題が存在していたため、十分に効率的なハードウェアの利用が困難であった。それに対して GA は多点探索であり、かつ、種々の遺伝的操作を含んでいるため、アルゴリズム自体に並列性が内在していると言われている。⁴⁾ この特性により、多くの GA の並列モデルが提案され、検討が行われて来ており、その有効性が確認されてきた。^{5, 6, 7)} 今後、さらに多くの CPU を使用する環境が身近になり、大規模問題を対象にするようになると、GA の持つ並列性という長所が繰り返し計算を多く必要とするという欠点を補うようになるものとも考えられる。

そこで本研究では、まず、GA の並列分散モデルのうち代表的な 3 つのモデルについて、その概要および特徴を述べる。さらに、本研究では、その中でも PC クラスタシステム上でよく利用されている 2 つのモデル、マスター・スレーブモデルと分散 GA モデルの基本的な性能の違いを数値計算を通じて検討する。その結果を踏まえて、PC クラスタシステムに適したモデルである「2 個体分散遺伝的アルゴリズム (Dual Individual Distributed Genetic Algorithm: DuDGA) の提案を行い、その有効性を検討する。

2 並列分散遺伝的アルゴリズム

GA は多点探索アルゴリズムの一つであり、いくつもの遺伝的操作が含まれるために、基本アルゴリズムに並列性が内在する。そのため、並列化を行うことは、他の最適化アルゴリズムと比較すると容易であり、いくつも

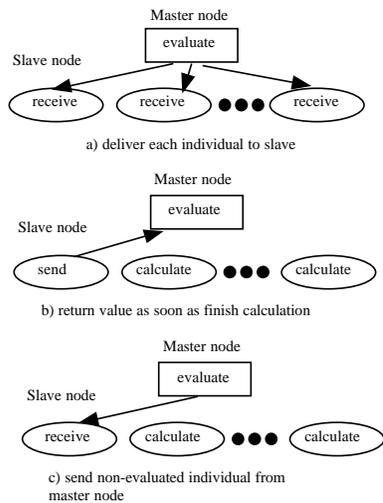


図 1: The flow of fine grained model

のモデルがこれまでに提案されている．それらは下記の 3 種に大別できる．

- マスタ・スレーブモデル⁸⁾
- 分散 GA モデル^{8, 9, 10, 11, 12, 13, 14, 15)}
- Cellular GA モデル¹⁶⁾

GA において，各個体の適合度の評価部分が計算のホットスポットとなる場合が多い．そのため，この部分を並列処理にて高速化を行うことが考えられる．このモデルがマスタ・スレーブモデルである．マスタ側では，適合度値を求める以外の操作を行い，適合度値を求める際には，個体情報をスレーブに送信し，スレーブは求めた値をマスタに送信する．マスタはすべての個体の適合度値が求まるまで，個体情報を順次スレーブに送信する．

このモデルは基本的には逐次モデルと同一であり，そのため，実装は非常に容易である．通常は個体数の方がプロセッサ数よりも多く存在すると考えられ，実問題では適合度値を求めるのに時間が多く必要となるため，このモデルは有効であると考えられている．ネットワークの粒度の点から本モデルは Fine Grained Model と呼ばれる場合がある．マスタ・スレーブモデルの概要を図 1 に示す．

先に述べた通り，GA は多点探索手法であり，各探索点は個体と呼ばれる．また，個体集合は母集団と呼ばれる．分散 GA (Distributed Genetic Algorithm: DGA) モデルはこの母集団をいくつかのサブ母集団に分割し，各母集団の中で GA の遺伝的操作を繰り返す．このサブ母集団は島と呼ばれる．数世代の遺伝的操作の繰り返しの後，サブ母集団内でいくつかの個体を選択され，他の島への移動が行われ，この操作が移住と呼ばれるからである．そのため，本モデルは島モデルと呼ばれる場合がある．移住が行われる世代数を移住間隔，移住の際選択される個体

