

法政大学学術機関リポジトリ

HOSEI UNIVERSITY REPOSITORY

PDF issue: 2024-07-29

ノイズを含む信号解析問題へウイルス感染GA を適用した場合のトロピズムの有効性について

三浦, 潤一 / MIURA, Junichi

(発行年 / Year)

2008-03-24

(学位授与年月日 / Date of Granted)

2008-03-24

(学位名 / Degree Name)

修士(理学)

(学位授与機関 / Degree Grantor)

法政大学 (Hosei University)

ノイズを含む信号解析問題へウイルス感染 GA を
適用した場合のトロピズムの有効性について

**Efficiency of GA involving Viral Infection with Tropism
for Noisy Signals**

三浦 潤一

Junichi Miura

法政大学大学院

情報科学研究科情報科学専攻

Abstract

This paper reports on the influence that the tropism gives searching performance in GA involving viral infection with tropism. Previous research shows that applying GA is efficiency to analyze the noisy signals. Moreover, we proposed applying viral infection operation to GA so as to improve accuracy of analysis of noisy signals and showed the possibility that the searching performance improves by introducing tropism into the viral infection operation. On the other hand, a more detailed investigation is necessary for the influence that the tropism exerts on the searching performance of GA involving viral infection. This paper reports on the result of investigating the influence which GA involving viral infection with tropism exerts on the searching efficiency in terms of convergent error in analysis of noisy signals in GA in detail. As a result, in GA involving viral infection, the probability of converging local optima tends to increase while a local searching performance improves. Moreover, GA that uses the viral infection operation introducing tropism has the effect of decreasing probability of converging local optima.

1 はじめに

これまで、進化的計算は多目的最適化問題、時系列予測問題、ノイズを含むデータ解析問題など様々な問題に対して適用されてきた。一方、現実世界の問題の多くは複数の課題要素を含んでいる問題であることが多い。よって、進化的計算を適用するに当たり、単独の課題要素を含む問題に加えて、複数の課題要素を兼ね備えた問題について考えることも重要となってくる。これまで、多目的最適化問題、時系列予測問題、ノイズを含むデータ解析問題などの課題要素を単独、または複数含む問題に対して、進化的計算を適用した例についてはそれぞれ複数の研究成果が報告されている[1-6]。一方、上記全ての課題要素を含む問題に対しGAを適用した例は、わずかしか報告されておらず、これまで海上を航行する移動体の経路の推定[7]や時系列信号の特徴解析にGAを適用した研究が報告されている[8]。また、時系列信号の特徴解析を行うGAの精度向上のための改善案として、ノイズを含む信号の特徴解析にトロピズムを用いたウイルス感染操作を適用した研究について報告されている[9]。具体的には、計測機器より得られたノイズを含む時系列信号から、フーリエ級数の特性要素をGAによって探索し、未知の要素が多く含まれる信号を解析した。その結果、信号を海水中の音と仮定した場合には音源位置の解析も可能である可能性が示された。また、信号にはノイズが含まれていることを仮定しているため局所的な探索能力が重要となる。この問題に対応するため、ウイルスが解候補となる個体群の部分遺伝情報を運ぶウイルス感染操作をGAに導入することで、局所探索能力の向上に効果があり、また、ウイルスに宿主個体への感染を制御するトロピズムを持たせることで、ウイルス感染を効率的に制御できる可能性が示された[9, 10]。本研究ではトロピズムがウイルス感染に与える効果についてより詳細な分析を進めるために、収束誤差に注目してトロピズムの有効性について調査、検証を行った結果を報告する。

以下、2章で信号解析問題の定式化について、続いて3章で時系列信号問題へのGAの適用方法について記述する。4章では、トロピズムがウイルス感染を伴うGAに与える影響の調査の方針を記述し、本研究で用いるトロピズムについて説明する。5章で、実験方法と実験結果を示し、最後に考察とまとめを述べる。

2 信号解析問題の定式化

従来、コンピュータで信号を解析する場合、一般的にフーリエ変換に基づく解析手法である離散フーリエ変換(DFT)が用いられている。DFTを用いる場合、無限長の信号から窓関数によって有限の時間区間の信号を取り出して解析が行われる。信号を取り出した区間が信号の基本波周期の整数倍の場合、取り出した区間を繰り返して得られる周期信号は正しい結果となる。一方、信号を取り出した区間が信号の基本波周期の整数倍でない場合、取り出した区間を繰り返すときに境界で不連続点が生じてしまい、DFTによって得られた信号にはスペクトル漏れが発生する[11]。この漏れ誤差を小さくするためには、時間区間の境界での不連続性を小さくするなどの工夫を窓関数に施す必要がある。現在までにハミング窓やハニング窓、ブラックマン窓など様々な窓関数が提案されているが、周波数分解能などにそれぞれ差があり、一般的に信号の特性や目的に応じてふさわしい窓関数を選択する必要がある。この様なことを考えることなく信号の解析を行うために、GAが信号解析に適用された。

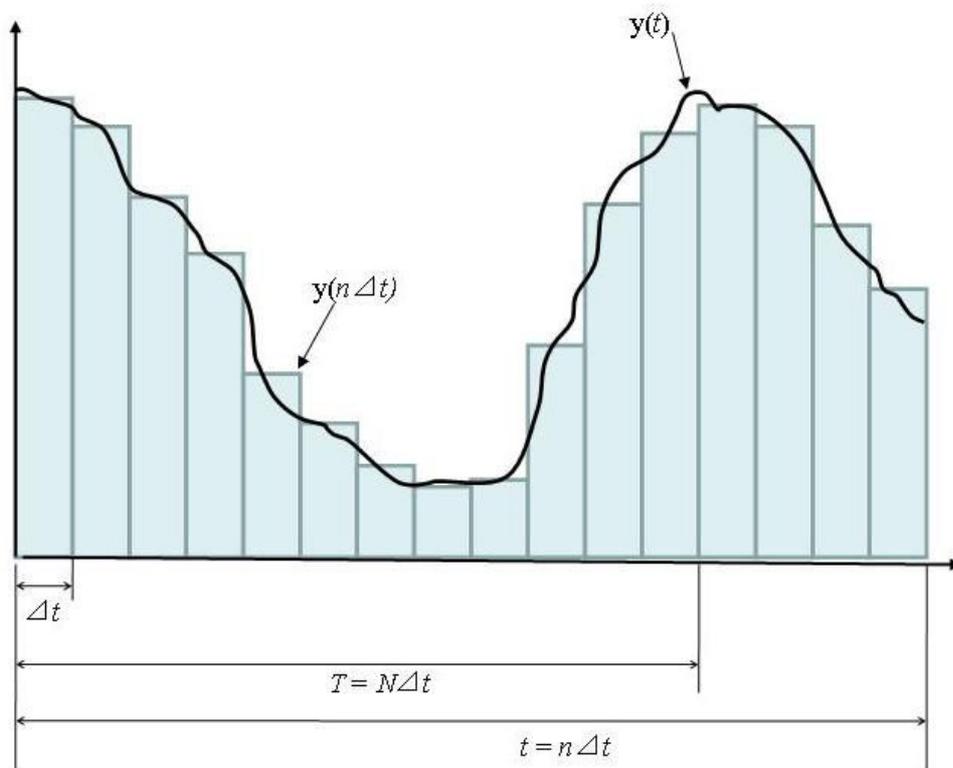


図1 連続信号の量子表現

図1の $y(t)$ は観測機器によって観測された信号の時間的変化の一例である。この信号は周期が T 、角速度を ω_0 とすれば、式1のフーリエ級数展開によって表すことができる。

$$y(t) = A_0 + \sum_{k=1}^{\infty} \sqrt{a_k^2 + b_k^2} \sin(k\omega_0 t + \varphi_k)$$

where

$$A_0 = \frac{1}{T} \int_0^T y(t) dy,$$

$$a_k = \frac{2}{T} \int_0^T y(t) \cos k\omega_0 t dt,$$

$$b_k = \frac{2}{T} \int_0^T y(t) \sin k\omega_0 t dt,$$

$$\varphi_k = \tan^{-1}(a_k / b_k)$$
(1)

