

混合システムの視点に基づく遺伝的アルゴリズムのモデリング

今井 順一[†] 塩谷 浩之^{††} 栗原 正仁[†]

遺伝的アルゴリズム(GA)を理論的に解析する従来の研究は、対象となるGAを限定したうえで、その振舞いを精密に記述している。これに対して本論文では、GAを、現世代の個体群を入力、次世代の個体群を出力とした入出力データを発生する情報源と見なし、データからの学習を行う。GAの入出力関係のみに注目し、その内部構造をあえて捨象することで、多様なGAを共通の形式で記述するとともに、従来にない視点からの解析が可能となる。本論文では、この入出力関係の表現として混合モデルを採用する。GAを混合システムの視点からモデリングすることで、多くの要因が複雑に絡み合うGAシステム全体を複数の部分システムに分解した表現が得られる。本論文では2種類の混合モデルを取り上げ、これらを利用したGA解析の有効性を数値実験を通じて検証する。

Modeling of Genetic Algorithms Based on the Viewpoint of Mixture Systems

JUN-ICHI IMAI,[†] HIROYUKI SHIOYA^{††} and MASAHITO KURIHARA[†]

Some mathematical models have been proposed for theoretical analyses of genetic algorithms (GAs). However, these works have limited their objects to a few kinds of GAs in order to formulate them accurately. In this paper, we regard a GA as an information source that generates input-output data. That is, we regard a population and its next population generated by the GA as input and output respectively. Then we model the GA by learning from these data. Since this method uses only the input-output relations of data and ignores interior structures, we can describe a variety of GAs in a common form, and analyze them from a new point of view. We use some mixture models for a representation of these input-output relations in this paper. By using a mixture model for modeling a GA, we can represent the GA system as a combination of some partial systems. In this paper, we treat two types of mixture models, and investigate how these models are effective for analyzing GAs through some numerical experiments.

1. ま え が き

遺伝的アルゴリズム (genetic algorithm; GA)¹⁾ は生物進化に着想を得た確率的探索手法であり、様々な最適化問題に対して良好な結果が報告されてきた。それにともない、その基礎的な性質を探る研究が重要性を増している。

GAを理論的に解析するための手法として、微分方程式²⁾や差分方程式³⁾、Markov連鎖^{4)~6)}、離散力学系^{7)~9)}などを用いる手法が提案されている。また、Markov連鎖と離散力学系を関連させた解析も行われ

ている^{10),11)}。これらの研究はいずれも、対象となるGAを限定したうえで、その振舞いをそれぞれの形式で定式化する。最も汎用性が高いと考えられるVoseモデル¹¹⁾においても、遺伝子型分布による個体群の表現や、個体群サイズが有限の場合と無限大の場合との関係など、その枠組みに関しては一般のGAに広く適用できるものの、主要な構成要素である写像の定式化はsimple GA (SGA)に対してのみ行われている。

これに対して本論文では、GAを、現世代の個体群を入力、次世代の個体群を出力とした入出力データを発生する情報源と見なし、データからの学習を行う。GAの入出力関係のみに注目し、その内部構造をあえて捨象することで、多様なGAを共通の形式で記述するとともに、従来にない視点からの解析が可能となる。

本論文では、この入出力関係の表現として混合モデルを採用する。混合モデルは複数の部分システムを組み合わせることで全体を表現する手法である。GAを

[†] 北海道大学大学院工学研究科システム情報工学専攻
Division of Systems and Information Engineering,
Graduate School of Engineering, Hokkaido University

^{††} 室蘭工業大学工学部情報工学科
Department of Computer Science and Systems Engineering,
Faculty of Engineering, Muroran Institute of
Technology

混合システムの視点からモデリングすることで、多くの要因が複雑に絡み合う GA システム全体を複数の部分システムに分解した表現が得られる。本論文では、異なる 2 つのアプローチによる混合モデルを取り上げ、これらを利用した GA 解析の妥当性、有効性を数値実験を通じて検証する。

2. 準備

以降の議論の準備として、Vose が提案した GA モデルの枠組みを概説する。

2.1 Vose モデルの概要

P をサイズ n の個体群とし、各個体は長さ l のビット列で表現されるとする。このとき P は集合 Ω の要素を n 個含む多重集合として表現される。ここで $\Omega = \{0, 1\}^l$ は長さ l の遺伝子型の集合を表す。各遺伝子型は整数 $0, 1, \dots, 2^l - 1$ の 2 進表現と対応するので、以後、遺伝子型を整数で表す。

P は Ω 上の分布 $x = (x_0, x_1, \dots, x_{2^l-1})^T$ で特徴付けられる。ここで、 $x_k = n_k/n$ ($k = 0, 1, \dots, 2^l - 1$)、 n_k は遺伝子型 k の個体の数、 T は転置を表す。 n が既知のとき、 P は対応する分布 x から一意に定まる。 Ω 上の分布の集合を

$$\Lambda = \left\{ x \mid \sum_{k=0}^{2^l-1} x_k = 1, x_k \geq 0 (\forall k) \right\} \quad (1)$$

とする。また、サイズ n の個体群に対応する分布の集合を $\Lambda_n (\subset \Lambda)$ とする。

$n \rightarrow \infty$ の極限において、ある個体群の遺伝子型分布から次世代の個体群の遺伝子型分布への推移は大数の法則により決定論的となる。現世代の個体群の分布が x のとき、その次世代の個体群の分布を、写像 $\mathcal{G}: \Lambda \rightarrow \Lambda$ を用いて $\mathcal{G}(x)$ と書く。このとき GA は

$$x(t+1) = \mathcal{G}(x(t)) \quad (2)$$

という差分方程式の反復による離散力学系で記述される。ここで $x(t)$ は第 t 世代の個体群の分布を表す。

n が有限のとき、遺伝子型分布 x を持つ個体群から次世代の個体群を生成する GA の動作は、分布 $\mathcal{G}(x)$ に従って Ω から n 個のサンプルを抽出することと同等と見なせる。現世代の個体群の分布が x のとき、次世代の個体群の分布が y となる条件付き確率 $p(y|x)$ は次の多項分布で与えられる。

$$p(y|x) = n! \prod_{k=0}^{2^l-1} \frac{(\mathcal{G}(x)_k)^{ny_k}}{(ny_k)!} \quad (3)$$

