

6Q-7

ベイジアンネットワークにおける インタラクティブモデル構築手法の検討

吉見将太[†] 黒川悦子[‡] 橋本和夫[†]東北大学大学院 情報科学研究科[†] 東北大学大学院 医工学研究科[‡]

1 はじめに

近年、生活習慣に起因する疾患が増加しており、早期発見や予防のために、健康状態を評価する健康診断の受診やフィードバックへの関心がより高まっている。筆者らは、生活習慣改善支援のためのシステム開発を行っているが、より効果的な支援を行うことを目的として、ベイジアンネットワーク (BN)[1][2] という確率モデルを用いた健康診断情報の分析結果の利用を考えている。BN は有向非循環グラフであり、循環や相関構造を持たない。また、変数間の網羅的な学習を行いモデル構築・推論を行い、かつ、明確な仮説がないところから分析が可能である。

しかし、BN ではデータの統計的偏りを情報学的基準に基づき分析してパラメータ間の依存関係をグラフとして抽出するが、その結果の因果関係としての妥当性は必ずしも保証されない。妥当なモデル構築のためには、専門的知識も利用するインタラクティブなモデル構築手法が必要である。

本論文では、標本データのみから学習を行うフェーズと専門的知見を導入するフェーズからなるインタラクティブなモデル構築手法について検討する。

2 専門的知見導入の必要性

筆者らは BN を用いて生活習慣の分析 [3] を行い、歯磨き回数とメタボリックシンドロームの関連性を抽出した。一方 [4] は、口腔衛生の不良と心血管疾患のリスク上昇の因果関係を示した。このことは、筆者らが抽出した依存関係の妥当性を支持するものである。しかし、疫学分析の専門家からは、抽出されたグラフの中には、因果関係における原因と結果のリンクの向きに逆向する部分が含まれているとの指摘を受け、100% の妥当性を確認できない場合が存在した。これは、BN におけるリンクの向きは不安定なものであることを示している。与えられたデータからのリンクの向きを決定するのは不可能であり、専門家の最終判断が必要である。

A Study of an Interactive Structural Modeling in Bayesian Networks

[†]Shota YOSHIMI, Graduate School of Information Sciences, Tohoku University

[‡]Kazuo HASHIMOTO, Graduate School of Information Sciences, Tohoku University

[‡]Etsuko KUROKAWA, Graduate School of Biomedical Engineering, Tohoku University

依存関係を因果関係と見なすための基準は Hill の基準 [5] や Evans の条件 [6] などいくつか存在する。Hill の基準では、9 つの視点を挙げ、それらがすべて満たされた場合に因果関係として妥当であるとする。

したがって、BN により因果関係導出を行うためには、抽出された依存関係を専門家の知見によって検証する必要があるためインタラクティブなモデル構築が必須となる。

3 既存の専門的知見導入を行うモデル構築手法

BN を用いた因果関係のモデル構築の際に、専門的知見の導入を提案している論文がいくつかあるが、ここでは 2 つの既存手法についてまとめた。

[7] は、最初に領域データの収集を行い、これを MedKnow を用いてデータ構造を洗練する。MedKnow は、専門家に疾患とその後の経過、それらの相互作用や、周辺確率、条件付確率などを示すツールであり、専門家はデータ構造の改変を行う。最後に、Knowledge-Compiler により、先ほどの知識構造を用いた BN を構築している。この手法は、専門家の知識とデータから得られた知識を効率よく統合していると言える。しかし、2 節で述べた通り、BN のリンクの向きは不安定であり、BN が構築された後に、その構造について検証しなければ誤った因果関係を採用してしまう可能性がある。

[8] では、SFOBE (Smallest Forward-Backward Expert-Based) Model を提案している。この手法は専門家の知識によって構築されたモデルに関して再計算を行い、洗練されたモデルにする手続きであり、データから得られた依存関係に強い優先度を持たせるのが特徴である。しかし、データからの学習を重視しすぎると、抽出されるグラフは専門家の考える因果関係としての妥当性を満たさない危険性が高まる。専門的知識からではなくデータからの学習に大きく偏ってしまった場合、疫学的妥当性を満たさない可能性が高まる。また、最初に専門家の知識だけでモデルを構築しているが、専門家にも知識量の限界があるため、有用な情報の取りこぼしが十分考えられる。

以上より、BN のリンクの向き不安定性の考慮、また、疫学的妥当性の保証と知識量の多さが必要と言える。特に後者は専門的知見とデータの学習結果から偏りなく知識を得ることが手掛かりとなる。

