確率的グラフィカルモデル
—ベイジアンネットワークとその周辺—

安田 宗樹，片岡 駿，田中 和之

ベイジアンネットワークは人の知識を確率で表現する一つの表現形式であり，複数の事象間の因果関係に対する知識を条件付き確率という確率表現で定式化し，それらをつなぎ合わせることで構成される．ベイジアンネットワークを利用することにより逆に結果から原因を確率的に推論するいわゆる確率推論が可能となり，さまざまな診断システムにおける強力なツールとなっている．

本稿ではベイジアンネットワークへの導入のための基礎的な概念と数理について解説し，簡単な例題を用いてベイジアンネットワークへの直感を深めていく．

キーワード：ベイジアンネットワーク，確率推論，マルコフ確率場

1. はじめに

ベイジアンネットワーク (Bayesian network: **BN**) は人の知識を確率で表現する一つの表現形式である [1]. 複数の事象（または命題）間の因果関係（原因と結果）に対する知識を条件付き確率という確率表現で定式化し，それらをつなぎ合わせることで構成される．構成されたBNを利用することにより，われわれは逆に結果から原因を推論する**確率推論**(probabilistic inference) を行うことが可能となる．確率推論は人工知能における重要なツールの一つであり，システムの故障診断や危険予測システム，医療診断システムなどさまざまな応用が見込まれる．

推論に確率を介在させることにより，現象の不確実さ（あいまいさ）を系統的に扱えるようになる．それにより，あり得そうなことと，あまりあり得なさそうなことの両方を完全に考慮に入れつつ，最終的な結果にたどり着くという複雑な推論が可能となるのである．あまりあり得なさそうなことでもそれらが積もり積もって最終的な結果に大きく影響を与えてるということも十分考えられ，それを数理的にきちんと扱えるというところがBNのメリットの一つである．これは従来の決定論的な推論法ではなかなか成し得ない方法論であり，それゆえBNが注目を集めてきているわけである．

本稿ではBNの確率推論を中心とした基礎数理を中心に解説し，次いでもう一つの応用上重要なグラフィカルモデルであるマルコフ確率場について触れる．

やすだ むねき
山形大学大学院理工学研究科
〒 992–8510 山形県米沢市城南 4–3–16
かたおか しゅん，たなか かずゆき
東北大学大学院情報科学研究科
〒 980–8579 宮城県仙台市青葉区荒巻字青葉 6–3–09

2. ベイジアンネットワークの基礎と応用

BNは，因果関係 (causality) を確率により記述する**グラフィカルモデル**(graphical model) の一種で，複雑な因果関係の推論を**非循環有向グラフ**(directed acyclic graph: **DAG**)[1]により表し，個々の変数間の関係を条件付き確率 (conditional probability) で表す確率推論のモデルである．図1は5ノードからなるDAGの例である．リンクには矢印で示したような方向性があるので，このグラフは有向グラフである．またこの有向グラフは循環をもたないのでDAGであると言える．あるノードに注目し，そのノードから矢印がのびているノードを**子ノード**(children) と呼び，またそのノードへの矢印をもっているノードを**親ノード**(parents) と呼ぶことにする．図1でいえば，ノード3の子はノード4と5であり，ノード4の親はノード2と3である．

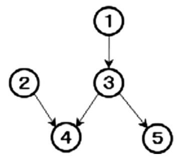

図1 5ノードのDAG
ノードに書かれた数字はそのノードの番号を表している．

[1] 有向グラフ (directed graph) とはグラフのリンク（辺または弧とも呼ばれる）に方向性が定義されているものであり，非循環有向グラフとはリンクを方向どおりにたどっていったときループ（循環）がない有向グラフのことである．

