

グローバルベストを向上させる可能性を考慮した PSO を用いた遺伝子ネットワーク推定手法の提案

村田 裕章^{†1} 木村 春彦^{†1}

本研究では遺伝子ネットワークの S-system モデル推定問題を対象としたグローバルベストを向上させる可能性を考慮した PSO の提案を行う。PSO は全粒子の再評価を行うことで、グローバルベストを向上させ最適化を行っている。しかしながら全粒子がグローバルベストを向上させる可能性を同程度に有しているとは考えにくい。そこで遺伝子ネットワークの S-system モデル推定手法の一種である PUNN モデルを用いた推定手法を用いて、グローバルベストを向上させる可能性を推定し、その可能性が高い粒子のみを再評価対象とする手法の提案を行う。評価実験の結果、従来手法に比べ実行時間、推定精度ともに良い結果を得ることができたことを示す。

PSO Using Possibility to Improve the Global Best for Inference of the Genetic Networks

HIROAKI MURATA^{†1} and HARUHIKO KIMURA^{†1}

In this paper, we proposed the PSO using possibility to improve the global best for inference of S-system models in the genetic networks. In order to find an optimal solution, PSO improves the global best, which is the position of the best solution found by past optimization, by re-evaluating all particles. However, it is not likely that all particles have possibility to improve the global best at the same level. Therefore, the proposed method will estimate the possibility to improve the global best using PUNN model's fitness function, and limit re-evaluation particles. The results show the proposed method is able to obtain a better solution compared to the previous method which has not used the possibility to improve the global best.

^{†1} 金沢大学
Kanazawa University

1. はじめに

生体内では様々な遺伝子が互いに制御し合い、発現量を調節することで生命活動の維持が行われている。そのような制御関係は遺伝子ネットワークと呼ばれ、新薬の開発や病気の治療に応用できることから、遺伝子ネットワークの解析や推定を行う研究が行われている^{1),2)}。特に DNA マイクロアレイ技術の進歩により、細胞全体レベルでの遺伝子発現量の解析が可能になってきており、これらの情報を用いた遺伝子ネットワーク推定が注目されている。

遺伝子発現量の時系列データから、遺伝子ネットワークを推定することは、遺伝子ネットワーク推定問題と呼ばれ、生物学的実験により得られた遺伝子発現量の時系列データを再現する数理モデルを求めることが目的となる。遺伝子ネットワークを示す数理モデルとして、ブリアンネットワークやベイジアンネットワーク、微分方程式モデルなどが提案されている^{1),2)}。それらのモデルの中で本研究では、微分方程式モデルの一種である S-system モデル³⁾ と Product Unit based NN (PUNN) モデル⁴⁾ に注目する。

S-system モデルは環境や未知の転写制御因子など、多くの要素を省略したモデルであるが、遺伝子の制御関係について有意義なヒントを得る手法として期待されている。しかしながらネットワーク推定の際に、微分方程式を繰り返し解く必要があり膨大な計算量を必要とする。なお一般に生物学的実験により得られる遺伝子発現量の時系列データには実験誤差が多く含まれることから、S-system モデルを用いて遺伝子ネットワーク推定を行う際には、遺伝子ネットワークの大まかな構造を求めようとする研究も多く行われている。しかしながら、本研究では実験誤差を考えていないため、生物学的実験により得られた遺伝子発現量の時系列データを正確に再現できる。そのため本研究では、時系列データを正確に再現することを目標とする。PUNN モデルはニューラルネットワークの一種である Product Unit based NN (PUNN)⁵⁾ を用いて S-system モデルを表現したモデル⁴⁾ である。PUNN モデルを用いた推定手法では遺伝子発現量と発現変化量の関係を PUNN の学習によって得るため、S-system モデルを用いた推定手法とは異なり、微分方程式を解く必要がなく大幅な計算量の削減ができる。しかしながら遺伝子発現量の時系列データから、正確に発現変化量を求めることは難しいため、推定精度は S-system モデルを用いた推定手法に比べ劣ってしまう。

本研究では数理モデルではなく最適化手法、特に Particle Swarm Optimization (PSO)^{6),7)} に注目し、新しい遺伝子ネットワーク推定手法の提案を行う。PSO は鳥や魚などの群れをなして移動する生物のそのような行動パターンから着想を得た最適化手法で

あり、各粒子がそれまでに各自が発見した最も良い解候補の位置（ローカルベスト）を記憶し、かつ各粒子が持つローカルベストの中で最も良い解候補の位置をグローバルベストとして群れ全体が共有しており、粒子がその方向に移動することで最適化を行う。PSO は集団探索戦略であるため、進化的計算手法の 1 つに含まれるが、遺伝的アルゴリズム（Genetic Algorithm: GA）⁸⁾ のように選択、交叉、突然変異は行わず、各粒子がグローバルベストに向かって移動するため、GA に比べ収束性が良いという特徴を持ち、他のメタヒューリスティクスと比べて遜色ない性能を有することが報告されている^{7),9)}。しかしながら解候補の評価を幾度となく繰り返すため、設計などのシミュレーションを行い解候補の評価を行う問題などにおいては、さらなる高速化が求められている。

そこで本研究では評価関数の使用回数を減少させることで PSO の高速化を図る。PSO において粒子の評価が必要な理由の 1 つとして、グローバルベストの向上があげられる。PSO はグローバルベストに粒子を集めることで最適化を行っており、最適解を示す問題空間上の位置にグローバルベストが近づいていくことで最適解を求めている。この際グローバルベストをより近付けるために、すべての粒子の現在位置が示す解候補の評価値を求めグローバルベストを向上させている。これはグローバルベストが向上する可能性を少しでも上げるために、すべての粒子を再評価しているのとらえることもできる。しかしながらすべての粒子が“グローバルベストを向上させる可能性”を同程度に有するとは考えにくく、可能性の高い粒子や低い粒子などが混在していることは想像に難くない。そのためグローバルベストを向上させる可能性が低い粒子に対して、再評価を行わないようにすることで、再評価回数の軽減を実現し、計算時間の短縮を図るのではないかと考え、グローバルベストを向上させる可能性を考慮した Particle Swarm Optimization を用いる遺伝子ネットワーク推定手法の提案を行う。

2. 遺伝子ネットワーク推定方法

2.1 S-system モデルを用いた推定手法

2.1.1 S-system モデル

S-system モデル³⁾ は環境や未知の転写制御因子など、多くの要素を省略したモデルであるが、遺伝子の制御関係について有意義なヒントを得る手法として期待されている。S-system モデルはすべての遺伝子についてその関与を意味するパラメータを含むモデルであり、 N 個の遺伝子から構成される遺伝子ネットワークを S-system モデルで表現すると、遺伝子 i の発現量 X_i の発現変化量 $\frac{dX_i}{dt}$ （転写速度に相当）を用いて、次のようになる。

$$\begin{cases} \frac{dX_1}{dt} = \alpha_1 \prod_{j=1}^N X_j^{g_{1,j}} - \beta_1 \prod_{j=1}^N X_j^{h_{1,j}} \\ \vdots \\ \frac{dX_N}{dt} = \alpha_N \prod_{j=1}^N X_j^{g_{N,j}} - \beta_N \prod_{j=1}^N X_j^{h_{N,j}} \end{cases} \quad (1)$$

ここで i, j は遺伝子ネットワークに含まれる遺伝子を示しており、 $i = 1, \dots, N$ および $j = 1, \dots, N$ となる。 $g_{i,j}$ は X_i の発現過程に關する X_j の相互作用係数、 $h_{i,j}$ は X_i の抑制過程に關する X_j の相互作用係数、 α_i, β_i はそれぞれ X_i の発現項、抑制項に乘じる係数を示す。このモデルは複数の発現・抑制過程からなる反応経路を近似した表現となっており、詳細な構造が明らかになっていない遺伝子ネットワークの記述に有効であると考えられている¹⁰⁾。S-system モデルを用いることで、 N 個の遺伝子からなる遺伝子ネットワークを $2N(N+1)$ 個の S-system パラメータ $\alpha_i, \beta_i, g_{i,j}, h_{i,j}$ で表現できる。また $g_{i,j}$ と $h_{i,j}$ は遺伝子 i と遺伝子 j 間の制御関係を表し、これらの値によって遺伝子ネットワークが大きく変化することから、キネティックオーダー（kinetic order）と呼ばれている^{10),11)}、*1。

