解説

関数最適化のための分布推定アルゴリズム†

半田 久志*

1. はじめに

分布推定アルゴリズム (Estimation of Distribution Algorithms: EDA) は交叉演算子や突然変異演算子の代わりに確率モデルを用いた進化計算である。ここで用いる確率モデルは、各世代において、良好な個体群内の遺伝分布を推定したものである。この確率モデルを用いて、次世代の個体群をサンプリングすることにより探索を行う。EDAは、統計学で得られている知見を積極的に進化という枠組みに取り込むことにより、従来の進化計算と比べて探索性能が向上している。特に、設計変数の依存関係を明示的に探索するEDAでは、スキーマを積極的に探索しており、このことが探索性能の改善に大きく貢献していると考えられる。また、通常の進化計算では探索の結果として近似解を得ることができる。これに加えて、EDAでは、探索中に得られた確率モデルを解析することにより、対象問題に内在する設計変数の依存関係や問題の構造を知ることができる。

計算は図1に示す手順で行う。まず、通常の進化計算と同様に、初期個体群をランダムに生成し、個体評価を行う。つぎに個体群から良好な個体を選択し、選択された個体群から確率モデルを推定する。推定した確率モデルから遺伝情報をサンプリング(sampling)することにより子個体を生成し、これを個体群サイズまで繰り返すことにより次世代の個体群を形成する。そして、個体群内の各個体の適応度を評価する。

本稿では、連続変数を対象としたEDAを紹介する。当初から研究されている離散変数を対象としたEDAは、[1]に詳しく記述しているので参照されたい。連続変数を対象としたEDAでは、正規分布などのパラメトリックな確率モデルを表現することが多い。ガウシアン以外の表現として、4章に示しているように、ノンパラメトリックなアプローチとしてヒストグラムを使う方法もある。本稿では、設計変数間の

図1 EDAの処理手順

依存関係を考慮しない、基本的なアルゴリズムを2章で紹介し、共分散を用いることにより設計変数間の依存関係を考慮したEDAを3章で説明する。さらに、ノンパラメトリックなアプローチとしてカーネル密度推定法を用いたIDEAの他に、ヒストグラム用いたEDA(HEDA)を4章で紹介する。最後に、CMA-ES[2]や実数値GA(REX*)[3]で行われていて成果を挙げているMean Shiftや、多目的最適化においてクラスタリングを併用したEDAについて5章で紹介する。

2. 設計変数間の依存関係を考慮しないEDA

設計変数間の依存関係を考慮しないEDAでは平均 X 、分散 σ^2 の正規乱数を用いて子個体を生成する。設計変数間の依存関係を考慮しない、つまり、各設計数において、平均 X_i 、分散 σ_i^2 のパラメータを保持し、独立してサンプリングを行う。設計変数間の依存関係を考慮しないEDAとして、SHCLVND, PBIL_c, UMDA_cが存在する。これらの違いは、平均 X_i 、分散 σ_i^2 の更新方法にある。

Rudlof, S.らのSHCLVND(stochastic hill climbing with learning by vectors of normal distributions)では、サンプリングした個体の中から適応度の高い B 個体を選出し、その平均ベクトル X_B とする[4]。 X_B を用いて、正規乱数の平均 X を次式で更新する。

† Estimation of Distribution Algorithms for Function Optimization
Hisashi HANDA

* 岡山大学大学院自然科学研究科
Graduate School of Natural Science and Technology, Okayama University

図2 典型的な関数最適化のための分布推定アルゴリズムの振舞い
(灰色の丸：初期個体；黒丸：選択された個体群)

$$X \leftarrow X + \alpha(X_B - X)$$

偏差 σ^2 は、上式の平均 X の更新ごとに、一定割合 σ_{reduce} だけ減少させる。

$$\sigma^2 \leftarrow \sigma_{reduce} \cdot \sigma^2$$

SHCLVNDではすべての設計変数について同じ分散を用いている。

一方、Sebag, M. らのPBIL_c (population-based incremental learning for continuous search spaces) では、サンプリングした個体の中で、最も適応度が高い個体 $X^{Best,1}$ 、二番目に適応度が高い個体 $X^{Best,2}$ 、と最も適応度が低い個体 X^{Worst} を用いて正規乱数の平均 X を更新する [5, 6]。

