

遺伝子集団の多様性維持のための 疑似減数分裂型遺伝アルゴリズム

吉田由起子 安達統衛

富士通研究所 情報社会科学研究所

本研究では、遺伝アルゴリズム (GA) における遺伝子集団の局所解落ちを回避する一手法として、十分長い世代にわたって遺伝子集団の多様性を保ち続けることのできる二倍体染色体型 GA を提案する。従来の二倍体型 GA の多くが、対立遺伝子に関する優性遺伝により表現型を作ってきたのに対し、相同染色体間で生物の減数分裂に似た操作により表現型を作ることから、この GA を疑似減数分裂型 GA (pseudo-Meiosis GA, psM GA) と名付けた。

psM GA は、巡回セールスマン問題 (TSP) などに用いられる order-coding のような、通常の優性遺伝を用いると致死表現型を作ってしまう coding に対しても、適用可能である。環境変動型 TSP に psM GA を適用して、集団の多様性を維持しつつ、変動に素早く追随することが確認された。

A Genetic Algorithm for Preserving Population Diversity — pseudo-Meiosis GA

YOSHIDA Yukiko ADACHI Nobue

Institute for Social Information Science, Fujitsu Laboratories Ltd.

The paper proposes a diploid genetic algorithm called pseudo-Meiosis Genetic Algorithm (psM GA) for the purpose of preserving population diversity. The psM GA makes a phenotype from a pair of chromosomes by means of meiosis procedure, while other diploid GAs mostly use conventional dominance mechanism.

The psM GA is applied to a time-variant TSP which the conventional diploid GAs are hard to handle, and the results confirm the ability of psM GA to preserve population diversity and to follow changes of the TSP quickly.

1 はじめに

遺伝アルゴリズム GA は、集団が収束し、遺伝子の多様性が失われると、それ以後交叉などによる遺伝子操作の効果が失われてしまい、たとえば、解空間が多峰的であったり、時間的に変動するような問題において、集団が局所解に収束した場合にそこから抜け出せなくなるといった不都合を生じる。

集団の多様性維持を目的とする単純 GA (SGA) からの拡張としては、多倍体染色体の導入が考えられている。これは、SGA が 1 本の文字列 (染色体) により通常は一意的に解を表現していたものを、複数の文字列の組により解の表現に冗長性を持たせたものである。倍数体型 GA (多くは二倍体が用いられている) は大抵、異なる文字列 (遺伝子型) から同一の解 (表現型) を表すための方法として、自然界の倍数体生物のように何らかの優性遺伝法を導入している。

倍数体型 GA の優性遺伝法の多くは、自然界の優性遺伝にならい、遺伝子座毎に対立遺伝子の優性・劣性をコントロールしている。さらに、遺伝子座毎の優性法則自体も GA によって進化させようとするモデルも存在する。そのようなモデルでは表現型に直接関与する遺伝子座・対立遺伝子以外に、優性法則をコードする遺伝子座・対立遺伝子も染色体に組み込まれている ([1] 参照)。

このような優性遺伝法の場合、適用できる問題は限られていて、たとえば関数最適化問題 (FOP) のように変数域つまり解空間が binary や gray といった bit-coding で表現された問題には適用できる。このとき染色体上の各遺伝子座間でとくに制約条件はなく、またどの遺伝子座上の対立遺伝子も共通の対立遺伝子集合 $\{0, 1\}$ から選ばれることに注意されたい。

ところがこのような優性遺伝法はどの coding タイプにも適用できるわけではない。たとえば、巡回セールスマン問題 (TSP) において都市を巡る順序

を (いつものように) order-coding で表し、それを二倍体化する場合を考える。TSP の order-coding を遺伝子座毎に切り離して扱うのはほとんど無意味であるばかりか、遺伝子座毎に独立に対立遺伝子を選んでいくと、TSP の制約条件 (ひとつの染色体上に同じ対立遺伝子が 2 つ以上あってはならない) に反するような表現 (致死表現) を生じやすい。