4 提案手法

BNのリンクの不安定性を専門家の知見を用いて解消するためのインタラクティブなモデル構築手法について検討する。

Algorithm インタラクティブモデル構築手法

```

1: input グラフ  $G$ 
2:  $G' \leftarrow G$ 
3: 専門家にグラフ  $G'$  を提示
4: while  $G'$  でリンク修正の必要がある do
5:   if 不安定なリンクを検出 then
6:      $G'$  の他の部分は保持したままでリンクを修正
7:   else if 不要なリンクを検出 then
8:      $G'$  の不要なリンクを削除
9:   else if 因果関係の妥当性が不明なリンクを検出 then
10:     $G'$  のリンクの妥当性の検証
11:   end if
12: end while
13: output グラフ  $G'$ 

```

図1:インタラクティブモデル構築手法

提案手法の擬似コードを図1に示す。事前準備として、分析する健康情報のデータからBNを構築しておき、1行目ではそのグラフ G を入力とする。次に G' に G を初期値として与え、4-12行目をグラフ G' に修正の必要がある限り繰り返し実行する。この際、専門家が提示されたグラフを読み取り、修正が必要かどうかの判断を下す。修正が必要な場合5-10行目の操作が行われる。5,6行目は、 G' において専門家が不安定だと判断したリンクがあれば、 G' の他の部分はその構造を保持したまま不安定箇所のみを修正する。7,8行目は、妥当性がない因果関係を検出した場合はそれを削除する。この操作も、 G' の他の部分はその構造を保持したまま行う。9,10行目は、因果関係の妥当性が確認がされていない場合、そのリンクの妥当性の検証を行う。最終的に、入力したグラフ G は、グラフ全体において因果関係が示された因果構造グラフ G' として出力される。

次に、グラフの操作方法について説明する。5,6行目において、専門家が疫学的知見に基づきリンクの向きが不適切だと判断した場合、適切だと判断したリンクの向きに修正する。この際、再計算は行わない。7,8行目における操作は、依存関係を表すリンクを消去するのみである。例えば、ノード A, B, C, D において $A \rightarrow B \rightarrow C \rightarrow D$ というリンクが抽出され、専門家が $B \rightarrow C$ のみ因果関係の妥当性がないと判断した場合、 $A \rightarrow B$ と $C \rightarrow D$ というリンクだけが残る。9,10行目の因果関係の妥当性の検証は、2で挙げたHillの基準やEvansの条件に基づいた検証を行う。

以上が提案手法の概要である。既存手法と比較すると、提案手法は専門的知見の導入を重視しており、グラフにおけるリンクの不安定性の解消を図っている。

5 まとめ

本論文では、標本のデータのみから学習を行うフェーズと専門的知見を導入するフェーズのインタラクティブなモデル構築手法について提案した。この手法では、BNにおけるリンクの向きの不安定性を専門的知見の導入により改善する。また、BNで構築されたモデルの因果関係の妥当性の検証という手続きにより、2,3節で述べた疫学的妥当性の低さ、専門家の知識量の不足を補う。

今後は、具体的な事例に即して、今まで明らかにされてこなかった依存関係の発見と、疫学的な妥当性を持った生活習慣と疾患のモデル構築を実証する。

謝辞

本研究の一部は、文部科学省の平成19年度知的クラスター創成事業(第II期)の助成を受けて実施したものである。

同事業の研究統括である後藤順一教授、東北大学大学院医工学研究科永富良一教授には、本論文をまとめる際に様々なアドバイスを頂いた。ここに感謝する。

参考文献

- [1] David Heckerman, "A Tutorial on Learning With Bayesian Networks", *Technical Report*, MSR-TR-95-06 (1995).
- [2] Judea Pearl, "統計的因果推論", 共立出版株式会社 (2009).
- [3] 吉見将太, 黒川悦子, 橋本和夫, "ベイジアンネットワークを用いた生活習慣の分析", 第9回情報科学技術フォーラム, G-029 (2010).
- [4] Cesar de Oliveira, Richard Watt and Mark Hamer, "Toothbrushing, inflammation, and risk of cardiovascular disease: results from Scottish Health Survey", *BMJ*, 340:c2451 (2010).
- [5] Austin Bradford Hill, "The Environment and Disease: Association or Causation?", *Proceedings of the Royal Society of Medicine*, 58, 295-300 (1965).
- [6] ALFRED S. EVANS, "Causation and disease: The Henle-Koch postulates", *Yale J Biol Med*, 49: 175-194 (1976).
- [7] J. Horn, T. Birkholzer, O. Hogl, M. Pellegrino, R. Lupas Scheiterer, K.-U. Schmidt, and V. Tresp, "Knowledge Acquisition and Automated Generation of Bayesian Networks for a Medical Dialogue and Advisory System", *Artificial Intelligence in Medicine*, Springer, 199-202, (2001).
- [8] Ruxandra Lupas Scheiterer, Dragan Obradovic and Volder Tresp, "Tailored-to-Fit Bayesian Network Modeling of Expert Diagnostic Knowledge", *Journal of VLSI Signal Processing*, 49, 301-316 (2007).