講　座

ベイジアンネットワーク入門 (1)
Introduction to Bayesian Network (1)

須鎗　弘樹*
Hiroki SUYARI

要　旨

近年，不確実な情報の環境下において，ユーザの意志決定を支援する知能システムの一つとして，ベイジアンネットワークが注目を集めている．本小文では，ベイジアンネットワークの基礎について，できるかぎり平易に解説する．

キーワード：ベイジアンネットワーク，確率変数，有向非循環グラフ，条件付確率，確信度

1. ベイジアンネットワーク

具体例から始める．定期健診で血液検査をしたとする．血液検査の結果にはさまざまな項目があり，それぞれ正常値の範囲がある．仮に，血液検査のある 1 つの項目で異常値を示していたとしても，その原因となる病気の候補は複数あり，通常，その原因を特定することはできず，より詳細な検査が必要になってくる．しかし，血液検査の複数の項目で異常値を示していた場合，先の 1 つの項目だけで異常値を示していた場合よりも，疑うべき原因となる病気に対する確信度は高まると考えられる．そこで，原因となる病気を $D_1, \cdots, D_4$ とし，その結果である血液検査の項目を $T_1, \cdots, T_3$ とし，これらの関係を次のような図（**Fig. 1**）で表すことにする．**Fig. 1** は，7 つの頂点（ノードともいう）と 8 つの有向辺（矢印のこと）からなる有向グラフであり，とくに，どのノードも自分自身へ戻ってくる経路（サイクル）が存在しないので，有向非循環グラフという．さらに，各病気 D_i が発症しているとき $D_i = 1$，発症していないとき $D_i = 0$，同様に，血液検査の各項目 T_j が異常値を示しているとき $T_j = 1$，

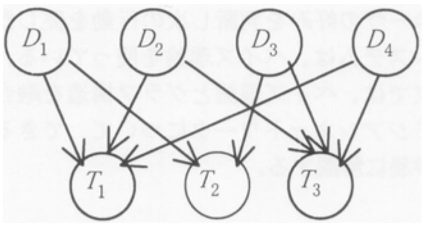

Fig. 1 One example of Bayesian Network for medical diagnosis.

正常値であるとき $T_j = 0$ として，それぞれの場合（事象）に確率を対応させることにより，これら $D_1, \cdots, D_4, T_1, \cdots, T_3$ を確率変数とみなすことができる．つまり，**Fig. 1** は，確率変数をノードとする有向非循環グラフである．**Fig. 1** において，各矢印は，病気（原因）が血液検査の結果に直接的に影響を及ぼすことを示している．その影響の度合いは，条件付確率で定量化される（**Fig. 1** では，スペースの関係上，省略した）．このように，(i) 各ノードがそれぞれ確率変数で表され，(ii) ノード間の矢印が因果関係を表し，(iii) その因果関係が条件付確率で定量化されている有向非循環グラフをベイジアンネットワークという〔1～3〕（信念ネットワークなどの他の呼称もあるが，最近は，この呼称がもっともよく使われている．なお，有向辺ではなく無向辺のグラフで，それら無向辺が相関関係を表すモデリングをグラフィカルモデリング〔4〕といい，ベイジアン

* 千葉大学工学部情報画像工学科〔〒263-8522 千葉県千葉市稲毛区弥生町 1-33〕: Department of Information and Image Sciences, Faculty of Engineering, Chiba University.
e-mail:suyari@faculty.chiba-u.jp

ネットワークと区別する）．**Fig. 1** は単純な因果関係を例にしたが，ある病気 D_i が別の病気 D_j を誘発し，その結果が血液検査 T_k に現れている場合も，同様に考えることができる（**Fig. 2**）．ただし，これら因果関係が循環するような経路が存在する場合には，ベイジアンネットワークの議論は使えない．

ここでは，医療診断の例をあげたが，これ以外にも，故障診断・ロボティックス・認識・ユーザモデリング・データマイニングなど，不確実性を含む対象のモデリングは，非常に多い〔5, 6〕．ユーザモデリングの有名な例として，マイクロソフト社のOfficeを起動すると現れるOfficeアシスタント（イルカのこと．Windows 版 Office では XP のバージョンから消えた）やアマゾン・コム社のお勧め本など，ユーザの使用を助けたり，ユーザの好みを判断し次の行動を促したりするシステムは，ベイズ理論を使っている．この小文では，ベイズ理論とグラフ構造を融合したベイジアンネットワークについて，できるかぎり平易に解説する．

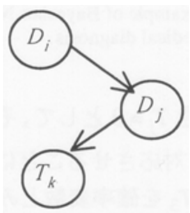

Fig. 2 One example of Bayesian Network.