二倍体型 GA の表現型を作る方法の中には、自然界の優性遺伝法と趣の異なるものがある。それは、2 本の染色体のうち一方を何らかの法則にしたがって (たとえば、ランダムとか、適応度を計算して高い値の方とか) 選んで、その 1 本のみで表現型を作るというものである。この方法ならば TSP の order-coding でも致死表現を避けることができる。しかし、1 世代につき一方の染色体の性質しか表現型に反映されない。

本研究ではこれらの問題点を踏まえ、二倍体染色体が致死表現を作らず、かつ両方の染色体の性質が表現型に反映されるように工夫した変換方法を持つ GA を提案する。

2 疑似減数分裂型遺伝 GA

— pseudo-Meiosis GA

2.1 表現型の構成

本研究で提案する疑似減数分裂型 GA (pseudo-Meiosis GA, psM GA) は、従来の優性遺伝法のような遺伝子座毎に優性・劣性をコントロールする手間を省き、様々なタイプの問題に対し容易に適用できる二倍体染色体から表現型への変換方法として、疑似減数分裂 (pseudo-Meiosis, psM) を用いている。

生物の減数分裂は染色体数が半減するような相同染色体間の対合・分離のことであるが、psM が行なう減数分裂は表現型を作るための見かけ上の操作であり、相同染色体自体は変化を受けないことを明示するために、“疑似”と名付けた。

疑似減数分裂

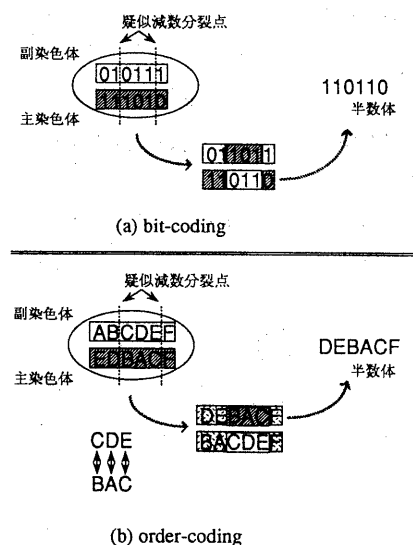


図 1: 2つの coding タイプによる疑似減数分裂

psM は次のような特徴を持つ：

- 各個体は、単純 GA 型染色体を 2 本一組にして持ち、これら 2 本の染色体（相同染色体）はそれぞれ 主染色体、副染色体 として特徴づけられている。
- 単純 GA の交叉を利用して、2 本の染色体から 1 本の染色体（半数体）が作られる。その半数体を用いて表現型が作られる。

psM では、二倍体から表現型を構成する際に、SGA の交叉を利用する。ここで、GA の代表的な 2 つの coding について、psM が表現型を構成する様子を示す。

bit-coding は、解を 2 値 $\{0, 1\}$ の列で表現する coding であり、関数最適化問題・ナップザック問題などで用いられている。図 1 (a) に 2 点交叉を利用した psM の例を示す。

order-coding は、解を文字の“並び”として表現する coding であり、TSP・スケジューリング問題などで用いられている。図 1 (b) に、order-coding

でしばしば用いられる交叉方式 Partially Matched Crossover (PMX) を利用した psM の例を示す。

これらの交叉方式では、2 本の染色体から 2 本の染色体が作られるが、そのうち一方（例えば、主染色体の性質をより多く保持している方）を選んで半数体とする。

2.2 アルゴリズム

GA の世界は集団のサイズや選択ルールなどによる遺伝的浮動の影響が強いので、単に冗長性のある遺伝子型から表現型への対応を工夫するだけでは、長い世代を経るうちに結局集団全体の多様性が失われやすいといえる。そこで、遺伝子集団の多様性維持をより効果的にするために、遺伝子変換法 psM を組み込んだアルゴリズム 疑似減数分裂型 GA (psM GA) を提案する（図 2 参照）：