現世代の分布が x のとき、次世代の分布の条件付き期待値ベクトルは $\mathcal{G}(x)$ である¹²⁾。有限の個体群サ

イズ n を持つ GA は、 Λ_n を状態集合、式 (3) を状態 $x (\in \Lambda_n)$ から状態 $y (\in \Lambda_n)$ への推移確率とする Markov 連鎖で記述される。

Vose は選択、交叉、突然変異からなる SGA について写像 \mathcal{G} を定式化した。 \mathcal{G} は選択オペレータに関する写像 $\mathcal{F}: \Lambda \rightarrow \Lambda$ と遺伝子組換えに関する写像 $\mathcal{M}: \Lambda \rightarrow \Lambda$ の合成として次のように記述される。

$$\mathcal{G} = \mathcal{M} \circ \mathcal{F} \quad (4)$$

たとえば選択オペレータとして適応度比例選択を用いた場合、 \mathcal{F} は次のような写像となる。

$$\mathcal{F}(x) = \frac{Fx}{\mathbf{1}^T Fx} \quad (5)$$

ここで、 $\mathbf{1}$ はすべての成分が 1 の 2^l 次元ベクトル、 F はその成分が

$$F_{i,j} = \begin{cases} \text{fitness}(i), & \text{if } i = j \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

と定義される 2^l 次対角行列、 $\text{fitness}(i)$ は遺伝子型 i の適応度を表す。また、 \mathcal{M} はその第 k 成分が

$$\mathcal{M}(x)_k = (\sigma_k x)^T M(\sigma_k x) \quad (7)$$

と記述される写像となる ($k = 0, 1, \dots, 2^l - 1$)。ここで M は交叉率 p_c と突然変異率 p_m から定まる 2^l 次対称行列であり、その成分 $M_{i,j}$ は遺伝子型 i と j の個体から遺伝子型 0 の個体が生成される確率を表す (行列 M の定義の詳細は付録に示す)。また、 σ_k はその成分が

$$(\sigma_k)_{i,j} = \begin{cases} 1, & \text{if } i \oplus j = k \\ 0, & \text{otherwise} \end{cases} \quad (8)$$

と定義される 2^l 次置換行列であり ($i, j = 0, 1, \dots, 2^l - 1$)、 $(\sigma_k x)^T M(\sigma_k x)$ は遺伝子型分布が x のとき遺伝子型 k の個体が生成される確率を表す。ここで \oplus はビットごとの排他的論理和を表す。

2.2 GA を特徴付ける写像

Vose モデルにおいて重要な役割を果たすのが写像 \mathcal{G} である。 \mathcal{G} は $n \rightarrow \infty$ の極限での決定論的な個体群推移の記述であり (式 (2))、サイズ n が有限な個体群の推移確率 (式 (3)) の中にも現れる。これを知らばあらゆるサイズの個体群の振舞いを記述できるという意味で、 \mathcal{G} は対象の GA を特徴付ける。すなわち、離散力学系、Markov 連鎖のいずれにおいても、異なる条件下での振舞いの相違はすべて \mathcal{G} の性質の相違に起因する。

逆にいえば、 \mathcal{G} の性質 (不動点の位置など) はすべて GA の振舞いの特徴を反映する。ここで振舞いの特徴とは、どのような傾向の個体群になりやすいのか、

表 1 GA の設定例

Table 1 An example of parameter settings.

最適化問題	0 → 4.1, 1 → 2.2, 2 → 1.3, 3 → 3.6 (genotype → fitness value)
使用オペレータ (パラメータ値)	適応度比例選択 一点交叉 (交叉率 $p_c = 0.6$) 突然変異 (突然変異率 $p_m = 0.001$)

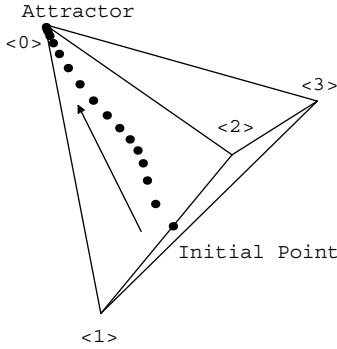


図 1 Λ における軌道の例
Fig. 1 An orbit in Λ.

どのような変化を経て最適解へ到達するのかなどであり、これらは解く最適化問題やオペレータ、パラメータの値などに複合的に依存する。これらの要素はすべて \mathcal{G} に集約されるため、これを通じて GA の振舞いに与える影響を探ることができる。したがって、各種パラメータ設定における \mathcal{G} の性質の差異を解析することにより、パラメータの設定指針が得られる可能性もある。

ここで注意すべきは、SGA 以外に対しても写像 \mathcal{G} は (陽に定式化されないが) 存在する点である。したがって、 \mathcal{G} の定式化以外の Vose モデルの枠組み (遺伝子型分布による個体群の表現や、 n が有限なら Markov 連鎖で、無限大の極限では離散力学系で表現されることなど) は一般の GA に対しても広く適用できる。

2.3 写像とベクトル場

2.1 節で述べたとおり、 $n \rightarrow \infty$ の極限における GA は写像 \mathcal{G} のなす離散力学系として記述される。簡単な場合として、表 1 のような設定の $l = 2$ の GA を表す力学系の軌道の例を図 1 に示す。状態空間 Λ は \mathbb{R}^4 内の 3 次元部分空間であり、図 1 では正四面体として表されている。 Λ の頂点は同じ遺伝子型の個体で占められた個体群の遺伝子型分布を表す。図 1 ではその遺伝子型 k を “ $\langle k \rangle$ ” の形で示した。

力学系の初期点は GA の初期個体群に、アトラクタは最適解 (局所最適解) が多数を占める個体群に相当する。この例では、遺伝子型 0 (= “00”) が最適解、

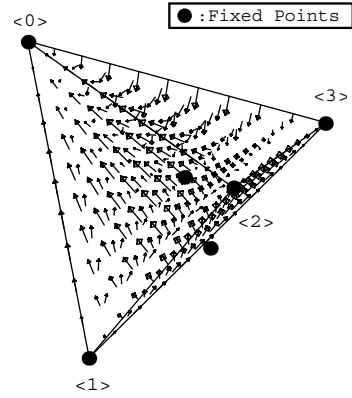


図 2 Λ 上のベクトル場
Fig. 2 A vector field on Λ.

