

ベイジアンネットワークへの役割期待

佐々木 健佑, 岡田 幸彦

ベイジアンネットワークは、統計的機械学習の手法の一つである。確率的な依存関係を表現するグラフィカルモデルの特性をもつことから、説明可能性が高い機械学習手法として注目されている。本稿では、ベイジアンネットワークの基礎用語や構築方法を解説するとともに、応用研究として、疾病発症に応用した例を紹介する。また、それらを踏まえて、ベイジアンネットワークへの役割期待を議論する。

キーワード：統計的機械学習，グラフィカルモデル，ベイジアンネットワーク

1. ベイジアンネットワーク

ベイジアンネットワークは、1985年にJudia Pearlによってその概念が整備された確率的グラフィカルモデルの一つである [1]。不確実性を含む事象の予測や意思決定、具体的には、疾病の診断 [2]、政策効果の測定 [3] など、分野を問わず幅広く応用されている手法である。まずグラフィカルモデルにおける基礎用語を説明する。グラフィカルモデルとは、確率変数間の条件付き独立関係をグラフで表現する数理モデルのことである [4]。グラフを構成する確率変数 $X = X_1, \dots, X_N$ をノード (node) と呼ぶ。ノード同士の確率的な依存関係は、アーク (arc) と呼ばれる矢印で結ばれることによって表現される。原因を示すノード X_1 と、 X_1 の結果のノード X_2 の関係を $X_1 \rightarrow X_2$ と表現し、アークの元になるノード (X_1) を X_2 にとっての親ノード、アークの先になるノード (X_2) を X_1 にとっての子ノードと呼ぶ。ベイジアンネットワークは、方向をもつアークのみで構成されたグラフである有向グラフを用いて確率的な依存関係を表現する。またベイジアンネットワークは、閉路をもたない有向非循環グラフ (directed acyclic graph: DAG) として表される。例として、ノード $X = X_1, X_2, X_3$ で構成されたベイジアンネットワークを図 1 に示す。

ベイジアンネットワークは、グラフを構成する確率変数の同時確率分布を、確率変数の条件付き確率の積で表現することができる。図 1 の場合、それぞれの確率変数の依存関係は条件付き確率を用いて、 X_2 につ

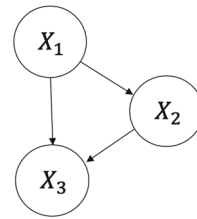


図 1 ベイジアンネットワークの例

いては $P(X_2 | X_1)$ 、 X_3 については $P(X_3 | X_1, X_2)$ と定義される。したがって、図 1 のベイジアンネットワークにおける確率変数の同時分布は、以下のように表現することができる。

$$P(X_1, X_2, X_3) = P(X_3 | X_1, X_2)P(X_2 | X_1)P(X_1) \quad (1)$$

2. ベイジアンネットワークのモデル構築

ベイジアンネットワークのモデル構築は、知識駆動型とデータ駆動型の 2 種類の方法が存在する。専門家の知見から手動で構築する方法が知識駆動型であり、データから構築する方法、さらには専門家の知見をデータからの構造学習に反映させる方法がデータ駆動型の方法である。知識駆動型によるベイジアンネットワークのモデル構築は、分析者が事前に指定した形でグラフ構造を決定し、そのベイジアンネットワークモデルを用いた確率推論を行うことによって確率変数間の関係を把握するようなアプローチである [5]。他方、データ駆動型によるベイジアンネットワークのモデル構築は、大きく分けて構造学習とパラメータ推定の二つのステップで行われる。以下では、データ駆動型によるベイジアンネットワークのモデル構築の方法について説明する。

2.1 データ駆動型の構造学習

構造学習のアルゴリズムには、制約ベース、スコアベース、ハイブリッドの 3 種類のアプローチがある [6]。

ささき けんすけ
筑波大学大学院システム情報工学研究科
〒 305-8573 茨城県つくば市天王台 1-1-1
sasaki.kensuke.rw@alumni.tsukuba.ac.jp
おかだ ゆきひこ
筑波大学システム情報系、筑波大学人工知能科学センター
〒 305-8573 茨城県つくば市天王台 1-1-1
okayu@sk.tsukuba.ac.jp