1. (初期集団の設定) 単純 GA と同様に各染色体をそれぞれランダムに作る。
各個体に 2 本ずつ染色体を振り分ける。ここで各個体に対し、2 本の染色体の一方を主染色体、他方を副染色体としておく。このように決定された相同染色体間の主・副関係は、以後各世代を通して不変である。
2. (表現型の構成) 各世代の初めに、確率 $p_{meiosis}$ で相同染色体間で psM を行う。psM を行わない個体は、主染色体のコピーを半数体とする。
次の段階からは、この半数体集団に対し GA の遺伝子操作（選択・交叉・突然変異）を行っていく。
3. (選択) 表現型が持つ適応度にしたがって半数体集団の 選択/増殖を行う。その際あらかじめ、集団内でもっとも優秀な半数体から若干数 N_r のコピーを無条件で作っておく（この半数体を Reserve とよぶことにする）。残

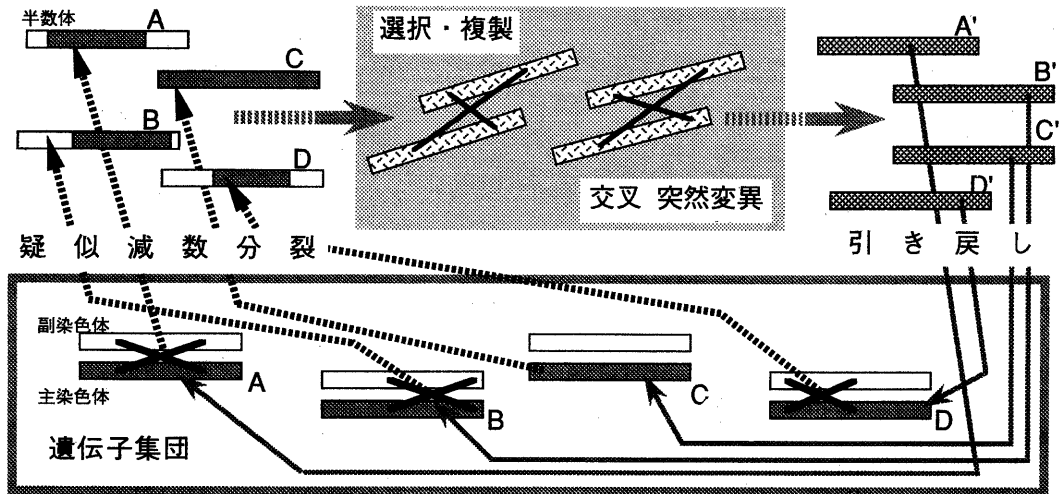


図 2: pseudo-Meiosis GA の遺伝操作

りの選択は通常の GA に準じ、たとえばルーレット選択で行なう。

4. (交叉・突然変異) 半数体集団内でランダムにペアを作り、確率 p_c で交叉を行う。

その後、各半数体について確率 p_m で突然変異を行う。

5. (引き戻し) 半数体集団を各個体の新しい主染色体として引き戻す。すなわち、選択時に生き残った半数体は元の個体の主染色体に戻す。

選択時に消滅した半数体の元の個体には、増殖した（優秀な）半数体を代わりにする。

確率 p_{ms} で各副染色体の突然変異を行なう。

6. 段階 2-5 を一世代とし、これを打ち切り条件を満たすまで繰り返す。

psM GA の長所は、遺伝子の多様性を副染色体の形で確実に次世代へ受け継がせることである。また、この遺伝子の多様性は効果的に表現型へ反映される。というのは、選択の段階で生き残った半数体は元の個体の主染色体として引き戻され、常に同じ副染色体と psM を行うことで優れた表現型にお

ける近傍探索を行う一方、選択で消滅した半数体の代わりになったものは、ランダムな分布を持つ副染色体と psM を行うことで多様な表現型を作ることができるからである。

