

Bayesian Network を用いた確率モデル遺伝的アルゴリズム 中村 康昭

1 はじめに

一般的な遺伝的アルゴリズムでは、選択、交叉、および突然変異といった遺伝的操作によって効率的な探索を図る。この中で特に、交叉、突然変異は新たな探索点 (個体) を生成するために用いられる操作である。交叉には親個体のもつ良好な部分解を交換するという利点があるが、同時に部分解を破壊する可能性もあることが知られている¹⁾。近年、このような交叉の問題点を解決するための新しいアプローチとして、交叉を用いることなく、母集団内の良好な個体の統計的な情報により新しい個体を生成する確率モデル GA に関する研究がさかに行われている²⁾。本稿では、設計変数間の依存関係を確率的にとらえる Bayesian Network と呼ばれるグラフ構造を用いて分布の推定を行う方法について述べる。

2 確率モデル遺伝的アルゴリズム

GA における中心的な探索オペレータである交叉には親個体の持つ良好な部分解を交換するという利点があるが、同時に部分解を破壊する可能性があることも知られている¹⁾。このような交叉の問題点を解決するための新しいアプローチとして、交叉を用いることなく、母集団内の良好な個体の統計的な情報を用いて新しい個体を生成する確率モデル GA に関する研究がさかに行われている²⁾。このような手法は一般に、分布推定アルゴリズム (Estimation of Distribution Algorithms: EDAs) と呼ばれる⁴⁾。

EDA では GA と同様にランダムに生成された個体群のなかから良好な解が選択される。選ばれた個体の確率分布が推定され、それによって新しい探索点が生成される。こうして生成された新しい探索点は母集団内の個体と置き換えられる。これを終了条件を満たすまで繰り返す。EDA の概念図を Fig. 1 に示す。

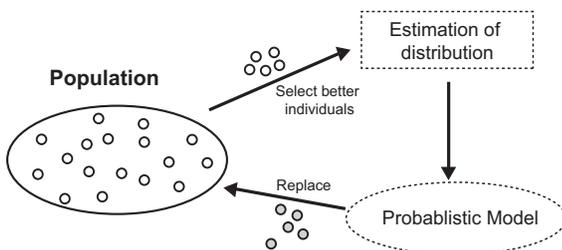


Fig. 1 Estimation of distribution algorithms

このような EDA のアルゴリズムは GA における交叉と突然変異を

1. 選択された個体群の分布の推定 (モデルの構築)
2. 構築したモデルに従った新しい探索点の決定

という 2 つの操作と置き換えたものであると考えられる。このため EDA は確率モデル GA (Probabilistic model-building GA) と呼ばれる。本稿でも EDA を確率モデル GA と称する。良好な個体群の分布を推定することは、対象問題のビルディングブロック構造をモデル化することにつながる。このため、確率モデル GA はランダム性の高い従来の交叉法と比較して、ビルディングブロックを効果的に組み換え、生成することができると期待される。

3 Bayesian Network

3.1 不確定性モデリング

完全に観測することのできない事象を扱う確率的な枠組みについて考える。GA において個体はまず、ランダムに発生されるため、ユーザの意図のように確定値を得ることは難しい。これらを体系的に扱うため、確率変数を用いることができる。複雑な要因やノイズの影響によって不確定さを含む対象を確率変数として大文字 X で表わし、その変数がとりえる具体値は小文字 x_1, x_2, \dots, x_n の様を示す。ここで変数 X は確率分布を考えることができ、その傾向はその統計量によって特徴づけることができる。

3.2 Bayesian Network とは

3.1 節で述べた確率的な枠組みによって、ある変数間の不確定のために漠然とした因果関係は定量的な条件付き依存性として表現することができる。また、因果関係が複雑な問題領域を表わすため、複数の変数間の因果関係を順次結んでいき、グラフ構造を持つネットワークとして表現することを考えたものが Bayesian Network であり、設計変数間の依存関係を表現する。Bayesian Network は、以下のような性質を有するグラフ構造として提示される。

- 変数の集合がネットワークのノードを形成する。
- リンクまたは矢印の集合がノード対を結ぶ。ノード X からノード Y への矢印の直感的な意味は X が Y

に直接的影響を与える (Y が X に依存する) という
ことである .

- グラフは矢印の方向にサイクルを持たない (有向非循環グラフである) .

Bayesian Network の例を Fig. 2 に示す .