制約ベースアルゴリズムは、Pearl [1] の理論の応用であり、条件付き独立性検定を用いてベイジアンネットワークの構造を推定する。まず、条件付き独立について簡単に説明する。はじめに、二つの確率変数の独立について考える。以下の式が成り立つとき、確率変数 X と Y は独立であるという。

$$P(X, Y) = P(X)P(Y) \quad (2)$$

これは、 X についての情報がわかっていても、 Y についての情報は得られないことを意味する (X と Y を入れ替えても同じことが言える)。以下の式が成り立つとき、確率変数 X 、 Y は Z を与えられたもとで条件付き独立であるという。

$$P(X, Y | Z) = P(X | Z)P(Y | Z) \quad (3)$$

これは、 Z についての情報がわかった上で新たに X がわかっていても、 Y に関する情報は得られないということの意味する (X と Y を入れ替えても同じ)。

典型的な制約ベースアルゴリズムによるアプローチは、条件付き独立性の検定を含む以下の三つの手順でグラフ構造を推定する。

1. 完全無向グラフを作成する (すべてのノード同士が方向をもたないアークで結ばれたグラフ)。
2. 完全無向グラフから、条件付き独立性を用いてアークを削除する。
3. オリエンテーションルール (DAG の制約を満たすように矢印の方向付けを行う規則) によってアークの方向付けを行う。

制約ベースアルゴリズムの例として、PC (prototypical constraint-based) アルゴリズム [7]、HITON-PC (HITON parents and children) アルゴリズム [8, 9]、RAI (recursive autonomy identification) アルゴリズム [10] が挙げられる。

スコアベースアルゴリズムは、スコア関数が最大となるような DAG を探索するアルゴリズムであり、候補となる複数の構造の DAG に対して、スコア関数に応じたスコアが割り当てられる。スコア関数には、対数尤度や BIC [11] などが用いられる。データセット D から得られるグラフ構造 g のスコアを $Score(g | D)$ とすると、以下の式を満たすようなグラフ構造 G^* を探索するアプローチである。

$$G^* = \arg \max Score(g | D) \quad (4)$$

代表的なスコアベースアルゴリズムの例として、Greedy Hill Climbing アルゴリズム [12] が挙げられる。

ハイブリッドアルゴリズムによるアプローチは、スコアベースと制約ベースのそれぞれの欠点をカバーした方法である [6]。スコアベースアルゴリズムは、確率変数の数が増えるほど探索候補の DAG の数が増えるため、計算量が指数関数的に増加するという問題点がある [13]。また、制約ベースアルゴリズムには、独立性検定の有意水準を決める必要がある点や、独立性検定のエラーがグラフ構造に不安定性をもたらす可能性があるなどの問題点がある [14]。ハイブリッドアルゴリズムでは上記の問題点に対して、探索範囲の絞り込みに制約ベースアルゴリズムの特徴である条件付き独立性検定を用い、その範囲内におけるモデル探索に、スコア関数を用いるという方法で対処するアプローチである [6]。ハイブリッドアルゴリズムには、SC (sparse candidate) アルゴリズム [15] や、MMHC (max-min hill-climbing) アルゴリズム [16] などがある。

2.2 事前情報の導入

ベイジアンネットワークの構造学習では、専門家の知見や事前に明らかになっている因果関係 (causal assumptions) を事前情報としてモデルに組み込むことができる。たとえば、ある疾病の発症に関するベイジアンネットワークのモデルを、性別、年齢、喫煙習慣、飲酒習慣、疾病発症を確率変数として構築することを考える。この場合、喫煙習慣が年齢に影響することは考えにくい。そのため、構造学習の際に喫煙習慣 \rightarrow 年齢のアークが出現ないように制約条件を加えることが可能である。これにより、得られるグラフ構造をより実際の事象に即した解釈しやすいモデルに近づけることができる。制約条件の表現方法の例として、表 1 のように表すことができる。表中の横線で示された箇所は、 X_{from} から X_{to} のアークを制限することを意味する。

