

ベイジアンネットワークを用いた確率モデル GA の検討

The examination of probabilistic model-building GA using bayesian network

中村 康昭

Yasuaki NAKAMURA

Abstract: Recently, a number of evolutionaly algorithms that guide the exploration of the search space by building probabilistic models of promising solutions have been proposed. These algorithms have shown good performance. In this paper we examine the efficiency of probabilistic model-building genetic algorithms using bayesian network as a probabilistic model.

1 はじめに

最適化アルゴリズムの一つである遺伝的アルゴリズムは、選択、交叉、および突然変異といった遺伝的操作により効率的な探索を行う。特に交叉、突然変異は新たな探索点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の報告がされている¹⁾。近年、この問題点を解決する探索点の生成アプローチとして、母集団内の良好な個体の統計的な情報を用いて探索点を生成する確率モデル遺伝的アルゴリズムに関する研究がさかんに行われている²⁾。本稿では、設計変数間の依存関係を確率的にとらえるベイジアンネットワークと呼ばれるグラフ構造を用いて分布の推定を行う最適化手法における探索について検討を行う。対象問題としては、だまし問題および巡回セールスマン問題を考える。

2 ベイジアンネットワーク

2.1 ベイジアンネットワークとは

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ベイジアンネットワーク (Bayesian Network: BN) である。BN では設計変数間の依存関係を有向グラフにより表現する。BN 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. 1 に示す。

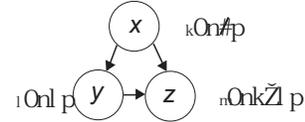


Fig. 1 ベイジアンネットワークの例

Fig. 1 において、ネットワークを構成する x, y, z という 3 つのノードがそれぞれ変数を表わしており、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) を親ノードと呼び、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。

2.2 ベイジアンネットワークの信頼度

BN の構築時には、設計変数間に存在する依存関係を認識する必要がある。つまり、あるネットワークが示す変数の依存関係をどの程度信じてよいかを知るため、その指標が必要となるのだ。本稿では、依存関係を示すネットワークの品質の指標としては、K2 metric を用いる。これは、対象問題に関して与えられたデータセットの統計的なデータをからネットワークの信頼度を導き出す。構築されたネットワークを B とすると、K2 metric は式 (1) により与えられる。

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

式 (1) において、 m と m' はネットワークが構築されたときと仮定したときに、そのネットワークに適合するサンプルの数を示す。ただし、K2 metric では、 $m'(X_i, \pi_{X_i}) = 1$ として扱う。式 (1) では、前半部分は親ノードとなる変数は様々な値を取るにより大きくなり、後半部分は親ノードが特定されたときには子ノードの変数が偏った時に大きくなる。様々なネットワークについて $p(B)$ を

計算し、最大値をもつものが最も信頼度の高いネットワークとなる。

3 数値実験

3.1 だまし問題への適用

BN を用いて、次数 3 のだまし問題³⁾を解く。この問題は全ビットを 3 ビットずつのサブ問題に分けたものであり、各サブ問題は式 (2) によって評価値が与えられる。全体の評価値は各サブ問題の評価値の総和となる。

$$f_{3deceptive}(u) = \begin{cases} 0.9 & \text{if } u=0 \\ 0.8 & \text{if } u=1 \\ 0 & \text{if } u=2 \\ 1.0 & \text{if } u=3 \end{cases} \quad (2)$$

総ビット長を 30 とする次数 3 のだまし問題に対して、BN を用いて最適化を行った。本実験では BN が認識する依存関係が有効性を確認するため、個体に対する突然変異操作は行わないものとする。また、パラメータとしては、優良個体の抽出率と世代交代の際の個体の置換率をそれぞれ 50% とし、個体数は 900 とする。

3.1.1 実験結果

Fig. 2 に、解探索の履歴と母集団内に存在する同一個体の割合を示す。結果は 30 試行の平均および中央値である。

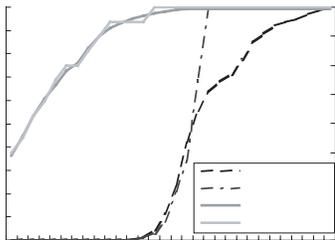


Fig. 2 評価値と母集団内にある同一個体の割合の履歴

Fig. 2 において横軸は総ビット数、縦軸は最良個体の関数評価値と母集団中の同一個体の割合を示している。結果より、エリート個体の適合度が最適解にほぼ到達した時から急激に母集団内の個体が収束していくことが分かる。

3.1.2 考察

次数 3 のだまし問題に対して BN を用いた確率モデル GA を適用したが、30 試行中 26 試行で最適解を得ることができた。残りの 4 試行については、1 つのサブ問題で局所解に陥っており、評価値が 9.9 の状態で母集団が収束していた。しかし、構築されたネットワークを調べたところ局所解に陥っているサブ問題についてもそのビット間の依存関係が認識されていることが確認され

た。従って、今回は用いていない突然変異操作を用いることによって 1 ビットが反転すると最適解を得るよう他のビットも反転するようになると考えられる。

3.2 TSP への適用

BN を用いた確率モデル GA の巡回セールスマン問題 (Traveling Salesman Problem: TSP) への適用を試みた。この際、遺伝子の交叉方法としては EX を使い、バイナリコードへの変換にはグレイコードを用いた。EX では、未訪問都市のうちどの都市を訪れるかということを選択情報として持つ。従って、遺伝子の最後尾の設計変数は必ず 0 となる。このような環境で実験を行ったが、探索が全く進まないという結果となった。

このような結果になった原因について調査すると、まず構築されたネットワークを見たときに、ある特徴が認められた。それは整数値をグレイコードに変換する際に生じる依存関係のみが認識されている点である。例えば最後尾から 3 つ目の設計変数では、その時点で未訪問都市が 3 都市となっているため値は 0~2 のいずれかとなる。この値をグレイコードに変換すると {0,0}, {0,1}, {1,1} となる。つまり、先頭ビットが 1 となると必ず最終ビットは 0 となるのである。このような依存関係は BN により認識されているが、設計変数間の依存関係を認識していないはなかった。したがって、そのビット自体に意味のあるように TSP という問題を変換するなどして対応する必要がある。これが、今後の課題となる。

4 おわりに

本稿では、ベイジアンネットワークを用いた確率モデル遺伝的アルゴリズムを用いてだまし問題および TSP を解くことによりその探索に対する検討を試みた。だまし問題を対象問題としたときには設計変数間の依存関係を認識しており、良好な探索が確認されたが、TSP を解く際にはその都市間の依存関係を見つけることはできなかった。TSP のような問題を解くため、バイナリコードではなく整数値の偏りを見るようにプログラムを変更する必要がある。

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.