

確率的選択法を用いた分布推定アルゴリズム

Estimation of Distribution Algorithm with a Probabilistic Selection

武田 典久†
Michihisa Takeda

末松 伸朗†
Nobuo Suematsu

林 朗†
Akira Hayashi

1 はじめに

最適化手法の一つに、遺伝的アルゴリズム (Genetic Algorithm : GA) [1] がある。GA は生物の進化と自然淘汰を工学的に模倣した最適化アルゴリズムである。GA の交叉と呼ばれる操作は、適応度の高い解を生成するためのものだが、優良な部分解 (スキーマ) を破壊してしまう可能性があるという問題点がある。近年、この問題点を解決する手法として分布推定アルゴリズム (Estimation of Distribution Algorithm : EDA) [2] が提案されている。EDA では、GA の交叉の代わりに解集団から確率分布を推定し、その分布に従って新たな解を生成するためスキーマ破壊が起こらない。

しかし、EDA は解集団が急速に1種類の解に収束し、解の探索が不十分になり優れた解が発見できない場合がある。そこで本研究では、解集団に多様性を持たせつつ十分に探索を行うための手法を提案し、EDA の問題点の改善を目指す。

2 問題設定

最適化問題とは、制約条件の下で目的関数 $f(\mathbf{x})$ を最大 (最小) にするような解 \mathbf{x} を求める問題のことである。本研究では、組み合わせ最適化問題を扱う。GA や EDA では、一般に問題の解 (個体) を2進数のビット列 $\mathbf{x} = (x_1, x_2, \dots, x_n) \in \{0, 1\}^n$ でコーディングする。以降では、このような解のコーディングが行われているものとする。

3 EDA

EDA とは、個体集団に確率モデルを当てはめて分布を推定し、その分布に従って新たな個体集団を生成する確率的探索手法である。1回の探索を1世代と呼び、世代数を増やしながら最適解の発見を目指す。EDA の基本的手順を以下に示す。

1. 個体数が $2N$ の初期集団をランダムに生成する。
2. 集団から、適応度の高い個体 (優良な個体) を上位 N 個選択する。(選択されない個体は淘汰される)
3. 選択された集団に確率モデルを当てはめ、その分布を推定する。
4. 推定された分布から新たな N 個の個体を生成し、それと前世代の集団とで新たな $2N$ 個の個体からなる集団を生成する。
5. 終了条件が満たされなければステップ2に戻る。

手順のステップ2では、適応度の高い個体ばかりが選択されるため、個体集団が急速に1種類の個体に収束し、個体の探索が不十分になり優れた個体が発見できない場合がある。

4 シミュレーテッド・アニーリング

EDA の問題点を改善するため、EDA とシミュレーテッド・アニーリング (Simulated Annealing : SA) の比較からアイデアを得る。そこで、まず SA について述べる。

SA とは、高温で融解状態にある物質を徐々に冷却することによって、欠陥の少ない結晶を生成するアニーリングと呼ばれるプロセスを計算機上で模倣した手法である。SA の手順を以下に示す。

1. 初期温度の設定と解の初期状態の生成。
2. アニーリング。
3. クーリング。
4. 終了条件が満たされなければステップ2に戻る。

4.1 アニーリング

SA では、マルコフ連鎖モンテカルロ (MCMC) に基づき、一つの候補解の存在確率がギブス分布

$$P(\mathbf{x}|\beta) = \frac{\exp\{\beta f(\mathbf{x})\}}{\sum_{\mathbf{y} \in \{0,1\}^n} \exp\{\beta f(\mathbf{y})\}} \quad (1)$$

に従うようにする。ただし、 $\beta = 1/T$ 、 T は温度パラメータと呼ばれる。現状態から次の候補解 \mathbf{x}' が生成されると、その受理判定 (採択・棄却) が行われる。すなわち、確率

$$p = \min\left\{1, \frac{P(\mathbf{x}'|\beta)}{P(\mathbf{x}|\beta)}\right\} = \min\{1, e^{\beta(f(\mathbf{x}') - f(\mathbf{x}))}\} \quad (2)$$

で \mathbf{x}' を受理し、棄却された場合は \mathbf{x} にとどまる。これは、 $f(\mathbf{x}') < f(\mathbf{x})$ であっても、ある確率で \mathbf{x}' が受理されることを意味する。このため、目的関数が複数の局所解を持つ問題において、解が局所解に陥ったとしてもそこから抜け出すことが可能である。

4.2 クーリング

一定期間アニーリングを行ったあとにクーリングを行う。クーリングとは、アニーリングで用いられた温度 T を与えて、

$$T \leftarrow \gamma \cdot T \quad (3)$$

により温度を低下させ、これを次の温度とする。ただし、 $0 < \gamma < 1$ とし、温度が急冷にならないように γ を設定する。アニーリングにより、状態の存在確率が概ねギブス分布に従う状態を維持しつつ、クーリングにより温度を下げていくのが SA の戦略で

† 広島市立大学大学院情報科学研究科
〒731-3194 広島市安佐南区大塚東 3-4-1
Email: takeda@robotics.im.hiroshima-cu.ac.jp

ある。 $T \rightarrow 0$, つまり, $\beta \rightarrow \infty$ では, ギブス分布は目的関数が最大となるところへ集中した分布となるため, 理想的には大域的な最大化が達成される。

5 提案手法

5.1 SA から見た EDA

SA では, 終了条件が満たされるまでアニーリングとクーリングを繰り返し, 最適解発見を目指す。温度が高温状態の場合では改悪となる解の受理確率が高く, 局所解にある場合でもそこから抜け出すことが可能である。そうして探索領域を広く構えておき, 徐々に温度を下げるるとともに改悪となる解の受理確率が低下する。最終的に, ごく低温の状態では探索が山登り法と同じ振る舞いをする。