BN では各ノードに各事象を対応させ，有向リンクがそれぞれの事象間の因果関係（簡単に言えば，親が子の直接の原因であり，子がその結果であるという関係）を表す．まず，図 1 の $i = 1, \ldots, 5$ までの各ノードにそれぞれ事象を表す確率変数 $\boldsymbol{X} = \{X_1, \ldots, X_5\}$ を対応させよう．すると確率変数 $\boldsymbol{X}$ のすべての状態を表す確率はそれらの結合確率 $P(X_1, \ldots, X_5) = P(\boldsymbol{X})$ で表すことができる．

2.1 因果独立の仮定

図 1 の DAG 上に定義された確率 $P(\boldsymbol{X})$ を条件付き確率の定義を用いて変形してみよう．グラフの子から順々に分解していくことにより一般に

$$\begin{aligned} P(\boldsymbol{X}) &= P(X_5 \mid X_1, X_2, X_3, X_4) P(X_4 \mid X_1, X_2, X_3) \\ &\quad \times P(X_3 \mid X_1, X_2) P(X_2 \mid X_1) P(X_1) \end{aligned} \tag{1}$$

と因数分解することができる．BN では**因果独立の仮定**(causal independence) というものを課す．つまり，**直接の親ノード以外は因果的に独立であると仮定**する．図 1 をみると例えば X_5 の直接の親は X_3 のみであるから，それ以外からは独立と考えることにより

$$P(X_5 \mid X_1, X_2, X_3, X_4) = P(X_5 \mid X_3)$$

と表せる．同様に

$$\begin{aligned} &P(X_4 \mid X_1, X_2, X_3) = P(X_4 \mid X_2, X_3), \\ &P(X_3 \mid X_1, X_2) = P(X_3 \mid X_1), \\ &P(X_2 \mid X_1) = P(X_2) \end{aligned}$$

となる．以上の 4 つの書き換えを式 (1) に代入することにより

$$\begin{aligned} P(\boldsymbol{X}) &= P(X_5 \mid X_3) P(X_4 \mid X_2, X_3) \\ &\quad \times P(X_3 \mid X_1) P(X_2) P(X_1) \end{aligned} \tag{2}$$

という表現を得る．式 (2) が図 1 に対応した BN の確率モデル（結合確率）である．式 (2) の右辺にある 5 個の条件付き確率の具体的な値が与えられれば，結合確率を完全に把握できるので，この BN を用いて確率推論を行うことができる．

例えば n 個のノードで構成されるある DAG があるとして i 番目のノードに確率変数 X_i を対応させるとする：$\boldsymbol{X} = \{X_i \mid i = 1, 2, \ldots, n\}$．またノード i の直接の親のノードの番号の集合を Ω_i とする．するとこの DAG 上の BN の確率モデルは一般に

$$P(\boldsymbol{X}) = \prod_{i=1}^{n} P(X_i \mid \{X_j \mid j \in \Omega_i\}) \tag{3}$$

と書くことができる．ただし $P(X_i \mid \emptyset) = P(X_i)$ と約束する．図 1 で見てみよう．図 1 ではそれぞれ $\Omega_1 = \emptyset$, $\Omega_2 = \{\emptyset\}$，$\Omega_3 = \{1\}$，$\Omega_4 = \{2, 3\}$，$\Omega_5 = \{3\}$ なので式 (3) は

$$\begin{aligned} P(\boldsymbol{X}) &= P(X_1) P(X_2) P(X_3 \mid X_1) \\ &\quad \times P(X_4 \mid X_2, X_3) P(X_5 \mid X_3) \end{aligned}$$

と表され，式 (2) と一致することがわかる．

2.2 ベイジアンネットワークの簡単な例題

ここで，BN についての直感をより深めるために，BN の簡単な例題をみてみることにする．

図 2 は Family Out Problem と呼ばれる BN の例題で，家の中に家族がいるかどうかを外からわかる状況から推論する問題となっている．【家に誰もいない (Family Out)】，【飼い犬のお腹の調子が悪い (Bowel Problem)】，【玄関先の明かりがついている (Light On)】，【飼い犬が外にいる (Dog Out)】，【飼い犬の鳴き声が聞こえる (Hear Bark)】の 5 つの事象からなる BN であり，それぞれ 1〜5 まで番号をつけ，確率変数 X を割り当てている．それぞれの事象に対して YES なら 1 を No なら 0 を割り当てることとすると，例えば $X_3 = 0$ は玄関先の明かりがついていないということを表し，$X_5 = 0$ は犬の鳴き声が聞こえないということを表すこととなる．