3 (= “11”) が局所最適解である。図 1 の軌道は最適解が多数を占める個体群のアトラクタへ収束している。同じ例について、点 $x \in \Lambda$ にベクトル

$$\mathcal{V}(x) = \mathcal{G}(x) - x \tag{9}$$

を対応させたベクトル場を、 \mathcal{G} の不動点と合わせて図 2 に示す。この例では $\langle 0 \rangle$ と $\langle 3 \rangle$ の頂点付近にアトラクタが存在する。ベクトル場 \mathcal{V} は写像 \mathcal{G} と 1 対 1 に対応するため、本論文では以後、これらを同一視する。

3. 学習による GA のモデリング

写像 \mathcal{G} が GA を特徴付け、これを通じて GA を解析できることを 2.2 節で述べた。しかし、対象が SGA 以外であったり、適応度関数や使用オペレータなどが未知だったりする場合には \mathcal{G} を得ることができない。この問題を解決するために、本論文では「入出力データからの学習」によるモデリング手法を利用する。

3.1 モデリングの手順

対象となる GA (ビット長 l) に対し、可変なパラメータ $\theta = (\theta_1, \theta_2, \dots, \theta_u)^T \in \mathbb{R}^u$ を持ち、入力 $x \in \mathbb{R}^{2^l}$ に対して $f(x; \theta) \in \mathbb{R}^{2^l}$ を出力する入出力モデルを用意する。以下の手順に従って適切なパラメータ $\hat{\theta}$ を決定する。

- (i) 学習に用いるデータを対象の GA から作成する。任意に選んだ個体群 P に対し、対象の GA を実際に動作させてその 1 世代先の個体群 Q を生成する。この個体群の組 (P, Q) を遺伝子型分布の組 (x, y) に変換し、1 組の学習データとする。これを d 組作成し、まとめて $S = \{(x_1, y_1), (x_2, y_2), \dots, (x_d, y_d)\}$ とする。
- (ii) S を用いて学習を行い、 $\hat{\theta}$ を決定する。学習には一般的によく用いられる最小 2 乗法を

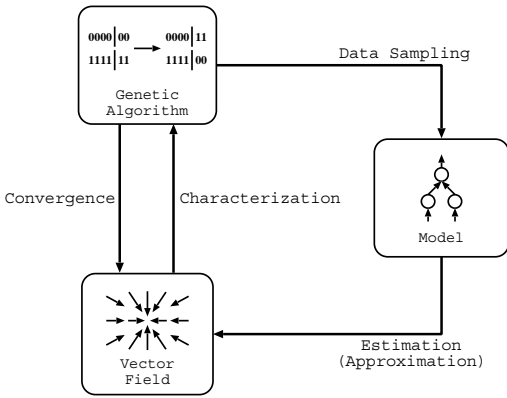


図 3 GA, モデル, ベクトル場の関係

Fig. 3 The relation among GAs, models and vector fields.

用いる。すなわち

$$\hat{\theta} = \arg \min_{\theta} E(\theta) \quad (10)$$

$$E(\theta) = \frac{1}{2d} \sum_{j=1}^d \|y_j - f(x_j; \theta)\|^2 \quad (11)$$

とする。ここで $\|\cdot\|$ はユークリッド ノルムを表す。パラメータ推定には最急降下法を用いる。

3.2 Vose モデルとの関係

現世代の遺伝子型分布 x に対する次世代の遺伝子型分布の期待値ベクトルは $G(x)$ であるから、モデルの表現能力とデータサイズ d が十分であれば、上述の手順に従って得た $f(x; \hat{\theta})$ は GA を特徴付ける写像 $G(x)$ の近似となる。図 3 に GA, モデル, G に対応するベクトル場の関係をまとめた。GA の個体群の確率的な推移は、 n の増加とともにベクトル場が記述する決定論的な推移へ収束する。一方、ベクトル場は $n \rightarrow \infty$ の極限における GA の記述であると同時に、有限なサイズの GA の振舞いを特徴付ける。モデルは有限なサイズの GA を観測することでデータを得、学習を通じてベクトル場を推定、近似する。

学習によるモデリングは、現世代の個体群を入力、次世代の個体群を出力としたときの入出力関係のみを利用する。したがって、対象の内部構造に依存せず、設定の詳細が未知の場合や SGA 以外に対しても G を求められる。このため、この手法は Vose モデルの枠組みにおいて対象についての写像 G を求める部分を補完する。

3.3 モデルの記述形式を利用した GA 解析

学習によるモデリング手法を用いるもう 1 つの目的は、 G の記述に共通の形式を与え、その視点から GA を比較、解析することである。GA の内部構造をあえて捨象するため、十分な表現能力さえあればモデ

ル $f(\cdot; \theta)$ の記述形式に制限はない。形式によっては GA の直接的な表現とならない反面、従来の枠にとられない解析が可能となる。

たとえば、階層型ニューラルネットワークや動径基底関数 (radial basis function; RBF) ネットワークによるモデリングは、 G をシグモイド関数展開、あるいは RBF 展開することに相当する。また、2 章で述べた Vose モデルの写像 $G (= M \circ F)$ を利用することもできる。この場合のモデリングは、対象の GA に「式 (11) の意味で」最も近い SGA のパラメータ (各遺伝子型の適応度や交叉率、突然変異率) を求めることに相当する。

4. GA の混合モデル

混合モデルは複数の部分システム (以後、エキスパートと呼ぶ) の組合せで全体のシステムを表現する手法であり、汎化性能の向上を目的として複数の学習機械を混合するアンサンブル学習¹³⁾ や、互いに異なる役割を持つエキスパートを入力に依存した比率で混合する mixture of experts¹⁴⁾ などが知られている。また、RBF ネットワークや正規化ガウス関数ネットワーク¹⁵⁾ など、基底となる局所ユニットをエキスパートとした混合モデルといえる。

本論文では、GA のモデリングに混合モデルを適用する。混合モデルは GA の直接的な表現ではないが、多くの要因が複雑に絡み合う GA を混合システムと見なし、複数の部分に分解して表現することが可能となる。