$$X \leftarrow (1 - \alpha)X + \alpha(X^{Best,1} + X^{Best,2} - X^{Worst})$$

SHCLVNDと異なり、PBIL_cでは設計変数毎に偏差を調節している。具体的には、適応度の上位 K 個体の偏差 σ_i^K を用いて次のように偏差を更新する。

$$\sigma_i \leftarrow (1 - \alpha)\sigma_i + \alpha\sigma_i^K$$

SHCLVNDやPBIL_cでは前世代までの平均値や分散の値を考慮し、Hebb則と同様の方法で新たな値に更新していたが、UMDA_cでは平均値や分散の値をサン

プリングされた個体の上位個体の平均・分散に置き換えてしまう [5]。すなわち、UMDA_cは他の2つのアルゴリズムと比べて、GAでいうところの世代交代モデルが異なる。また、UMDA_cでは正規乱数以外にも、t分布、ベータ分布、ガンマ分布などの他の密度関数を用意しておき、上位個体のデータを用いて検定を行うことによりモデル選択を行い、選択された密度関数を用いて次世代のサンプリングを行うことも提案されている。

本章で紹介したSHCLVND、PBIL_c、UMDA_cの典型的な振舞いを図2に示す。以降、適時、同図を参照されたい。

3. 設計変数間の依存関係を考慮したEDA

EGNA (estimation of Gaussian network algorithms) では、図3に示すようなガウシアンネットワークを用いる [7]。EGNAでは各設計変数 X_i に対して、平均 $m_i + \sum_{X_j \in \Pi_i} b_{ji}(x_j - m_j)$ 、分散 v_i の正規乱数を用いて値の更新を行う。

$$\begin{aligned} f(x) &= \prod_{i=1}^N N(m_i + \sum_{X_j \in \Pi_i} b_{ji}(x_j - m_j), v_i) \\ &= \prod_{i=1}^N \frac{1}{\sqrt{2\pi v_i}} e^{\frac{1}{2v_i}(x_i - m_i - \sum_{X_j \in \Pi_i} b_{ji}(x_j - m_j))^2} \end{aligned}$$

図3 ガウシアンネットワークの例

Π_i は設計変数 X_i が依存する設計変数の集合とする。図3の場合、 $\Pi_1 = \Pi_2 = \emptyset$ 、 $\Pi_3 = \{X_1, X_2\}$ 、 $\Pi_4 = \{X_3\}$ となる。 m_i は設計変数ごとの平均、 b_{ji} は設計変数 X_i と X_j との関係の強さを表すパラメータを表す。また v_i は、依存する設計変数の集合 Π_i が与えられたときの条件付分散 (conditional variance) を表す。

これらのパラメータは、離散的な問題のためのEDAの時と同様に、サンプリングされた個体群を用いて選択された上位 N 個体の遺伝情報からネットワークの構造 S を推定する。この推定は、ネットワーク構造に対してスコアを与え、局所探索を行うことにより推定される。

そして上位 N 個体の遺伝情報と推定された Π_i からパラメータ m_i 、 b_{ji} 、 v_i を決定する。これらは、下式の対数尤度関数を最大化するものが選ばれる。

$$\ln L(D|S, \theta) = \sum_{r=1}^N \sum_{i=1}^n [-\ln(\sqrt{2\pi v_i}) - \frac{1}{2v_i} (x_{ir} - m_i - \sum_{x_{jr} \in \Pi_i} b_{ji}(x_{jr} - m_j))^2]$$

文献[7]ではこの最大化の結果として以下のパラメータが選ばれることが示されている。

$$\begin{aligned} \hat{m}_i &= \bar{X}_i \\ \hat{b}_{ji} &= \frac{S_{X_j X_i}}{S_{X_j}^2} \\ \hat{v}_i &= S_{X_i}^2 - \sum_{X_j \in \Pi_i} \frac{S_{X_j X_i}}{S_{X_j}^2} \\ &\quad + \sum_{X_j \in \Pi_i} \sum_{\substack{X_k \in \Pi_i, \\ X_k \neq X_j}} \frac{S_{X_j X_k} S_{X_j X_i} S_{X_k X_i}}{S_{X_j}^2 S_{X_k}^2} \end{aligned}$$

ここで $S_{X_j}^2$ と $S_{X_j X_i}$ はそれぞれ分散と共分散を表しており、次式で計算される。

$$\begin{aligned} S_{X_j}^2 &= \frac{1}{N} \sum_{r=1}^N (x_{jr} - \bar{X}_j)^2 \\ S_{X_j X_i} &= \frac{1}{N} \sum_{r=1}^N (x_{jr} - \bar{X}_j)(x_{ir} - \bar{X}_i) \end{aligned}$$