2.1.2 S-system モデルを用いた評価関数

S-system モデルを用いた遺伝子ネットワーク推定問題を最適化問題として考えた場合、生物学的実験により得られた遺伝子発現量の時系列データを再現する S-system パラメータを求める問題として、以下のように定式化される。

$$f = \sum_{i=1}^N \sum_{t=1}^T \left(\frac{X_{cal,i,t} - X_{exp,i,t}}{X_{exp,i,t}} \right)^2 \quad (2)$$

ここで $X_{exp,i,t}$ は生物学的実験により得られた時刻 t における遺伝子 i の発現量、 $X_{cal,i,t}$ は連立微分方程式 (1) を解くことによって得られる遺伝子発現量、 N はネットワークに含まれる遺伝子数、 T は生物学的実験により得られた時系列データのタイムポイント数である。

しかしながら遺伝子ネットワーク推定問題の問題空間は $2N(N+1)$ 次元であるため、既存の最適化手法を用いて大規模な遺伝子ネットワークを推定することは困難である。そのため問題をいくつかのサブ問題（各遺伝子に相当）に分割し、遺伝子ネットワークの推定を行

*1 正確にはキネティックオーダーとは化学反応論/反応速度論における反応次数を指すため、 $g_{i,j}$ や $h_{i,j}$ はキネティックオーダーに相当する量となるが、本研究では文献 10), 11) にないキネティックオーダーと呼ぶ。

う方法が提案されている¹⁰⁾。遺伝子 i に相当するサブ問題は次のように定式化される。

$$f_i = \sum_{t=1}^T \left(\frac{X'_{cal,i,t} - X_{exp,i,t}}{X_{exp,i,t}} \right)^2 \quad (3)$$

$$\frac{dX_i}{dt} = \alpha_i \prod_{j=1}^N Y_j^{g_{i,j}} - \beta_i \prod_{j=1}^N Y_j^{h_{i,j}} \quad (4)$$

$$Y_j = \begin{cases} X'_{cal,i} & \text{if } j = i \\ \hat{X}_j & \text{if } j \neq i \end{cases} \quad (5)$$

ここで $X'_{cal,i,t}$ は微分方程式 (4) を解くことで得られる時刻 t における遺伝子 i の発現量を示し、 \hat{X}_j は生物学的実験により得られた時系列データから補間することで求める。補間方法として生物学的実験により得られた遺伝子発現量の時系列データに実験誤差などのノイズが含まれない場合はスプライン補間を用い、ノイズが含まれると考えられる場合は局所線形回帰などの平滑化手法を用いて補間を行う¹⁰⁾。このようにサブ問題に分割することで、 $2N(N+1)$ 次元の遺伝子ネットワーク推定問題を、 N 個の $2(N+1)$ 次元のサブ問題に変換できる。

S-system モデルは自由度が高いため複数の解が存在する可能性があり、また一般的に与えられる遺伝子発現量の時系列データには実験誤差が含まれるため、正確な S-system パラメータを求めることは困難である。そこで遺伝子ネットワークの結合は疎であるという特徴を導入した評価関数が提案されている¹²⁾。S-system モデルにおいて $g_{i,j}$ は X_i の発現過程に関与する X_j の相互作用係数であり、 $g_{i,j} = 0$ とは遺伝子 j は遺伝子 i の発現過程に関与していないことを意味する。同様に $h_{i,j}$ は抑制過程の相互作用係数であり、 $h_{i,j} = 0$ とは遺伝子 j は遺伝子 i の抑制過程に関与していないことを意味する。このことから遺伝子ネットワークの結合が疎になる場合、多くのキネティックオーダー ($g_{i,j}$ や $h_{i,j}$) が 0 になることが分かり、以下のペナルティ項が提案されている。

$$P_i = P_c \sum_{j=1}^{2N-I} (|K_{i,j}|) \quad (6)$$

ここで $K_{i,j}$ は遺伝子 i のキネティックオーダーの絶対値を昇順にソートしたときの j 番目の値を意味する。 P_c はペナルティ係数、 N は遺伝子数、 $2N$ はキネティックオーダー数、 I は最大入り次数を示す。ここで、最大入り次数とは遺伝子 i に影響を及ぼす遺伝子の最大数を規

定するパラメータである。

また統計モデルの良さを評価する際に多く用いられる赤池情報量基準 (Akaike's Information Criteria: AIC) を用いた評価関数も Noman らにより提案されている¹¹⁾。Noman らは生物学的実験により得られた遺伝子発現量 $X_{exp,i,t}$ と推定したパラメータから計算した遺伝子発現量 $X_{cal,i,t}$ との絶対誤差 $\epsilon = (X_{cal,i,t} - X_{exp,i,t})$ が正規分布に従うとし、式 (7) の AIC を導入した評価関数を定義している。ここで Φ はパラメータ数を表す。

$$f_i^S = -2\Lambda + 2\Phi + P_i \quad (7)$$

$$\Lambda = -\frac{1}{2\sigma_i^2} (X_{cal,i,t} - X_{exp,i,t})^2 - \frac{T}{2} \ln(2\pi\sigma_i^2) \quad (8)$$

$$\sigma_i^2 = \frac{1}{T} \sum_{k=1}^T (X_{cal,i,t} - X_{exp,i,t})^2 \quad (9)$$

同じ遺伝子ネットワークを対象とした遺伝子ネットワーク推定問題において、文献 10) および文献 12) よりも文献 11) の方が良い S-system パラメータが求められていたために、本研究では S-system モデルを用いた評価関数として式 (7) を用いる。

2.2 PUNN モデルを用いた推定手法

2.2.1 PUNN モデル

Product Unit based NN (PUNN)⁵⁾ は入出力関係が総乗で定義される Product Unit を用いたニューラルネットワークであり、中間層は Product Unit が用いられ、出力層では入出力関係が総和で定義される一般的なユニット (Summation Unit と呼ぶ) で構成されている。 k 番目の中間層ユニットの出力 y_k^H および唯一の出力層ユニットの出力 y^O は、それぞれ式 (10) および式 (11) となる。

$$y_k^H = \prod_{j=1}^N (X_j)^{w_{j,k}^{IH}} \quad (10)$$

$$y^O = \sum_{k=1}^{N_H} w_k^{HO} y_k^H - \theta^O \quad (11)$$

ここで入力層のユニット数を N 、中間層のユニット数を N_H 、出力層のユニット数を 1 とし、 X_j は j 番目の入力層ユニットの入力値、 $w_{j,k}^{IH}$ は j 番目の入力層ユニットと k 番目の中間層ユニット間の結合荷重、 w_k^{HO} は k 番目の中間層ユニットと出力層ユニット間の結合荷重、 θ^O は出力層ユニットの閾値パラメータを示す。

PUNN を次の設定にすることで、S-system モデルの遺伝子ネットワークを表現する PUNN

モデルを定義することができる⁴⁾。入力層のユニット数を遺伝子数 N 、中間層のユニットを 2 個、出力層のユニットを 1 個に設定し、 j 番目の入力層ユニットと 1 番目の中間層ユニット間の結合荷重 $w_{j,1}^{IH}$ を $g_{i,j}$ 、 j 番目の入力層ユニットと 2 番目の中間層ユニット間の結合荷重 $w_{j,2}^{IH}$ を $h_{i,j}$ 、1 番目の中間層ユニットと出力層ユニット間の結合荷重 w_1^{HO} を α_i 、2 番目の中間層ユニットと出力層ユニット間の結合荷重 w_2^{HO} を $-\beta_i$ 、出力層のユニットの閾値パラメータ $\theta^O = 0$ とする。この PUNN の入出力関係は、式 (12) で表され、この式は S-system モデルにおける遺伝子 i の発現量に関する微分方程式 (13) と同じになる。

$$y_{i,t}^O = \alpha_i \prod_{j=1}^N X_j^{g_{i,j}} - \beta_i \prod_{j=1}^N X_j^{h_{i,j}} \quad (12)$$

$$\frac{dX_{i,t}}{dt} = \alpha_i \prod_{j=1}^N X_{j,t}^{g_{i,j}} - \beta_i \prod_{j=1}^N X_{j,t}^{h_{i,j}} \quad (13)$$