4.1 2 種類の混合モデル

本論文では次の 2 種類の混合モデルを取り上げる (図 4)。

G -混合モデル エキスパートとして、互いに異なる設定の SGA を特徴付ける Vose モデルの写像を m 個用意する。これを $G^i(x)$ ($i = 1, 2, \dots, m$) とし、次のように混合する。

$$f(x; t) = \sum_{i=1}^m t_i G^i(x) \quad (12)$$

ここで、 $t = (t_1, \dots, t_m)^T$, $\sum_{i=1}^m t_i = 1$, $t_i \geq 0$ ($i = 1, 2, \dots, m$) とする。この形式はアンサンブル学習のそれと同じであるが、同一対象のモデルである学習機械を複数用意するアンサンブル学習に対し、互いに異なる SGA のモデルであるエキスパートを用意する点が異なる。

F -混合モデル SGA に対する Vose モデルの選択オペレータに関する写像 F と遺伝子組換えに関す

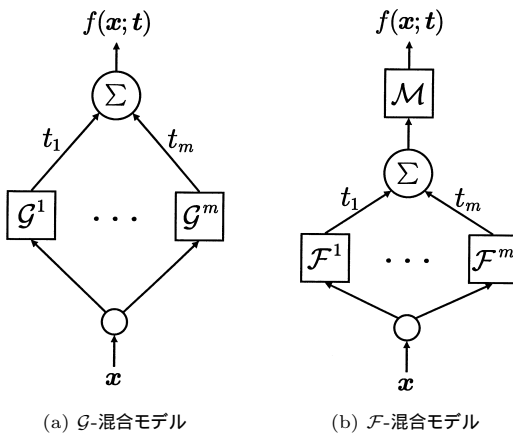
(a) G -混合モデル(b) F -混合モデル

図 4 2 種類の混合モデル

Fig. 4 Two types of mixture models.

る写像 \mathcal{M} を利用する．互いに異なる最適化問題に対応する m 個のエキスパート $\mathcal{F}^i(x)$ ($i = 1, 2, \dots, m$)，および遺伝子組換えに関する写像 $\mathcal{M}(x)$ を用意し，次のように混合する．

$$f(x; t) = \mathcal{M} \left(\sum_{i=1}^m t_i \mathcal{F}^i(x) \right) \quad (13)$$

ここで， $t = (t_1, \dots, t_m)^T$ ， $\sum_{i=1}^m t_i = 1$ ， $t_i \geq 0$ ($i = 1, 2, \dots, m$) とする． $\mathcal{F}^i(x)$ は線形混合されるが， $\mathcal{M}(x)$ が非線形変換であるから(式(7))，式(13)は混合率 t_i に関して非線形となる．

4.2 モデルの意味付け

G -混合モデルは，それぞれが GA を特徴付けるエキスパートを組み合わせ，これらとは異なる GA を表現する．これはベクトル場 \mathcal{V} をエキスパートのベクトル場 \mathcal{V}_i の重ね合わせで近似することに相当する．あるいは混合率 t_i をエキスパート i の選択確率と考え， \mathcal{V}_i の確率的な切換えとも見なせる．これらはいずれも，エキスパートが特徴付ける GA のシステムを基底と見なし，既存の構造を利用して対象の GA を表現することを意味する．

F -混合モデルでは，SGA に対する Vose モデル同様，選択オペレータに関する部分 (\mathcal{F}) と遺伝子組換えに関する部分 (\mathcal{M}) とを分けて考えた．これは「対象の GA が解く最適化問題」に依存する部分，しない部分ともいい換えられる．混合率 t_i が \mathcal{M} によって非線形的に変換されるため， F -混合モデルに「ベクトル場の重ね合わせ」という意味付けはできないが，選択オペレータに関する部分を混合する形式は「最適化問題の混合」という概念を自然な形で導入する．

このような意味付けを行うことで，各エキスパート

表 2 対象の GA が解く最適化問題 (実験 1)

Table 2 Optimization problems that the GA solves (Experiment 1).

	適応度関数 (genotype \mapsto fitness value)	だまし性
f_1	0 \mapsto 4.1, 1 \mapsto 2.2, 2 \mapsto 1.3, 3 \mapsto 3.6	Type II
f_2	0 \mapsto 3.6, 1 \mapsto 3.2, 2 \mapsto 1.3, 3 \mapsto 2.9	Type I
f_3	0 \mapsto 3.6, 1 \mapsto 3.2, 2 \mapsto 2.9, 3 \mapsto 1.3	—

表 3 実験設定 (実験 1)

Table 3 Parameter settings (Experiment 1).

個体群サイズ	$n = 10$
使用オペレータ (パラメータ値)	適応度比例選択 一点交叉 (交叉率 $p_c = 0.6$) 突然変異 (突然変異率 $p_m = 0.001$)
データサイズ	$d = 200$
学習係数	$\eta = 0.01$
学習回数	2,000 回

の役割を明確にし，対象の GA を特徴付ける写像 G の特徴を理解しやすい形で表現できる．また，写像どうしを比較する有効な手段になると考えられる．逆に，このように混合モデルを構成することで，混合率 t ，すなわち基底上の分布から GA の特徴を抽出できると期待される．

5. 数値実験

3 つの数値実験を通じて，混合モデルを利用した GA 解析の妥当性，有効性を検証する．

5.1 実験 1

5.1.1 設定

表 2 に示す $l = 2$ の最適化問題 $f_1 \sim f_3$ を解く GA をモデリングする．これらは Juliany ら⁸⁾ が用いた最適化問題をもとにしている． f_1 は Type II の， f_2 は Type I の最小だまし問題¹⁶⁾ である．個体群サイズやオペレータに関する設定は表 3 のとおりとする．いずれの最適化問題についても，無作為に生成した $d = 200$ 個の個体群の遺伝子型分布を入力とした．対象の GA に各入力に対する次世代の個体群を生成させ，その遺伝子型分布を出力とする．このデータに対して表 3 の条件で学習を行い，混合率を推定する．また，入力に対して真の G による像を出力とする 200 組のテストデータ S_i を用意した．

エキスパートとして，表 4 に示す最適化問題を解く 24 種類の GA に対応する写像 G^i ， \mathcal{F}^i を用意する．これらの最適化問題は，適応度によって遺伝子型に付けられるすべての順序を網羅するように選んだ．最適化問題を表す 4 桁の数字は遺伝子型を適応度の高い順に並べたものであり，以後，対応するエキスパートをこの数字で表す．また，最適化問題以外の設定は表 3

表 4 エキスパートに対応する最適化問題 (実験 1)

Table 4 Optimization problems that correspond to experts (Experiment 1).