ここで、 a_k 、 b_k をフーリエ係数という。

コンピュータは、演算を逐次的に行うため、連続した信号を処理することができない。コンピュータで計算を行うためには、信号をデジタル化する必要がある。デジタル信号に変換するためには、時間方向の離散化である標本化(サンプリング)と振幅方向の離散化である量子化を行う必要がある[12]。図1の $y(n\Delta t)$ は信号 $y(t)$ を一定時間間隔 Δt でサンプリングし、量子化した波形である。 Δt はサンプリング周期で、 N は周期 T の分割回数である。つまり、 T は $N\Delta t$ で表される。 n が標本化の次数なら、量子化を行った時間は、 $n\Delta t$ で表される。信号 $y(t)$ は $n \geq N$ の条件を満たす場合、量子化表現の式2の $y(n\Delta t)$ で近似することができる。

$$\begin{aligned} y(t) &= y(n\Delta t) \\ &= A_0 + A_1 \sin(2\pi f_p n\Delta t + \phi_1) \\ &\quad + A_2 \sin(4\pi f_p n\Delta t + \phi_2) \\ &\quad \vdots \\ &\quad + A_k \sin(2k\pi f_p n\Delta t + \phi_k), \end{aligned}$$
(2)

where

$$A_0 = \Delta t f_p \sum_{n=0}^{1/\Delta t f_p} y(n\Delta t)$$

$$A_k = \sqrt{a_k^2 + b_k^2}$$

$$\phi_k = \tan^{-1} \frac{a_k}{b_k}$$

また、本稿で扱う問題は先の報告[9, 10]同様、観測される信号は海水中の音波と仮定する。海水中の音信号の特徴解析の場合、低周波フィルターが信号の高周波を取り除くため、高

い周波数は低い周波数に比べ有用でなく、特徴解析の妨げとなることもある。そのため、基本周波数を数 100kHz と想定し k の最大値を 6 とした。つまり、6 次フーリエ部分和によって信号 $y(n\Delta t)$ を近似することとなる。上記の式において、 A_0 はサンプリングしたデータを基に計算している。サンプリング周期が信号周期より長い場合、 A_0 を求めることができる。一方、フーリエ係数 a_k, b_k を求めておらず、周波数も分からない。つまり、時系列の周期的な信号は、式 2 のフーリエ係数と周波数を求めることで、表すことができると言える。

本稿で対象とする問題は、複雑な間接関数を含む逆問題として定式化される。ここでは、観測機器から得られた $y(n\Delta t)$ のノイズのある時系列周期のサンプルデータから逆算することによって時系列周期信号のパラメータを発見する必要がある。

3 時系列信号解析問題への GA の適用方法

3.1 染色体の定義

前章より、観測機器から観測された信号を解析するためには、フーリエ級数の特性要素を探し出す必要がある。本研究では 6 次フーリエ部分和によって信号を近似しているため、フーリエ級数の特性要素として基本周波数 f_p と 6 次までのフーリエ係数を探索対象とする。そこで、GA の染色体は $f_p, a_1, b_1, a_2, b_2, a_3, b_3, a_4, b_4, a_5, b_5, a_6, b_6$ の 13 の部分染色体から構成する。次に各部分染色体のビット長を定義する。ビット長は長ければ長いほど精度が高くなるが、ビット長が長いほど探索空間も広がるため処理時間が増大する。よって、染色体に必要なとされる精度を考慮してビット長を決定する必要がある。本研究は、先の研究同様、観測誤差と A/D 変換の誤差から求められる最大の誤差を 0.105% と仮定する [9]。そのため、LSB の重みは 0.105% より小さくする必要がある。よって、部分染色体の長さは最低 14 ビット必要となる。本稿では、各部分染色体は 14 ビットで構成されるものとする。図 2 に本稿で扱う染色体の構成を示す。

3.2 評価関数

評価関数とは、個体が保持している染色体がどの程度正解に近付いたかを計算するための関数である。評価関数の数値が大きいくほどその個体は正解に近い。本研究では、観測機器によって観測された信号と GA によって得られた周波数、及び 6 次までのフーリエ係数から推測される信号との差によって表される。ある経過時間に観測された振幅を $oy(n\Delta t)$ 、推測される振幅を $esy(n\Delta t)$ 、観測回数を J とすると、評価関数は式 3 によって定義される。この評価関数では、 $oy(n\Delta t)$ と $esy(n\Delta t)$ の差の絶対値の平均値が小さい場合に評価値が高くなる。

$$fitness = \frac{J}{\sum_{n=0}^{J-1} |oy(n\Delta t) - esy(n\Delta t)|} \quad (3)$$

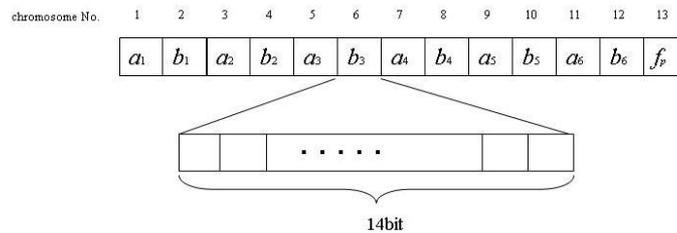


図 2 染色体の構成

3.3 子個体生成

子個体の生成は選択，交叉，突然変異によって行われる．まず，エリート保存選択によって評価値が上位の個体を一定数そのまま子個体として次世代に残す．不足した個体数は次の操作を繰り返すことで補充する．まず，ルーレット選択を用いて親となる個体を決定する．次に，親となった個体を基に交叉，突然変異の遺伝的操作を施し新たな子個体を生成する．これを不足した個体数が補充されるまで繰り返す．以下に，各遺伝的操作について説明する．

3.3.1 エリート保存選択

エリート保存選択とは，個体群の中で最も適合度が高い個体を無条件でそのまま次世代に残す選択方法である．後述するルーレット選択は確率に基づいた選択手法であるため，非常に優れた個体が現れても次世代に選ばれない可能性がある．また，ルーレット選択で選ばれたとしても交叉，突然変異によって破壊されてしまう可能性が考えられる．上記現象を避けるため，エリート選択を用いることで個体群の適合度の最大値が下がらないことを保証することができる．一方，エリート個体数が多すぎると，エリート個体の遺伝子が集団の中に広まりすぎて解の多様性が失われるという恐れがある．そのため，本研究では適合度をもとに個体群を降順にソートし，上位 60 個体のみをエリートとして次世代に残すことにする．

3.3.2 ルーレット選択

次世代に残す個体(子個体)を生成するための親となる個体(親個体)を決定する手法としてルーレット選択を適用する．ルーレット選択とは確率に基づいた選択手法であり，各個体の適合度に比例した割合で個体を選択する手法である．本研究は，実数値の多目的最適化

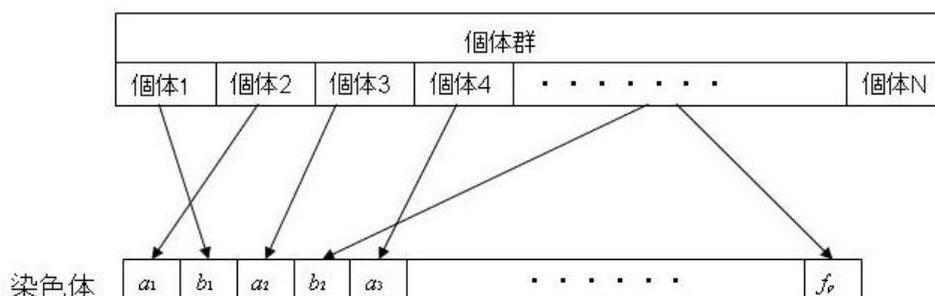


図 3 ルーレット選択の説明図

問題であるため、選択の際に、同じ個体から全ての部分染色体を選択すると探索空間が広がりやすく、大域的な探索が行われにくくなると考えられる。そこで、図 3 に示すように、部分染色体毎に適合度に基づくルーレット選択を行う。よって、ある個体の各部分染色体において、親となる個体は基本的に異なってくるといえる。部分染色体毎に選択を行うことにより、個体群が収束し始める前の探索の初期段階において、広い空間を探索できると考えられる。一方、個体群が収束し始めた後においては、部分染色体毎に選択を行うことで探索空間を不用意に広げることがなくなると考えられる。

3.3.3 交叉

交叉とは、選択された 2 つの親個体間で遺伝子の組み換えを行うことにより新しい個体を生成する操作である。本研究では交叉に一点交叉を適用する。一点交叉とは、交叉させる位置を染色体上のある一点に決定し、その交叉点の前後で 2 つの親個体の遺伝情報を入れ替えることで新しい個体を生成する手法である。本研究において 1 つの個体は 13 個の部分染色体から構成されているため、ある個体に対する交叉の処理方法として、13 個の部分染色体の遺伝子を直列につなげたものに対して交叉を行う方法と、各部分染色体毎に独立に一点交叉を行う方法が考えられる。しかしながら、各部分染色体は個別に意味をもち、部分染色体毎に収束性が異なっている。そこで、同種の部分染色体毎に独立に交叉を行うことで、全体の収束性の向上が期待できると考えられる。