のサブ母集団サイズに対する割合を移住率と呼ぶ．分散 GA の流れの概念図を図 2 に示す．

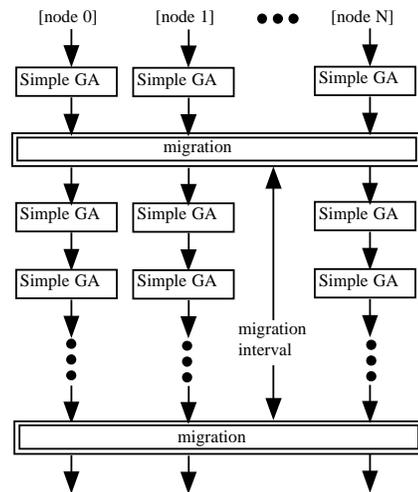


図 2: The flow of coarse grained model

このモデルは逐次モデルとは異なった並列モデルであり，通信を必要とする移住は毎世代行われるわけではないので，通信コストが低いことが特徴である．ネットワークの粒度から Coarse Grained Model と呼ばれる場合がある．

Cellular GA モデルは例えば図 3 で示すような 2 次元に個体を配置し近傍を定義して，近傍のみの個体で交差や選択を行うモデルである．図 3 では灰色で囲まれた部分が中央の cell の近傍を示している．

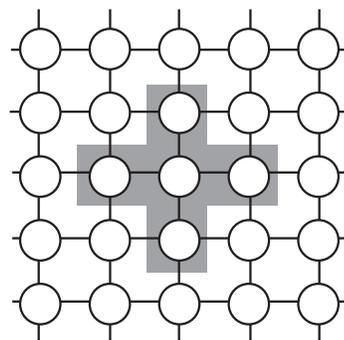


図 3: Cellular GA model

本モデルは近傍で通信を多く発生するためネットワークの粒度から Fine Grained Model と呼ばれる場合がある．近傍でのネットワークの頻度が多いため，通常，超並列計算機で使用され，1CPU に 1 島が割り当てられる．探索能力が高いモデルであるが，PC クラスタ型並列計算機のような，超並列計算機と比較するとネットワークが遅いアーキテクチャでは通信の部分がボトルネックとなる．本モデルも，逐次モデルとは異なり並列処理用のモデルである．

3 マスタ・スレーブモデルと分散 GA モデルの比較

前章では GA の並列モデルとして、マスタ・スレーブ、分散 GA、セルラー GA の 3 つのモデルを説明した。セルラー GA は通信を多く必要とするモデルであり、超並列計算機向けのモデルである。本章では、近年利用が進んでいるクラスター型並列計算機向けのモデルであると考えられるマスタ・スレーブモデルと分散 GA モデルの基本性能を数値計算を通じて検討を行う。

3.1 テスト関数と使用したクラスターシステム

検討を行うために使用したテスト関数は下記に示すような 4 種類の関数である。

$$F_1 = 10N + \sum_{i=1}^N \{x_i^2 - 10\cos(2\pi x_i)\} \quad (-5.12 \leq x_i < 5.12) \quad (1)$$

$$F_2 = \sum_{i=1}^N \{100(x_{2i-1} - x_{2i}^2)^2 + (x_{2i} - 1)^2\} \quad (-2.048 \leq x_i < 2.048) \quad (2)$$

$$F_3 = 1 + \sum_{n=i}^{10} \frac{x_i^2}{4000} - \prod_{n=1}^N \cos\left(\frac{x_i}{\sqrt{i}}\right) \quad (-512 \leq x_i < 512) \quad (3)$$

$$F_4 = \sum_{i=1}^N \left(\sum_{j=1}^i x_j\right)^2 \quad (-64 \leq x_i < 64) \quad (4)$$

使用した設計変数 (ND) およびビット数 (NB) は 1 に示す通りである。

表 1: Test Functions

	Function Name	ND	NB
F_1	Rastrigin	20	200
F_2	Rosenbrock	5	50
F_3	Griewank	10	100
F_4	Ridge	10	100

使用した GA パラメータを表 2 に示す。

また、使用した計算機環境は表 3 に示すような 16CPU からなる PC クラスターシステムである。

3.2 モデルの比較

図 4 にテスト関数 F_1 の総個体数と得られる解の値の関係を示す。これらの値は 30 試行の平均として得られたものである。また、島数は 1, 4, 8, 16 の場合を示し、マスタ・スレーブモデルは 1 島に対応している。