PUNN を用いた遺伝子ネットワーク推定問題は、遺伝子の発現変化量の時系列データ（生物学的実験により得られた遺伝子発現量の時系列データから求める）を再現する S-system パラメータを求める問題となり、次のように定式化される。

$$f_{PUNN} = \sum_{i=1}^N \sum_{t=1}^T (y_{i,t}^O - y_{i,t})^2 \quad (14)$$

ここで $y_{i,t}^O$ は式 (12) を用いて求められた時刻 t における遺伝子 i の発現変化量を示し、 $y_{i,t}$ は生物学的実験により得られた遺伝子発現量の時系列データの傾きを計算することで得た時刻 t における遺伝子 i の発現変化量を示す。

2.2.2 PUNN モデルを用いた評価関数

遺伝子 i に相当するサブ問題を PUNN モデルを用いて解くことは、時刻 t における全遺伝子の発現量ベクトル $\mathbf{X}_t = (X_{1,t}, \dots, X_{N,t})$ に対して、時刻 t における遺伝子 i の発現変化量 $y_{i,t}$ を出力する PUNN を学習させることになる。PUNN モデルを用いた評価関数は、学習の際の誤差関数となり、遺伝子の発現変化量を一致させる目的関数項と、キネティックオーダの個数制限に由来するペナルティ項で構成され、目的関数項は式 (15) で定義される⁴⁾。

$$E'_{P,i} = \begin{cases} -\infty & \text{if } e_{P,i} = 0 \\ \log(e_{P,i}) & \text{if } 0 < e_{P,i} < 1 \\ e_{P,i} & \text{if } 1 \leq e_{P,i} \end{cases} \quad (15)$$

$$e_{P,i} = \sum_{t=1}^T (y^O(\mathbf{X}_t) - y_{i,t})^2 \quad (16)$$

「遺伝子ネットワークの結合は疎である」という遺伝子ネットワークの特徴は、S-system モデルを用いた評価関数と同様に、非零のキネティックオーダの個数を制限することでペナルティ項に導入される。しかしながら S-system モデルを用いた評価関数では、それらの制限を固定（定数パラメータ）としていたが、PUNN モデルを用いた評価関数では制限を動的に決めている⁴⁾。基本的な考えは、キネティックオーダを絶対値を基準に、2 つのクラスに分け、絶対値が小さい方のクラスに属するキネティックオーダを 0 にするようにペナルティを与えるというものである。なお非零のキネティックオーダのみを対象にクラスタリングを行う。しかし数個のキネティックオーダの値が 0 となり、それ以外の多くのキネティックオーダの絶対値が比較的大きな値になる場合、結合が疎の遺伝子ネットワークを表現していないにもかかわらず、ペナルティ値は小さくなってしまふ。このような事態を避けるために、値が 0 のキネティックオーダ数が多いほどペナルティ値を小さくするようにしている。クラスタリングを用いてキネティックオーダを値が 0 であるクラス K_{c0} 、絶対値の小さなクラス K_{c1} 、絶対値の大きなクラス K_{c2} に分け、 K_{nz} を K_{c0} の要素数とすると、クラスタリングを用いたペナルティ項は以下のようになる。

$$P_{P,i} = \begin{cases} \infty & \text{if } K_{nz} < 2 \\ 0 & \text{if } K_{nz} = 2 \\ \sum_{x \in K_{c1}} |x| + (2N - K_{nz}) & \text{otherwise} \end{cases} \quad (17)$$

N は遺伝子数であり、 $2N$ はキネティックオーダ数となる。 $\sum_{x \in K_{c1}} |x|$ は絶対値が小さいと判断されたキネティックオーダの絶対値の合計値を表しており、絶対値が小さなキネティックオーダをより絶対値の小さな値にする働きがある。 $(2N - K_{nz})$ は値が 0 のキネティックオーダの数が多いほど小さな値となり、結合が疎の状態といえない、数個のキネティックオーダが 0 になっている場合にペナルティ項を大きな値にする働きがある。値が 0 のキネティックオーダの個数 (K_{nz}) を 2 と比較して場合分けを行う理由として、文献 4) には次のように書かれている。「非零のキネティックオーダ数が 0 個の場合は遺伝子発現量が一定値となり、1 個の場合は発現過程もしくは抑制過程のみとなり遺伝子発現量が単調増加もしくは単調減少になると考えられる。そのため、発現過程と抑制過程のバランスによって遺伝子発現量が決まるとする S-system モデルを再現するためには、非零のキネティックオーダが 2 個以上

表 1 設定した S-system パラメータ
Table 1 Set S-system parameters.

j	α_j	β_j	$g_{1,j}$	$g_{2,j}$	$g_{3,j}$	$g_{4,j}$	$g_{5,j}$	$h_{1,j}$	$h_{2,j}$	$h_{3,j}$	$h_{4,j}$	$h_{5,j}$
1	5	10	0	0	1	0	-1	2	0	0	0	0
2	10	10	2	0	0	0	0	0	2	0	0	0
3	10	10	0	-1	0	0	0	0	-1	2	0	0
4	8	10	0	0	2	0	-1	0	0	0	2	0
5	10	10	0	0	0	2	0	0	0	0	0	2

必要であると考えペナルティ項の設計を行った」。

ペナルティ項が付加されている目的関数の問題を解く際には、しばしば本来の目的関数の項とペナルティ項にそれぞれ動的に変化する係数を乗じることで、効率的に最適化を行う方法^{13),14)} が用いられ、PUNN を用いた評価関数においても、動的に変化する係数を導入している。ネットワーク推定開始時刻を 0 として終了時刻を t_{end} 、現在の時刻を t とすると、PUNN モデルを用いた評価関数は式 (18) となる。

$$f_i^P = c_P(t) \times E'_{P,i} + (1 - c_P(t)) \times P_{P,i} \tag{18}$$

$$c_P(t) = \left(1 - \left(1 - \frac{t}{t_{end}}\right)^2\right) \times 0.4 + 0.5 \tag{19}$$

2.3 S-system モデルと PUNN モデルの比較

S-system モデルを用いた評価関数による解の評価精度と、PUNN モデルを用いた評価関数による解の評価精度の違いについて以下の方法で検証を行った。5 個の遺伝子で構成される S-system モデルの遺伝子ネットワーク (表 1 に示す値を S-system パラメータとして設定した) を用意し、連立微分方程式 (1) を解くことによって遺伝子発現量の時系列データを得る。その後、設定した S-system パラメータの各値を中心にして $\pm 3, \pm 2, \pm 1, \pm 0.1, \pm 0.01, \pm 0.001$ の 6 つの範囲を設定し、その範囲内で一様乱数を用いてそれぞれ 5 万個、計 30 万個の解候補を求め、各解候補に対してパラメータ誤差 (設定した S-system パラメータとの絶対誤差の総和)、S-system モデルを用いた評価関数値、PUNN モデルを用いた評価関数値を計算した。

まず S-system モデルを用いた評価関数値とパラメータ誤差の関係を図 1 に示す。図中の各点が解候補を表している。この図からパラメータ誤差が約 0.01 以下の解候補については、評価関数値が一定値になっており評価できないことが分かる。このような現象はサブ問題に分割せずに評価関数を求めた場合 (式 (2) を用いた評価関数) では発生しないことから、サブ問題に分割しているためだと考えられる⁴⁾。

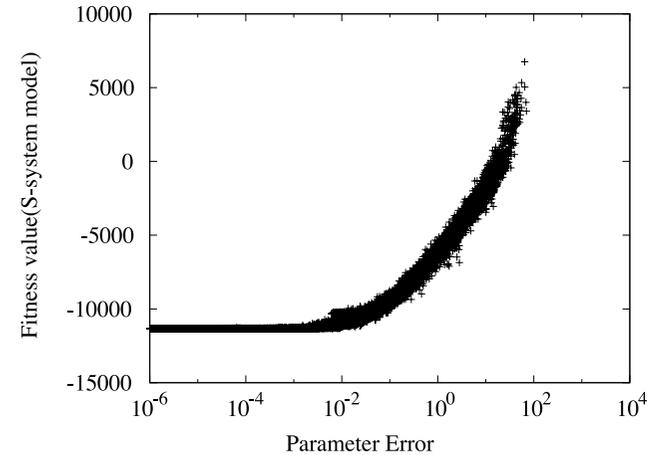


図 1 S-system モデルを用いた評価関数とパラメータ誤差の関係

Fig. 1 The relation between S-system model's fitness function value and parameter error.