3.3.4 突然変異

突然変異は個体の遺伝子の一部を変化させる操作である。突然変異を用いることで親個体に存在しない遺伝情報を生成する効果がある。本研究で用いる突然変異は、あらかじめ設定された確率に従って全ての遺伝子座に対して突然変異を行うか否かの判定を行い、突然変異する遺伝子座の遺伝子を対立遺伝子に置き換えるものとする。部分染色体 f_p は他の部分染色体よりも早く収束を始める傾向にあるため、部分染色体 f_p の突然変異率は他の部分染色体よりも低い確率に設定する。また、全ての部分染色体において、10 世代前後から収束が始まる傾向にあるため、7 世代以降の突然変異率を低く設定する。

3.4 ウイルス感染操作の導入

本研究では、観測された信号のデータにノイズが含まれていることを仮定している。そのため、信号の解析に GA を適用する際に、GA の大域的探索能力に加えて局所探索能力についても考慮する必要がある。前節では、一点交叉やビット突然変異など単純な遺伝的操作

作を備えた GA によって信号を解析する方法について述べた。一方、単純な遺伝的操作のみでは、局所的な部分に関しては十分な探索が実現できていない可能性がある。そこで、局所的な探索能力の向上を図るために、ウイルス感染操作を GA に適用している。

ウイルス感染操作は、ウイルス進化論に着想を得た遺伝的操作である。ウイルス進化論によれば、生物は突然変異によってではなく、ウイルスによって運ばれた遺伝子が生物間を伝播することで、生物が進化してきたと考えられる[13]。例えば、レトロウイルスと呼ばれるウイルスは細胞に感染すると逆転写酵素により、宿主個体 DNA に組み込まれる。そして、自己増殖後に宿主細胞を抜け出し、別の細胞へと感染するが、このときレトロウイルスは宿主細胞の遺伝子の一部を自分の遺伝子に加えて持ち出し、新たな細胞に感染する。その結果、ある細胞の遺伝子が別の細胞の遺伝子に運ばれたことになる。すなわち、ウイルスの遺伝子を運ぶ性質が生物を進化させてきたと考えられる。ウイルス進化論の考えをコンピュータ計算に適用した研究例についてはすでに複数報告されており[14-21]、その研究例において局所情報の迅速な生成、伝播、計算時間の削減に効果があることが報告されている。以下、本研究における GA 上でのウイルス感染操作について説明する。

3.4.1 GA 上のウイルス染色体の定義

GA 上のウイルスは、解候補として表わされる個体の遺伝情報の一部を保持する。(以降、解候補として表わされる個体を宿主個体、部分的に遺伝情報を保持する個体をウイルス個体と呼ぶことにする。) ウイルス個体は、宿主個体の 13 の部分染色体の中から確率的に複数の部分染色体を選択する。ウイルス個体の染色体は、選択された 14 ビットの部分染色体の一部から構成される。図 4 にウイルスの染色体の例を示す。この図において、ウイルス個体は部分染色体の内、第 2, 第 7, 第 10, 第 11 部分染色体上に遺伝情報を持っている。部分染色体の中の第 10 部分染色体を見ると、第 1 から第 3, 第 6 から第 8, 第 11, 第 12 遺伝子座に 0, 1 のバイナリの遺伝情報を持っている。遺伝情報を保持していない部分染色体の遺伝子座は、全て '*' (Don't care)である。

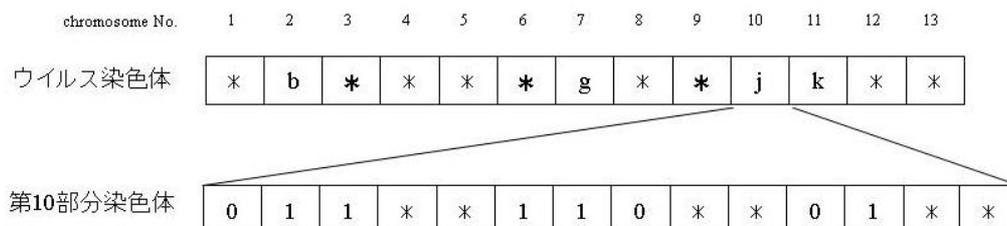


図 4 ウイルス染色体の構成

3.4.2 GA 上のウイルスの振舞い

ウイルス個体を媒介にして、ある宿主個体から別の宿主個体へ遺伝情報を伝達する機能を実現するため、GA 上のウイルスには感染操作として取込み操作と逆転写操作が実装されている。ウイルス個体は、取込み操作によって宿主個体の遺伝子列の一部を取り込み、逆転写操作によって取り込んだ遺伝子列を他の宿主個体に転写する。

- ・ 取り込み操作

ウイルス個体は取込み操作によって宿主個体の遺伝情報の一部を獲得する。取込み操作を図 5 に示す。ウイルス個体は取込み対象となる宿主個体から、取込みを行う部分染色体を任意に複数選択する。図 5 において、太線で囲われた宿主個体の第 3、第 7 部分染色体がウイルス個体に取り込まれる。ウイルス個体の第 3 部分染色体は、'*' から 'c' へ、第 7 部分染色体は 'B' から 'g' へと宿主個体の部分遺伝情報を取り込んでいる。

また、ウイルス個体が宿主個体の部分遺伝情報を取込むときに、ウイルス個体が持つ部分遺伝情報の一部を失う場合がある。部分遺伝情報を失う取込み操作を図 6 に示す。部分遺伝情報を失う取り込み操作は、宿主個体とは関係なく行われる。始めに、部分遺伝情報を失う部分染色体が無作為に選択される。この図においては、第 3 部分染色体と第 7 部分染色体が選択されている。取り込み操作の結果、第 3 部分染色体は操作前において、'*' であったため変化しないが、第 7 部分染色体 'B' は '*' へと変化している。その結果、図のウイルスが保持する部分遺伝情報は、'A' と 'D' となる。

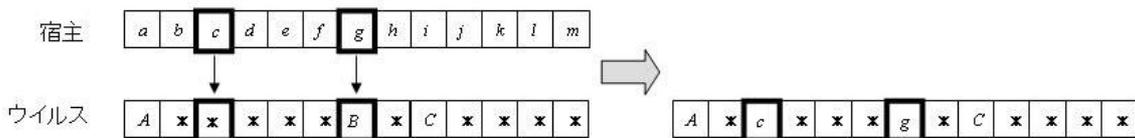


図 5 取込み操作

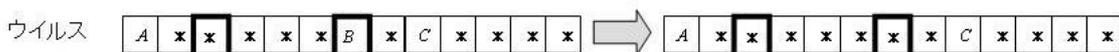


図 6 遺伝情報を失う取込み操作

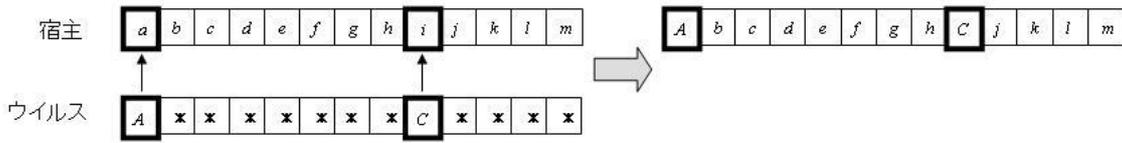


図 7 逆転写操作

・逆転写操作

ウイルス個体は逆転写操作によって自身が保持する局所遺伝情報を宿主個体に転写する．逆転写操作を具体的に図 7 に示す．ウイルス個体は，それぞれ感染率を持っており，それに応じて宿主個体に感染し，逆転写を行う．図 7 において，ウイルス個体は第 1，第 9 部分染色体上に遺伝情報を保持している．このウイルスが宿主個体に感染することで，宿主個体の太線の四角に囲われた第 1，9 部分染色体がウイルス個体の 'A'，'C' に置き換わっている．

3.4.3 GA 上のウイルスのパラメータ

ウイルス個体は評価値，感染率，生命力をパラメータとして持っている[16]．各パラメータはウイルス個体によって異なり，その値は逆転写の成否によって変動する．以下，各パラメータの計算方法について説明する．