Ahn, C.W. らによる rBOA (real-coded Bayesian optimization algorithms) でも EGNA と同様にガウシアンネットワークが用いられている[8]。計算手順も類似している。すなわち、個体をサンプリングしてから選択し、ネットワークの構造推定を行い、パラメータ推定後、サンプリングを行い、次世代の計算に移る。彼らのアプローチでは、各ノードにおいて正規分布の混合分布を用いている。パラメータを決定する前に個体を c 個のクラスタに分割する。 $f_i(x)$ を i 番目のクラスタに対して式(1)と同じ方法で得られた確率モデルとする。

$$f(x) = \prod_{j=1}^N \sum_{i=1}^c \beta_i f_i(x)$$

$\beta_i (0 < \beta_i < 1, i=1, \dots, N)$ は混合分布を形成する確率モデル $f_i(x)$ の重みで、これは各クラスタに含まれる個体群サイズに比例して決められる。

4. ノンパラメトリックな確率モデルを用いたEDA

4.1 カーネル密度推定に基づいたEDA

Bosman, P.A.N. らはIDEA (iterated density estimation algorithms) という名前でいくつかのEDAを提案しており、前節で紹介したパラメトリックな確率モデルを用いたEDAだけでなくノンパラメトリックな確率モデルを用いたEDAも提案している[9,10]。ここでは、カーネル密度推定 (kernel density estimation) を用いたIDEAを紹介する。

カーネル関数を用いた確率モデルでは選択された個体のそれぞれに対してカーネル関数を割り当てることにより確率分布を推定する。下記のガウシアンカーネルを例にする。

$$g(\mathbf{X}, \mathbf{M}, \Sigma) = \frac{(2\pi)^{-\frac{d}{2}}}{(\det \Sigma)^{\frac{d}{2}}} e^{-\frac{1}{2}(\mathbf{X}-\mathbf{M})^T \Sigma^{-1}(\mathbf{X}-\mathbf{M})}$$

N 個の個体 $\mathbf{X}^i (i=1, \dots, N)$ があつたとき、次式の確率密度関数を用いて子個体の生成を行う。

$$f(\mathbf{X}) = \frac{1}{N} \sum_i g(\mathbf{X}, \mathbf{X}^i, \Theta)$$

ここで、 Θ は対角行列で、各対角成分が対応する設計変数における分散を表している。

このガウシアンカーネルのIDEAは (μ, λ) 進化戦略と類似した振舞いを示す[1]. 相違点は (μ, λ) 進化戦略では各個体が分散パラメータを内包しているのに対して, IDEAでは選択された個体の分散を計算している. 子個体生成についても少し異なり, (μ, λ) 進化戦略では各個体が決められた個体群サイズだけサンプリングを行うのに対して, IDEAでは選択された個体群で一つの確率モデルを形成し, この確率モデルに基づいてサンプリングを行う. 仮に (μ, λ) 進化戦略の分散をIDEAのように各世代ごとの親個体から計算すれば, これらはほとんど同じ振舞いをすることが予想される.

また, 3.2節のUNDX+MGGによる実数値GAも, ガウシアンカーネルを用いたEDAの一つと見なすことができる[11]. 本章で紹介したIDEAとは, 1) 世代交代モデルと2) 実数値GAではカーネルの分散や平均のパラメータがいくつかの親個体で推定されているという点異なる.

4.2 ヒストグラムを用いたEDA

周辺分布ヒストグラムを用いた進化アルゴリズム(evolutionary algorithm using marginal histogram models)がTsutsui, S.らによって提案されている. 文献[12]では, 図4に示されているような等幅ヒストグラム(fixed width histogram: FWH)や等高ヒストグラム(fixed height histogram: FHH)を用いたEDAが提案されている.

FWHでは各設計変数の値域が H 個のビンに分割されており, 各ビン h の推定確率を(ビンに含まれる個体数)/ N で表す. 個体のサンプリングはE-SUS(extended stochastic universal sampling)が提案されており, これは, 各ビンについて, 期待サンプリング個数の整数部分にあたる個数を固定的にサンプリングを行い, 残りの個体について, 小数部分を用いて確率的にサンプリングするものである.

FHHでは, 各ビンは等確率でサンプリングされる. しかしながら, ビンの値域がそれぞれ異なっており, 解が密集している領域ではビンの幅が狭くなっている. 解が密集している領域にたくさんの個体がサンプリングされることになる.