	適応度関数 (<i>genotype</i> \mapsto <i>fitness value</i>)
f_{0123}	0 \mapsto 4.0, 1 \mapsto 3.0, 2 \mapsto 2.0, 3 \mapsto 1.0
f_{0132}	0 \mapsto 4.0, 1 \mapsto 3.0, 2 \mapsto 1.0, 3 \mapsto 2.0
f_{0213}	0 \mapsto 4.0, 1 \mapsto 2.0, 2 \mapsto 3.0, 3 \mapsto 1.0
\vdots	\vdots
f_{3201}	0 \mapsto 2.0, 1 \mapsto 1.0, 2 \mapsto 3.0, 3 \mapsto 4.0
f_{3210}	0 \mapsto 1.0, 1 \mapsto 2.0, 2 \mapsto 3.0, 3 \mapsto 4.0

表 5 実験 1-(i) の結果 (平均 2 乗誤差)

Table 5 Experimental results 1-(i) (mean squared error).

	G-混合モデル	F-混合モデル
f_1	0.70×10^{-3}	1.44×10^{-3}
f_2	0.32×10^{-3}	0.25×10^{-3}
f_3	0.70×10^{-3}	0.98×10^{-3}

表 6 実験 1-(i) の結果 (混合率の高いエキスパート)

Table 6 Experimental results 1-(i) (the first six experts in order of the mixture rate).

G-混合モデル	
遺伝子型の順序	混合率の高いエキスパート
f_1	0 > 3 > 1 > 2 0321 0312 3012 0132 3102 1032
f_2	0 > 1 > 3 > 2 0132 3012 1032 0312 3102 1302
f_3	0 > 1 > 2 > 3 0213 2013 2103 1203 0132 0123
F-混合モデル	
遺伝子型の順序	混合率の高いエキスパート
f_1	0 > 3 > 1 > 2 3012 0312 0321 0132 3102 1032
f_2	0 > 1 > 3 > 2 0132 3012 1032 0312 3102 1302
f_3	0 > 1 > 2 > 3 0213 2013 2103 0132 1203 1032

と共通とする。これらのエキスパートのうち、24 個すべてを用いた場合を実験 1-(i) とする。また、だまし性がなくベクトル場が比較的単純と考えられる 0123, 0213, 1032, 1302, 2031, 2301, 3120, 3210 の 8 個のエキスパートのみを用いた場合を実験 1-(ii) とする。

5.1.2 結果と考察

実験 1-(i) の結果を表 5, 表 6 に示す。表 5 には学習後のモデルの S_t に対する平均 2 乗誤差を示した。どちらのモデルも真の入出力関係をよく表現できている。表 6 には、高い混合率を得たエキスパートを左から値が高い順に 6 個示した。下線を引いたエキスパートは、対応する最適化問題の遺伝子型の順序が、対象の GA が解く最適化問題のそれと同じことを示す。表 6 から、いくつかの例外はあるが、次の傾向が読み取れる。

- 遺伝子型の順序が対象の最適化問題と一致するエキスパートが高い混合率を得ている。
- $f_1 \sim f_3$ において適応度の値に近い遺伝子型どうしの順序を網羅するようにエキスパートが選ば

表 7 実験 1-(ii) の結果 (平均 2 乗誤差)

Table 7 Experimental results 1-(ii) (mean squared error).

	G-混合モデル	F-混合モデル
f_1	8.26×10^{-3}	7.22×10^{-3}
f_2	2.04×10^{-3}	1.43×10^{-3}
f_3	0.77×10^{-3}	1.00×10^{-3}

ている。たとえば f_1 では遺伝子型 0 と 3, 1 と 2 がそれぞれ適応度の値に近いが、“03” または“30”, “12” または“21” の組合せのエキスパートが上位で選ばれている。 f_2 では、適応度が最も低い遺伝子型 2 の位置が最後尾であるようなエキスパートが網羅的に選ばれている。 f_3 でも f_2 と同様に遺伝子型 3 が最後尾となるエキスパートが上位で選ばれているが、例外 (0132, 1032) も生じている。

続いて表 7 に実験 1-(ii) の結果を示す。実験 1-(i) に比べて精度は劣るものの、真の入出力関係を十分よく表現できている。

実験 1-(i) では、GA が解く最適化問題の遺伝子型の順序という特徴に注目した。G-混合モデルは「複数の GA の混合」、F-混合モデルは「複数の最適化問題の混合」とアプローチ方法は異なるが、どちらのモデルも対象の GA、最適化問題の特徴をよく反映して、主要な役割を果たすエキスパート、微調整に必要なエキスパートが選ばれている。遺伝子型の順序は、適応度関数が未知な問題の最適解、準最適解を知るには十分な情報である。それを GA の振舞いの観察、および写像 g の推定を通じて抽出できたことは、 g をもとにした GA 解析の妥当性、および混合モデルによるモデリングの有効性を示している。

一方、遺伝子型の順序が一致、もしくは類似するエキスパートがなければ対象の GA を表現できないわけではない。実験 1-(ii) では、少数のエキスパートから十分な精度の混合モデルを構成し、必要な基底が揃えば対象の GA を表現できることを示した。これは少ないコストでの学習や GA の表現の簡素化に関連して重要である。

また、実験 1-(i) の f_3 に見られたエキスパート選択の混乱については、学習データ S のサイズ d が小さいことが原因と考えられる。しかし、データサイズの増加は学習に要する計算コストを増大させることにもなる。少ないコストでの高精度の解析に関しては、今後、機械学習における「汎化」の議論と関連させながら検討する必要がある。

表 8 実験 2 の結果 (混合率の高いエキスパート)

Table 8 Experimental results 2 (the first six experts in order of the mixture rate).

