自由エネルギー最小化によるベイジアンネットワークのロバストな構造推定

Robust Structure Inference of Bayesian Networks by Minimizing Free Energies

磯崎 隆司　植野 真臣
Takashi Isozaki　Maomi Ueno

電気通信大学 大学院情報システム学研究科
Graduate School of Information Systems, the University of Electro-Communications

We propose a new independence testing method for constraint-based structure inference in Bayesian networks. It is based on minimum free energy principle with "data temperature" assumption that we recently proposed. This method shows the effectiveness for small and medium data size.

1. はじめに

不確実性を伴う知識表現形態であるベイジアンネットワークを観測データから推定する構造推定には大きく分けて二つのアプローチがあり, ひとつはスコア & サーチ・アプローチと呼ばれるもので, 残るは制約に基づく (以下制約ベースと呼ぶ) アプローチと呼ばれるものである.

制約ベース・アプローチは計算効率が高い点で魅力のある方法である. このアプローチにおいては条件付き独立性検定を行なうことが正統的な手法であるが, ベイジアンネットワークは通常多くの変数を持つため, 検定のためのデータが不足する状況が発生することが多い. この場合には通常, 条件付き依存性を仮定せざるを得ないため, エッジの有無についての推定エラーが発生する可能性が高くなる. またこの情報はエッジの有向化にも利用されるためその推定エラーの可能性も高くなる.

我々は少数データに起因するこのような問題に対して自由エネルギー最小原理を利用してその改善を図る. すでに我々はベイジアンネットワークのパラメータ推定において熱物理学とのメタファーから "データ温度" を提案し少数データでの推定における有効性を確認しており [Isozaki 08], 構造の推定に対してもこの手法を適用する. そして漸近領域では通常の独立性検定手法に一致することが望まれるが, 本手法はこの要請を満たすことを示す.

また, 本手法の現実的な有効性を確認するため, 代表的な制約ベースのアルゴリズムである PC アルゴリズム [Spirtes 00] に適用しシミュレーション実験を行ない, その有効性を確認する.

2. ベイジアンネットワークとその構造推定

ベイジアンネットワークは結合確率分布のコンパクトな表現形態であり, 確率変数をノードとする非循環有向グラフ (DAG) である. ノード X に入射する有向エッジに関する親ノード群を Π とすれば, n 変数のベイジアンネットワークの結合確率分布は以下のように表わされる:

$$P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i|\Pi_i). \tag{1}$$

我々は条件付き独立性を $Ind(X;Y|\boldsymbol{Z})$ と表わす. ここで X と Y は変数で $\boldsymbol{Z}$ は条件となる変数の組を表わす. そしてこの条件付き独立性は DAG 構造における d-分離 [Spirtes 00] と等価であるという忠実性 [Spirtes 00] を仮定する. 本研究ではその他に離散確率分布と欠損データがないことを仮定する.

本研究で用いる制約ベース・アプローチの基本的なアルゴリズムは以下の通りである: (i) 変数 X と Y の組に対して $Ind(X;Y|\boldsymbol{Z})$ となる組 $\boldsymbol{Z}$ を探索する. 見つかれば X と Y 間のエッジを除去し見つからなければ無向エッジを張る. (ii) 無向エッジのない組 X と Y に対して両ノードとの間に無向エッジのある W が $Ind(X;Y|\boldsymbol{Z})$ なる $\boldsymbol{Z}$ に対して $W \in \boldsymbol{Z}$ であるかを調べ, そうでなければ $X \rightarrow W \leftarrow Y$ という有向エッジを張る. これは v-structure と呼ばれる. (iii) 有向化規則 [Verma 92] から可能な限り有向化された部分的な DAG を得る.

3. 自由エネルギー最小化による構造推定

我々は制約ベース・アプローチによるベイジアンネットワーク構造推定に対して有効な検定方法を提案する. この手法は熱物理学に起源をもつ自由エネルギーの最小化原理と我々が既に提案しているデータ温度モデルに基づく.

自由エネルギー F は内部エネルギーを U, エントロピーを H, 逆温度を β_0 とすれば

$$F := U - H/\beta_0 \tag{2}$$

と表わされる. 我々は過学習を避けるために逆温度 β_0 がデータ数に対して単調増加関数であるという仮説を提案した. これに従えば内部エネルギーをデータとモデルとの近さと捉えることができる.