この図からもわかるように、分散 GA モデルでは、島数がいずれの場合にも、マスタ・スレーブモデルと比較

表 2: Used parameters of GA

gene length	design variable*10
coding	gray coding
crossover	one point crossover
crossover rate	0.6
mutation	0>>1 1>>0
mutation rate	1/gene length
topology of migration	ring that is constructed random at every migration chance
migration interval	5
migration rate	0.15

表 3: Speck of 16 PC Cluster System

CPU	Pentium II (400 MHz) × 16
Memory	128 Mbytes
OS	Linux 2.2.10
Network	FastEthernet Switching HUB TCP/IP
Communication library	MPICH 1.1.2

して、同一の総個体数であれば、解の精度が高いことがわかる。すなわち、分散 GA モデルでは、同程度の解が少ない計算量で求めることができると言い、解の探索能力は高い。この結果は他のテスト関数の最適化を行った場合にも同様な傾向が得られる。

次に、テスト関数 F_1 の経過時間と得られる適合度値の変遷の関係を図 5 に示す。

この図では、8 および 16CPU を使用した際の、マスタ・スレーブモデル (MGA と表記) および分散 GA モデル (CGA と表記) によって得られる適合度値の変遷を示している。

これからも明らかのように、分散 GA モデルの方がマスタ・スレーブモデルよりも高速に解が得られる。この F_1 関数の場合には、8CPU を使用した分散 GA モデルの方が 16CPU を使用したマスタ・スレーブモデルよりも高速に解が得られており、明らかに分散 GA モデルの方が使用した計算機環境に適したモデルであると言える。この結果が得られた理由としては次の 4 点が考えられる。

- 先に述べたように解の探索能力が分散 GA の方が高い
- 分散 GA では通信がほとんど生じない
- マスタ・スレーブモデルではマスタが必要であり、特に小さな規模のクラスターでは得られる並列化効率が低い
- マスタ・スレーブモデルでは、理論上はすべてのスレーブプロセッサが使用されるが、実際には適合度計算の処理と個体の分配のスケジューリングの問題

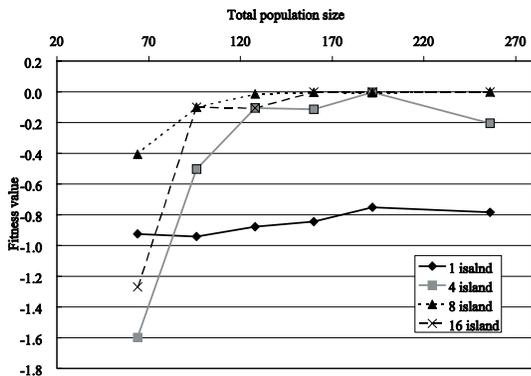


図 4: Population size and fitness value (F1)

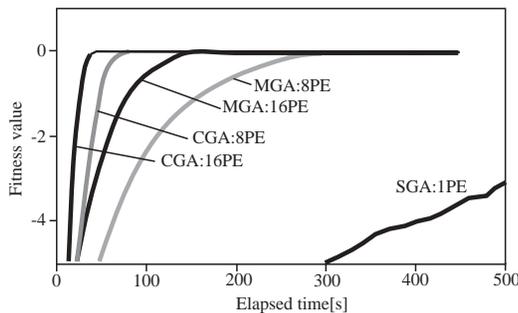


図 5: Elapsed time and fitness value (F1)

が発生し，すべてのスレーブプロセッサを効率良く利用することは難しい

そのため，マスタ・スレーブモデルを使用する際には，効率的にスレーブを使用するスケジューリングや通信の隠蔽などを行う実装が必要となる。

4 2 個体分散遺伝的アルゴリズム

前章では，クラスタ型並列計算機では，分散 GA モデルが有効であることが明らかとなった。本章では，分散 GA モデルの一つである「2 個体分散遺伝的アルゴリズム (Dual Individual Distributed Genetic Algorithm: DuDGA)」の提案を行い，数値計算を通じてその有効性を検討する。