G-混合モデル	
	遺伝子型の順序 混合率の高いエキスパート
f_1	0 > 3 > 1 > 2 1230 2013 2103 1203 1032 3210
f_2	0 > 1 > 3 > 2 1320 1032 0132 0213 3210 1230
f_3	0 > 1 > 2 > 3 0213 2013 2301 0321 0132 1230
F-混合モデル	
	遺伝子型の順序 混合率の高いエキスパート
f_1	0 > 3 > 1 > 2 2013 1230 0213 3210 2301 1320
f_2	0 > 1 > 3 > 2 1320 1230 3210 0213 1032 2013
f_3	0 > 1 > 2 > 3 2013 0213 2301 2103 1230 0321

5.2 実験 2

5.2.1 設定

実験 1 と同様、表 2 に示す $l = 2$ の最適化問題 $f_1 \sim f_3$ を解く GA をモデリングする。個体群サイズやオペレータに関する設定も実験 1 と共通 (表 3) とするが、突然変異率のみ $p_m = 0.3$ に設定を変更する。

エキスパートも実験 1-(i) と同様、表 4 に示す最適化問題を解く 24 種類の GA に対応する写像 g^i, f^i を利用する。最適化問題以外の設定は表 3 のとおりとし、こちらの突然変異率は $p_m = 0.001$ のまま変更しない。

5.2.2 結果と考察

実験 2 の結果 (混合率の高いエキスパート) を表 8 に示す。実験 2 では、突然変異率が不適切に高く、正常な探索が困難な GA に対してモデリングを行った。解く最適化問題は共通であるが、これらの GA に対応する写像 g は実験 1 におけるそれとは性質が大きく異なると考えられる。実際、表 8 に示されたエキスパートは表 6 のものとは大きく異なっている。

突然変異率が極端に大きい場合、通常よりもはるかに多くの個体が別の遺伝子型に変化する。したがって、選択オペレータによって多く選ばれた適応度の高い遺伝子型の個体が突然変異によって破壊され、結果的に適応度の低い遺伝子型の割合が増加しやすい。表 8 に示されたエキスパートは、この「突然変異によって生じた遺伝子型の割合の増減」を、対象の GA とはかけ離れた最適化問題を解く GA の振舞いで近似的に表現するために選ばれたと考えられる。すなわち、あるパラメータの設定が GA の振舞いに与える影響を別のパラメータの影響に置き換えて表現したととらえることができる。これは GA を解析するための新たな視点を提供する。

5.3 実験 3

5.3.1 設定

表 9 に示す品物数 5 のナップザック問題を解く GA

表 9 対象の GA が解くナップザック問題 (実験 3)

Table 9 A knapsack problem that the GA solves (Experiment 3).

品物	1	2	3	4	5
価値	15	9	21	42	18
重量	12	14	6	26	13
制限重量	50				

表 10 実験設定 (実験 3)

Table 10 Parameter settings (Experiment 3).

個体群サイズ	$n = 100$
使用オペレータ (パラメータ値)	適応度比例選択 一点交叉 (交叉率 $p_c = 0.6$) 突然変異 (突然変異率 $p_m = 0.001$)
データサイズ	$d = 500$
学習係数	$\eta = 0.001$
学習回数	500 回

表 11 エクスパートに対応するナップザック問題 (実験 3)

Table 11 Knapsack problems that correspond to experts (Experiment 3).

k_1	
品物	1 2 3 4 5
価値	17 3 22 40 22
重量	12 14 6 26 13
制限重量	50
k_2	
品物	1 2 3 4 5
価値	9 27 18 48 6
重量	12 14 6 26 13
制限重量	50

($l = 5$) をモデリングする。ナップザック問題は制約条件付き組合せ最適化問題である。品物 j に 2 値変数 a_j を割り当て、その品物をナップザックに詰めるとき $a_j = 1$ 、そうでないとき $a_j = 0$ とする。遺伝子型 $k = (a_1 a_2 \dots a_l)_2$ の適応度は次のように与えられる。

$$fitness(k) = \begin{cases} \sum_{j=1}^l v_j a_j, & \text{if } \sum_{j=1}^l w_j a_j \leq W \\ 0, & \text{otherwise} \end{cases} \quad (14)$$

ここで、 v_j, w_j はそれぞれ品物 j の価値と重量、 W は制限重量を表す。また $(\cdot)_2$ は 2 進表現であることを表す。個体群サイズやオペレータ、学習に関する設定は表 10 のとおりとする。

エキスパートとして、表 11 に示すナップザック問題を解く 2 種類の GA に対応する写像 g^i, f^i を用意する。これらの問題は、各品物の重量、制限重量は表 9 の問題と等しく、各品物の価値が $v_j = 0.75 v_j^1 + 0.25 v_j^2$ となるよう設定した。ここで v_j^1, v_j^2 はそれぞれ k_1, k_2 における品物 j の価値を表す。最適化問題以外の

表 12 実験 3 の結果 (混合率)
Table 12 Experimental results 3 (mixture rate).

G-混合モデル	F-混合モデル
(0.763, 0.237)	(0.762, 0.238)

設定は表 10 と共通とする。

5.3.2 結果と考察

表 12 に実験 3 の結果 (推定された混合率 $t = (t_1, t_2)$) を示す。実験 3 では、あらかじめ適切なエキスパートを用意し、振舞いを観察して得た入出力データからの学習によって対象の GA の分解表現を獲得した。4.2 節で述べたように、F-混合モデルは「最適化問題の混合」を意識した形式である。したがって、学習で得られた F-混合モデルは「ナップザック問題の分解」を表現する。また、この実験設定のように最適化問題以外の要素がすべて等しいエキスパートを用いた G-混合モデルも、F-混合モデルと記述形式こそ違うものの、同様に「ナップザック問題の分解」を表すと考えられる。学習の結果、両モデルとも真の混合率 $t^* = (0.75, 0.25)$ に近い値を推定したことから、この「最適化問題の分解」という意味付けは妥当と考えてよい。

このように GA の混合モデルを介して得られる「最適化問題の分解」は、GA が解く対象としての最適化問題に新しい関係を導入する。この関係は、Vose モデルの写像 G や式 (3) からは導出されず、混合モデルを用いることで初めて導入される。

6. 議 論

6.1 エキスパートの個数

Vose モデルは最も微視的な GA モデルであるため、ビット長 l の値が大きくなるにつれて次元が指数関数的に増大し、その取扱いが非常に困難となる。本論文における 2 つの混合モデルについても、Vose モデルと同様の制約を受ける。