通常の GA の選択・交叉は、結果として主染色体に作用する。一方、副染色体への作用は突然変異のみである。これは、（突然変異部分を除いては）変化しない副染色体を親子代々受け継いでいくことを意味する。

3 実験

psM GA の特長のひとつは、従来の優性遺伝法では致死表現型を生じてしまうような order-coding でも二倍体化が可能なことである。そこで、order-coding を用いる代表的な問題 巡回セールスマン問題 (TSP) に対して psM GA を適用してみた。

TSP は指定された都市群を巡るためのもっともコストのかからないコースを見つける問題である。ここでは、遺伝子の多様性維持の必要性が高いと予想される、時間とともにコストが変動する環境変動型 TSP を用いた。この問題は現実世界での交通渋滞などを反映したものと考えることができる。

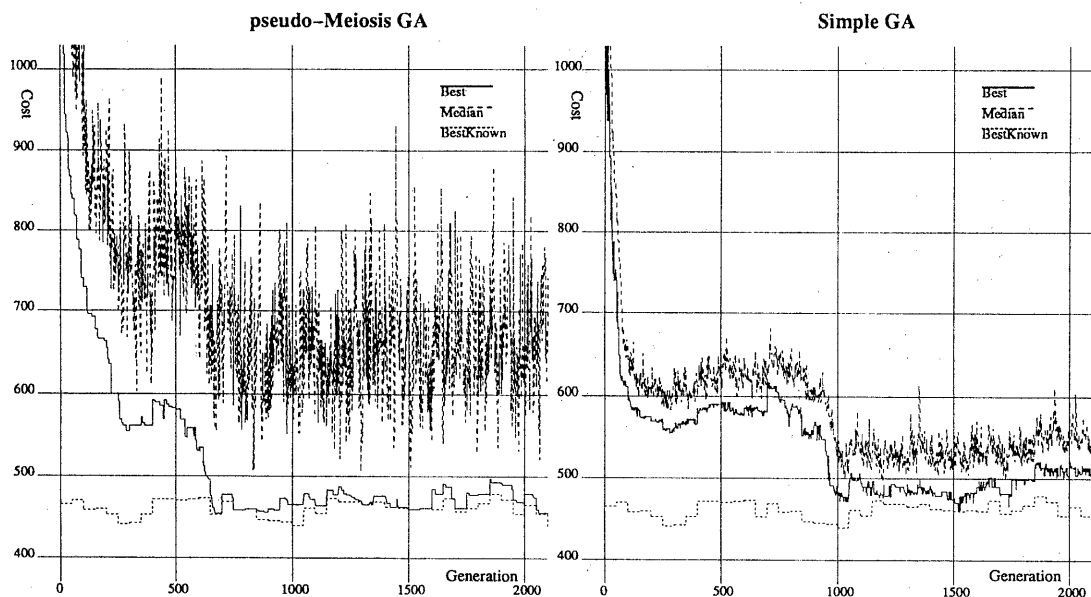


図 3: 集団内の最良値とメディアン値の世代変化 —pseudo-Meiosis GA, Simple GA

具体的には、初期状態として縦 100 横 100 の正方格子上にランダムに 30 都市を配置する。コストは都市間の直線距離とする。その後 50 世代 (1 Phase) を経るごとに環境変動を起こさせる。それには各都市を、ある確率分布に従い、縦横最大 5 以内でそれぞれの位置をずらす。これにより、コストの変化を視覚的にとらえることができる。

3.1 SGA との比較実験

まず、psM GA と従来の SGA との比較のための実験を行なった。

集団全体での染色体の総数を一致させるために、集団サイズを psM GA では 100、SGA では 200 に設定した。また、選択の段階で SGA も psM GA と同様に Reserve を作るものとした。Reserve 数は $N_r = 4$ 、交叉率は $p_c = 0.6$ である (SGA, psM GA 共に)。

psM GA については、疑似減数分裂率 $p_{meiosis} = 0.4$ 、半数体の突然変異率 $p_m = 0.01$ 、副染色体の突然変異率 $p_{ms} = 0.1$ に設定した。SGA について