図4 等幅ヒストグラム(左); 等高ヒストグラム(中); ランキングを用いたヒストグラム(右)

YuanらはFWHの拡張としてPBIL_Hを提案している. 各世代の選択された個体群から推定したヒストグラムの各ビン i の度数を T_i とすると, ヒストグラムの各ビン i の度数 H_i を次式で更新する[13]:

$$H_i(t+1) = (1-\alpha)H_i(t) + \alpha T_i,$$

ここで, T_i は正規化されているものとする($\sum T_i=1$).

Dingらは, PBIL_Hでなされた拡張に加えて, ランキングを用いたヒストグラム推定や, 突然変異を導入したHEDA(Histogram EDA)を提案している[14]. ビンに属する個体が観測された時, 通常のヒストグラムではビンの度数を一つ増加させる. このランキングを用いたヒストグラム推定では, 増加幅 Δh_k^i を次式のランキングに基づいた式に従って増大させる:

$$\Delta h_k^i = \frac{N-k+1}{\sum_{l=1}^N l} = \frac{2(N-k+1)}{N(N+1)}$$

ここで k は, 対象とする個体のランクを表わしており, N は個体群サイズである. 彼らの論文では, これらの改良を施すことにより, 性能の改善が見られていると報告されている.

5. 実数値EDAの最近の進展

AMaLGaM(Adapted Maximum Likelihood Gaussian Model)ではMeanShiftを行い性能改善を図っている[15, 16]. これは, [17]でも指摘されているように, 標準的な実数値EDAの枠組み(切り取り選択+各世代ごとの最尤推定)では, 単純なスロープのような関数であっても途中で停留してしまうことがあることに拠る. AMaLGaMでは, 平均ベクトル, 共分散ベクトルの計算を以下のようにする:

$$\begin{aligned}\hat{\mu}(t) &= \frac{1}{N} \sum_i^N \mathbf{X}^i \\ \hat{\Sigma}(t) &= (1-\eta^\Sigma) \hat{\Sigma}(t-1) + \eta^\Sigma \frac{1}{N} \sum_i^N (\mathbf{X}^i - \hat{\mu}(t))(\mathbf{X}^i - \hat{\mu}(t))^T\end{aligned}$$

通常のガウシアンを用いた実数値EDAでは平均ベクトル $\hat{\mu}(t)$ をそのまま用いる. AMaLGaMでは, MeanShiftを行った $\hat{\mu}^{shift}(t)$ をガウシアンのパラメータとして用いる.

$$\begin{aligned}\hat{\mu}^{shift}(t) &= (1-\eta^{shift}) \hat{\mu}^{shift}(t-1) + \eta^{shift}(\hat{\mu}(t) - \hat{\mu}(t-1))\end{aligned}$$

また, 多目的最適化の分野において, クラスタリン

図5 目的関数空間(左)と設計変数空間(右)における
パレート最適解付近の個体群

グを用いることにより、集団を分割する試みがなされている。代表的なアルゴリズムとして、ZhangらのRM-MEDA(regularity model-based multiobjective estimation of distribution algorithms) [18]や先のAMaLGaMを多目的最適化に拡張したアルゴリズムがある [19]。RM-MEDAは特定の確率モデルを用いたアルゴリズムという訳ではない、図5に示すように、左図にある目的関数空間のパレート集合付近の個体群が、設計変数空間では右図のようにパレート集合付近に分布していると仮定し、パレート集合付近の個体をクラスタリングすることによって、いくつかの集合に分割してから確率モデルを推定するということを提案している。

6. むすび

本稿では、関数最適化のための分布推定アルゴリズムを解説した。筆者が分布推定アルゴリズムが良いと思うのは、確率モデルを用いることにより、進化計算のad hocなイメージを払しょくした点である。解がベクトルで表記される問題については、離散・連続の両方について一通りの提案がなされているように思う。しかしながら、木構造などの構造化されたデータや順列表現を用いないといけない問題については発展の余地がある。連続領域におけるMean Shiftのように積極的に探索方向にバイアスをかけつつ探索をするという意味でEDA版のMemetic Algorithmが興味深い。