次に PUNN モデルを用いた評価関数値とパラメータ誤差の関係を図 2 に示す。この図から、パラメータ誤差が 1 付近において評価関数値が 10 になる可能性があること、パラメータ誤差が 10^{-2} においても評価関数値が 10 になることが分かる。これよりパラメータ誤差が約 1 以下の解候補に関しては、評価できないことが分かる。この値は S-system モデルを用いた評価関数に比べて大きな値となっており、PUNN モデルを用いた評価関数は S-system モデルを用いた評価関数に比べ、パラメータ誤差が小さい、良い解候補の評価が行えないことが分かる。PUNN モデルを用いた評価関数では、生物学的実験により得られた遺伝子発現量の時系列データからその発現変化量を計算し、その値との差を評価関数値に用いる。しかしながら得られる遺伝子発現量の時系列データのサンプリング間隔が広いため、遺伝子の発現変化量を正確に計算することができず、学習の目標となる値に多くの計算誤差が含まれる。そのため良い解候補の評価を行うことができないと考えられる⁴⁾。

また各関数の実行時間に関しては、文献 4) から分かり、S-system モデルを用いた推定手法は 19,891 秒であることに対し、PUNN モデルを用いた推定手法は 123.43 秒となっている。2 つの推定手法は評価関数が異なるだけで、同じ最適化手法を使っており、評価関数の計算時間の差が推定手法の実行時間の差になっている。

以上のことから、S-system モデルを用いた評価関数は、計算量は多いがより正確に評価

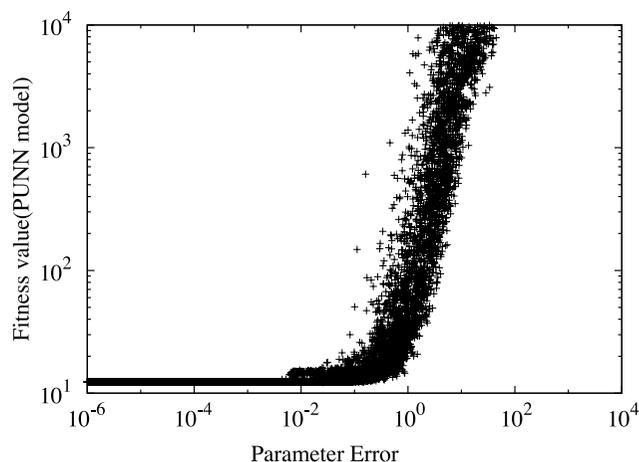


図 2 PUNN を用いた評価関数とパラメータ誤差の関係

Fig. 2 The relation between PUNN model's fitness function value and parameter error.

できることが分かる．一方，PUNN モデルを用いた評価関数は，評価の正確性に関しては S-system モデルを用いた評価関数に劣ってしまうが，計算量が少なく抑えられていることが分かる．

3. Particle Swarm Optimization

本研究では最適化手法，特に PSO に注目して遺伝子ネットワーク推定問題に取り組む．そのため，ここで PSO について簡単に紹介する．鳥の群れや魚の群泳における個体は自身の情報だけでなく群れの情報も用いて行動しており，PSO はそのような群れの中における個体の振舞いを模倣した最適化手法である⁶⁾．PSO において個体は粒子 (particle) と呼ばれ，与えられた問題の解候補を表す問題空間上の座標 (位置) と移動方向を表す速度を持っており，群れをなして D 次元の問題空間を移動することで最適化が行われる．そのため PSO は各粒子の速度を制御し，粒子の位置を移動させることで，与えられた問題の最適解を見つけようとする手法ともいえる．各粒子における速度の制御には，自身の情報として過去の最適化における最良の位置を示すローカルベスト \mathbf{p}_i および，群れ全体の過去の最適化における最良の位置を示すグローバルベスト \mathbf{p}_g が用いられる． \mathbf{p}_i および \mathbf{p}_g は次のように定義される．

$$f(\mathbf{p}_i) = \min (\{f(\mathbf{x}_i^k) \mid k = 1, \dots, k_n\}) \quad (20)$$

$$f(\mathbf{p}_g) = \min (\{f(\mathbf{p}_i) \mid i = 1, \dots, N_p\}) \quad (21)$$

ここで $f(\mathbf{x})$ は解候補 \mathbf{x} の評価関数値を表し，また評価関数値が最も小さい解候補が最適解になるように設定されているとする． \mathbf{x}_i^k は粒子 i の k 回移動後の位置， k_n は現在の繰返し回数， N_p は粒子数を表す． $\min(\cdot)$ は与えられた集合の中で最も小さい要素を返す関数である．これらの情報を用いて基本的な速度 $\mathbf{v}_i^k = (v_{i,1}^k, \dots, v_{i,D}^k)$ および位置 $\mathbf{x}_i^k = (x_{i,1}^k, \dots, x_{i,D}^k)$ の更新式は次のように定義される．

$$v_{i,d}^{k+1} = wv_{i,d}^k + c_1 \text{rand}_1 \times (p_{i,d} - x_{i,d}^k) + c_2 \text{rand}_2 \times (p_{g,d} - x_{i,d}^k) \quad (22)$$

$$x_{i,d}^{k+1} = x_{i,d}^k + v_{i,d}^{k+1} \quad (23)$$

\mathbf{v}_i^k は粒子 i の k 回移動後の速度， c_1, c_2 は，それぞれ $\mathbf{p}_i, \mathbf{p}_g$ が $(k+1)$ 回移動後の速度 \mathbf{v}_i^{k+1} に及ぼす影響を示す係数である．なお $w = 0.730, c_1 = c_2 = 1.496$ が良い値であるといわれている¹⁵⁾． $\text{rand}_1, \text{rand}_2$ は 0 から 1 の一様乱数を表し，毎回更新時に各粒子，各次元ごとに独立に乱数を発生させる．また速度には上限速度 $v_{max,d}$ と下限速度 $-v_{max,d}$ が定められており，速度がその範囲内に収まるように，速度を次のように修正する．

$$v_{i,d}^{k+1} = \begin{cases} v_{max,d} & \text{if } v_{i,d}^{k+1} \geq v_{max,d} \\ -v_{max,d} & \text{if } v_{i,d}^{k+1} \leq -v_{max,d} \\ v_{i,d}^{k+1} & \text{else} \end{cases} \quad (24)$$

本研究で対象とする遺伝子ネットワーク推定問題では，決定変数である S-system パラメータに上下制限約が設けられている (詳しくは 5.1 節を参照)．決定変数 d の上下制限約を $[R_{min,d}, R_{max,d}]$ とすると，本研究では $v_{max,d}$ を文献 4) と同様に $v_{max,d} = (R_{max,d} - R_{min,d})/2.0$ と定義する．初期粒子の生成方法は，位置，速度とともにランダムで生成する．PSO のアルゴリズムを図 3 に示す．3~11 行目では各粒子の再評価を行い，グローバルベストおよびローカルベストを向上させている．12~14 行目では粒子の移動および速度変更を行っている．

上記は基本的な PSO について述べたにすぎず，速度の更新式などに様々な改良を加え，より高精度の解候補を求めることができる手法が提案されている．しかしながら本研究ではグローバルベストを向上させる可能性を用いて，再評価する粒子を決める手法を提案しており，速度の更新式などにはいっさい変更を加えない．そのため本提案手法に新たに速度更

```

(1) 各粒子の位置と速度をランダムに初期化
(2) Repeat 終了条件を満たすまで do
(3)   for 全ての粒子  $i$  に対して do
(4)      $f(x_i^k)$  を計算する
(5)     if  $f(x_i^k) < f(p_i)$  then
(6)        $p_i$  を向上させる
(7)     end if
(8)     if  $f(p_i) < f(p_g)$  then
(9)        $p_g$  を向上させる
(10)    end if
(11)  end do
(12) for 全ての粒子  $i$  に対して do
(13)   式 (22) 及び (23) により, 粒子  $i$  の位置  $x_i^k$  と速度  $v_i^k$  を変更
(14) end do
(15) end do

```

図 3 PSO のアルゴリズム
Fig.3 PSO algorithm.