(1) 評価値

ウイルスの評価値は，逆転写することによって得られる宿主個体の評価値の変化量(改善度)で表される．ウイルス*i*の感染する個体群を*S* とし，宿主*j*の逆転写前の評価値を fit_{host_j} ，逆転写後を $fit_{host_j'}$ とするとウイルス*i*の評価値 fit_{virus_i} は以下の式4によって定義される．

$$fit_{virus_i} = \sum_{j \in S} fit_{virus_{i,j}} \tag{4}$$

where

$$fit_{virus_{i,j}} = fit_{host_j'} - fit_{host_j}$$

この改善度の和が，各世代におけるウイルス個体の評価値となる．従って，評価値が高いほどウイルスは宿主個体群の評価値を改善，すなわち，宿主個体の進化に貢献していることになる．

(2) 感染率

感染率は宿主個体に感染を行う確率を表しており，感染率が高いウイルス个体ほど宿主個体に感染を行う機会が増える．宿主个体群の進化に有効なウイルス个体の感染を促すために，逆転写に成功したウイルス个体は感染率を増加させ，逆転写に失敗したウイルス个体は感染率を減少させる．感染率は定率で変化し，以下の式 5 によって定義される．

$$inf\ rate_{i,t+1} = \begin{cases} (1 + \alpha) \times inf\ rate_{i,t} & \text{if } fitvirus_i > 0 \\ (1 - \alpha) \times inf\ rate_{i,t} & \text{otherwise} \end{cases} \quad (5)$$

(3) 生命力

ウイルス个体の中には，宿主个体を改悪するウイルスも存在する．従って，ウイルス自体の宿主个体群に対する長期的な適応度を考慮する必要がある．そこで，各ウイルスは $fitvirus_i$ に依存するパラメータとして生命力 $life_{i,t}$ をもつ．生命力はウイルス个体が次世代に生き残れるかを表しており，生命力が 0 以下になったときそのウイルス个体は破棄されて新しいウイルス个体が生成される．感染率と同様，逆転写に成功したウイルス个体の生命力は高く維持される．生命力は以下の式 6 によって定義される． r は生命力減衰率を表し， $0 < r < 1$ の間で定義される．生命減衰率は過去の改善度に重み付けを与えるための係数である． r の値が低いほど前の世代の改善度は考慮されず， r の値が高いほど前の世代の改善度を重視する．

$$life_{i,t+1} = r \times life_{i,t} + fitvirus_i \quad (6)$$

4 トロピズムがウイルス感染 GA に与える影響の調査

4.1 基本的方針

ウイルス感染を用いた GA において、感染によって宿主個体の適応度を高めたウイルスは高い感染確率を有するようになる。一方、高い感染確率を有するウイルスが必ずしも宿主個体群に対して有効に働くとは限らない。そこで、先の研究ではウイルスの感染を効率的に制御するため、自然界のウイルスがもつトロピズムと呼ばれる仕組みを GA 上のウイルス個体に導入した。トロピズムとはウイルスがどの臓器、組織、細胞に感染するかを決定する仕組みである[22]。自然界のウイルスは DNA の周りにリガンドと呼ばれるたんぱく質を持ち、そのリガンドと引き合う受容体をもつ細胞にのみ感染する。受容体とは細胞外からの情報や刺激を受け取る器官である[23]。このような自然界のウイルスが持つトロピズムの仕組みを GA 上のウイルス感染操作に導入しウイルス個体の感染を制御することで、探索能力の向上に効果がある可能性が示された[9]。具体的には、先の研究では、トロピズムによって高い感染率を有するウイルス個体の不要な感染を防ぐことで、ウイルス個体の生存率を高めた。その結果、部分遺伝情報が保持されやすくなり、探索能力が宿主個体数に影響されにくくなる可能性を示した。本研究では、トロピズムの有効性について検証するため、局所解に着目してトロピズムを適用したウイルス感染を伴う GA が探索能力にどのような影響を与えているのかより詳細に調査を行う。具体的には、ウイルス感染を用いない GA とウイルス感染を用いた GA、トロピズムを導入したウイルス感染を用いた GA による探索能力の違いを収束誤差の観点から比較し、トロピズムを用いたウイルス感染 GA が有効に働いているかどうか検証を行う。続いて、トロピズムの閾値に着目する。閾値はトロピズムを決定付ける重要な要素であり、閾値の違いが GA の探索能力やウイルス感染操作に及ぼす影響について調査を行う。

4.2 トロピズムの実装方法

ウイルス感染を用いた GA では感染される宿主個体の遺伝子が局所解に近いと、ウイルスの局所探索によって宿主個体が局所解に導かれやすくなる可能性がある。また、ウイルスが局所解の部分遺伝情報を取り込んだ場合、逆転写によって局所解へ導く部分遺伝情報を他の宿主個体群に広める可能性がある。さらに、局所解の遺伝情報によって一時的に宿主個体の評価値を上昇させたウイルスは感染率を上昇させ、より多くの宿主個体に局所解へ導く部分遺伝情報を広めてしまうおそれがある。すなわち、GA にウイルス感染操作を導入することで宿主個体群が局所解へ導かれ、初期収束に陥る危険性も増していると考えられる。

よって、本稿では大域的探索が十分行われた後にウイルス感染が行われるようにトロピズムを設計する。本稿におけるトロピズムの機能は、ある程度収束した部分染色体に対してのみウイルスによる詳細探索を行うことで初期収束の危険性を減らし、より効率的な探索を目指す。具体的には、受容体を宿主個体の部分遺伝情報として、ウイルスにリガンドを持たせることでGA上のウイルスにトロピズムを実現する。図8にリガンドの取込み方法を示す。図8において*はドントケア、アルファベットは実数値で表現されたウイルスの遺伝情報を表す。感染可否の判定基準となるリガンド情報には真値に近い値を期待しているので、宿主個体群の中で最も評価値の高いエリート個体群の遺伝情報を用いてリガンドを生成する。図9にトロピズムを用いたウイルス感染の方法を示す。ウイルスは宿主個体に感染するときリガンドと受容体を比較し、その差が閾値以内であれば感染を行う。閾値に小さい値を設定することで宿主個体群がある程度収束してからウイルス感染の頻度が増加するようにする。

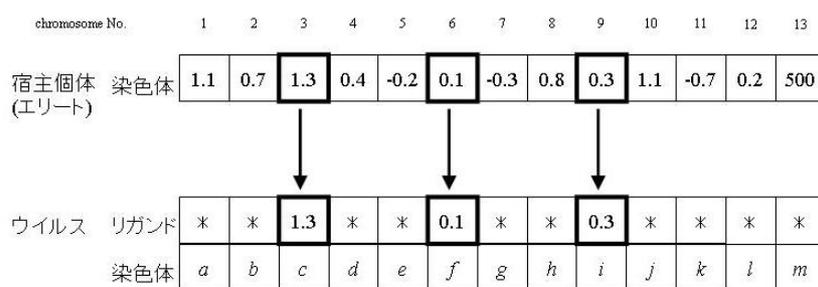


図8 リガンドの取込み。

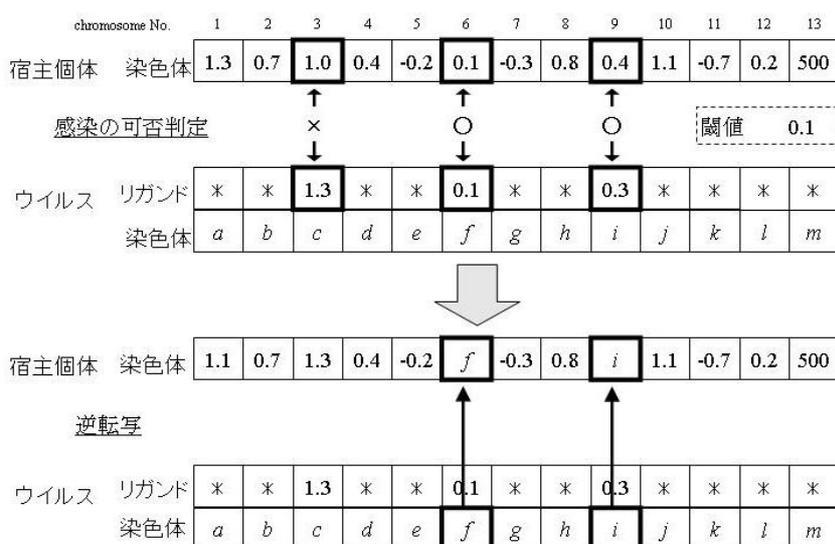


図9 トロピズムによるウイルス感染。

5 実験

5.1 実験方法

シミュレーション上でノイズを含んだ信号に対して GA を用いて信号解析を行い、トロピズムが探索能力に及ぼす影響について調査する。シミュレーションは 100 世代を 1 セットとして 100 セット行う。まず、GA 手法にウイルス感染を用いない GA(Standard GA), ウイルス感染を伴う GA(Virus GA), ウイルス感染にトロピズムを適用した GA(Tropism virus GA) を用いて実験を行い、トロピズムが収束誤差に与える影響について調べる。続いて、トロピズムの閾値がウイルス感染操作に与える影響について調べるために、トロピズムの閾値を変化させて実験を行う。また、局所探索法として、ウイルス感染操作の代わりにタブーサーチ[24, 25]を用いて実験を行い、ウイルス感染操作を用いた GA と探索能力を比較する。本稿では、従来 GA(Standard GA)による大域的探索によって得られた解に対して、タブーサーチ(TS)による局所探索を行う。近傍探索は任意の遺伝子座を 1 ビット対立遺伝子に置き換えることで行う。