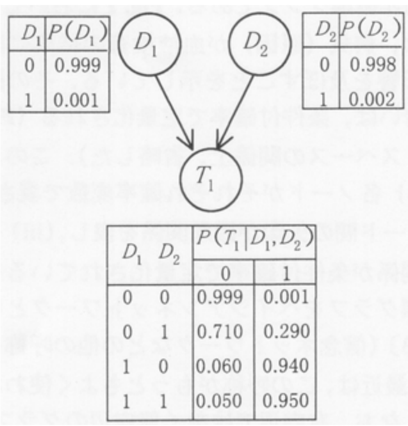

Fig. 3 One example of Bayesian Network (Subgraph of Fig.1).

2. ベイジアンネットワーク上の確率計算

Fig. 1 にあげた例では少々複雑なので，**Fig. 1** から D_1, D_2, T_1 の部分を取り出した **Fig. 3** で話を進めることにする（グラフ理論では，**Fig. 3** のグラフは **Fig. 1** の部分グラフという）．**Fig. 3** では，**Fig. 1**で省略した各ノードに確率を付してあることに注意していただきたい．

Fig. 3 のグラフにおいて，ノード D_1, D_2 については，これらを矢印の終点とするノードが存在しない．これを D_1, D_2 の親が存在しないという．このように親が存在しないノードには，それらの確率変数で定まる確率が付随する．これを事前確率という．これに対して，ノード T_1 は 2 つの親 D_1, D_2 をもつので，条件付確率 $P(T_1| D_1, D_2)$ が付随する．**Fig. 3** のノード T_1 の下のような条件付確率を要素とする表を条件付確率表（CPT: Conditional Probability Table）という．たとえば，

$$P\,(T_1 = 0 \mid D_1 = 0, D_2 = 0) = 0.999 \qquad (1)$$

は，2 つの病気 D_1, D_2 が発症しないとき $(D_1 = 0, D_2 = 0)$，血液検査の結果が正常値の範囲にある $(T_1 = 0)$ 確率は，0.999 であることを示している．したがって，条件付確率表の各行の和が 1 になることがわかる．

このように，ベイジアンネットワークにおいて，グラフ構造が確率変数間の定性的な関係を表しているのに対して，各ノードに付随する条件付確率が定量的な関係を表している．

ベイジアンネットワークのように複数の確率変数をもつシステムにおいて，ある条件付確率を求めたい場合，結局のところ，それらすべての確率変数の同時分布がわかればよい．たとえば，**Fig. 3** の場合，D_1, D_2, T_1 の同時分布