4.1 DuDGA のコンセプトと遺伝的操作

DuDGA は各島内の個体数を 2 とした分散 GA モデルである。そのコンセプトを図 6 に示す。

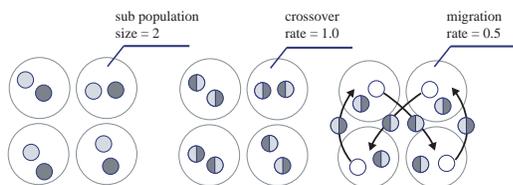


図 6: Dual Individual Distributed Genetic Algorithms

DuDGA では島内の解の多様性を維持するために，次のような遺伝的操作を行っている。

選択 交差を行うと 2 つの親個体から 2 つの子個体が生成される。このうち，親個体の中で高い適合度値を有するもの，子個体の中で高い適合度値を有するものを選択する。

移住 移住のトポロジは踏み石モデルであり，移住先は移住機会ごとにランダムに決定される。移住する個体はランダムに決定するが，島から島への移動ではなく，コピーを他島へ移住し，移住先の島において低い適合度値を持つ個体と交換が行われる。

DuDGA では下記のようにいくつかのパラメータが自動的に決定され，ユーザがパラメータ調整を行う必要はない。

- 総個体数が決定されれば島数が決定される (島数 = 総個体数 / 2)
- 交差率は 1.0
- 移住率は 0.5

これにより，本モデルは通常の分散 GA モデルと比較して，実際の利用の際に有効なモデルと言えよう。

4.2 DuDGA の探索能力

DuDGA の探索能力を検討するために，先に使用したテスト関数の最適化を行う。検討は通常の分散 GA の結果と比較することにより行う。使用した計算機環境は，表 3 に示す前章と同様のものであり，使用したパラメータを表 4 に示す。

表 4: Used Parameters

	DGA	DuDGA
Crossover rate	1.0	1.0
Population size	240	240
Mutation rate	1/L	1/L
Number of islands	4, 8, 12, 24	120
Migration rate	0.3	0.5
Migration interval	5	5

L : Chromosome length

図 7 および図 8 に解の発見率 (Reliability) と最適解を発見した際に要した評価関数計算数を示す。これらの結果は 20 試行行った平均値である。解の発見率とは 20 試行中，どの程度の割合で真の最適解が発見されたかの割合を示す。すなわち，発見率が 1.0 の場合，20 試行中すべての試行で真の最適解が発見されたことを示す。

これらの図からもわかるように DuDGA は他の分散 GA モデルと比較し，どのテスト関数においても高い解の発見率を示し，かつ，それらの解は，少ない評価回数計算で得られている。この結果，DuDGA は高い探索能力を有していることがわかる。