しかし、混合モデルの構造を考えると、あらかじめ適切な個数のエキスパートを用意できれば、学習によって決定する必要のあるパラメータはエキスパート数と同次元の混合率のみとなる。エキスパートのモジュール計算を固定オーダーとし、混合率の学習計算が主要なコストと考えると、学習スピードは混合モデルにおけるエキスパートの個数に依存する。したがって、どんなビット長においてもエキスパートの個数が学習計算のコストに関する主要な要因であり、これをいかに適切に選択するかが現実的な問題に適用する際のキーポイントになると考えられる。実験 1-(i) で主

要な役割を果たすエキスパートが選択的に高い混合率を与えられている点や実験 1-(ii) の結果から、モデリングに必ずしも多くのエキスパートが必要とされるわけではない。最適なエキスパート数の決定に関しては、AIC¹⁷⁾ や BIC^{18),19)}、MDL^{20),21)} などの情報論的モデル選択を利用した解決が有望と考えられる。

6.2 他手法との関連

混合モデルの学習法については、たとえば混合ガウス分布のモデル的仮定が GA からの遺伝子型分布の遷移データ (入出力データ) に対して妥当な適合がある場合、EM アルゴリズム²²⁾ などは有効な手段になると考えられる。

また、Vose モデルは遺伝子型分布空間 Λ にとどまらない広い空間における離散力学系までも扱っており⁸⁾、こうした拡張空間における解析も重要な課題である。このような拡張空間においては、たとえば Markov Chain Monte Carlo (MCMC) 手法^{23),24)} などの適用は難しい。一方、本論文における混合モデルの学習は一般のベクトル化された入出力学習機械を基盤とした学習方式であるため、こうした場合においても対応可能であるという特徴がある。

6.3 2 つの混合モデルの使い分け

本論文の実験では G-混合モデルと F-混合モデルとの間に顕著な差異が見られなかったが、実際に GA 解析に適用するにはこれら 2 つのモデルの特徴を把握しておくことが重要である。

F-混合モデルはその内部で SGA に対する写像 F 、 M を利用しており、比較的 SGA の影響が強いモデルといえる。したがって、対象が SGA、あるいはそれに近いとの情報が得られ、かつ最適化問題に関する解析を行いたい場合に最も有効に活用できると考えられる。一方、G-混合モデルはその構造が比較的単純であり、各エキスパートがそれぞれ 1 つの GA に対応するという高いモジュール性を有している。また、本論文ではエキスパートとして SGA に対する Vose モデルを利用しているが、モデルの構造上は必ずしも SGA である必要はなく、各エキスパート G^i を学習機械でも用意できる。したがって、F-混合モデルよりも汎用的な利用が可能である。

7. む す び

本論文では、混合システムの視点に基づいた GA のモデリング、および解析について述べた。

まず、対象の GA から入出力データを作成し、それに対する学習によって Vose モデルにおける写像 G を近似するモデリング手法について述べた。入出力関係

のみに注目し内部構造を捨象することで、従来にない視点からの GA 解析が可能となる。本論文では、2種類の混合モデルを GA のモデリングに適用した。それぞれのモデルに対して意味付けを行うとともに、これらを利用した GA 解析の有効性を示した。

今後の課題として、5.1.2 項で述べた汎化能力や 6.1 節で述べた最適なエキスパート数に関する考察のほか、GA の巨視的なモデリングへの適用があげられる。6.1 節でも述べたように、個体群の表現として遺伝子型分布を用いる場合、ビット長 l が大きくなるにつれて分布の次元が指数的に増大するという問題がある。これを避けるために、個体群の表現として遺伝子型分布よりも巨視的な分布を採用する方法が考えられる。たとえば、van Nimwegen らは個体群の表現に適応度分布を採用し、その推移を表現するモデルを構築した¹⁰⁾。この表現を学習によるモデリングに適用するには、3.1 節の手順 (i) において学習データを作成する際、個体群の組を遺伝子型分布ではなく適応度分布の組に変換すればよい。これ以外の手順は変更せずに学習を行うことで、得られたモデルは適応度分布の推移を表現する。この枠組みを利用し、巨視的表現と混合モデルとを組み合わせた GA 解析について検討する予定である。

参 考 文 献

- 1) Holland, J.H.: *Adaptation in Natural and Artificial Systems*, University of Michigan Press (1975).
- 2) Naitoh, K.: Macroscopic Kinetic Equation for a Genetic Algorithm, *Japan Journal of Industrial and Applied Mathematics*, Vol.15, No.1, pp.87-133 (1998).
- 3) 高橋祥兼: 遺伝的アルゴリズムの 2 ビット問題 (タイプ II) に対する収束性, *電子情報通信学会論文誌 A*, Vol.J78-A, No.1, pp.88-93 (1995).
- 4) Davis, T.E. and Principe, J.C.: A Markov Chain Framework for the Simple Genetic Algorithm, *Evolutionary Computation*, Vol.1, No.3, pp.269-288 (1993).
- 5) Rudolph, G.: Convergence Analysis of Canonical Genetic Algorithms, *IEEE Trans. Neural Networks*, Vol.5, No.1, pp.96-101 (1994).
- 6) Suzuki, J.: A Markov Chain Analysis on Simple Genetic Algorithms, *IEEE Trans. Systems, Man and Cybernetics*, Vol.25, No.4, pp.655-659 (1995).
- 7) Vose, M.D. and Liepins, G.E.: Punctuated Equilibria in Genetic Search, *Complex Systems*, Vol.5, No.1, pp.31-44 (1991).
- 8) Juliany, J. and Vose, M.D.: The Genetic Algorithm Fractal, *Evolutionary Computation*, Vol.2, No.2, pp.165-180 (1994).
- 9) Bertoni, A., Campadelli, P., Carpentieri, M. and Grossi, G.: Genetic Model: Analysis and Application to MAXSAT, *Evolutionary Computation*, Vol.8, No.3, pp.291-309 (2000).
- 10) van Nimwegen, E., Crutchfield, J.P. and Mitchell, M.: Finite Populations Induce Metastability in Evolutionary Search, *Physics Letters A*, Vol.229, No.3, pp.144-150 (1997).
- 11) Vose, M.D.: Modeling Simple Genetic Algorithms, *Evolutionary Computation*, Vol.3, No.4, pp.453-472 (1996).
- 12) Vose, M.D. and Wright, A.H.: Simple Genetic Algorithms with Linear Fitness, *Evolutionary Computation*, Vol.2, No.4, pp.347-368 (1995).
- 13) Hansen, L.K. and Salamon, P.: Neural Network Ensembles, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.12, No.10, pp.993-1001 (1990).
- 14) Jacobs, R.A., Jordan, M.I., Nowlan, S.J. and Hinton, G.E.: Adaptive Mixtures of Local Experts, *Neural Computation*, Vol.3, No.1, pp.79-87 (1991).
- 15) 石井 信, 佐藤雅昭: 正規化ガウス関数ネットワーク, Mixture of experts と EM アルゴリズム, *日本神経回路学会誌*, Vol.6, No.1, pp.30-40 (1999).
- 16) Goldberg, D.E.: *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley (1989).
- 17) Akaike, H.: A New Look at the Statistical Model Identification, *IEEE Trans. Automatic Control*, Vol.19, No.6, pp.716-723 (1974).
- 18) Schwarz, G.: Estimating the Dimension of a Model, *The Annals of Statistics*, Vol.6, No.2, pp.461-464 (1978).
- 19) Akaike, H.: Likelihood and the Bayes Procedure, *Bayesian Statistics*, Bernardo, J.M., De Groot, M.H., Lindley, D.U. and Smith, A.F.M.(Eds.), pp.141-166, University Press, Valencia (1980).
- 20) Rissanen, J.: Modeling by Shortest Data Description, *Automatica*, Vol.14, pp.465-471 (1978).
- 21) Rissanen, J.: Stochastic Complexity, *Journal of the Royal Statistical Society B*, Vol.49, No.3, pp.223-239 (1987).
- 22) Dempster, A.P., Laird, N.M. and Rubin, D.B.: Maximum Likelihood from Incomplete Data via the EM Algorithm, *Journal of the Royal Statistical Society B*, Vol.39, No.1, pp.1-22 (1977).