GA に使用したパラメータを表 1 に示す。特に断らない限り、実験には表 1 のパラメータを使用している。infection rate はウイルス個体が逆転写を行う染率を示し、Incorporation rate はウイルス個体が取込み操作を行う確率を示す。Copy rate と Cut rate は、ウイルス個体の生成時、または取込み操作時に、ウイルス個体の各部分染色体が遺伝情報を入力、または失う確率を示す。ウイルス個体の適応度が 0 より大きいとき、Copy rate の確率でウイルス個体は宿主個体から部分遺伝情報を取込み、ウイルス個体の適応度が 0 以下のとき、Cut rate の確率でウイルス個体は自身の部分遺伝情報を失う。Ligand incorporation はウイルス個体がリガンド情報を入力する確率を示し、Threshold はトロピズムにおける閾値を示す。

表 1 パラメータ設定.

GA parameters	Maximum generation	100
	Population size	1000
	Crossover rate	0.8
	Mutation rate for f_p	0.0001
	generation ≥ 7	0.00005
	Mutation rate for a_k, b_k	0.0005
generation ≥ 7	0.0001	
Virus parameters	Virus population size	Population size*30%
	Initial infection rate	0.005
	Maximum infection rate	0.1
	Copy rate	0.2
	Cut rate	0.15
	Incorporation rate	0.1
Tropism parameters	Ligand incorporation	0.3
	Threshold	0.001

5.2 実験結果

5.2.1 トロピズムが探索能力に与える影響

図10にウイルス感染を用いない従来GA, ウイルス感染を用いたGA, トロピズムを適用したウイルス感染を伴うGA, それぞれの手法を信号解析問題に適用したときのエラー値の推移を示す. 図10におけるエラー値の推移は100回実験を行ったときのエラー値の平均を表している. ここでエラー値とは, GAによって推測された基本周波数及び12のフーリエ係数と, 真値として設定された基本周波数及び12のフーリエ係数との差の平均のこととする. 図10より, エラー値が0.1より低くなるまではGA手法の違いによる探索能力の差はほとんど見られない. 一方, エラー値が0.1を下回ると, ウイルス感染操作をGAに適用することで若干ではあるがウイルス感染を用いない従来GAよりも高い探索精度を示した. また, error値が0.01より低くなるまではトロピズムの有無に関わらずウイルス感染を用いたGAは同様の探索能力を示しているが, error値が0.01を下回ると, ウイルス感染操作のみを用いたGAは探索速度が低下しているのに対して, トロピズムをウイルス感染操作に適用したGAはより詳細な探索精度を示すという結果が得られた.

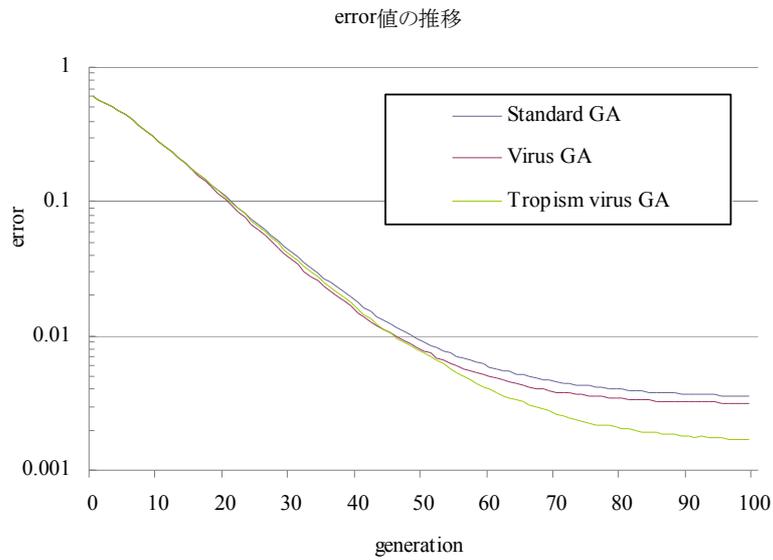


図 10 ウイルス感染を伴う GA と一般的な GA の探索能力の比較

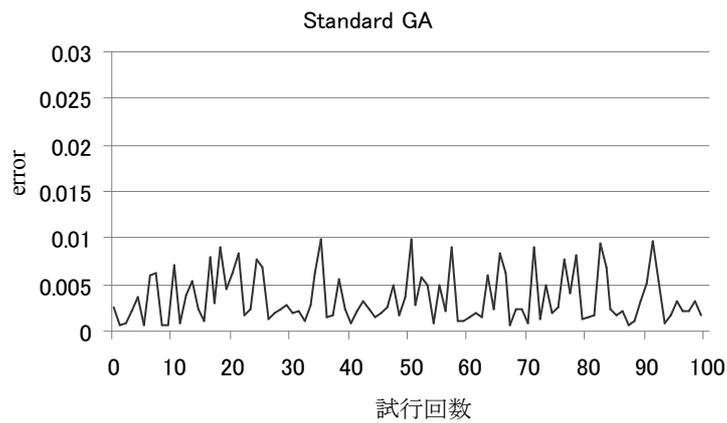


図 11 従来 GA を用いたときの 100 世代目のエラー値

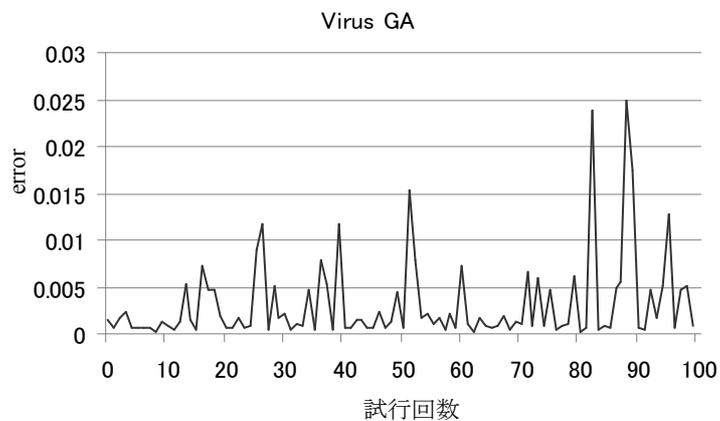


図 12 ウイルス感染を伴う GA を用いたときの 100 世代目のエラー値

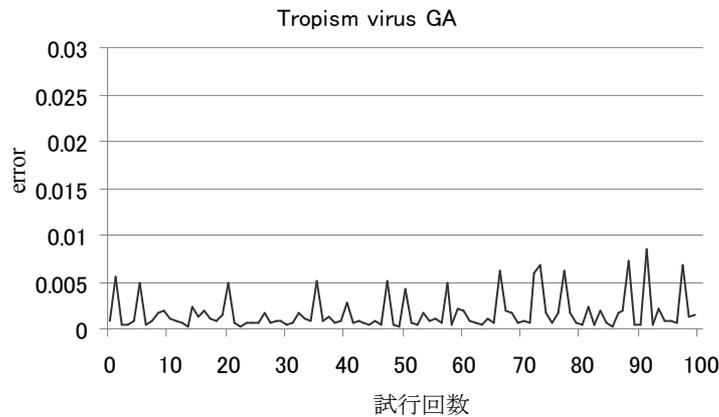


図13 トロピズムを適用したウイルス感染を伴う GA を用いたときの100 世代目のエラー値

図11, 図12, 図13に, 100回シミュレーションを行った際, 100世代目のエラー値が収束した値を1試行毎に示す. 図11に, 従来GAを適用したときの100世代目のエラー値を1試行毎に示す. エラー値は0 から0.01 の間を一様に散らばっており探索の精度はあまり高くない. 図12にウイルス感染を伴うGAを適用したときの100世代目のエラー値を1試行毎に示す. 従来GA に比べて, エラー値がより低くなる回数が増加した一方, 高いエラー値に収束する回数も増えており探索精度にばらつきがみられる傾向にある. 図13にトロピズムを適用したウイルス感染を伴うGAを適用したときの100世代目のエラー値を1試行毎に示す. 他の手法に比べてよりエラーの値が減少し, 試行ごとのばらつきも少なくなっていることが分かる. 100世代目のエラー値がどのような値に収束する傾向があるか比較するため, 図11, 図12, 図13における100世代目のエラー値の分散の割合を図14に示す. 図14から, ウイルス感染操作をGAに適用することで, 従来GAに比べてより低いエラー値に達する回数が増えている. すなわち, ウイルス感染操作によって局所探索能力が向上しているといえる. 一方, 従来GAよりも高い値に収束する回数も増えている. すなわち, 初期収束に陥る可能性が高くなっていることが分かる. ウイルス感染操作にトロピズムを用いることで, ウイルス感染操作のみの探索に比べて低い値に収束する回数が増え, さらに, ウイルス感染操作のみの探索で見られるような高い値に収束することもなくなっている. すなわち, ウイルス感染操作にトロピズムを用いることで初期収束に陥りにくくなっているといえる.