は、突然変異率を調整し、成績の良かった値 $p_m = 0.3$ を用いている。

集団内の最良値の世代変化を図 3 に示す。Best-Known は、各 Phase 毎で現在知られている最良値である。また集団の多様性の目安として、集団内のメディアン値の世代変化も一緒に示す。図 4 は、SGA のいくつかの突然変異率に対する最良値の世代変化である。図 5 は、psM GA と SGA について各世代の最良値と BestKnown の差を Phase 毎に総和をとったものである (各世代を通じてどれだけ安定して良い解が得られるかの目安になるだろう)。

psM GA は、長い世代に渡って最良値とメディアン値がかなり離れつづけており、しかもメディアン値が世代毎に大きく変動しているのとは対照的に最良値の変化は単調的で安定している。また、環境変動後かなり早いうちに最良値を改善させることも分かる。

SGA は、psM GA と同様に Reserve を作っているにもかかわらず、最良値が安定しない。その原因は、突然変異率の高さが Reserve の効果を上回っ

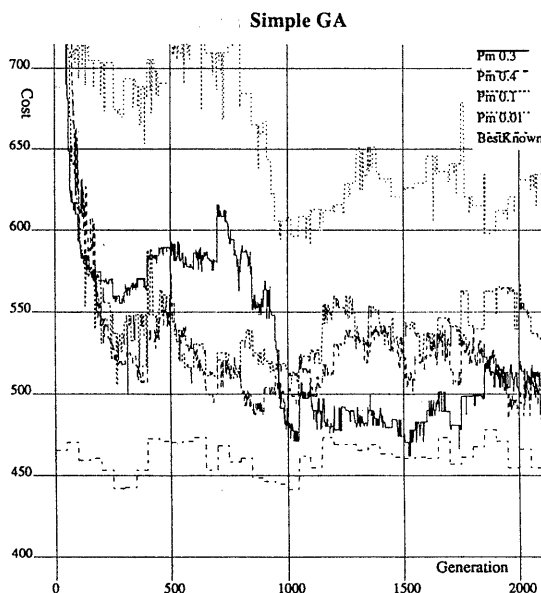


図 4: SGA の突然変異率 p_m の影響。

ているためと思われる。逆に、突然変異率を下げると、集団が局所解に落ちてしまうらしく、BestKnown からのずれは大きいままであった（図 4 の $P_m = 0.01$ の例）。

図 7 は、連続する 2 つの Phase 間で SGA および psM GA が経路を変えていく様子を表したものである。図 7 (a) は、Phase 32 の最終世代（1649 世代）における経路である。つづく 1650 世代目に環境変動が起こり Phase 33 に移る（図 7 (b)）。そのため前世代と同じ経路では明らかに遠回りになるような部分を生じている。3 都市 A, B, C は同一直線上になったために、一度通った道を戻らなくてはならない。図 7 (c) は、Phase 33 の最終世代（1699 世代）である。psM GA は、都市 A, B, C を一直線に進む経路に改良したが、SGA はその部分を改良していない。