参 考 文 献

- [1] 進化技術応用調査専門委員会編. 進化技術ハンドブック：第1巻基礎編，近代科学社，2010.
- [2] N.Hansen and A. Ostermeier. Completely derandomized self-adaptation in evolution strategies. *Evolutionary Computation*, 9 (2) : 159-195, 2001.
- [3] 小林重信. 実数値 GA のフロンティア. 人工知能学会論文誌, 24 (1) : 147-162, 2009.
- [4] S.Rudlof and M. Köppen. Stochastic hill climbing with learning by vectors of normal In *Proc. of 1st Online Workshop on Soft Computing*, pages 8-9, 1996.
- [5] P.Larrañaga and J. A. Lozano. *Estimation of Distribution Algorithms : A New Tool for Evolutionary Computation*. Kluwer Academic Publishers, Norwell, MA, USA, 2001.
- [6] M.Sebag and A.Ducoulombier. Extending population-based incremental learning to continuous search spaces. In *PPSN IV : Proc. of the 4th Inter. Conf. on Parallel Problem Solving from Nature*, pages 418-427, 1996.
- [7] E.Bengoetxea. *Inexact Graph Matching Using Estimation of Distribution Algorithms*. PhD thesis, The University of the Basque Country, 2002.
- [8] C.W. Ahn, R.S. Ramakrishna, and D.E. Goldberg. Real-coded bayesian optimization algorithm : Bringing the strength of boa into the continuous world. In *Proc. of 2004 Genetic and Evolutionary Computation Conf.*, volume 1, pages 840-851, 2004.
- [9] P.A.N. Bosman and D.Thierens. Expanding from discrete to continuous estimation of distribution algorithms : The idea. In *PPSN VI : Proc. of the 6th Inter. Conf. on Parallel Problem Solving from Nature*, pages 767-776, 2000.
- [10] P.A.N. Bosman and D.Thierens. Ideas based on the normal kernels probability density function. Technical report, Utrecht University technical report UU-CS-2000-11, 2000.
- [11] J.Sakuma and S.Kobayashi. Real-coded crossover as a role of kernel density estimation. In *Proceedings of the 2005 conference on Genetic and evolutionary computation*, GECCO '05, pages 703-710, 2005.
- [12] S.Tsutsui, M.Pelikan, and D.E. Goldberg. Evolutionary algorithm using marginal histogram models in continuous domain. In *Proc. of the Optimization by Building and Using Probabilistic Models OBUPM Workshop at the 2001 Genetic and Evolutionary Computation Conf.*, pages 230-233, 2001.
- [13] B.Yuan and M.Gallagher. Playing in continuous spaces : some analysis and extension of population-based incremental learning. In *Proc. of 2003 IEEE Congress on Evolutionary Computation*, pages 443-450, 2003.
- [14] N.Ding, S.Zhou, and Z.Sun. Optimization continuous problems using estimation of distribution algorithms based on histogram model. In *Proc. of the 6th Conference on Simulated Evolution and Learning*, pages 542-552, 2006.
- [15] P.A.N. Bosman, J.Grahl, and D.Thierens. Enhancing the performance of maximum-likelihood gaussian EDAs using anticipated mean shift. In *Proc. 10th International Conference on Parallel Problem Solving from Nature-PPSN X*, pages 133-143, 2008.
- [16] P.A.N. Bosman. On empirical memory design, faster selection of bayesian factorizations and parameter-free gaussian EDAs. In *Proc. Genetic and Evolutionary Computation Conference (GECCO 2009)*, pages 389-396, 2009.
- [17] 秋本洋平, 永田裕一, 小野功, 小林重信. 実数値進化計算の極値収束性の挙動解析. In 第3回進化計算フロンティア研究会, pages 73-80, 2010.
- [18] Q.Zhang, A.Zhou, and Y.Jin. RM-MEDA : A regularity

model-based multiobjective estimation of distribution algorithm. *IEEE Trans. Evolutionary Computation*, 12 (1) : 41–63, 2008.

- [19] P.A.N. Bosman. The anticipated mean shift and cluster registration in mixture-based edas for multi-objective optimization. In *Proceedings of the 12th annual conference on Genetic and evolutionary computation (GECCO 2010)*, pages 351–358, 2010.
(2011年1月4日 受付)

[問い合わせ先]

〒700-8530 岡山市北区津島中3-1-1
岡山大学大学院自然科学研究科

半田 久志

E-mail: handa@sdsc.it.okayama-u.ac.jp

著者紹介

はん だ ひさし
半田 久志 [非会員]

1998年京都大学大学院工学研究科博士後期課程退学, 同年より岡山大学工学部情報工学科助手, 2005年より同大学自然科学研究科助手. その間, 英国バーミンガム大学訪問研究員(2004–2005), 同大学 Honorary Research Fellow (2006–2011). 2009 GECCO Best Paper Award, 2009年度人工知能学会研究会優秀賞を受賞. IEEE CISの進化計算技術委員会委員, GECCO 1999–2011, CEC 2005–2010, PPSN IX–XI 等のプログラム委員等を務める. 進化計算・強化学習の研究に従事. 博士(情報学)