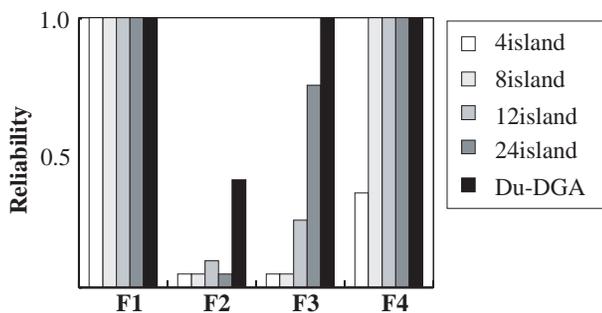


図 7: Comparison of Reliability of DGA and DuDGA

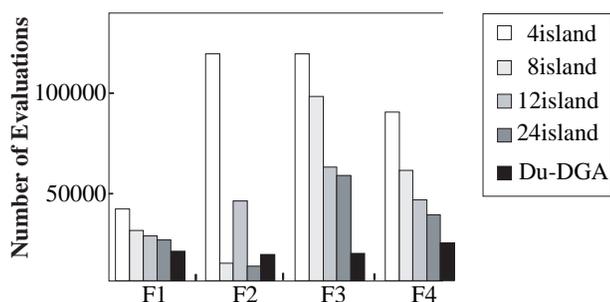


図 8: Comparison of Number of Function Evaluations Required by DGA and DuDGA

図 9 は $F1$ 関数の最適化の際の適合度値の変遷の 20 試行平均を示す。

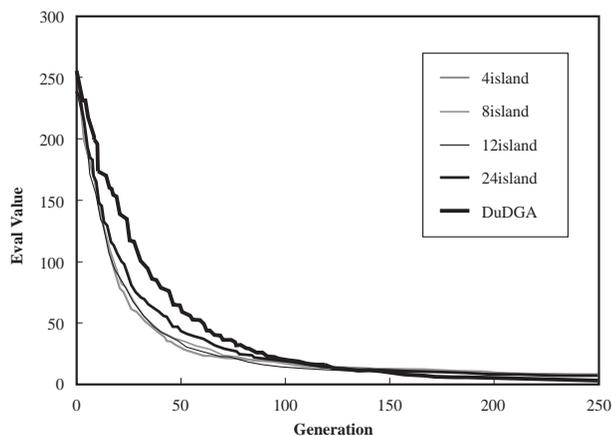


図 9: Iteration History of the Objective Function Value

この図からも DuDGA がどのようにして解探索を行っているかがわかる。DuDGA は他の分散 GA モデルと比較して、探索初期で得られる適合度の値が悪く、解の終盤で良好な適合度値が得られている。これは、DuDGA は解探索初期では、多様性を維持し、終盤では解の収束が急速に行われ、局所探索が行われていると考えられる。この探索状況は遺伝的アルゴリズムでの解探索としては理想的であり、多様性維持のフェーズから局所探索のフェーズへの移行が、パラメータ設定などの人為的に行われているのではなく、モデルの特性として行われている点が重要である。

4.3 Parallel Implementation of DuDGA

前節の結果から、DuDGA は非常に探索能力の高い分散 GA モデルであることが明らかとなった。しかしながら、一般に GA において個体数は非常に多く存在するため、DuDGA では必然的に島数も多く存在し、CPU 数はそれに比較して少ないため、PC クラスシステム上に実装するには多少のモデルの変更が必要となる。

図 10 に DuDGA を並列に実装するためのモデルの概念図を示す。ここでは 2 プロセッサの場合を示している。

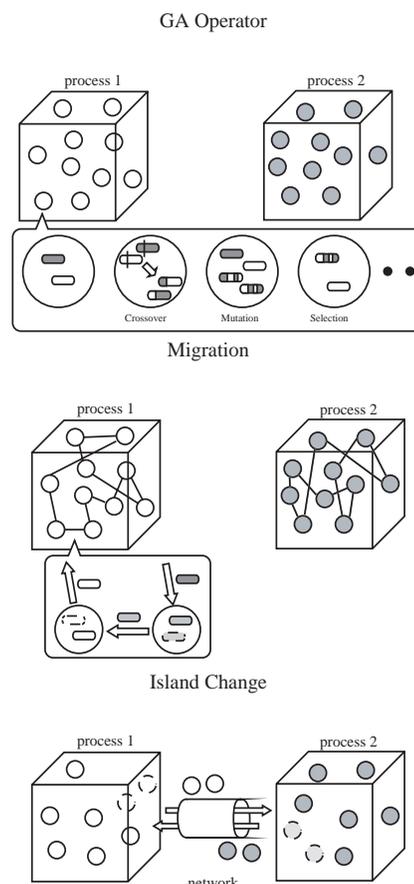


図 10: Parallel Implementation of DuDGA

このモデルは各プロセッサ内で DuDGA を行い、数世代後に、いくつかの島をランダムに選択し、他のプロセッサに移住するというモデルである。すなわち、ハイブリッド型となっており、2 種類の移住が存在する。

図 11 に $F1$ 関数の場合のプロセッサ数とスピードアップの関係を示す。

この図において、スピードアップとは、1 プロセッサを使用した際に解を求めるのに必要な時間を複数プロセッサを使用した際に必要な時間で除した値である。プロセッサ数が増加すると、DuDGA を行っているグループ数も増加するため、純粋な同一モデルによるスピードアップの比較とはならず、並列効果と同時に分散効果も含んだ結果である。