2.3 モデル平均化による構造学習

ベイジアンネットワークの応用に関する研究では、よりロバストなベイジアンネットワークのモデルを得ることを目的とし、モデル平均化 [17] (model averaging)

表 1 制約条件の表現方法の例

from \ to	性別	年齢	喫煙習慣	飲酒習慣	疾病発症
性別	—				
年齢		—			
喫煙習慣	—	—	—		
飲酒習慣	—	—	—	—	
疾病発症	—	—	—	—	—

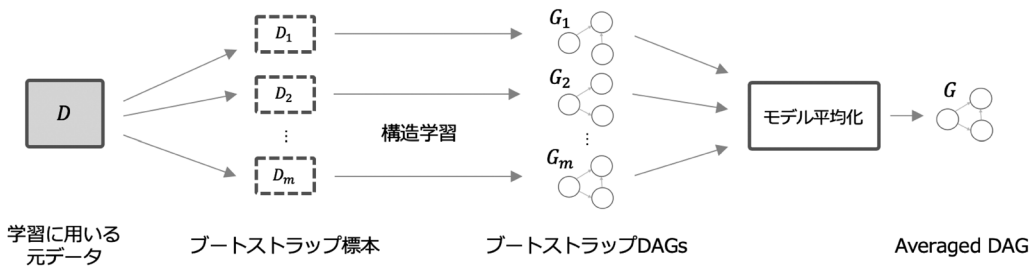


図2 モデル平均化による構造学習のプロセス

という方法が用いられている [18, 19]. モデル平均化によるベイジアンネットワークモデルの構造学習のプロセスを図2に示す. モデル平均化は, ブートストラップによって生成された標本で構造学習を行い, 構築した複数の DAG の平均をとる形でベイジアンネットワークのモデルを構築する方法である. モデル平均化による構造学習のプロセスは以下のとおりである.

1. 元データから, n 個の標本を復元抽出し, ブートストラップ標本を m 回作成する (n は元データの標本数).
2. ブートストラップ標本を用いて, 指定したアルゴリズムによって DAG (ブートストラップ DAG) を学習する.
3. ブートストラップ DAG に出現したアークの数を記録し, それぞれのアークについて, m 回のうち何回出現したかの割合を算出する.
4. 出現割合が一定の閾値を上回るアークを平均化 DAG として採用する.

上記のプロセスによって, 元データをそのまま利用して構築したベイジアンネットワークのモデルよりも高い予測性能をもつモデルが構築できるとされている [6]. また, ベイジアンネットワークのグラフ構造において, どのアークが信頼性の高いアークであるか判断するための方法としても, 上記のブートストラップによるアプローチが利用可能である. 具体的には, ブートストラップ DAG の中で, 上のプロセスにおいて算出したアークごとの出現割合をアークの信頼性指標とし, その値が大きければ大きいほど, アークの信頼性が高いと評価する方法である. ブートストラップ DAG におけるアークの出現割合は, arc strength と呼ばれ, モデル平均化を用いてベイジアンネットワークを構築する研究において利用されている [19, 20]. arc strength のイメージについて, 表2に示す. 表2では, X_1 から X_2 へのアークは strength の値が 1.00 であり, ブート

表2 arc strength のイメージ

from	to	arc strength
X_1	X_2	1.00
X_1	X_3	0.82
X_2	X_3	0.65

ストラップ標本に依存せず, すべてのブートストラップ DAG において出現したアークであることがわかる. つまり, X_1 から X_2 へのアークはほかのアークと比較して, 信頼性の高いアークであると考えられる.

上記のモデル平均化による構造学習の計算は, 並列に実行することが可能である [21]. そのため, 元データをそのまま利用してベイジアンネットワークモデルを構築する場合と比較した際に懸念される計算コストへの対処がある程度可能である.