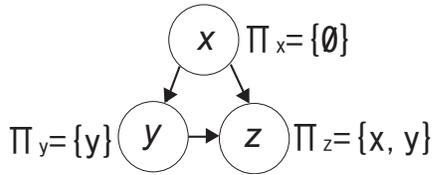


Fig. 2 Bayesian Network の例

Fig. 2 では , x, y, z という 3 つの変数がネットワークにおけるノードを構成しており , 矢印は依存関係を示す . つまり , z は x, y に依存し , y は x に依存することを示す . 影響を与える変数 (矢印の始点となるノード) を親ノードと呼ぶ . また , 影響を受ける (矢印の終点となるノード) を子ノードと呼ぶ .

3.3 Bayesian Network を用いた確率モデル GA

Bayesian Network を用いた確率モデル GA を Fig. 3 に示す .

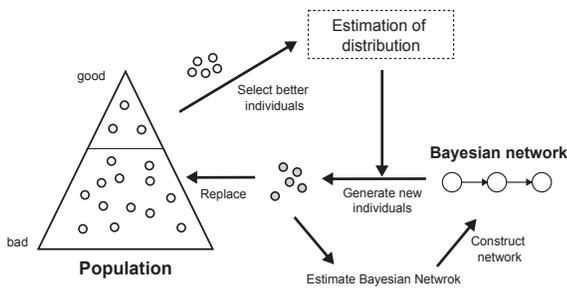


Fig. 3 Bayesian Network を用いた確率モデル GA

アルゴリズムの流れとしては ,

1. ランダムに個体を発生し , 評価を行う
2. 優良な個体を抽出し , Bayesian Network の構築を行う .
3. 構築されたネットワークに従い , 条件付き確率の推定を行う
4. 確率推論の実行 (子個体の生成) をし , 生成個体を母集団に戻す
5. 終了判定をし , ケイゾクする際は (2) へ戻る

以上を繰り返すことにより解探索を進める . 優良個体からネットワークを構築する手順について次節で述べる .

3.4 Bayesian Dirichelet metric

ネットワークを構築する際には , どの設計変数間に依存関係があるかを認識する必要がある . この際 , ネットワークの品質の指標としては , Bayesian Dirichelet (BD)metric が用いられる . これは , 対象問題に関する前の (prior) 知識と与えられたデータセットの統計的なデータを結合するものである . 与えられたデータセットを D , それによって得られたネットワークを B , 前の知識を ξ とすると , BD metric は

$$p(D, B|\xi) = p(B|\xi) \prod_{i=0}^{n-1} \prod_{\pi_{X_i}} \frac{m'(\pi_{X_i})!}{(m'(\pi_{X_i}) + m(\pi_{X_i}))!} \cdot \prod_{X_i} \frac{(m'(X_i, \pi_{X_i}) + m(X_i, \pi_{X_i}))!}{m'(X_i, \pi_{X_i})!} \quad (1)$$

によって示される .

$p(B|\xi)$ については , $p(B|\xi) = c\kappa^\sigma$ とする手法が提案されており , c は正規化定数 , κ は既知のものとは異なるネットワークに与えるペナルティを , σ は既知のネットワークと異なる矢印の数を示す .

式 (1) において , m と m' はネットワークが構築されたと仮定したときに , そのネットワークに適合するサンプルの数を示す .

様々なネットワークについて , $p(D, B|\xi)$ を計算し , 値が最大となるものが最良のネットワークとなる .

4 おわりに

Bayesian Network など , 学習メカニズムを用いることにより , 依存関係のある対象問題に対し , 効率的な探索を目指して . Bayesian Network を利用した GA についてはその性能が高いと報告されており ³⁾ , この実装を行うことで , 今後新たなメカニズムを搭載するときの参考 , および比較対象となる .

参考文献

- 1) Robert K. Lindsay Annie S. Wu and Rick L. Riolo. Empirical observation on the roles of crossover and mutation. *Proc. 7th International Conference on Genetic Algorithms*, pp. 362-369, 1997.
- 2) David E. Goldberg M. Pelikan and Fernando Lobo. A survey of optimization by building and using probabilistic models. Technical Report 99018, IlliGAL Report, 1999. University of illinois at Urbana-Champaign, Illinois Genetic Algorithms Laboratory.
- 3) David E. Goldberg Martin Pelikan and Erick Cantu-Paz. Boa: The bayesian optimization algorithm. Technical Report 99003, IlliGAL Report, 1999.
- 4) H. Muhlenbein and G. Paab. From recombination of genes to the estimation of distributions. *Parallel Problem Solving From Nature IV*, pp. 178-187, 1996.