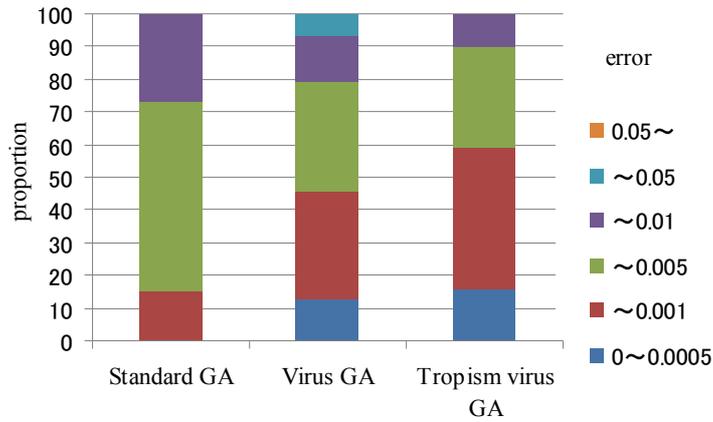


図 14 100 世代目のエラー値の割合

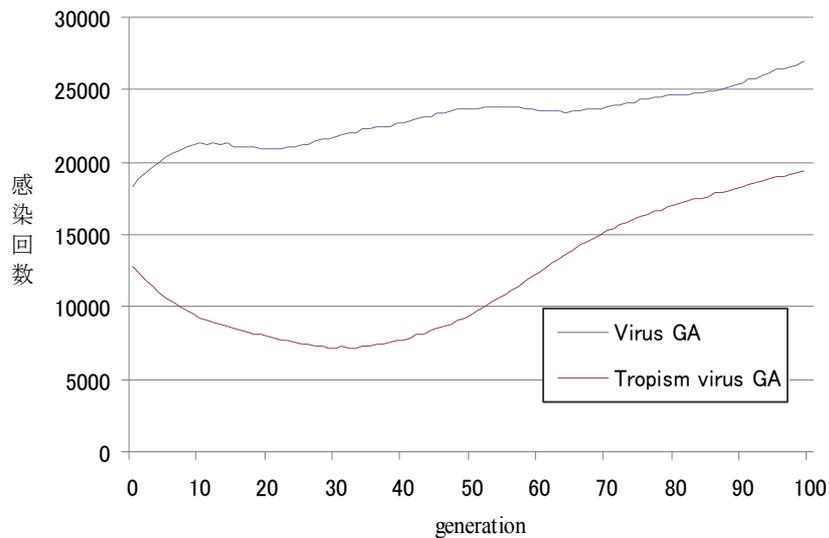


図 15 ウイルスの感染回数の推移

図 15 にウイルス個体の感染回数の推移を示す。感染回数はウイルス個体の部分染色体が宿主個体の部分染色体に逆転写するごとにカウントする。例えば、あるウイルス個体がある宿主個体に対して逆転写を行ったとき、宿主個体が持つ 13 個の部分染色体のうち 3 つの部分染色体の遺伝情報がウイルス個体によって書き換えられたとする。この場合、感染回数は 3 とカウントされる。図 15 より、トロピズムを適用しない場合に比べてトロピズムを適用した場合の方が、同世代におけるウイルス個体の感染回数が少ない。このことから、ウイルス個体の宿主個体への感染がトロピズムによって制限されていることが確認できる。また、トロピズムを適用しない場合、ウイルス個体の感染回数は増加とわずかな減少を繰

り返しながら、世代を経るにつれて徐々に増加するという結果を示した。一方、トロピズムを適用した場合、ウイルス個体の感染回数は30世代付近までは緩やかに減少し、その後、緩やかに単調増加するという結果を示した。

5.2.2 トロピズムの閾値がウイルス感染操作に与える影響

図16にウイルス感染のみを用いたGAと閾値の異なるトロピズムを適用したウイルス感染を伴うGA、それぞれの手法を信号解析問題に適用したときのエラー値の推移を示す。図16より、トロピズムを適用したGAは閾値に関係なくウイルス感染のみ適用したGAよりも低いエラー値に収束していることがわかる。また、トロピズムを適用したGAの中では、

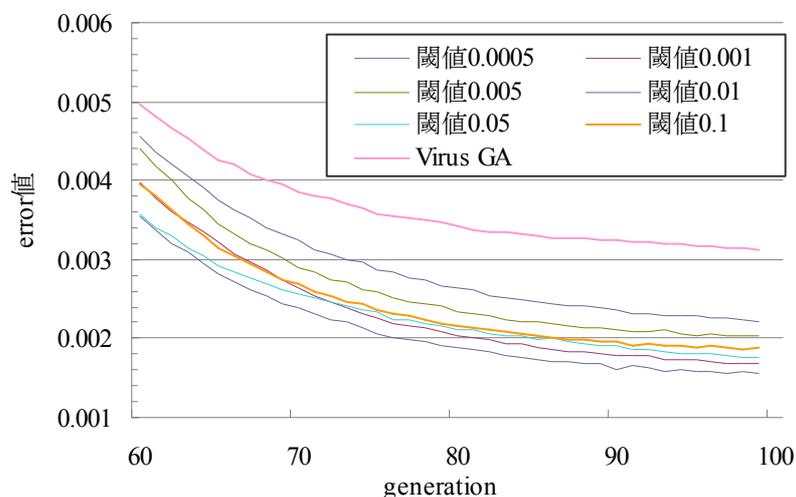


図16 トロピズムの閾値とGAの探索能力

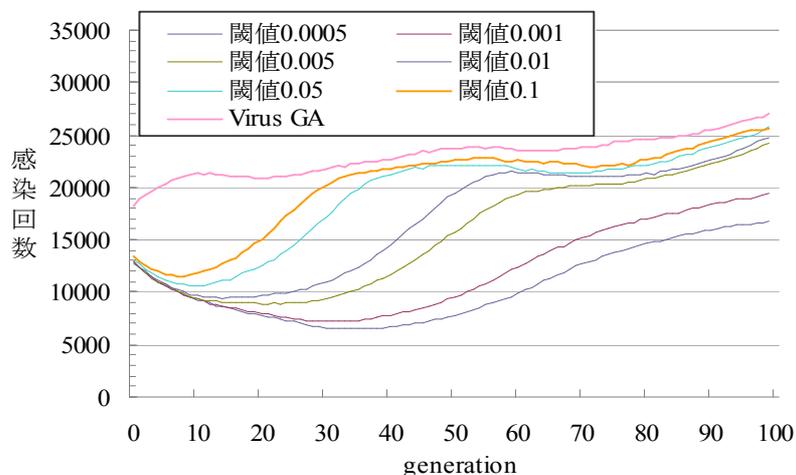


図17 トロピズムの閾値とGA上のウイルスの感染回数

閾値が 0.01 のときに最も低いエラー値に収束し、閾値が 0.0005 のときに最も高いエラー値に収束する結果となった。

図 17 にウイルス感染を伴う GA とトロピズムを適用したウイルス感染を伴う GA、それぞれのウイルス個体の感染回数の推移を示す。図 17 より、ウイルス感染操作のみを用いた場合に比べて、トロピズムを適用した場合は閾値に関係なくウイルス個体の感染回数が少なくなっていることがわかる。このことから、トロピズムによってウイルスの感染が制限されていることが確認できる。また、トロピズムを適用した場合、閾値が大きくなるに従ってウイルス個体の感染回数が多くなっていることがわかる。また、トロピズムを適用した場合は閾値に関係なく、初期世代で一度ウイルス個体の感染回数が減少し、その後、徐々に増加するという結果を示した。ウイルス個体の感染回数が減少する世代数は、閾値が小さくなるに従って長くなる傾向にある。