パラメータ推定同様, 構造推定においてもデータ数無限大の極限で最尤推定に近づきデータ数ゼロの極限で最大エントロピー原理が支配的となるように内部エネルギーを定義する. 一方で制約ベース・アプローチにおいて通常採用されるように条件付き独立性を帰無仮説とし条件付き依存性を対立仮説とし, それぞれのモデルを 1, 2 として内部エネルギーは Kullback-Leibler (KL) 情報量を用いて

$$\begin{aligned} U_1 &:= -D(\hat{P}_1(X,Y|\boldsymbol{Z})||P_0(X,Y|\boldsymbol{Z})) \\ U_2 &:= -D(\hat{P}_2(X,Y|\boldsymbol{Z})||P_0(X,Y|\boldsymbol{Z})) \end{aligned} \tag{3}$$

と定義する. ここで $\hat{P}$ はそれぞれの構造で最尤推定される分布を表し P_0 は共通の真の分布とする. エントロピーはそれぞれの構造で推定される Shannon エントロピーを用いる. すると自由エネルギーの差をとれば, 自由エネルギー最小 (MFE) 原

連絡先: 磯崎 隆司, E-mail: t-isozaki@ai.is.uec.ac.jp

理より条件付き独立性の条件は相互情報量を用いて次のように表せる.

$$g_\beta^2 := \hat{I}(X,Y|\boldsymbol{Z}) - \frac{1-\beta}{\beta} I(X,Y|\boldsymbol{Z}) < 0. \qquad (4)$$

ここで g_β^2 は後のために定義している. また温度パラメータを新たに $\beta := \beta_0/(1+\beta_0)$ と定義した. ここで式 (4) の不等式における左辺第 1 項の $\hat{I}$ は経験相互情報量であり, 第 2 項はデータ温度によって推定される確率分布によって計算される相互情報量である. それはパラメータ推定で提案された β のデータ数 N に陽に依存するモデル:

$$\beta = 1 - \exp(-N/\gamma N_c) \qquad (5)$$

と MFE によるパラメータ推定の式:

$$P(\boldsymbol{x}) = \frac{\exp[\beta(\log \hat{P}(\boldsymbol{x}))]}{\sum_{\boldsymbol{x}} \exp[\beta(\log \hat{P}(\boldsymbol{x}))]} = \frac{[\hat{P}(\boldsymbol{x})]^\beta}{\sum_{\boldsymbol{x}} [\hat{P}(\boldsymbol{x})]^\beta} \qquad (6)$$

を用いて計算される [Isozaki 08]. ここで γ は自由度に関係する量であり N_c は β に対するハイパーパラメータである.

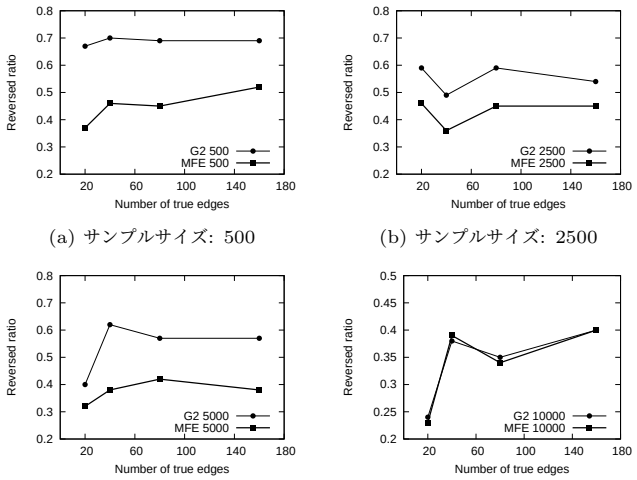
条件付き独立性に基づくベイジアンネットワークの構造推定では所謂「オッカムの剃刀」の原理にも従っている. すなわち取り除いても構わない有向エッジは取り除きより簡単な構造を採用する考えである. この原理も考慮すれば, データ数が多くなるにつれ我々の提案手法は通常の仮説検定に近づくことが望ましい.

