

ベジアンネットワーク型診断システムに対する高効率学習法

高橋隼汰[†]山形大学大学院理工学研究科[†]安田宗樹[‡]山形大学大学院理工学研究科[‡]

1. はじめに

症状の発生からその原因を推測を行うプロセスは重要である。例えば機械の故障診断、農作物の生育障害に対する原因推定などの例が挙げられる。本研究では、原因と症状の各項目をそれぞれ確率変数で表し、各確率変数同士の関連性を図1に示すような有向完全2部グラフによって表現したベジアンネットワーク (Bayesian network (BN)) [1, 2] を用いる。このBNを学習データ集合を用いて学習し、事後分布を求めることで、症状から原因の自動診断が可能となる。

学習データ数が少ない場合や学習データ集合の中に強いバイアスが入っている場合など、学習データ集合の質が低い場合には過学習の問題で診断システムの性能が劣化する恐れがある。本研究では、マルコフ確率場 (Markov random field (MRF)) [1, 2] を経由した、原因から症状への条件付き分布に対する新しい学習法を提案する。

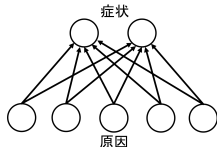


図1: 症状層と原因層からなる有向完全2部グラフ型BN。

2. ベジアンネットワーク型診断システム

症状層の K 個のノード $\mathcal{Y} := \{1, 2, \dots, K\}$ と原因層の n 個のノード $\mathcal{X} := \{1, 2, \dots, n\}$ からなる図1に示すような有向完全2部グラフ型のBNを考える。原因ノード間の直接的な因果関係はないものとしている。各症状ノードと各原因ノードにバイナリ確率変数 $\mathbf{y} := \{y_j \in \{0, 1\} \mid j \in \mathcal{Y}\}$, $\mathbf{x} := \{x_i \in \{0, 1\} \mid i \in \mathcal{X}\}$ を割り当てる。ここで、1を異常状態、0を正常状態としてい

る。原因層から症状層への条件付き分布を

$$P_{\text{BN}}(\mathbf{y} \mid \mathbf{x}) \propto \exp \left(\sum_{j \in \mathcal{Y}} c_j y_j + \sum_{j \in \mathcal{Y}} \sum_{i \in \mathcal{X}} w_{ji} y_j x_i \right) \quad (1)$$

の形で定義する。 $\mathbf{c} := \{c_j \in \mathbb{R} \mid j \in \mathcal{Y}\}$ と $\mathbf{w} := \{w_{ji} \in \mathbb{R} \mid j \in \mathcal{Y}, i \in \mathcal{X}\}$ はそれぞれ症状層バイアスと層間結合を表わす学習パラメータである。式(1)の条件付き分布と原因ノードの事前分布 $P_{\text{prior}}(\mathbf{x}) = \prod_{i \in \mathcal{X}} P_i(x_i)$ を用いて、事後分布はベイズの定理より

$$P_{\text{BN}}(\mathbf{x} \mid \mathbf{y}) = \frac{P_{\text{BN}}(\mathbf{y} \mid \mathbf{x}) P_{\text{prior}}(\mathbf{x})}{\sum_{\mathbf{x}} P_{\text{BN}}(\mathbf{y} \mid \mathbf{x}) P_{\text{prior}}(\mathbf{x})} \quad (2)$$

の形で得られる。式(2)の事後分布を用いることにより、症状 \mathbf{y} が与えられた際の各原因 x_i の異常確率を求めることができる。なお、本研究では原因ノードの事前分布 $P_{\text{prior}}(\mathbf{x})$ は何かの方法で事前に設計されているとする。学習パラメータ \mathbf{c} と \mathbf{w} は最尤法を用いて学習する。すなわち、学習データ集合 $D := \{\{\mathbf{y}^{(\mu)}, \mathbf{x}^{(\mu)}\} \mid \mu = 1, 2, \dots, N\}$ に対する対数尤度

$$\ell_{\text{BN}} := \frac{1}{N} \sum_{\mu=1}^N \ln P_{\text{BN}}(\mathbf{y}^{(\mu)} \mid \mathbf{x}^{(\mu)}) \quad (3)$$