新式などを導入することは可能であり, それにより性能向上は同程度に期待することができる. 以上のことから, 本研究では基本的な PSO のみを対象にする.

4. グローバルベストを向上させる可能性を考慮した PSO

4.1 基本アイデア

PSO はグローバルベスト付近に粒子を集めることで最適化を行っている. そのため最適化を進めていくうえで, グローバルベストが最適解の位置に近づいていくことが重要であり, グローバルベストがそれ以外の位置に近づくことは, うまく最適化が行われていないことを意味している. そのようなことから, グローバルベストを向上させる作業は重要であり, そのため毎回すべての粒子の位置を再評価しているといっても過言ではないと考えられる. これはすべての粒子がグローバルベストを向上させる可能性を秘めており, グローバルベストが向上する可能性を最大にするための手段とも考えることができる. しかしながらグローバルベストを向上させる可能性がすべての粒子で等しいという保証はなく, 可能性の高い粒子や低い粒子などが混在することは想像に難くない. それを示すデータとして, 図 4 を提示する. これは S-system モデルを用いた評価関数を用いて PSO を実行した際に, 群れ内における各粒子の現在位置の評価関数値の平均値, 最大値, 最小値の推移を示している. これから, つねに最小値と最大値に大きな開きがあり, 粒子の評価値が分散していることが分かる. グローバルベストは過去の最適化において最も評価値の良い位置を示すため, 群れ内

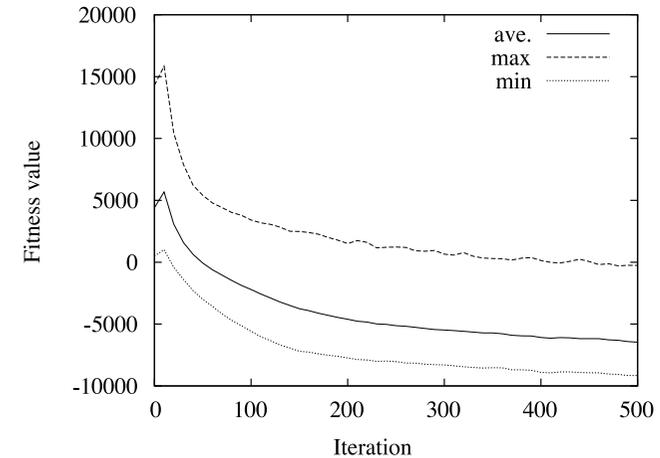


図 4 S-system モデルを用いた評価関数値の分布の推移

Fig.4 The progress of distribution of S-system model's fitness function value.

において評価値が悪い (値が大きい) 粒子はグローバルベストを向上させる可能性はないと考えることができる*1.

以上のことより, グローバルベストを向上させる可能性の低い粒子を再評価しないようにすることで, 再評価に必要な評価関数の使用回数を抑制し, 高速に最適化を行えるのではないかと考えられる. このような考えのもと, 本研究ではグローバルベストを向上させる可能性を考慮した PSO を提案する.

4.2 可能性推定関数

提案手法は PSO をベースとし, グローバルベストを向上させる可能性の低い粒子の再評価を行わないようにしているため, 可能性の求め方が重要な要素となる. しかしながらグローバルベストを向上させる可能性を多くの計算量を用いて厳密に計算することはできない. もしグローバルベストを向上させる可能性を, 評価関数よりも多くの計算量を用いて厳密に求めてしまうと, 評価関数を使用する回数を減らしたとしても最適化に必要な計算時間

*1 同時に群れ内に存在する 2 つの粒子 A と B を考えた場合, 評価値が低い粒子 A もグローバルベストを向上させる可能性があるが, より評価値の良い粒子 B によってただちにグローバルベストが向上するため, 粒子 A によってなされるグローバルベストの向上には意味はないと見ることができる. したがって, 粒子 A が持つグローバルベストを向上させる可能性を 0 と見なしても問題ないと考えられる.

は逆に多くなってしまふ．そのため評価関数に比べて十分小さな計算量で，グローバルベストを向上させる可能性を求める必要がある．本研究では可能性推定関数 $f_p(\cdot)$ を用いて，粒子が有するグローバルベストを向上させる可能性を求められるとする．ここで本来なら粒子の位置 \mathbf{x}_i^k とグローバルベスト $\bar{\mathbf{p}}_g$ を用いて可能性推定関数 $f_p(\mathbf{x}_i^k, \bar{\mathbf{p}}_g)$ と定義する必要があるが，本研究では計算量を削減するためにグローバルベスト $\bar{\mathbf{p}}_g$ を使わず $f_p(\mathbf{x}_i^k)$ と定義する．なお，位置 \mathbf{x}_i^k の粒子 i がグローバルベストを向上させる可能性が高いほど，可能性推定関数値 $f_p(\mathbf{x}_i^k)$ は小さな値になるように設定されているものとする．

4.3 再評価割合

本提案手法のもう 1 つの重要な要素として，再評価されない粒子の決め方がある．これは再評価されない粒子はグローバルベストを向上させることがないために，最適化の結果に大きな影響を与えるからである．本研究ではグローバルベストを向上させる可能性の高い粒子のみ再評価対象とすることで，評価する粒子を減らす方法を採用し，再評価対象の粒子が全粒子に占める割合を表すパラメータとして再評価割合 $R_{rf} \in (0, 1)$ を導入する．

4.4 グローバルベストおよびローカルベストの再定義

提案手法ではグローバルベストを向上させる可能性を用いて，一部の粒子のみ再評価（評価関数値を計算する）を行っているため，評価関数値が計算されていない候補が存在する．そのため粒子 i の過去の最適化における最良の位置を示すローカルベスト \mathbf{p}_i および，群れ全体の過去の最適化における最良の位置を示すグローバルベスト \mathbf{p}_g を求めることができない．そのため提案手法におけるローカルベスト $\bar{\mathbf{p}}_i$ およびグローバルベスト $\bar{\mathbf{p}}_g$ をそれぞれ式 (25) および式 (26) と定義する．

$$f_p(\bar{\mathbf{p}}_i) = \min \{ \{ f_p(\mathbf{x}_i^k) \mid k = 1, \dots, k_n \} \} \quad (25)$$

$$f(\bar{\mathbf{p}}_g) = \min \{ \{ f(\mathbf{x}_i^k) \mid i = 1, \dots, N_p, k = 1, \dots, k_n \} \} \quad (26)$$

少なくともすべての候補の可能性推定関数 $f_p(\cdot)$ は計算されていることから，粒子 i の過去の最適化において，グローバルベストを向上させる可能性が最も高い位置をローカルベスト $\bar{\mathbf{p}}_i$ と定義する．グローバルベスト $\bar{\mathbf{p}}_g$ は，群れ全体の過去の最適化において，評価関数値が計算された位置（候補）の中で最も評価関数値が良い位置と定義する．なお評価関数値が計算されていない候補 \mathbf{x} の評価関数値 $f(\mathbf{x})$ を ∞ と定義する．また従来ではローカルベスト \mathbf{p}_i の中で最も良い評価値を持つ位置がグローバルベスト \mathbf{g}_i であったが，提案手法ではそのような関係はない．

```

(1) 各粒子の位置と速度をランダムに初期化
(2) Repeat 終了条件を満たすまで do
(3)   for 全ての粒子  $i$  に対して do
(4)      $f_p(\mathbf{x}_i^k)$  を計算する
(5)     if  $f_p(\mathbf{x}_i^k) < f_p(\mathbf{p}_i)$  then
(6)        $\mathbf{p}_i$  を向上させる
(7)     end if
(8)   end do
(9)   for  $f_p(\mathbf{x}_i^k)$  が上位  $R_{rf}$  の粒子  $i$  に対して do
(10)     $f(\mathbf{x}_i^k)$  を計算する
(11)    if  $f(\mathbf{x}_i^k) < f(\mathbf{p}_g)$  then
(12)       $\mathbf{p}_g$  を向上させる
(13)    end if
(14)  end do
(15)  for 全ての粒子  $i$  に対して do
(16)    式 (22) 及び (23) により，粒子  $i$  の解  $\mathbf{x}_i^k$  と速度  $\mathbf{v}_i^k$  を変更
(17)  end do
(18) end do

```

図 5 提案手法のアルゴリズム
Fig. 5 Proposed method's algorithm.