我々の手法はこの考えに適合し, 漸近領域では G^2 検定 [Spirtes 00] に近づくことを示す. 変数 X と Y の $\boldsymbol{Z}$ を条件とした場合の独立性に関する G^2 統計量は $G^2 = 2N\hat{I}(X;Y|\boldsymbol{Z})$ と表わすことができ, これは漸近的に χ^2 分布に近づく [Kullback 68]. そこで擬似的な統計量として $G_\beta^2 = 2Ng_\beta^2$ を考えれば, 我々のデータ温度モデルを用いると漸近的に $G_\beta^2 \to G^2$ となる. 従って提案手法は漸近的に G^2 検定と等価になる. Spirtes らはこの G^2 検定が χ^2 分布近似で実施できる条件として検定にかかわる総セル数の 10 倍のデータが必要であるとし至らない場合は依存性を仮定している [Spirtes 00]. 我々の方法はこの制限を取り除き任意のデータ数で検定を実施できる. すなわち有意水準を α, 自由度を df として α と df による χ^2 分布の閾値 $\chi^2_{\alpha,df}$ を用いて $G_\beta^2 < \chi^2_{\alpha,df}$ なる条件付き独立性が棄却されない条件を見出す. これは式 (4) の条件も含んでいる.

4. 実験

本手法を制約ベース・アプローチの代表的アルゴリズムである PC アルゴリズムに組み込み, 通常の PC アルゴリズムと比較してその効果を調べた. 実験は変数の数を 20, 40, 80, 160 と変化させエッジの数をその 2 倍として 1 つの構造とそれに対する 5 つのパラメータセットをランダムに生成する. そのベイジアンネットワークからランダムサンプリングして構造を復元させるシミュレーションを実施した. 内部状態は全て 4 とし, 有意水準は 5%, ハイパーパラメータ N_c は 2 とした.

実験の結果, 我々の方法は図 1 に示すように特にエッジの向きの検出において従来手法に対して優位性を示した. これは条件付き独立性が成り立つような条件セット $\boldsymbol{Z}$ を相対的に正しく検出できたことを示している. しかも構造推定では通常小さいとは言えないサンプル数 5000 においても有意な差が得られている. エッジの向きを正しく検出することは特に因果的な構造の知識獲得に寄与すると考えられる. 今回の実験ではエッジ数に対しては偽陰性・偽陽性ともそれほど大きな差が得られなかったが, 条件セットを正しく認識していることから, この理由は本実験では忠実性の実現が難しかったためだと考えられる. 従って忠実性の高いデータ構造であればエッジ数もより正しく検出できるはずである.



(a) サンプルサイズ: 500 (b) サンプルサイズ: 2500

(c) サンプルサイズ: 5000 (d) サンプルサイズ: 10000

図 1: 横軸はエッジ数を示し縦軸が逆向きに認識されたエッジ数の検出エッジ数に対する比を示す. 図中の G2 は標準的な PC, MFE は提案手法によるもの. 5 つのパラメータセットでの平均値を示す.

5. まとめ

計算効率の高い制約に基づくアプローチによるベイジアンネットワークの構造推定精度を高めるためには少数データにおいても条件付き独立性を正しく認識できることが重要である. 我々はデータ温度モデルと自由エネルギー最小化による独立性検出方法を提案し, 漸近的には通常の G^2 検定に近づくことを示した. またシミュレーション実験により構造推定の精度, 特にエッジの向きを従来より正しく認識できることを確認した. これは因果的な構造の知識獲得に資すると考えられる.

謝辞

著者の一人 (磯崎) は情報理論についてアドバイスを下さった電気通信大学の小川朋宏准教授に感謝します.

参考文献

[Isozaki 08] Isozaki, T., Kato, N., and Ueno, M.: Minimum Free Energies with "Data Temperature" for Parameter Learning of Bayesian Networks, in *Proc. of IEEE International Conference on Tools with Artificial Intelligence (ICTAI-08)*, pp. 371–378 (2008)

[Kullback 68] Kullback, S.: *Information Theory and Statistics*, Dover Publications, Mineola, NY (1968)

[Spirtes 00] Spirtes, P., Glymour, C., and Scheines, R.: *Causation, Prediction and Search*, MIT Press, Cambridge, MA, second edition (2000)

[Verma 92] Verma, T. and Pearl, J.: An Algorithm for Deciding If a Set of Observed Independencies Has a Causal Explanation, in *Proc. of Conference on Uncertainty in Artificial Intelligence (UAI-92)*, pp. 323–330 (1992)