5.2.3 ウイルス感染操作とタブーサーチの比較

局所探索手法にタブーサーチを用いた場合と、ウイルス感染操作を用いた場合の実験結果を比較する。実験はそれぞれ 100 回行った。表 2 に各手法を用いて実験を行ったときの収束値と改善度を示す。ここで改善度は、Standard GA を基準として error 値がどの程度、改善されたかを示している。Standard GA, Virus GA, Tropism virus GA のデータは 5.2.1 章の実験で得られたデータと同じである。表 2 より、タブーサーチを用いた場合のエラー値は、ウイルス感染のみを適用した GA より低く、トロピズムを有するウイルス感染を伴う GA より高いエラー値に収束している。図 18 に 100 回実験を行ったときに収束したエラー値の分散の割合を示す。Standard GA, Virus GA, Tropism virus GA の実験結果は図 14 と同じである。タブーサーチを用いた場合、従来 GA と比較すると、エラー値が 0.0005 から 0.001 の範囲に収束した回数と、0.001 から 0.005 の範囲に収束した回数がほとんど同じとなった。また、ウイルス感染を伴う GA と比較すると、エラー値が 0.005 以下に収束した回数は共に約 80% となっている。一方、0.001 以下に収束した回数は、ウイルス感染を伴う GA が約 45% であるのに対して、タブーサーチを用いた GA は約 25% とウイルス感染を伴う GA より低い値を示している。しかしながら、0.01 以上に収束した回数はウイルス感染を伴う GA の方が高い値を示している。トロピズムを有するウイルス感染を伴う GA と比較すると、タブーサーチを用いた場合の方が 0.01 より高いエラー値に収束する回数は多く、0.01 より低いエラー値に収束する回数は少ないという結果を示した。

表 2 各 GA 手法の収束値と改善度.

	Standard GA	Virus GA	Tropism virus GA	GA+TS
error 値の収束値	0.003529	0.00313	0.001673	0.002579
改善度		11.32%	52.58%	26.92%

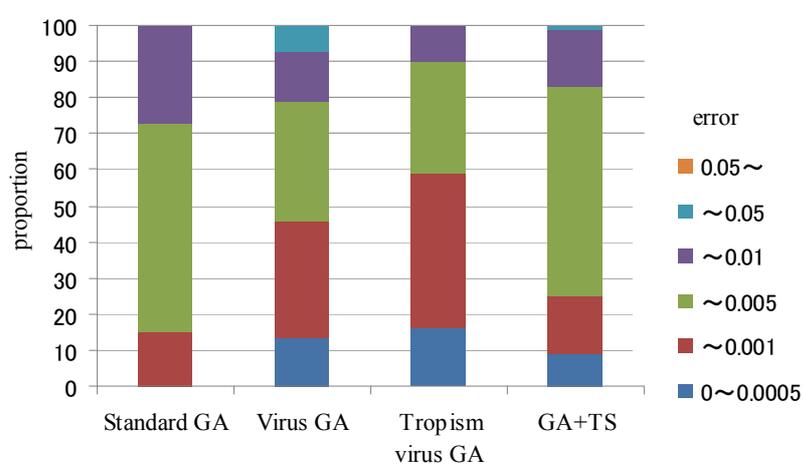


図 18 エラー値の収束値の割合.

6 考察

6.1 トロピズムが収束誤差に与える影響

まず、ウイルス感染を伴う GA を、ウイルス感染を用いない従来の GA と比較し、ウイルス感染操作が収束誤差に与える影響について考察を行う。図 14 から、ウイルス感染を伴う GA を用いた場合、従来 GA よりも低いエラー値に収束するようになる。また、0.001 よりも低いエラー値に収束する回数も増加している。このことから、ウイルス感染操作によって GA の局所探索能力が向上していることがわかる。一方、ウイルス感染を伴う GA は従来 GA には見られなかった高いエラー値にも収束するようになる。すなわち、初期収束に陥っていると考えられる。このことより、ウイルス感染操作によって初期収束に陥る可能性も高くなっていると言える。初期収束に陥りやすくなった要因は、ウイルス個体が局所解に関する遺伝情報を取り込み、宿主個体群に広めてしまったためと考えられる。例えば、初期世代において、ウイルス個体が局所解に関する遺伝情報を宿主個体から取込み、取込んだ遺伝情報を他の宿主個体に逆転写したとする。局所解によって宿主個体の評価値を向上させたウイルス個体は式 5 より感染率が上昇する。感染率が上昇したことで、局所解に関する遺伝情報をもつウイルス個体はますます他の宿主個体に感染しやすくなる。その結果、局所解に関する遺伝情報は急速に宿主個体群に広まってしまい、宿主個体群は多様性を失うこととなり、局所解に収束しやすくなったと考えられる。以上の考察より、局所探索能力の向上と初期収束へ陥る可能性の増加という 2 つの要因によって、ウイルス感染操作を用いた GA は収束誤差が大きくなる傾向があると考えられる。

次に、トロピズムを適用したウイルス感染を伴う GA を、トロピズムを適用しないウイルス感染のみの GA と比較し、トロピズムが収束誤差に与える影響について考察を行う。図 14 より、ウイルス感染操作にトロピズムを適用した場合、ウイルス感染操作のみを用いた GA に見られた高いエラー値への収束がなくなっている。このことから、初期収束に陥りにくくなっていると言える。これは、ウイルスのトロピズムによって感染が制限されたことが原因であると考えられる。図 15 より、トロピズムを用いない場合、ウイルスの感染回数はほとんど規則的に増加している。このことから、ウイルスは宿主個体群の収束状況にほとんど関係なく、常に感染による局所探索を行っていると考えられる。一方、トロピズムを用いた場合、ウイルスの感染回数は初期世代において減少し、中盤以降から徐々に増加している。このことから、ウイルスは大域的探索が十分行われていない初期世代においては感染が抑制され、ある程度、宿主個体群が収束するに従って、感染による局所探索を活発に行うようになったと考えられる。すなわち、ウイルス感染にトロピズムを適用することで、ウイルスの無駄な感染が抑制され、その結果、初期収束に陥りにくくなったと考えられる。また、初期収束に陥りにくくなった結果、図 16 より、トロピズムを用いた場合、

ウイルスのみの場合に比べて低いエラー値に収束する回数が若干増加したことがわかる。つまり、ウイルス感染操作によってもたらされた局所解に陥る可能性の増加を、トロピズムを用いることで低減することができる。言い換えるとトロピズムはウイルス感染による収束誤差の拡大を抑える効果があるといえる。

6.2 トロピズムの閾値がウイルス感染操作に与える影響

まず、トロピズムの閾値の違いが探索能力に及ぼす影響について考察を行う。図 16 より、トロピズムを適用した GA は閾値に関係なくウイルス感染のみ適用した GA よりも低いエラー値に収束している。このことから、閾値の値によらず、トロピズムによってウイルス個体の感染を制限することは有効である可能性が示された。一方、トロピズムを適用した場合を比較すると、閾値が 0.01 のとき最も低いエラー値に収束し、閾値が 0.0005 のとき最も高いエラー値に収束した。このことから、閾値には問題空間に応じてふさわしい値が存在し、トロピズムの性能を最大限発揮するためには、問題空間に応じた値を設定する必要があると考えられる。

次にトロピズムの閾値が感染回数に及ぼす影響について考察を行う。図 17 より、トロピズムを適用した場合、ウイルス個体の感染回数は必ず一度減少し、その後増加する傾向にある。ウイルス個体の感染回数が減少したということは、ウイルスの持つリガンドと宿主個体の部分染色体の差が閾値よりも大きい個体が増えてきたということである。感染回数の減少は初期世代に見られることから、宿主個体群が大域的探索を行っていることが原因と考えられる。また、ウイルス個体の感染回数が増加したということは、ウイルスの持つリガンドと宿主個体の部分染色体の差が閾値よりも小さい個体が増えてきたということである。これは宿主個体群が大域的探索によってある程度収束したためだと考えられる。以上から、トロピズムには宿主個体群が大域的探索を行っているときウイルス個体の感染を抑制し、宿主個体群が大域的探索によってある程度収束してからウイルス個体による局所探索を促進する効果があると考えられる。