4.5 提案手法のアルゴリズム

図 5 に提案手法のアルゴリズムを示す．従来の PSO（図 3 を参照）に比べ粒子を評価する部分（図 5 の 3 行目から 14 行目）が異なっている．3 行目から 8 行目は，各粒子のグローバルベストを向上させる可能性を可能性推定関数 $f_p(\cdot)$ を用いて求め，ローカルベストを向上させている．9 行目から 14 行目はグローバルベストを向上させる可能性の高い粒子を対象に，評価関数値を計算しグローバルベストを向上させている．

4.6 実行時間

提案手法ではグローバルベストを向上させる可能性の低い粒子の再評価を行わないようにすることで，評価関数の使用回数を少なくし，実行時間の短縮を実現できると考えられる．しかしながら毎回すべての粒子に対してグローバルベストを向上させる可能性を計算するため，可能性推定関数 $f_p(\cdot)$ の計算時間によっては，逆に実行時間が長くなる可能性もある．また再評価割合 R_{rf} も実行時間に影響を与えられ，必ずしも提案手法の実行時間が短くなるとは限らない．そこで提案手法の実行時間が短くなる条件を明らかにする．なお提案手法と従来手法では粒子の評価方法のみが異なるため，実行時間を考える際には評価関数と可能性推定関数の計算時間のみに注目する．今，粒子数を N_p ，最大繰返し回数 k_{max} ，評価関数の計算時間を T_f ，可能性推定関数の計算時間を T_{fp} で表す．従来手法

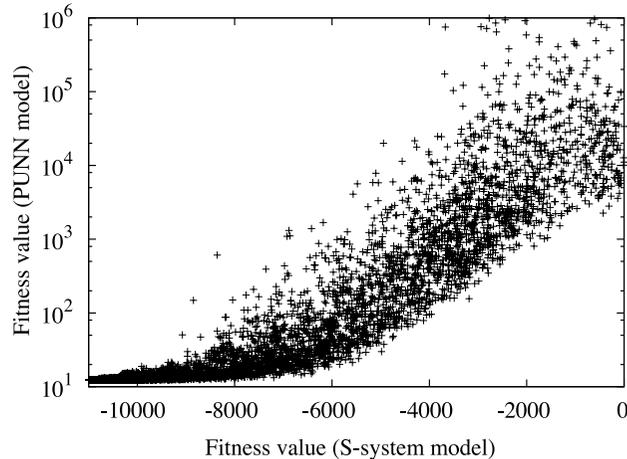


図 6 PUNN を用いた評価関数と S-system モデルを用いた評価関数の関係

Fig. 6 The relations between PUNN model's fitness function value and S-system model's fitness function value.

では、繰返しごとに評価関数を使って全粒子を評価するために $T_f \times N_p \times k_{max}$ の時間を必要とする。一方、提案手法では毎回粒子のグローバルベストを向上させる可能性を可能性推定関数で推定し、その後可能性の高い $R_{rf} \times N_p$ の粒子を評価関数で再評価している。そのため $(T_f \times R_{rf} \times N_p + T_{fp} \times N_p) \times k_{max}$ の時間を必要とする。これより従来手法より提案手法の実行時間が短くなるための条件は式 (27) になる。

$$(T_f \times R_{rf} \times N_p + T_{fp} \times N_p) \times k_{max} < T_f \times N_p \times k_{max} \quad (27)$$

ここで $T_{fp} = \alpha \times T_f$ とすると、式 (27) は式 (28) のようになる。

$$\alpha < 1 - R_{rf} \quad (28)$$

以上より提案手法が従来手法よりも実行時間が短くなる条件として、 $T_{fp} < T_f$ だけでなく、再評価割合 R_{rf} の値に応じて条件が変化することが確かめられた。

4.7 遺伝子ネットワーク推定問題への適用

提案手法を遺伝子ネットワーク推定問題に適用するためには、グローバルベストを向上させる可能性または、その可能性と相関がある値を求める可能性推定関数 $f_p(\cdot)$ を決める必要がある。そこで PUNN モデルを用いた評価関数と S-system モデルを用いた評価関数の関係を図で示すと図 6 になる。この図より PUNN モデルを用いた評価関数と S-system モ

デルを用いた評価関数ではおおむね大小関係が一致しており、関係性があることが分かる。またグローバルベストを向上させる可能性は、評価関数値が小さいほど高くなることから、PUNN モデルを用いてグローバルベストを向上させる可能性を推定することが可能だと考えられる。

実行時間に関しては、2.3 節で述べたように PUNN モデルを用いた評価関数は、S-system モデルを用いた評価関数に比べ実行時間は十分短く、またグローバルベストを向上させる可能性を大まかではあるが推定できることから、本研究では可能性推定関数 $f_p(\cdot)$ として PUNN モデルを用いた評価関数である式 (18) を使用することとする。

5. 評価実験

5.1 準備

評価実験で使用する遺伝子発現量は従来研究においてよく用いられている 5 つの遺伝子からなる遺伝子ネットワークを用いる。文献 10) と同様に実験に用いる遺伝子発現量の時系列データを各遺伝子の初期遺伝子発現量と S-system パラメータ (ともに文献 10) と同じ値であり、S-system パラメータは表 1 に示した値) を用いて、微分方程式 (1) を解くことによって得た。なお文献 3), 4), 10)–12) では与えられる遺伝子発現量の時系列データの量が十分でないと、モデルの自由度が高くなり遺伝子ネットワーク推定問題の解候補が増大するとの理由から初期遺伝子発現量の異なる 15 セットの時系列データを用いて評価実験が行われており、本研究でもそれにならい、初期遺伝子発現量の異なる 15 セットの時系列データを用いる。また時系列データは遺伝子ごとに 11 サンプルポイントになるように適当に間引き、これを 15 セット用意するので、各遺伝子上の時系列データは 165 サンプルポイントとなる。

実験条件は文献 10) と同様に以下のように設定した。S-system パラメータの上下制限約は、 α_i と β_i が $[0.0, 20.0]$ 、 $g_{i,j}$ 、 $h_{i,j}$ は $[-3.0, 3.0]$ とした。また計算コストを削減するため、 $g_{i,j}$ 、 $h_{i,j}$ が閾値 (1.0×10^{-3}) 以下の場合は 0 を返すようにした。粒子数は文献 4) と同様に 60 とし、実験環境は AMD Athlon (TM) 64 X2 Dual Core Processor 3800+ 2.01 GHz、1.00 GB のコンピュータを用いた。

比較対象として S-system モデルを用いた推定手法と PUNN モデルを用いた推定手法を実装した。なお両手法はそれぞれ S-system モデルを用いた評価関数、もしくは PUNN モデルを用いた評価関数のみを用いて最適化を行っていると思なすこともできるため、実験結果では、S-system モデルを用いた推定手法を $R_{rf} = 1$ 、PUNN モデルを用いた推定手法を

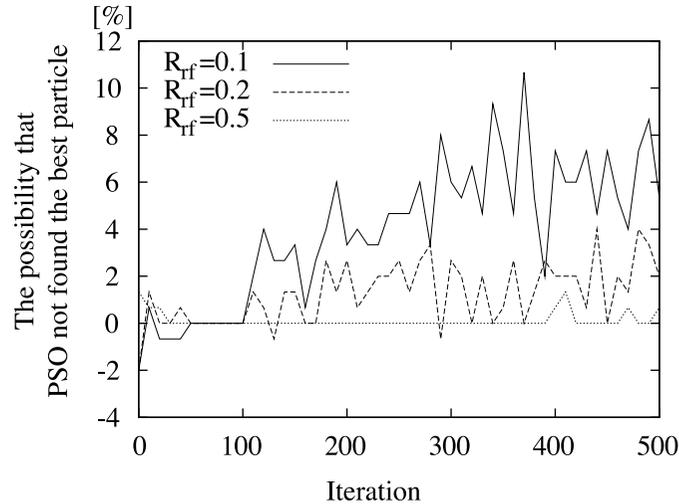


図 7 再評価対象を限定することで、最良の粒子が発見できなかった割合の推移
Fig.7 The progress of possibility that PSO not found the best particle.