を最大とする学習パラメータが学習解となる。

3. マルコフ確率場を用いた提案学習法

式(3)の最大化は、学習データ数が少ない場合や学習データ集合の中に強いバイアスが入っている場合など、学習データ集合の質が低いと過学習の問題が顕著になる。そのような場合は適切な正則化が必要である。本節では、条件付き分布 $P_{\text{BN}}(\mathbf{y} \mid \mathbf{x})$ に対する新しい学習法を提案する。提案法では次のMRFが基礎となる。

$$P(\mathbf{y}, \mathbf{x}) \propto \exp \left(\sum_{j \in \mathcal{Y}} c_j y_j + \sum_{i \in \mathcal{X}} b_i x_i + \sum_{j \in \mathcal{Y}} \sum_{i \in \mathcal{X}} w_{ji} y_j x_i \right) \quad (4)$$

ここで、 $\mathbf{b} := \{b_i \in \mathbb{R} \mid i \in \mathcal{X}\}$ は原因ノードに対するバイアスパラメータである。 $P(\mathbf{y}, \mathbf{x}) = P_{\text{BN}}(\mathbf{y} \mid \mathbf{x}) P(\mathbf{x})$ が成り立つ。ここで、 $P(\mathbf{x})$ はMRFの周辺分布 $P(\mathbf{x}) = \sum_{\mathbf{y}} P(\mathbf{y}, \mathbf{x})$ である。つまり、MRFの条件付き分布 $P(\mathbf{y} \mid \mathbf{x})$ は $P_{\text{BN}}(\mathbf{y} \mid \mathbf{x})$ と一致する。提案法では、式(3)に対す

Efficient learning method for Bayesian network-based diagnostic systems

[†] Syunta Takahasi; Graduate School of Science and Engineering, Yamagata University

[‡] Muneki Yasuda; Graduate School of Science and Engineering, Yamagata University

る最尤法の代わりにこの MRF の学習データ集合 D に対する (L_2 罰則付き) 対数尤度

$$\ell := \frac{1}{N} \sum_{\mu=1}^N \ln P(\mathbf{y}^{(\mu)}, \mathbf{x}^{(\mu)}) - \lambda \sum_{j \in \mathcal{Y}} \sum_{i \in \mathcal{X}} w_{j,i}^2 \quad (5)$$

に対する最尤法を実行する (この際, \mathbf{b} も学習パラメータとして扱う). ここで, $\lambda \geq 0$ は正則化係数である. ℓ の勾配はノード数が増加すると計算量爆発の問題で厳密に評価できなくなる問題があるが, その場合は例えば空間モンテカルロ積分法 [3] などの近似評価法を用いる. 式 (3) と (5) の対数尤度の間には次の関係がある.

$$\ell = \ell_{\text{BN}} + \frac{1}{N} \sum_{\mu=1}^N \ln P(\mathbf{x}^{(\mu)}) - \lambda \sum_{j \in \mathcal{Y}} \sum_{i \in \mathcal{X}} w_{j,i}^2 \quad (6)$$

つまり, ℓ は BN の対数尤度に \mathbf{x} に対する周辺尤度 (と重みに対する L_2 罰則) を正則化項として追加したものと解釈できる.

提案法ではここで一つの“トリック”を用いる. 式 (5) の対数尤度を勾配法を用いて上昇させるのだが, 学習解は ℓ を最大にするものではなく, 学習過程の中で ℓ_{BN} を最大にするものを選ぶ. つまり, 学習パラメータの更新ダイナミクスは MRF に対する対数尤度の勾配上昇に従うが, 学習の評価基準としては元々の目的の条件付き分布に対する対数尤度を用いるのである. ℓ_{BN} を学習の評価基準にすることにより, 正則化係数 λ も同基準内で最適化することができる. λ を変化させて個々に上記の手続きを実行し, その中で ℓ_{BN} を最大とする λ を最適な係数とすればよい. このトリックを用いて MRF を学習し, 得られた MRF の条件付き分布 $P(\mathbf{y} | \mathbf{x})$ を $P_{\text{BN}}(\mathbf{y} | \mathbf{x})$ と読み替えて式 (2) の事後分布計算に利用する.