- 23) Hastings, W.K.: Monte Carlo Sampling Methods Using Markov Chains and their Applications, *Biometrika*, Vol.57, No.1, pp.97-109 (1970).
- 24) Gelfand, A.E. and Smith, A.F.M.: Sampling-Based Approaches to Calculating Marginal Densities, *Journal of the American Statistical Association*, Vol.85, No.410, pp.398-409 (1990).
- 25) 今井順一, 塩谷浩之, 伊達 惇: 学習機械を利用した遺伝的アルゴリズムのモデリングに関する検討, 電子情報通信学会技術研究報告, NC 2001-115, pp.63-70 (2002).

付録 行列 M の定義

演算子 \oplus, \otimes をそれぞれビットごとの排他的論理和, 論理積とし, $\bar{\cdot}$ を $\bar{i} = (2^l - 1) \oplus i$ で定義される演算 (ビットごとの否定) とする ($i = 0, 1, \dots, 2^l - 1$). 演算の優先順位は $\bar{\cdot}, \otimes, \oplus$ の順とする. また $[\cdot]$ を次のように定義する.

$$[expr] = \begin{cases} 1, & \text{if } expr \text{ is true} \\ 0, & \text{otherwise} \end{cases} \quad (15)$$

ここで $expr$ は真または偽の値をとる論理式を表す.

ベクトル $\mu = (\mu_0, \mu_1, \dots, \mu_{2^l-1})^T$, $\chi = (\chi_0, \chi_1, \dots, \chi_{2^l-1})^T$ を次のように定義する.

$$\mu_i = (p_m)^{n(i)} (1 - p_m)^{l-n(i)} \quad (16)$$

$$\chi_i = \begin{cases} 1 - p_c + p_c c_0, & \text{if } i = 0 \\ p_c c_i, & \text{otherwise} \end{cases} \quad (17)$$

ここで $n(i)$ は整数 i の 2 進表現に含まれる “1” の数を表す. また, 交叉オペレータとして一点交叉を用いる場合,

$$c_i = \begin{cases} 1/(l-1), & \text{if } \exists k \in (0, l) \ i = 2^k - 1 \\ 0, & \text{otherwise} \end{cases} \quad (18)$$

である ($i = 0, 1, \dots, 2^l - 1$).

2^l 次対称行列 M を次のように定義する ($i, j = 0, 1, \dots, 2^l - 1$).

$$M_{i,j} = \sum_{n=0}^{2^l-1} \sum_{m=0}^{2^l-1} \mu_n \frac{\chi_m + \chi_{\bar{m}}}{2} \times [i \otimes m \oplus \bar{m} \otimes j = n] \quad (19)$$

たとえば, $l = 2$, $p_c = 0.6$, $p_m = 0.001$ のとき, $\mu = (0.998001, 0.000999, 0.000999, 0.000001)^T$, $\chi = (0.4, 0.6, 0, 0)^T$ となり,

$$M = \begin{pmatrix} 0.998001 & 0.4995 & 0.4995 & 0.2002 \\ 0.4995 & 0.000999 & 0.2998 & 0.0005 \\ 0.4995 & 0.2998 & 0.000999 & 0.0005 \\ 0.2002 & 0.0005 & 0.0005 & 0.000001 \end{pmatrix} \quad (20)$$

となる.

(平成 14 年 8 月 23 日受付)

(平成 14 年 10 月 11 日再受付)

(平成 14 年 10 月 26 日採録)



今井 順一 (正会員)

1974 年生. 1997 年北海道大学工学部情報工学科卒業. 1999 年同大学院工学研究科システム情報工学専攻修士課程修了. 2003 年同博士後期課程修了. 博士 (工学). 進化的計算等の研究に従事. 電子情報通信学会会員.



塩谷 浩之 (正会員)

1964 年生. 1990 年北海道大学理学部数学科卒業. 1992 年同大学院工学研究科情報工学専攻修士課程修了. 1995 年同博士後期課程修了. 現在, 室蘭工業大学工学部情報工学科助教授. 博士 (工学). 数理情報工学の研究に従事. 電子情報通信学会, 日本神経回路学会各会員.



栗原 正仁 (正会員)

1955 年生. 1978 年北海道大学工学部電気工学科卒業. 1980 年同大学院工学研究科情報工学専攻修士課程修了. 現在, 同大学院工学研究科システム情報工学専攻教授. 工学博士. 人工知能, 項書換え系等の研究に従事. 電子情報通信学会, 日本ソフトウェア科学会, 人工知能学会, EATCS (欧州理論計算機科学会) 各会員.