式 (3) に従うと，この BN の結合確率は

$$\begin{aligned} P(\boldsymbol{X}) &= P(X_1) P(X_2) P(X_3 \mid X_1) \\ &\quad \times P(X_4 \mid X_1, X_2) P(X_5 \mid X_4) \end{aligned} \tag{4}$$

と書ける．またそれぞれの条件付き確率の値は表 1〜3 のように設定されている．

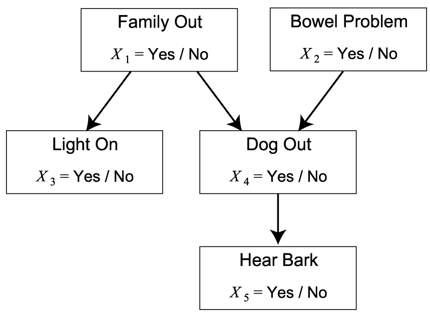

図 2 BN の例題 (Family Out)

表 1　$P(X_1)$（左）と $P(X_2)$（右）の値

$X_1 = 0$	0.85
$X_1 = 1$	0.15

$X_2 = 0$	0.99
$X_2 = 1$	0.01

表 2　$P(X_3 \mid X_1)$（上）と $P(X_5 \mid X_4)$（下）の値

	$X_1 = 0$	$X_1 = 1$
$X_3 = 0$	0.95	0.4
$X_3 = 1$	0.05	0.6

	$X_4 = 0$	$X_4 = 1$
$X_5 = 0$	0.99	0.3
$X_5 = 1$	0.01	0.7

表 3　$P(X_4 \mid X_1, X_2)$ の値

	$X_1 = 0$	
	$X_2 = 0$	$X_2 = 1$
$X_4 = 0$	0.7	0.03
$X_4 = 1$	0.3	0.97

	$X_1 = 1$	
	$X_2 = 0$	$X_2 = 1$
$X_4 = 0$	0.1	0.01
$X_4 = 1$	0.9	0.99

この表をみてみると，例えば“家に誰かがいる確率”は $P(X_1 = 0) = 0.85$ であるから 85%の確率で誰かがいることになり，“家に誰もいないときに玄関先の明かりがついている確率”は $P(X_3 = 1 \mid X_1 = 1) = 0.6$ であるから誰もいないときは 60%の確率で玄関先の明かりがついていることになる（防犯上の理由であろうか）．また，“犬のお腹が調子悪く，さらに家族が家にいないときに犬が外にいる確率”は $P(X_4 = 1 \mid X_1 = 1, X_2 = 1) = 0.99$ なので 99%の確率で犬が外に出されていることがわかる．表 1〜3 は**条件付き確率表**(conditional probability tabel: **CPT**) と呼ばれる．CBT の具体的な値は，現象を実際に観測し，得られた観測データをもとに決められる．

2.3　ベイジアンネットワークを利用した確率推論

ここでは BN の利用法について解説する．BN とは原因（親ノード）から結果（子ノード）への因果関係を矢印で表し，ネットワークを構成しており，各因果関係は条件付き確率の形で CPT により与えられている．大雑把に言えば，BN を利用することにより，逆に“結果から原因が何であったのかを推論（確率推論）する”ことが可能となる．例えば図 2 で，われわれは家の中に誰かがいるがどうかは知らないが，外観から玄関先に明かりがついていて，犬の鳴き声が聞こえることはわかっているとする．わかっているこの 2 つの情報（明かりがついている＆犬が吠えている）から家の中に誰かがいるかどうかを確率的に推論することが BN を利用することにより可能となる．