SA と EDA(GA) を対比させると, EDA では解集団を次々に低温のギブス分布に従わせていると解釈できる。よって, 適応度の高い解を優先して次世代に残す場合, **可能な限り急速に温度を下げて**いることになる。SA では温度を徐々に下げることが重要であるため, EDA の優良解の選択手続きを, **ゆっくりと温度を下げる**ような効果を持つものに変更すれば最適解を発見する能力が向上すると期待できる。

5.2 提案アルゴリズム

EDA の手順に, SA と同様の採択・棄却を用いた選択法を組み込むことを提案する。提案手法の探索手順を以下に示す。

1. 世代数 $t \leftarrow 0$, 温度 $T \leftarrow T_0$ とし, 初期集団 $S_t = \{x_1, x_2, \dots, x_N\}$, $S'_t = \{x'_1, x'_2, \dots, x'_N\}$ をランダムに生成する。
2. S_t と S'_t で採択・棄却を行う。すなわち, x_i と x'_i ($i=1, \dots, N$) について式 (2) の確率に従った受理判定を行い, 一方を選ぶ。選ばれた個体の集合を S_{t+1} とする。
3. S_{t+1} に確率モデルを当てはめ, その分布を推定する。
4. 推定された分布から新たな集団 S'_{t+1} を生成する。
5. 終了条件が満たされなければ, $t \leftarrow t+1$, $T \leftarrow T_t$ としてステップ 2 に戻る。

6 Optimal Dependency-Tree

現在, 分布の推定には様々な確率モデルを用いた EDA が提案されている [2]。本研究では, 確率モデルに Optimal Dependency-Tree[3] を使用する。

今, 仮に解が n ビットで表されているとする。このとき各ビットを一つの確率変数とみなし,

$$X_1, X_2, \dots, X_n \quad (4)$$

を考える。それぞれの確率変数に対応するノードを考え, 全ノード間の依存関係を非循環有向グラフで表し, 定量的な関係を条件付き確率で表したのがベイジアンネットワークである。これまで EDA で使われていた様々な確率モデルは, ベイジアンネットワークの構造に異なる制約を加えたものとして理解できる。今回使用するモデルは, ベイジアンネットワークが全域木 (spanning tree) [4] をなすという制約を課したものである。このとき, 計算量 $O(n^2)$ で尤度最大のモデルを学習できる。木ができあがると, それに基づいて次世代の解が生成される。

7 実験

FourPeaks 問題 [3] を用いて, EDA と提案手法を組み込んだ EDA との性能比較実験を行った。FourPeaks 問題とは

$$f(x) = \max\{\text{head}(1, x), \text{tail}(0, x)\} + R(L, x) \quad (5)$$

を最大化する問題である。ここで,

$$R(L, x) = \begin{cases} 100 & \text{if } \{\text{head}(1, x) > L\} \wedge \{\text{tail}(0, x) > L\}, \\ 0 & \text{otherwise} \end{cases}$$

ただし, $\text{head}(1, x)$ は個体 x の先頭から 1 が続いた数, $\text{tail}(0, x)$ は個体 x の末尾から 0 が続いた数を表している。パラメータ L は, $0 < L < n/2$ でなければならない。 $n=80$, $N=200$, 世代数の上限値を 5000 世代までとし, L の値を変えて各試行を 10 回ずつ行った。探索結果として最適解発見率を以下の表 1 に示す。

L	5	10	15	20	25	30	35
従来手法 [%]	100	90	60	30	20	0	0
提案手法 [%]	100	100	100	70	40	0	0

表 1 最適解発見率

表 1 から分かるように, 提案する確率的選択を入れることにより, 最適解の発見率は向上している。ただし, L の値が大きくなる (条件が厳しくなる) につれて改善が見られなくなった。これは, 世代数の上限値 5000 までに探索が終了しなかったため, また, 局所解に陥ってしまったためである。

温度をゆっくりと下げながら探索を十分に行う能力を組み込んだにも関わらず, このような結果となった。初期温度やクーリングスケジュールは性能を左右する重要なポイントであるが, それを適切に設定するのは容易ではないことが SA の研究において知られている。本実験結果は, それらの設定が不十分であったためかも知れない。SA のもう一つの欠点として, 最適解を得るのに非常に多くの計算量を必要とすることが挙げられる。問題の難易度が高くなれば, より広い探索が必要となり, 最適解の発見までにより多くのステップ数を要するのは自然なことである。しかし, このような問題があるものの, 従来手法より提案手法の方が最適解発見率が高いことが分かる。

8 まとめ

SA による採択・棄却に基づく選択法を EDA に組み込むことを提案し, 組み込んでいない EDA との性能比較実験を行うことで, その有効性を示すことができた。今回の実験結果よりも最適解発見率を上げるためには, 適切なパラメータ設定, 十分な計算時間の確保が必要になると思われる。

参考文献

- [1] 北野宏明ら, “遺伝的アルゴリズム”。産業図書, 1993
- [2] 倉橋節也ら, “ベイジアン最適化手法と分布推定アルゴリズムの動向”。人工知能学会誌 18 巻 5 号, pp. 487-494. 2003.
- [3] S.Baluja *et al.*, “Using Optimal Dependency-Trees for Combinational Optimization”. *Proceedings of the 14th International Conference on Machine Learning*, 1997
- [4] 西原 清一, “データ構造”, オーム社, 1993.