$$P\,(D_1, D_2, T_1) \qquad (1)$$

がわかれば，周辺分布：

$$P(D_1, D_2) = \sum_{t=0}^{1} P(D_1, D_2, T_1 = t) \qquad (2)$$

を求めることができる．さらに，条件付確率の定義より，

$$P(T_1|D_1, D_2) = \frac{P(D_1, D_2, T_1)}{P(D_1, D_2)} \qquad (3)$$

であるから，式 (1) と式 (2) を式 (3) に用いれば，$P(T_1|D_1, D_2)$ の条件付確率表をすべて埋めることができる．この例からわかるように，ベイジアンネットワークを構成するすべての確率変数の同時分布がわかれば，各ノードに付随する条件付確率あるいは事前確率を求めることができる．式 (1) の場合, 各確率変数が 2 つの値 (0 と 1) の値をとるので, 式(1)の値をすべて求めるには, $2^3 - 1 = 7$ 個の同時確率を求める必要がある（確率の総和が 1 であることから, 2^3 個ではなく, $2^3 - 1$ 個の確率がわかればよい）．より一般的に，ベイジアンネットワークを構成するノード（確率変数）の数が n 個あれば, $2^n - 1$ 個の確率が必要であり, n が大きくなると現実的な数ではなくなることが容易にわかる．これに対して，ベイジアンネットワークでは，同時確率の計算に必要な情報量を大幅に減らす方法がある．先に述べたベイジアンネットワークの定性的関係（グラフ構造）と定量的関係（同時分布）の間には，強固な関係が存在する．つまり, ベイジアンネットワークのグラフ構造が定まると，それを構成するすべての確率変数 $X_1, \cdots, X_n$ の同時確率は，次のように展開することができる．

$$P(X_1,\cdots,X_n) = P(X_1|\Pi(X_1)) \times \cdots \times P(X_n|\Pi(X_n)) \tag{4}$$

ここで, $\Pi(X_i)$ は，X_i の親の確率変数の集合を表す．たとえば，**Fig. 3** の場合，

$$\Pi(D_1) = \{\ \}, \Pi(D_2) = \{\ \}, \quad \Pi(T_1) = \{D_1, D_2\} \tag{5}$$

であるから，

$$P(D_1, D_2, T_1) = P(D_1)P(D_2)P(T_1|D_1, D_2) \tag{6}$$

を得る．逆に，ベイジアンネットワークを構成する確率変数の同時分布が式 (4) あるいは式 (6) のように与えられたとき，グラフ構造が決定される．

ここで，具体例を使って，先に述べた情報量の減少を確認する．**Fig. 3** の場合，各確率変数のとる値が二値（0 と 1）であり，確率の総和が 1 であることに注意すると，**Fig. 3** の各ノードの確率表より, 確率分布 $P(D_1)$ と $P(D_2)$ にはそれぞれ 1 個の確率の値が必要であり, $P(T_1|D_1, D_2)$ には 4 個の確率の値が必要であるから, 合計 6 個の確率の値が必要である．つまり，7 個必要だった確率の値が 6 個だけでよいことになる．ここでは，ノードの数が 3 個と少なく，その少ないノードの数に対して矢印が 2 個と多いために，情報量の減少の効果は少なかった．しかし，同様の考え方を **Fig. 1** について用いると，$2^7 - 1 = 127$ 個必要だった確率の値が 24 個の確率の値だけですむことになり, 大幅に減少することがわかる．この減少の効果は，ノードの数が多くなるにつれて，また，矢印の数が少なくなるに従って大きくなる．

3. 確信度

さて，**Fig. 3** のようにベイジアンネットワークが与えられたとき，実際に解きたい問題は，次のように与えられる．

問題：ベイジアンネットワーク上のいくつかの確率変数の値が観測できたとき，観測していない（できない）値を推定せよ．あるいは，その確率変数の分布を求めよ．

たとえば, **Fig. 3**の例では, 血液検査の項目T_1が異常値であったとき, つまり, $T_1 = 1$ のとき, その原因が病気 D_1 の発症なのか病気 D_2 の発症なのかを知りたいという問題である．これは，不確実な環境下である以上，明確にどちらであると断定できないので, 通常, 確信度 (belief) として,

$$BEL(x) := P(x|e) \tag{7}$$

で定義される量を物差しにする．ここで，eは観測値を表し，通常，evidence（証拠）の頭文字で表す．観測値は複数のときもあるので，eが複数になることもある．また，x は知りたい確率変数の値を表す．**Fig. 3** の例では，$T_1 = 1$ のとき，その原因が病気 D_1 の発症あるいは病気 D_2 の発症である確信度は，それぞれ

$$BEL(D_1 = 1) = P(D_1 = 1 | T_1 = 1),$$
$$BEL(D_2 = 1) = P(D_2 = 1 | T_1 = 1) \tag{8}$$

で与えられる．これは, 条件付確率あるいはより詳細にベイズの定理の言葉で言えば, 事後確率

に他ならない．上記の 2 つの確信度のうち，大きい方が原因の可能性が高いと言える．もちろん，両方が原因である確信度：

$$BEL\ (D_1 = 1, D_2 = 1)$$
$$= P\ (D_1 = 1, D_2 = 1 |\ T_1 = 1) \tag{9}$$

も考えることができる．

実際に,前節の方法で式(8)と式(9)を求めると,

$$BEL\ (D_1 = 1) \cong 0.37,$$
$$BEL\ (D_2 = 1) \cong 0.23, \tag{10}$$
$$BEL\ (D_1 = 1, D_2 = 1) \cong 0.00076$$