BN において確率変数のとる値は $\{0, 1\}$ の 2 値である必要はないが，以下の説明では簡単のためにすべての確率変数は $\{0, 1\}$ の 2 値をとるものとする．

2.3.1　周辺化とベイズの公式

確率推論を行うために以下の 2 つの数学が必要となる．一つは**周辺化**(marginalization) という技術である：

$$P(X_1) = \sum_{X_2=0,1} P(X_1, X_2)$$

X_1 と X_2 の結合確率 $P(X_1, X_2)$ を X_2 で和をとることにより X_1 のみの確率（周辺確率）を得ることができる．周辺化は確率変数がもっと多い場合も同様に可能で，一般に $\boldsymbol{X} = \{X_i \mid i = 1, 2, \ldots, n\}$ に対して，

$$\boxed{P(\boldsymbol{X}^*) = \sum_{\boldsymbol{X}^\dagger} P(\boldsymbol{X}) \qquad (5)}$$

により記述される．ここで $\boldsymbol{X}^*, \boldsymbol{X}^\dagger \subset \boldsymbol{X}$，$\boldsymbol{X}^* \cup \boldsymbol{X}^\dagger = \boldsymbol{X}$，$\boldsymbol{X}^* \cap \boldsymbol{X}^\dagger = \emptyset$ であり，和の記号 $\sum_{\boldsymbol{X}^\dagger}$ は $\boldsymbol{X}^\dagger$ の要素すべてに対する和を表す．例えば $\boldsymbol{X} = \{X_1, \ldots, X_5\}$ として，$\boldsymbol{X}^* = \{X_1, X_3, X_4\}$，$\boldsymbol{X}^\dagger = \{X_2, X_5\}$ とすると，式 (5) は

$$P(X_1, X_3, X_4) = \sum_{X_2=0,1} \sum_{X_5=0,1} P(X_1, \ldots, X_5)$$

となる．ある変数で和をとると，残りの変数についての周辺確率が計算されるというのが，上の周辺化の意味である．

もう一つは**ベイズの公式**(Bayesian formula) であり，これが BN を利用した推論のキーテクニックとなる．上で確率推論とは結果から原因を推論する技術であると述べた．つまり因果の矢印を逆にたどることが確率推論であり，そのためには矢印を逆転するための方法論が必要となる．それがこれから説明するベイズの公式である．　いま二つの事象 A と B があるとして，図 3 のように A が原因で B が結果であるとする．A と B に対する結合確率を BN の流儀で書き下せば $P(A, B) = P(B \mid A)P(A)$ となり，矢印に対応した条件付き確率 $P(B \mid A)$ は，A を知ったもとでの B の確率であると解釈できる．矢印を逆にたどるということはつまり，逆に，“B を知ったもとでの A の確率 $P(A \mid B)$”

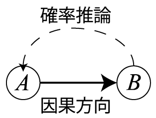

図 3 A が原因で B が結果の BN
確率推論とは因果の方向の逆向きをたどることである（図中の点線の矢印）．

を計算することである．図 3 の BN の結合確率は先にも述べたとおり $P(A,B)=P(B \mid A)P(A)$ であるので，CPT として与えられるのは $P(A)$ と $P(B \mid A)$ の値で，$P(A \mid B)$ は CPT には含まれていないことに注意．つまり，われわれは **CPT に含まれている $P(A)$ と $P(B \mid A)$ の情報のみから $P(A \mid B)$ を計算しなくてはならない**．それを可能にするのがベイズの定理である．結合確率 $P(A,B)$ は以下の 2 通りの表現が可能である．