2.4 パラメータ推定

構造学習によってベイジアンネットワークのグラフ構造を構築した後, 条件付き確率表を推定するパラメータ推定が行われる. 条件付き確率表 (conditional probability table: CPT) とは, 構築したベイジアンネットワークのモデルにおける子ノードと親ノードがとりうるすべての値 (状態) における確率を計算した表のことである. 条件付き確率表によって確率分布を表現することで, 線形から非線形な依存関係までを表現することができる [22]. パラメータ推定の方法には, 最尤推定 (maximum likelihood parameter estimation) とベイズ推定 (Bayesian parameter estimation) がある. ベイズ推定によるパラメータ推定は, 最尤推定によるパラメータ推定よりも頑強な推定結果が得られることが報告されている [23].

3. ベイジアンネットワークの解釈

ベイジアンネットワークは, 構築したモデルの解釈性に優れた手法であるといえる. 特に意思決定支援の目的でベイジアンネットワークによるモデルを利用する場合, なぜその結論に至ったのかを有向グラフによ

て示すことができる。たとえば疾病発症のベイジアンネットワークモデルの場合、疾病発症に直接的に関係する要因は何で、その背景にある要因は何かという情報を得ることができる。以下では、筆者らが過去に報告した記事 [24] の発展として構築したベイジアンネットワークのモデルを例にとって、解釈の方法について説明する。図 3 は、日本の自治体における住民の健康診断の項目を説明変数とし、糖尿病 3 年後新規発症を目的変数として構築したベイジアンネットワークのモデルを示している。また、利用データや各種設定について表 3 に示す。

図 3 のベイジアンネットワークのモデルは、糖尿病 3 年後新規発症、基本属性、生活習慣、検査値や既往歴、服薬や生活習慣の改善意志を確率変数として定義し、構造学習を行った結果である。なお、アークをもたないノード（貧血、20 歳の体重から 10 kg 以上増加している）は、構造学習の結果からほかのノードとの関係がみられなかったことを示している。

図 3 のグラフ構造の中で、糖尿病新規発症に関係するノードに着目する。糖尿病新規発症のノードの親ノードは、FPG (fasting plasma glucose, 空腹時血糖)、HbA1c (赤血球中のヘモグロビンのうち、糖と結合している割合を示す検査値) であること、FPG の親ノード

だが、過度な飲酒、メタボリックシンドロームであることなどが読み取れる。このことから、図 3 のベイジアンネットワークに利用したデータでは、空腹時血糖や HbA1c の値を下げるような介入が直接的に糖尿病 3 年後新規発症に影響を与えうること、さらに空腹時血糖の背景にある過度な飲酒やメタボリックシンドロームに対する介入が、間接的に糖尿病 3 年後新規発症に影響を与える可能性があることが読み取れる。

ベイジアンネットワークのグラフ構造を解釈する際に、注意すべき点がある。まず、ベイジアンネットワークのモデルが示す結果はあくまで利用した標本、確率変数の中での関係性を示しているという点である。未観測の確率変数や、利用したデータに存在するバイアスを踏まえた上で解釈する必要がある。また、構造学習に制約条件を利用した場合、その制約条件がグラフ構造に与える影響も考慮する必要がある。そのため、制約条件を検討する際は、モデル化の対象となる事象のメカニズムや専門家の知見などを十分に検討することが必要である。また、ノード同士の関係性を検討する際には、グラフ構造だけではなく、確率推論の結果も合わせて考察する必要があるという点も留意すべきである。確率推論では、構築したベイジアンネットワークのモデルに対して、あるノードの値を入力することで、ほかのノードの取りうる値の確率を、確率伝播法によって推定することができる。図 3 を例にとると、たとえば「該当、非該当」の二つの値をもつ過度な飲酒のノードに対して「該当」という値を入力することで、FPG のノードが取りうる値の確率（事後確率）がどのような値をとるか計算することができる。この結果を踏まえることで、ノード間の関係性を定量的に把握することが可能である。