4. 数値実験と考察

人工データを用いた数値実験により提案学習法の性能を調べる. $K = 3$, $n = 15$ とし, $P_{\text{BN}}(\mathbf{y} | \mathbf{x})P_{\text{prior}}(\mathbf{x})$ から伝承サンプリングを用いて N 個の学習データ点を生成する. ここで, $P_{\text{BN}}(\mathbf{y} | \mathbf{x})$ のバイアスパラメータは $c_j = 0$ に固定し, $w_{j,i}$ は平均ゼロ, 分散 $4/n$ のガウス分布からランダムに生成する. また, 事前分布 $P_{\text{prior}}(\mathbf{x})$ は $P_i(x_i) \propto \exp(\alpha_i x_i)$ としており, α_i は確率 β で $\alpha_i = -2.2$, 確率 $(1 - \beta)$ で $\alpha_i = 0$ となるようにランダムに決めている. $\alpha_i = -2.2$ となった x_i は約 90% の確率で $x_i = 0$ となり, 学習データ集合に強いバイアスをかける. $\beta = 0.1$ として実験を行う. 得られた人工的な学習データ集合を用いて, (1) ℓ_{BN} 最大化による学習 (BN), (2) $\lambda = 0$ に固定した提案法 (MRF), (3) λ の最適化も含めた提案法

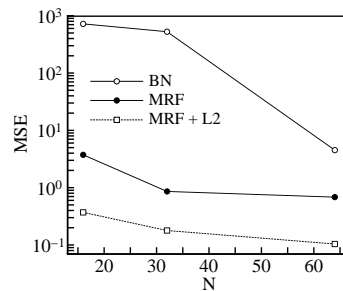


図 2: 重みパラメータの MSE. プロットは 200 回の実験の平均値である.

(MRF + L2), の 3 通りの方法で学習を行う. 学習パラメータはすべてゼロで初期化している. $N = 16, 32, 64$ の 3 通りの実験結果を図 2 に示す. 縦軸はデータ生成モデルの重み (真の重み) と学習重みとの間の平均 2 乗誤差 (mean square error (MSE)) である. すべての場合で提案法の方が高精度な重みを推定しているが, 特に N が小さいときその傾向が顕著である.

数値実験の結果から提案法の有用性が確認できる. しかしながら, やはり気になるのは提案法のトリックの意味である. 学習の過程をモニタリングすると, パラメータ更新の初期段階では ℓ と ℓ_{BN} の両方が単調に増加していくが, あるタイミングを堺に ℓ_{BN} が減少するといった挙動が見られる. ℓ_{BN} が減少する前の上がりきったところを切り出して学習解とするのが提案法のトリックである. 提案法の学習解は ℓ と ℓ_{BN} の両方から (ある一定のバランスの下で) 支持される解と解釈することができる.

5. まとめ

本研究では, 有向完全 2 部グラフ型の BN に基づく診断システムを考え, それに対する効率的な学習方法を提案した. 数値実験の結果から, 提案法は BN に対する直接的な最尤法より高品質な学習を達成できていることが確認できる. しかしながら, 提案法は正常な最尤法の枠組みからは説明が容易でない発見的“トリック”を用いている. 提案法のトリックに対する考察を深め, 理論的な背景を明確にすることが課題となる.

謝辞 本研究は科研費 (21K11778) の助成を受けたものである.

文献

- [1] D. Koller and N. Friedman: Probabilistic Graphical Models: Principles and Techniques, The MIT Press, 2009
- [2] K. P. Murphy: Machine Learning: A Probabilistic Perspective, The MIT Press, 2012
- [3] M. Yasuda and K. Uchizawa: A Generalization of Spatial Monte Carlo Integration, Neural Computation, Vol.33, issue 4, pp.1037-1062, 2021.