$$P(A,B)=P(A \mid B)P(B)$$
$$P(A,B)=P(B \mid A)P(A)$$

この 2 式より

$$\boxed{P(A \mid B)=\frac{P(B \mid A)P(A)}{P(B)} \qquad (6)}$$

を得る．周辺化 (5) を利用することにより

$$P(B)=\sum_{A=0,1} P(B \mid A)P(A) \qquad (7)$$

と表せるから，式 (6, 7) を組み合わせて

$$P(A \mid B)=\frac{P(B \mid A)P(A)}{\sum_{A=0,1} P(B \mid A)P(A)}$$

を得る．これがベイズの公式の基本形である（一般にベイズの定理と言ったら式 (6) のことを指す場合が多い）．CPT に含まれる情報のみから $P(A \mid B)$ が導出されていることがわかる．

2.3.2 確率推論の例 (Family Out)

周辺化 (5) とベイズの公式 (6) を応用して図 2 の BN に対する確率推論に挑戦する．玄関先に明かりがついていて（$X_3=1$），犬の鳴き声が聞こえる（$X_5=1$）ときに家の中に誰もいない確率を求めてみよう．目標となる確率は $P(X_1=1 \mid X_3=1, X_5=1)$ である．式 (6) より

$$P(X_1 \mid X_3, X_5)=\frac{P(X_1, X_3, X_5)}{P(X_3, X_5)} \qquad (8)$$

と表すことができる．また周辺化 (5) より分母・分子はそれぞれ

$$P(X_3, X_5)=\sum_{X_1, X_2, X_4=0,1} P(X_1, \ldots, X_5) \qquad (9)$$
$$P(X_1, X_3, X_5)=\sum_{X_2, X_4=0,1} P(X_1, \ldots, X_5) \qquad (10)$$

であり，右辺の $P(X_1, \ldots, X_5)$ は Family Out の結合確率 (4) である．式 (9, 10) を式 (8) の右辺に代入すると，

$$P(X_1 \mid X_3, X_5)=\frac{\sum_{X_2, X_4=0,1} P(X_1, \ldots, X_5)}{\sum_{X_1, X_2, X_4=0,1} P(X_1, \ldots, X_5)} \qquad (11)$$

となり，条件付き確率 $P(X_1 \mid X_3, X_5)$ が CPT（表 1〜3）に含まれている条件付き確率のみで表現されていることがわかる．表 1〜3 で与えられた CPT の値と式 (11) を利用して丹念に計算することにより $P(X_1=1 \mid X_3=1, X_5=1)=0.7577\cdots$ という結論を得る．つまり，玄関先に明かりがついていて犬の鳴き声が聞こえる場合，およそ 76%の確率で家の中には誰もいないという推論結果となる．

2.4 確率推論の難しさ

前節の方法論を利用することにより，BN の構造と CPT が与えられれば“原理的には”あらゆる条件付き確率を計算することができ，したがって確率推論システムを構成することができる．しかし，構成したシステムが実際に運用できるかと言われれば必ずしもそうではない．

例えば式 (11) の分母に注目してみよう．ここでは X_1，X_2，X_4 に関する和を計算している．それぞれが 2 値変数なので合計 $2^3=8$ 回の和の演算を行っていることとなる．今はたった 3 個の変数に関してのみの和なのでさほど問題とならないが，システムがより大きくなって（ノード数がより多くなって）きた場合，周辺確率を計算するために和をとらねばならない変数の数が多くなっていく．K 個の変数に関する和を計算するとしたら，和の演算は合計 2^K 回必要となる．つまり，計算量が K に関して指数関数的に増加することとなる．

仮に 1 秒間に 1 億回の演算が可能な計算機があったとして，その計算機を使って 2^K 回の演算するとする．表 4 に 2^K 回の演算に必要なおおよその計算時間を示している．K の増加に伴い計算時間が爆発的に増加していることがわかる．このような問題は**計算量爆発**などと呼ばれたりする．確率推論の際に必要な周辺化の