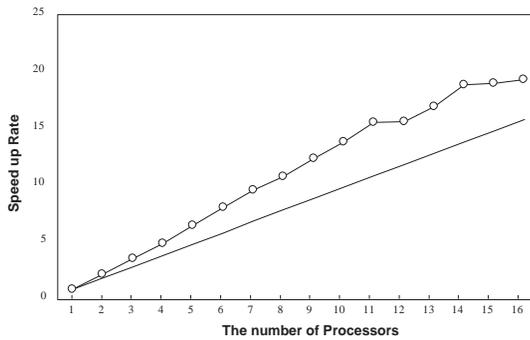


図 11: Speed Up

この図からもわかるように本モデルは非常に良いスピードアップを示しており、並列分散化効率の高いモデルであると言える。

5 結言

本研究では並列分散遺伝的アルゴリズムの代表的なモデルとして、マスタ・スレーブモデル、分散 GA モデル、セルラー GA モデルをあげた。そのうち、クラスター型並列計算機で広く使用されているマスタ・スレーブモデルと分散 GA モデルを数値計算例を通じて検討を行った。その結果、分散 GA モデルが、クラスター型並列計算機では優れたモデルであることが明らかとなった。

さらに、本研究では、分散 GA モデルの一つである、2 個体分散 GA モデル (Dual Individual Distributed Genetic Algorithm: DuDGA) の提案を行った。本モデルでは、各島内に 2 個体のみが存在するモデルであり、この設定により、遺伝的操作の設定が簡略化できる。また、他の分散 GA モデルと比較しても、解の探索能力が優れているといえ、クラスター型並列計算機に適したモデルである。

なお、本研究は、文部省科学研究費補助金研究 (基盤研究 C: 課題番号 10650104: 適応的分散並列遺伝的アルゴリズムに基づく構造システムの最適化) の一部として実施された。ここに記して謝意を表す。

参考文献

- 1) 北野編, "遺伝的アルゴリズム", 産業図書, (1993)
- 2) 北野編, "遺伝的アルゴリズム 2", 産業図書, (1995)
- 3) 北野編, "遺伝的アルゴリズム 3", 産業図書, (1997)
- 4) D. E. Goldberg, "Genetic Algorithms in search, optimization, and machine learning", Addison-Wesley, (1989)
- 5) E. Alba and J. M. Troya, "A Survey of Parallel Distributed Genetic Algorithms", Complexity, Vol. 4, No. 4, pp. 31-52 (1999)
- 6) J. Nang and K. Matsuo, "A Survey on the Parallel Genetic Algorithms", J. SICE, Vol. 33, No.6, pp. 500-509
- 7) E. Cantu-Paz, "A Survey of Parallel Genetic Algorithms", IlliGAL REport, No. 97003, (1997)
- 8) T. C. Fogarty and R. Huang, Proc of the 1st PPSN, pp. 145-149, (1991)
- 9) T.C. Belding, "The Distributed Genetic Algorithm Revised", Proc. of the 6th ICGA, pp. 249-256, (1991)
- 10) E. Cantu-Paz, "Topologies, Migration Rates, and Multi-Population Parallel Genetic Algorithms", Proc of GECCO, pp. 91-98, (1999)
- 11) R. Tanese, "Distributed Genetic Algorithms", Proc. of the 3rd ICGA, pp. 434-439, (1989)
- 12) V. S. Gordon and D. W. Whitley, "Serial and Parallel Genetic Algorithms as Function Optimizers", Proc. of the 5th ICGA, (1993)
- 13) D. Whitley, S. rana and R. B. Heckendorn, "Island Model Genetic Algorithms and Linearly Separable Problem", Proc. of the 7th ICGA, (1997)
- 14) 三木, 廣安, 中村, "遺伝的アルゴリズムの分散並列化に関する研究 (踏み石モデルによる分散遺伝的アルゴリズムの検討)", 機論 (A), 65-638, pp.2177-2183, (1999)
- 15) 三木, 廣安, 中村, "遺伝的アルゴリズムの分散並列化に関する研究 (ランダム移住型モデルによる分散遺伝的アルゴリズムの検討)", 機論 (A), 66-645, pp.972-977, (2000)
- 16) M. Gorges-Schleuter, "ASPARAGOS An Asynchronous Parallel Genetic Optimization Strategy", Proc. of the 3rd ICGA, pp. 422-427, (1989)

日本機械学会 [No. 00-27] 第4回最適化シンポジウム
講演論文集 [00-10-6,7,東京],pp. 305-310