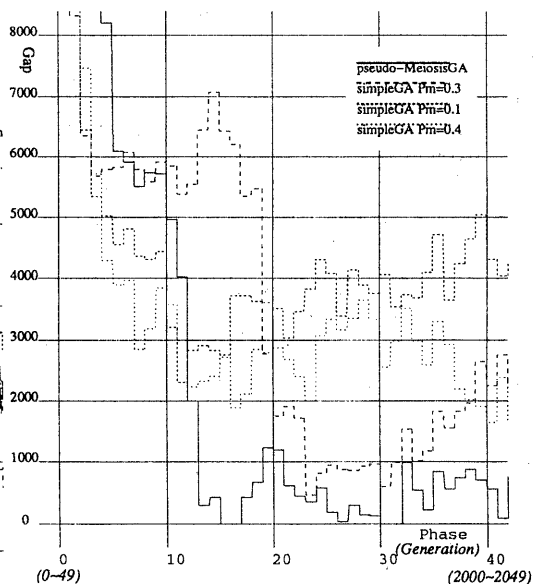


図 5: Phase 毎の BestKnown からの差の総和

3.2 psM GA のパラメータによる影響: 疑似減数分裂率、Reserve 数

つぎに、psM GA の構成において新たに加えられたパラメータによる影響について実験した。ひとつは疑似減数分裂率であり、 $p_{meiosis} = 0.3, 0.4, 0.5$ の 3 値で比較を行った。もうひとつは選択の際の Reserve 数について、 $N_r = 0, 3, 4, 5$ の 4 値で実験した。

各パラメータ値に対する集団内の最良値の世代変化を図 6 に示す。

疑似減数分裂率を上げれば、初期のランダムな分布を持った副染色体の性質がより多く表現型へ現れる。したがって $p_{meiosis} = 0.5$ による結果を見てもわかるように、かなり頻繁に優秀な表現型を見つけ出しているが、最良値の世代変化の安定性では疑似減数分裂率を下げたものに比べ劣っている。

Reserve 数 N_r は最良値の世代変化を安定させる目的で導入されている。ある世代で最良な表現型（半数体）が次の世代にも現れるように、交叉お

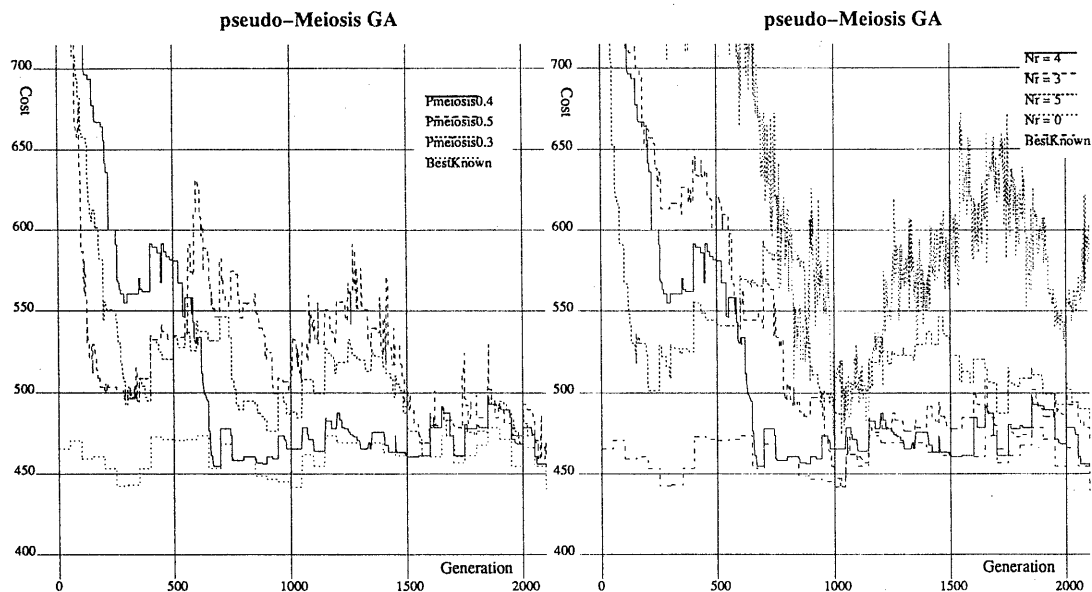


図 6: 疑似減数分裂率 $p_{meiosis}$ および Reserves 数 N_r の効果 — pseudo-Meiosis GA

および疑似減数分裂の作用を受けないコピーを用意しておくのである。しかし、前世代の最良半数体のコピーを一つでも残しておけば、環境の変わりめを除き単調的に改善させていくことができるから、解の多様性維持という psM GA の目的から見て、 N_r はあまり大きい値に設定するべきではない。実験結果からも、 $N_r = 0$ の場合最良値が安定せず、 $N_r = 5$ では環境変動への追従性が悪いことが分かる。