4. おわりに

本稿では、ベイジアンネットワークモデルの構築方法や、実際のベイジアンネットワークのグラフ構造に対する解釈例を紹介した。グラフ構造によってモデルを視覚的に表現することができる点や、arc strength の値や確率推論による事後確率の算出によって確率変数同士の関係を定量的に把握できる点から、ベイジアンネットワークは解釈性に優れた手法であると言える。ベイジアンネットワークはさまざまな現実の問題に適用されているが、先述したベイジアンネットワークの特徴が特に力を発揮する場面は、利害関係者への説明責任が伴う意思決定を支援する場面であると考えられる。たとえば地域健康政策を検討する場面では、政策の運営

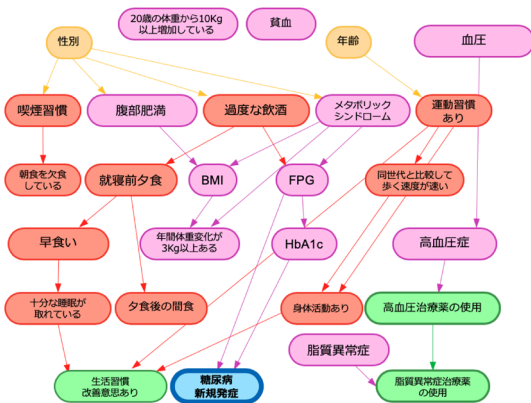


図 3 糖尿病 3 年後新規発症のベイジアンネットワークモデル

表 3 利用データ・各種設定

利用データ	・日本のある自治体における国民健康保険被保険者 1,007 人
説明変数	・基本属性、健康診断の問診票の項目、検査値の項目をそれぞれ離散化したもの
目的変数	・糖尿病 3 年後新規発症有無
構造学習アルゴリズム	・HITON-PC
	・ $m = 200$
モデル平均化の設定	・閾値は 0.25 (200 個のブートストラップ DAG のうち少なくとも 50 個に出現したアークを採用する)

主体である自治体がなぜその施策を行うのかについて、市民などに対して説明責任を果たす必要がある [25]。このような場面において、その自治体のデータを用いてベイジアンネットワークを構築し、結果を適切に解釈することで、エビデンスに基づく意思決定を支援することができるかも知れない。この点が、ベイジアンネットワークへの役割期待であると考える。

なお、ベイジアンネットワークのモデルを構築するためのソフトウェアには、統計ソフトウェア R のパッケージである bnlearn [26] や、Norsys Software 社が提供する Netica [27] などがある。特に bnlearn は、マニュアルやサンプルが充実しており [28]、ベイジアンネットワークに興味のある読者にとっての入り口として適していると考えられる。