$R_{rf} = 0$ と表記する .

5.2 実験 1

グローバルベストを更新する際に最も重要なことの 1 つは、群れ内の最良の粒子を発見することであり、そのために全粒子を再評価しているともいえる . 本研究ではグローバルベストを向上させる可能性を用いて再評価対象の粒子を限定することで、推定時間の短縮を目指している . しかしながら粒子を限定することで、最良の粒子を発見できなくなる可能性もある . そこでパラメータ誤差が最小の粒子を最良の粒子と定義し、移動回数を 500 回として 30 回試行を行い、粒子を限定することで最良の粒子が発見できなかった割合の推移を図 7 に示す . R_{rf} の値は 0.1 刻みに 0.1 から 0.9 の値を設定した . なお図中の値は平均値を示し、負の値は発見できなくなる可能性が減少した、つまり限定することで発見する可能性が向上したことを意味している . また $R_{rf} = 0.3, 0.4$ は $R_{rf} = 0.2$ に比べ若干小さな値になっているが、ほぼ同じ推移をしており、 R_{rf} が 0.5 以上はほぼ 0 であった . そのため図の視認性を考え R_{rf} の値は 0.1, 0.2, 0.5 のみを示す .

5.3 実験 2

移動回数を 500 回および 1,000 回に設定し、30 回の試行を行った結果をそれぞれ表 2 お

表 2 R_{rf} の変化による提案手法の性能変化 (移動回数 : 500)
Table 2 The results of proposed method in each R_{rf} (the number of iteration is 500).

R_{rf}	実行時間 (秒)	パラメータ誤差			
		平均値	標準偏差	最小	最大
0.0	30.73	2.4833	1.2875	0.4639	5.4777
0.1	296.76	0.7592	1.2335	0.0452	7.1666
0.2	560.23	0.3164	0.3049	0.0362	1.4201
0.3	821.01	0.3746	0.5377	0.0227	3.1755
0.4	1,081.36	0.2517	0.2245	0.0301	0.9486
0.5	1,335.79	0.2818	0.3528	0.0291	1.9544
0.6	1,593.98	0.2641	0.3224	0.0277	1.8645
0.7	1,850.65	0.2954	0.3665	0.0295	1.8582
0.8	2,101.74	0.1990	0.1595	0.0231	0.6994
0.9	2,349.94	0.2486	0.2418	0.0246	1.1278
1.0	2,753.63	2.6319	3.9998	0.0321	12.5604

表 3 R_{rf} の変化による提案手法の性能変化 (移動回数 : 1,000)
Table 3 The results of proposed method in each R_{rf} (the number of iteration is 1,000).

R_{rf}	実行時間 (秒)	パラメータ誤差			
		平均値	標準偏差	最小	最大
0.0	50.33	1.7294	0.8218	0.3703	3.6122
0.1	574.68	0.3805	0.4841	0.0333	2.5742
0.2	1,096.45	0.2062	0.1477	0.0293	0.8622
0.3	1,608.26	0.1685	0.1221	0.0268	0.5641
0.4	2,122.29	0.1826	0.2085	0.0284	1.3742
0.5	2,631.18	0.1574	0.1048	0.0231	0.4842
0.6	3,142.21	0.1832	0.1239	0.0194	0.4705
0.7	3,645.67	0.1604	0.1633	0.0275	1.0642
0.8	4,151.95	0.1491	0.0910	0.0315	0.3601
0.9	4,652.03	0.1694	0.1201	0.0301	0.5198
1.0	5,334.66	1.6059	3.2306	0.0290	14.6933

よび表 3 に示す . なお、表中の値は 30 回試行中の平均値、標準偏差、最小値、最大値を示している .

6. 考 察

6.1 実験 1

図 7 では視認性を確保するために省略しているが、 R_{rf} が 0.5 以上ではほぼ 0% になっており、再評価対象を限定することで最良の粒子が発見できなくなるという悪影響はほぼない

といえる。 R_{rf} が 0.4 以下では、最適化が進むに従って増加する傾向があることが分かるが、その割合は悪くても 10% であることが分かる。また少しではあるが、負数になるときもあり、再評価対象を限定することで発見できる割合が向上する場合もあることが分かる。

最適化が進むに従って、最良の粒子を発見できなくなる割合が増加する理由として、可能性推定関数として使用している PUNN モデルを用いた評価関数の評価精度の低さの影響が大きくなったと考えられる。PUNN モデルを用いた評価関数は図 2 に示したように、パラメータ誤差がある値以下の解候補を評価することはできない。そのため最適化に進むに従って、PUNN モデルを用いた評価関数では評価できないパラメータ誤差を持つ粒子が増加したために、最適化に与える影響が大きくなったためと考えられる。その結果、再評価対象の選択を誤ってしまい、最良の粒子を発見できなくなるという悪影響を及ぼし始めたと考えられる。なお R_{rf} が 0.5 以上では可能性推定関数の低い評価精度が影響を与えたととしても、それを補える分の粒子を再評価対象としているために、悪影響が現れなかったと考えられる。

再評価対象を限定することで最良の粒子が発見できなくなった割合が負数、つまり発見しやすくなるという現象については、最良の粒子だと誤って選択される粒子を可能性推定関数によって、再評価対象から外すことができたためだと考えられる。

6.2 実験 2

実行時間に関して R_{rf} の上昇にともない、長くなっていることが確認できる。これは R_{rf} の上昇にともない再評価する粒子が増え、S-system モデルを用いた評価関数の使用回数が多くなるためだと考えられ、予想どおりの結果となっている。

次にパラメータ誤差をみると、 $R_{rf} = 1$ のときにパラメータ誤差の平均値が悪くなっている。これは標準偏差が大きくなっているためであり、求められる解候補の良し悪しの差が大きいことを意味する。これはランダム性を有する多点探索においては仕方ないことだといえるが、 $R_{rf} \neq 1$ では標準偏差は小さくなっており、 $R_{rf} = 1$ が他に比べて、良し悪しの差が大きいことが分かる。また $R_{rf} = 0$ も $R_{rf} = 1$ ほどではないが、他の場合に比べると悪くなっていることが分かる。これよりグローバルベストを向上させる可能性を用いて再評価対象を決める提案手法は、それらの手法に比べ、より確実に解を求められることが分かる。PSO において群れ内の最良の粒子を選択することと、選択された最良の粒子の良さは、より良い解候補を発見する際に重要となる。実験 1 において、可能性推定関数を導入することで、最良の粒子を発見できなくなる可能性が負数になることは少ないことが確かめられている。つまり提案手法では効率的に最良の粒子を選択しているとはいえないことが分かっている。しかし実験 2 の結果では提案手法で得られる解のパラメータ誤差の平均値が小さ

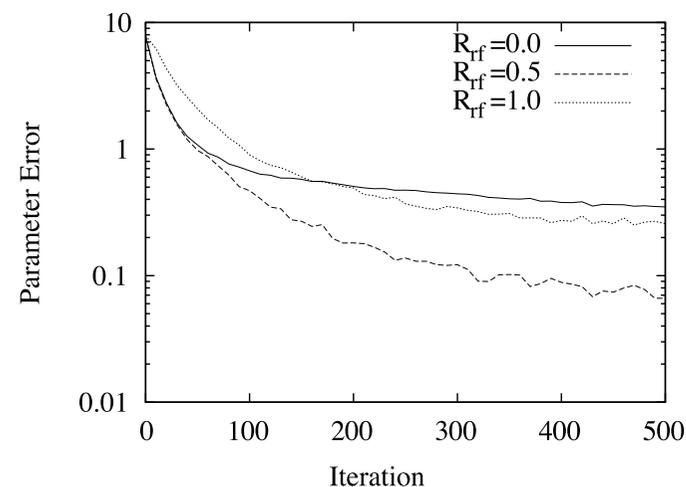


図 8 群れにおける最良の粒子のパラメータ誤差の推移

Fig. 8 The progress of best particle's parameter error in swarm.