次に、感染回数が減少する世代数の長さ注目すると、ウイルス個体の感染回数が減少から増加に移り変わるタイミングは、閾値が小さくなるに従って遅くなる傾向にある。言い換えれば、閾値が小さくなるに従って、ウイルス個体による局所探索が活発化するタイミングが遅くなっていると言える。つまり、閾値に設定する値によって、ウイルス個体による局所探索を活性化させるタイミングが制御できる可能性が考えられる。トロピズムの動作を決定付けるパラメータである閾値や、リガンドの定義率の違いによってウイルス感染がどのように制御されるかに関しては引き続き調査が必要である。

6.3 ウイルス感染操作とタブーサーチの比較

図 18 より，タブーサーチを用いた場合よりも，ウイルス感染を伴う GA の方が，エラー値が 0.001 以下に収束した回数が多い．一方，表 2 から，ウイルス感染操作のみを用いた GA よりも，タブーサーチを用いた場合の方がエラー値の改善度は高くなっている．ウイルス感染を伴う GA の改善度が低くなっている要因として，6.1 章で述べた初期収束が考えられる．言い換えると，タブーサーチを用いた場合の方が，100 回の実験を通して比較的，安定した探索結果が得られていると考えられる．すなわち，ウイルス感染操作のみを用いた GA は，タブーサーチに比べてより精度の高い探索結果を得やすくなるが，探索の安定性という点ではタブーサーチの方が優れていると考えられる．

一方，ウイルス感染操作にトロピズムを適用した GA は，タブーサーチを用いた場合に比べてエラー値の改善度は約 2 倍高く，エラー値が 0.001 以下に収束する回数も約 60% と高い値を示している．このことから，トロピズムをウイルス感染操作に適用した GA は，タブーサーチに比べて探索精度が向上し，安定した探索結果が得られているといえる．すなわち，トロピズムを適用したウイルス感染 GA は，局所探索法としてタブーサーチを用いるよりも有効である可能性を示している．

本稿では局所探索法の一例として単純なタブーサーチを採用し比較を行った．一方，タブーサーチの探索性能を向上させる工夫については様々な報告[25-27]があり，焼きなまし法[25]などの他の局所探索法による比較も考えられる．そのため，他手法との比較に関しては引き続き調査が必要である．

7 まとめ

本稿では、トロピズムを有するウイルス感染 GA の有効性を検証するために、ノイズを含む時系列信号の特徴解析に対してウイルス感染を用いない GA、ウイルス感染を用いた GA、トロピズムを有するウイルス感染を用いた GA を適用して探索能力への影響の違いについて比較評価を行った。その結果、ウイルス感染を用いた GA では、局所的な探索能力が向上する一方で、局所解に陥る危険性も増加する可能性を確認した。一方、トロピズムを有する GA では、局所解へ陥る危険性を減少させる効果があることを示した。また、トロピズムのパラメータである閾値の値によって、GA 上のウイルスによる局所探索を行うタイミングを制御できる可能性を示した。

文 献

- [1] A.L. Blumel, E.J. Hughes, and B.A. White, “Fuzzy Autopilot Design Using A Multiobjective Evolutionary Algorithm”, Proc. 2000 IEEE Congress on Evolutionary Computation, pp. 54-61, 2000.
- [2] H. Tamaki, H. Kita, and S. Kobayashi, “Multi objective optimization by genetic algorithms: A review”, Proc. 3rd IEEE Intl. Conf. on Evolutionary Computation, pp.517-522, 1996.
- [3] C.W. Reynolds, “Evolution of Obstacle Avoidance Behavior: Using Noise to Promote Robust Solutions”, in Advances in Genetic Programming, ed. K.E. Kinnear, Jr. , pp. 221-241, The MIT Press, 1994.
- [4] I. C. Parmee and A. H. Weston, “Preliminary Airframe Design Using Co-Evolutionary Multiobjective Genetic Algorithms”, Proc. The 1999 Genetic and Evolutionary Computation Conference, pp. 1657-1665, 1999.
- [5] I. Yoshihara, T. Aoyama and M. Yasunaga, ”GP-Based Modeling Method for Time Series Prediction with Parameter Optimization and Node Alternation.”, Proc. 2000 Congress on Evolutionary Computation, pp. 1475-1481, 2000.
- [6] H. G. Bayer and D. V. Arnold, ”Fitness Noise and Localization Errors of the Optimum in General Quadratic Fitness Models.”, Proc. the 1999 Genetic and Evolutionary Computation Conference, pp. 817-824, 1999.
- [7] R. Goto and Y. Sato, “Motion Analysis of Moving Objects with Genetic Algorithms”, Proc. The 2001 Genetic and Evolutionary Computation Conference, pp. 1276-1283, 2001.
- [8] R. Goto, and Y. Sato, “The Analysis for the Characteristics of Time Series Signals with Genetic Algorithms”, Proc. The 2004 International Conference on Computing, Communications and Control Technologies, Volume I, pp.267-272, 2004.
- [9] 安田勇太, ”トロピズムを実現したウイルス感染操作を伴う GA による時系列信号解析”, 法政大学大学院修士学位論文: 情報科学研究科; 2006 年度. 情報科学専攻, 2007.

- [10] Y. Sato, Y. Yasuda, and R. Goto. "Analysis of Noisy Time-series Signals with GA Involving Viral Infection with Tropism", Proc. of the 2007 Genetic and Evolutionary Computation Conference (GECCO-2007), ACM Press, pp.1396-1403, 2007.
- [11] 中村尚五, "デジタルフーリエ変換", pp.88-103, 東京電機大学出版局, 2003.
- [12] 山崎芳男 他, "音・音場のデジタル処理", コロナ社, 2002.
- [13] 中原, 佐川, "ウイルス進化論", 泰流社, 1989.
- [14] N. Kubota, T. Fukuda, "Scheme Representation in Virus Evolutionary Genetic Algorithm for Knapsack Problem", Proc. The 1998 IEEE International Conference of Evolutionary Computation, 1998.
- [15] 久保田, 福田, "ウイルス進化論に基づく遺伝的アルゴリズムの巡回セールスマン問題への応用", 計測自動制御学会論文集, Vol. 34, No.5, 1998.
- [16] N. Kubota and T. Fukuda, "ウイルス進化とファジィ理論", 日本ファジィ学会誌, Volume 10, No.4, pp. 626-636 , 1998.
- [17] N. Kubota, T. Fukuda, K. Shimojima, "Virus-Evolutionary Genetic Algorithm for A Self-Organizing Manufacturing System", Computer & Industrial Engineering Journal, Vol.30, No.4, pp. 1015-1026, 1996.
- [18] N. Kubota, K. Shimojima, T. Fukuda, "The Role of Virus Infection in Virus-Evolutionary Genetic Algorithm", Journal of Applied Mathematics and Computer Science, Vol.6, No.3, pp.415-429, 1996.
- [19] N. Kubota, T. Fukuda, K. Shimojima, "Trajectory Planning of Cellular Manipulator System Using Virus-Evolutionary Genetic Algorithm", Robotics and Autonomous Systems 19, pp.85-94, 1996.
- [20] 下島康嗣, 久保田直行, 福田敏男, "ウイルス進化論に基づく学習型ファジィコントローラ", 日本機械学会論文集 C 編, 63 巻 608 号, pp. 1261-1268, 1997.

- [21] 久保田, 福田, 小島, 森岡, ”ファジィコントローラのための進化的計算とウイルス感染”, 第14回ファジィシステムシンポジウム講演論文集, pp. 549-552, 1998.
- [22] 吉開 泰信 他, ”ウイルス・細菌と感染症がわかる”, ISBN 4-89706-960-2, 羊土社, 2004.
- [23] 加藤 茂明, “受容体がわかるーシグナル伝達を司る受容体の機能から多様な生命現象まで”, ISBN 4-89706-958-0, 羊土社, 2003.
- [24] Sadiq M. Sait, Habib Youssef, “組合せ最適化アルゴリズムの最新手法ー基礎から工学応用まで”, 丸善, 2002.
- [25] 北野 宏明 編, “遺伝的アルゴリズム”, 産業図書, 1993.
- [26] Kurahashi S. and Terano T, “A Genetic Algorithm with Tabu Search for Multimodal and Multiobjective Function Optimization“, Proc. of Genetic and Evolutionary Computation Conference (GECCO-2000), pp. 291-298, 2000.
- [27] 清水, 和田, “ハイブリッドタブサーチによる階層型ロジスティクス最適化“, システム制御情報学会論文誌, 17, pp. 241-248, 2004.