参考文献

- [1] J. Pearl, “Bayesian networks: A model of self-activated memory for evidential reasoning,” In *Proceedings of the 7th Conference of the Cognitive Science Society*, pp. 329–334, 1985.
- [2] M. Langarizadeh and F. Moghbeli, “Applying naive bayesian networks to disease prediction: A systematic review,” *Acta Informatica Medica*, 2016.
- [3] 上野真也, “地域政策の効果を予測する—ベイジアンネットワーク分析の応用—,” 熊本大学政策研究, 2010.
- [4] 鈴木謙, 植野真臣, 『確率的グラフィカルモデル』, 共立出版, 2016.
- [5] 宮内義明, “メタボリックシンドロームマネジメントのための特定健診対応ベイジアンネットワークの構築,” 博士論文, 兵庫県立大学大学院, 2016.
- [6] M. Scutari and J. B. Denis, *Bayesian networks: With examples in R*, Chapman and Hall/CRC, 2014.
- [7] P. Spirtes, C. N. Glymour, R. Scheines and D. Heckerman, *Causation, Prediction, and Search*, MIT press, 2000.
- [8] C. F. Aliferis, A. Statnikov, I. Tsamardinos, S. Mani and X. D. Koutsoukos, “Local causal and markov blanket induction for causal discovery and feature selection for classification part I: Algorithms and empirical evaluation,” *Journal of Machine Learning Research*, **11**, pp. 171–234, 2010.
- [9] C. F. Aliferis, A. Statnikov, I. Tsamardinos, S. Mani and X. D. Koutsoukos, “Local causal and markov blanket induction for causal discovery and feature selection for classification part II: Analysis and extensions,” *Journal of Machine Learning Research*, **11**, pp. 235–284, 2010.
- [10] R. Yehezkel and B. Lerner, “Bayesian network structure learning by recursive autonomy identification,” *Journal of Machine Learning Research*, **10**, pp. 1527–1570, 2009.
- [11] G. Schwarz, “Estimating the dimension of a model,” *The Annals of Statistics*, **6**, pp. 461–464, 1978.
- [12] D. Heckerman, D. Geiger and D. M. Chickering, “Learning Bayesian networks: The combination of knowledge and statistical data,” *Machine Learning*, **20**, pp. 197–243, 1995.
- [13] D. M. Chickering, D. Heckerman and C. Meek, “Large-sample learning of Bayesian networks is NP-hard,” *Journal of Machine Learning Research*, **5**, pp. 1287–330, 2004.
- [14] M. Gasse, A. Aussem and H. Elghazel, “An experimental comparison of hybrid algorithms for Bayesian network structure learning,” *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, Springer, Berlin, Heidelberg, pp.58–73, 2012.
- [15] N. Friedman, I. Nachman and D. Peer, “Learning bayesian network structure from massive datasets: The sparse candidate algorithm,” In *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence*, pp. 206–215, 1999.
- [16] I. Tsamardinos, L. E. Brown and C. F. Aliferis, “The max-min hill-climbing Bayesian network structure learning algorithm,” *Machine Learning*, **65**, pp. 31–78, 2006.
- [17] N. Friedman, M. Goldszmidt and A. Wyner, “Data analysis with Bayesian networks: A bootstrap approach,” In *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence*, pp. 196–205, 1999.
- [18] P. Caillet, S. Klemm, M. Ducher, A. Aussem and A. M. Schott, “Hip fracture in the elderly: A re-analysis of the EPIDOS study with causal Bayesian networks,” *PLoS One*, **10**, e0120125, 2015.
- [19] S. Xu, W. Thompson, J. Kerr, S. Godbole, D. D. Sears, R. Patterson and L. Natarajan, “Modeling interrelationships between health behaviors in overweight breast cancer survivors: Applying Bayesian networks,” *PloS One*, **13**, e0202923, 2018.
- [20] H. S. Sousa, F. Prieto-Castrillo, J. C. Matos, J. M. Branco and P. B. Lourenço, “Combination of expert decision and learned based Bayesian Networks for multi-scale mechanical analysis of timber elements,” *Expert Systems with Applications*, **93**, pp. 156–168, 2018.
- [21] M. Scutari, “Bayesian network constraint-based structure learning algorithms: Parallel and optimised implementations in the bnlearn R package,” *arXiv*, 2014.
- [22] 本村陽一, “ベイジアンネットワークによる確率の人間行動モデリング,” 博士論文, 電気情報通信大学大学院, 2008.
- [23] R. Nagarajan, M. Scutari and S. Lebre, *Bayesian Networks in R with Applications in Systems Biology*, Springer Verlag New York Inc, 2013.
- [24] 佐々木健佑, 久野譜也, 岡田幸彦, “ベイジアンネットワークによる地域健康予測 (特集 医療・医薬とヘルスケア・インフォマティクス),” *オペレーションズ・リサーチ: 経営の科学*, **64**, pp. 399–405, 2019.
- [25] 鳥海航, 生方裕一, 久野譜也, 岡田幸彦, “地域健康政策へのベイジアンネットワークの応用,” *統計数理*, **66**, pp. 267–278, 2018.
- [26] M. Scutari, “Learning Bayesian networks with the bnlearn R package,” *arXiv*, 2009.
- [27] Netica User’s Guide, Application for Belief Network and Influence Diagrams, <http://www.norsys.com/index.html> (2020年3月5日閲覧)
- [28] bnlearn-an R package for Bayesian network learning and inference, <https://www.bnlearn.com/> (2020年3月5日閲覧)