

Bayesian Network に関する調査
中村 康昭

1 現在の課題

BOA で用いられている Bayesian Network に関する調査が課題となっている .

2 研究の進捗状況

2.1 Bayesian Network

Bayesian Network とは , 変数間の依存関係を表現するためのデータ構造である . Bayesian Network は , 以下のような性質を有するグラフである .

- 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 .

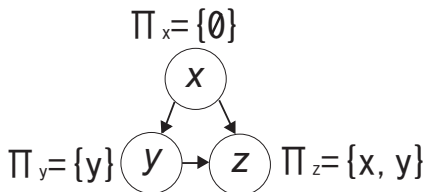


Fig. 1 Bayesian Network の例

Fig. 1 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 .

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 . m は新たに提示されたサンプルのうち , それぞれの状態を満たすものの数 (posterior) を示し , m' はあらかじめ持っている知識 (prior) から得られる数を示す .

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 .

ただし , BOA では BD metric の中で特殊なパターンである K2 metric を用いており , これは m' = 1 とすることで , 全てのネットワークを同様に見なし , 新たに得られたサンプルからネットワークを評価することを表す . すなわち , 予備知識を用いないのである .

ベイジアンネットワークの構造学習アルゴリズムとしては , K2 アルゴリズムが適用可能である . このアルゴリズムの流れは , 各ノードについて親ノードとなり得る候補を限定し , その候補からネットワークを作成 , 評価し , 評価が高くなったときにそのネットワークを採用する . この作業を全ての子ノードについて実行するというアルゴリズムである .

3 翌月への課題

Bayesian Network 構築の際の問題点としてネットワーク構築における計算付加が挙げられる . BOA(Bayesian Optimization Algorithm) では , 欲張り法 (greedy algorithm) が用いられている . どのように実装されているかをソースから分析する .

参考文献

- 1) Dan Geiger, David M. Chickering.
Learning Bayesian Networks: The Combination of Knowledge and Statistical Data (Microsoft Technical Report MSR-TR-94-09,1994)
- 2) Martin Pelikan, David E. Goldberg, and Erick Cantu-Paz.
BOA: The Bayesian Optimization Algorithm (IlliGAL Report No.99003, 1999)