となり，病気 D_1 の発症の可能性が病気 D_2 の発症の可能性よりも高いが，同時に発症している可能性は非常に低いことがわかる．

ここでは，確信度という値を求めたが，

$$(BEL\ (D_1 = 0), BEL\ (D_1 = 1)) \tag{11}$$

のような事後確率分布を求めることもできる．このように，観測値から知りたい確率変数の確率分布を計算できれば，その確率変数の期待値・事後確率最大値・エントロピーなどを求めることが可能になり，ユーザに合理的な意思決定を促すさまざまな物差しを定量化できる．

5. ベイジアンネットワークの応用にあたって

本小文では，簡単のため，ベイジアンネットワーク上の確率変数がすべて二値をとる場合を取り上げたが，三値以上の多値さらには連続値をとる場合も，計算がより複 雑になり計算コスト（計算量やメモリ容量）がかかるものの，考え方は同じである．さらに，ベイジアンネットワークのグラフ構造ならびに各ノードの意味も，応用分野 においてそれぞれ適切な形と意味に適用できることは容易に想像できるであろう．また，確信度の計算では，結果を観測値（証拠）として原因を推論する場合を考えたが，これとは逆に原因を観測値（証拠）として結果を推論したり，原因と結果の両方を観測値（証拠）として，その間を結ぶノードの値を推論する場合にも，同様に考えることができる．このようなベイジアンネットワークの枠組みから，その応用範囲は非常に広いことは容易に読み取れると思われる．それら応用について，文献〔3, 5, 6〕ならびにその文献を参照されたい．

ここまで読まれた読者は，各応用分野において，実際にベイジアンネットワークをどのように定めるのかという疑問をもつと思われる．そのためには，本小文から，次 の 3 つの要素の決定が不可欠であることがわかる．(i) ノードとなる確率変数 (ii) グラ フ構造 (iii) 各ノードに付随する条件付確率．これら 3 つの要素を決定する方法としてもっとも望ましいのは，収集したデータから決定することであるが，データが不完全あるいは不足しているために，確率変数や条件付確率の決定を難しくしたり，ノード数の増加に伴い，最適なグラフ構造を決定することを難しくしている．前者の場合は, EM アルゴリズムなどを使って最尤推定を行い，後者の場合は，アルゴリズムが提案されているものの得られるグラフに制約があるなどの問題も含んでおり，それら理論的研究が現在も進んでいる．後者のグラフ構造の決定に関しては, 実際には, その応用分野の専門家がその経験から決定することも少なくなく，一度決めたグラフ構造をデータが得られるたびに逐次修正するということも多い．そのような導入の容易さもまたベイジアンネットワークの利点である．このように，ベイジアンネットワークの決定は，その上の確率計算以上の計算コストがかかる場合も少なくなく，従来の学習理論で培った手法が積極的に使われている．

文　献

〔1〕 Pearl J：Probabilistic Reasoning in Intelligent Systems, Morgan Kaufmann, CA 1988
〔2〕 Russell S, Norvig P, 石塚　満（訳）古川康一（監訳）: エージェントアプローチ人工知能 15 章 確率推論システム．共立出版，1997，pp439-473
〔3〕 本村陽一，佐藤泰介：ベイジアンネットワーク－不確定性のモデリング技術－．人工知能学会誌 **15**(4): 1-8, 2000
〔4〕 宮川雅巳：グラフィカルモデリング. 朝倉書店, 1997
〔5〕 佐藤泰介，櫻井彰人 編：特集「ベイジアンネット」. 人工知能学会誌 **17**(5): 538-565, 2002
〔6〕 佐藤泰介，本村陽一，他：ベイジアンネットセミナー BN2002 予稿集．http:// staff.aist. go.jp/y.motomura/ bn2002/