4 おわりに

本研究では、GA における遺伝子集団の早期収束・局所解落ちを回避するための、独自の解の表現方法を持つ二倍体染色体型 GA — 疑似減数分裂 GA (psM GA) — を提案した。

環境変動を伴う TSP に対する実験によりその有効性を確認し、さらに、GA の各パラメータ、とくに psM GA に特有な減数分裂率、Reserve 数といったパラメータが解に及ぼす効果について、考察した。

今後の課題は、bit-coding を用いる問題に対す

る psM GA の有効性を確認すること、psM GA の各遺伝子操作が遺伝子集団に及ぼす効果を、平均適応度の推移などから解析することである。

謝辞

日頃ご指導・ご討論いただく当研究所の皆様にご感謝いたします。

参考文献

- [1] David E. Goldberg and Robert E. Smith, *Non-stationary Function Optimization Using Genetic Algorithms with Dominance and Diploidy*, in Proc. of the Second International Conference on Genetic Algorithms (1987).

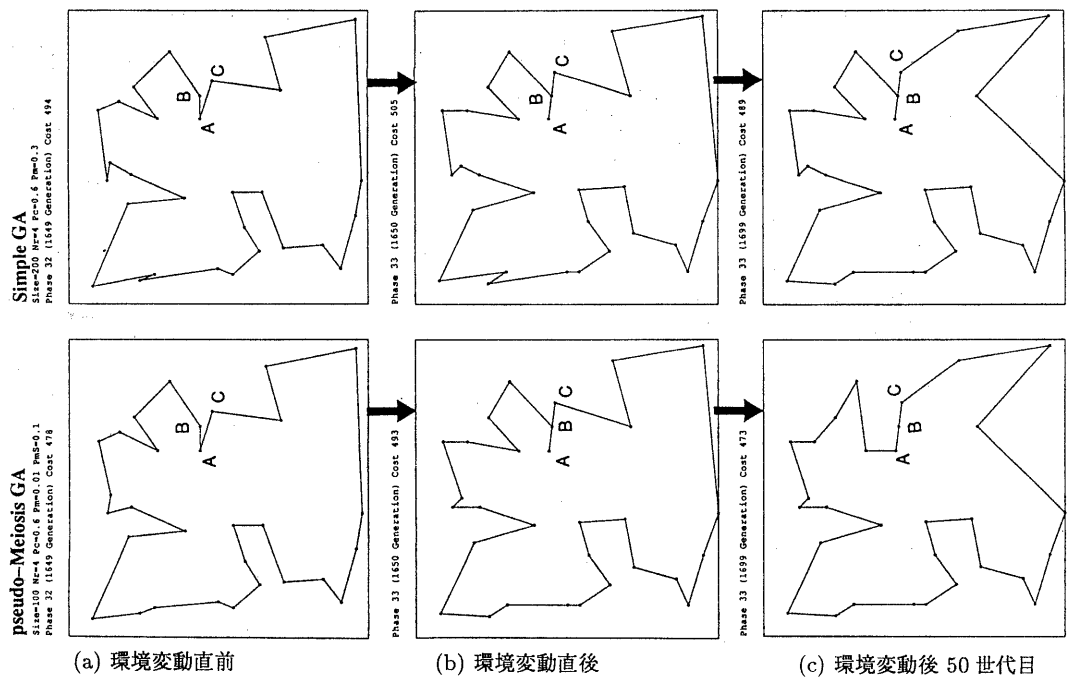


図 7: 環境変動前後での TSP の最良経路の変化 — simple GA および psM GA