くなっていることから、選択される最良の粒子の質が向上したと推測される。これを確認するために、群れ内の最良の粒子のパラメータ誤差の推移を図 8 に示す。なお R_{rf} が 0.1 ~ 0.9 はほぼ同じ推移になったため、図の視認性の確保のため $R_{rf} = 0, 0.5, 1$ の推移のみを示す。この図より $R_{rf} = 0, 1$ に比べて $R_{rf} = 0.5$ はパラメータ誤差が少なくなっており、提案手法は最良の粒子を選択することに関しては、それほど影響は与えないが、選択される粒子のパラメータ誤差が少なくなるために、従来手法に比べて確実に良い解を得ることができると考えられる。

群れ内の最良の粒子のパラメータ誤差が少なくなる要因としては、2 種類の評価基準を用いて最適化をしているためだと考えられる。つまり $R_{rf} = 0$ および $R_{rf} = 1$ では、グローバルベストを向上させる可能性もしくは、S-system モデルを用いた評価関数という 1 つの評価基準のみを用いて最適化をしている。そのため、もしその評価基準が間違った評価を下してしまうと、それを修正することができず、結果としてそれほど良い解を求められなくなる。これに対して $R_{rf} \in (0, 1)$ の場合、グローバルベストを向上させる可能性と S-system モデルを用いた評価関数という 2 つの評価基準を用いて探索を行うために、どちらか一方が間違ったとしても、他方で修正できる可能性が残っているために、1 つの評価基準しか使わない場合に比べて、確実に良い解を得ることができたと考えられる。

$R_{r,f} \geq 0.2$ において, $R_{r,f}$ の変化と平均パラメータ誤差の間には関係性があるとはいえない。これは $R_{r,f} = 0.2$ の時点で, グローバルベストをうまく向上させることができるためだと考えられ, $R_{r,f}$ を 0.2 以上に設定してもグローバルベストの向上にそれほど影響を与えていないと考えられる。しかしながら実行時間は $R_{r,f} = 0.2$ が最も短くなっていることから, 遺伝子ネットワーク推定問題において, PUNN モデルを用いた評価関数を用いてグローバルベストを向上させる可能性を推定した場合, 可能性の高い上位 20% の粒子だけを再評価対象とすることで, 効率良くネットワークの推定が行えると考えられる。

次に移動回数の違いについて考えてみると, 移動回数が多くなると, パラメータ誤差の平均値, 標準偏差は小さくなっている。しかしながらパラメータ誤差の最小値はそれほど変化はない。これから移動回数を多くすることで, より良い解が得られる確率は向上するが, 得られる解の精度はそれほど変化しないことが分かる。したがって, 移動回数を多くするより, 少ない移動回数で試行回数を増やす方が, 短時間でより良い解が得られるのではないかと考えられる。

7. おわりに

本研究では S-system モデルを用いた遺伝子ネットワーク推定問題において, グローバルベストを向上させる可能性を考慮した PSO の提案を行った。従来研究では, そのまま S-system モデルを用いる手法と遺伝子の発現変化量に注目した PUNN モデルを用いる手法が提案されている。S-system モデルを用いる手法では多くの推定時間を要するが比較的高い推定精度を得ることができ, PUNN モデルを用いる手法では推定時間は短い推定精度は低くなってしまおうという特徴を持っている。またそれらの手法の共通点として, 解候補である S-system パラメータをどのように評価するかという評価関数の方に重点を置いており, 最適化手法として既知の手法が使われていた。

そこで本研究では, 評価関数ではなく最適化手法に注目し, グローバルベストを向上させる可能性を考慮した PSO を提案した。PSO は粒子をグローバルベスト付近に集めることで最適化を行い, グローバルベストをより最適解の位置へと移動させ, 評価関数値を向上させていくことで, より良い解候補を見つけている。そのためグローバルベストの評価関数値を向上させることは重要であり, より効率的に向上させるために毎回全粒子の再評価を行っていると考えられる。しかしながらすべての粒子が, グローバルベストを向上させる可能性を等しく持っているとは考えにくく, 可能性の高い粒子や低い粒子が混在していることは想像に難くない。そこで本研究では各粒子が持つ “グローバルベストを向上させる可

能性” を推定し, その可能性に応じて再評価するか否かを決める, グローバルベストを向上させる可能性を考慮した PSO を提案した。グローバルベストを向上させる可能性の具体的な推定方法としては, PUNN モデルを用いた評価関数を用いた。

評価実験の結果, 従来手法 ($R_{r,f} = 0, 1$) に比べ実行時間, 推定精度ともに良い結果を得ることができたことを示した。またパラメータ誤差は $R_{r,f} \in [0.2, 0.9]$ でそれほど変化が見られないことも示し, 実行時間に関しては $R_{r,f}$ の設定によって調節できることも示した。

参 考 文 献

- 1) Jong, H.D.: Modeling and simulation of genetic regulatory system: A literature review, *Journal of Computational Biology*, Vol.9, No.1, pp.67–103 (2002).
- 2) Mason, J., Linsay, P.S., Collins, J.J. and Glass, L.: Evolving complex dynamics in electronic models of genetic networks, *Chaos*, Vol.14, No.3, pp.707–715 (2004).
- 3) Savageau, M.A.: Biochemical systems analysis: II. The steady state solution for an n-pool system using a power-law approximation, *J. Theor. Biol.*, Vol.25, No.3, pp.370–379 (1969).
- 4) 村田裕章, 越野 亮, 三田村公智, 木村春彦: Product unit based neural networks を用いた遺伝子ネットワークの S-system モデル推定手法の提案, 情報処理学会論文誌: 数理モデル化と応用, Vol.1, No.1, pp.74–87 (2008).
- 5) Durbin, R. and Rumelhart, D.: Product units: A computationally powerful and biologically plausible extension to backpropagation networks, *Neural Computation*, Vol.1, No.1, pp.133–142 (1989).
- 6) Kennedy, J. and Eberhart, R.C.: Particle swarm optimization, *Proc. IEEE International Conference on Neural Networks*, pp.1942–1948 (1995).
- 7) Kennedy, J., Eberhart, R. and Shi, Y.: *Swarm intelligence*, Morgan Kaufmann Publishers (2001).
- 8) Goldberg, D.: Genetic Algorithms in Search, Optimization, and Machine Learning (1989).
- 9) Angeline, P.J.: Evolutionary optimization versus particle swarm optimization: Philosophy and performance differences, *Lecture Notes In Computer Science*, Vol.1447, pp.601–610 (1998).
- 10) Kimura, S., Hatakeyama, M. and Konagaya, A.: Inference of S-system Models of Genetic Networks from Noisy Time-series Data, *Chem-Bio Info. J.*, Vol.4, No.1, pp.1–14 (2004).
- 11) Noman, N. and Iba, H.: Inference of Genetic Networks using S-system: Information Criteria for Model Selection, *Proc. GECCO-2006*, pp.263–270 (2006).
- 12) Kikuchi, S., Tominaga, D., Arita, M., Takahashi, K. and Tomita, M.: Dynamic

modeling of genetic networks using genetic algorithm and S-system, *Bioinf.*, Vol.19, No.5, pp.643-650 (2006).

13) Fletcher, R.: *Practical Methods of Optimization*, John Wiley (1987).

14) 福島雅夫: 数理計画入門, 朝倉書店 (1996).

15) Eberhart, R. and Shi, Y.: Comparing Inertia Weights and Constriction Factors in Particle Swarm Optimization, *Proc. Congress on Evolutionary Computation (CEC2000)*, pp.84-88 (2000).

(平成 22 年 4 月 28 日受付)

(平成 22 年 11 月 5 日採録)



村田 裕章 (学生会員)

平成 20 年金沢大学大学院博士前期課程修了。同年金沢大学大学院博士後期課程入学。平成 22 年 9 月金沢大学大学院博士後期課程修了。同年 10 月金沢大学大学院研究協力員, 同年 12 月金沢大学理工研究域博士研究員, 現在に至る。生命情報学に興味がある。



木村 春彦 (正会員)

昭和 54 年東北大学大学院工学研究科博士課程修了。工学博士。同年富士通(株)入社。昭和 59 年金沢大学経済学部助教授。現在, 同大学工学部情報システム工学科教授。最適コード変換やプロダクションシステムの高速化等の研究に従事。