表 4　2^K 回の演算に必要な計算時間

K	time
10	約 0.00001 秒
30	約 0.18 分
50	約 130 日
70	約 37 万 4000 年
100	約 4000 億年

計算は計算量爆発の問題を孕んでおり，システムが大きくなるにつれてその問題はより深刻になってくる．計算量爆発の問題は BN に限らず，確率的な計算システムの中ではしばしば現れる重要な問題である．

計算量爆発の問題を回避するために，実用上は厳密な周辺化を諦めて**近似計算による周辺化**に頼ることとなる．詳細は割愛するが，BN やその他さまざまな確率計算システムにおける近似計算技法として**確率伝搬法 (belief propagation: BP)** と呼ばれる近似計算法が広く利用されている（sum-product アルゴリズムとも呼ばれる）[1]．BN 上での確率伝搬法の入門書としては文献 [2] が詳しい．

2.5　最尤推定による CPT の学習

ここでは**完全データ**(complete data set) に対する CPT の統計的学習についての概要を述べる．完全データとはすべてのノードに対するデータが完全にそろっているデータセットのことであり，データの欠損などを全く含まないデータである．

いま n 個のノードを含む BN (3) に対して M 個の観測データセット $D := \{\boldsymbol{d}^{(\mu)} \mid \mu = 1, 2, \ldots, M\}$ が得られたとする．ここで μ 番目のデータ $\boldsymbol{d}^{(\mu)} = \{d_1^{(\mu)}, d_2^{(\mu)}, \ldots, d_n^{(\mu)}\}$ はすべてのノードに対するの 1 組の観測データであり，異なる番目のデータ同士は独立であるとする（例えば μ 番目のデータをある日に取って，$\mu+1$ 番目のデータはその数日後に取るなどと考える）．

BN (3) において，μ 番目のデータ $\boldsymbol{d}^{(\mu)}$ を生成する確率は

$$P(\boldsymbol{d}^{(\mu)}) = \prod_{i=1}^{n} P(d_i^{(\mu)} \mid \{d_j^{(\mu)} \mid j \in \Omega_i\})$$

と記述され，それぞれのデータは互いに独立であるから，“すべてのデータ D が生成される確率”はそれぞれの単純な積で表される．

$$l(D) := \prod_{\mu=1}^{M} \prod_{i=1}^{n} P(d_i^{(\mu)} \mid \{d_j^{(\mu)} \mid j \in \Omega_i\}) \qquad (12)$$

この $l(D)$ は**尤度関数**(likelihood function) と呼ばれ，この尤度関数を最大とする条件付き確率の値を求めることが**最尤推定**(maximum likelihood estimation: MLE) である（実際に観測したデータが尤も生成されやすいようなモデルを求めることに対応している）．

条件付き確率の規格化に注意しながら（ラグランジュの未定乗数法を利用して）尤度関数 $l(D)$ を最大化すると

$$P(X_i = k \mid \{X_j = r_j \mid j \in \Omega_i\}) = \frac{N_i(k; \{r_j\})}{M} \qquad (13)$$

なる結果を得る．ここで $N_i(k; \{r_j\})$ は観測データ D の中でノード i に対応するデータの値が k で，かつ同時に，i の親ノードのデータの値が $\{r_j\}$ となっている観測データの個数である（いまデータの個数は M なので $0 \le N_i(k; \{r_j\}) \le M$ である）．以上要するに完全データを持っている場合は，観測データの中での**標本条件付き確率が学習解**となっており，学習は非常に単純である．

観測データの中に欠損がある場合，すなわち**不完全データ**(incomplete data set) の場合はこのように単純にはいかず，2.3 節で紹介した確率推論と組み合わせた EM アルゴリズムが必要となり，少々入り組んでいるため本稿では割愛する．不完全データの場合の CPT の学習については文献 [4] を参照されたい．

3.　もう一つのグラフィカルモデル—マルコフ確率場—

BN は事象間に明確な因果関係を DAG の形で設け，グラフィカルモデルを構成している．対して，事象間に親子の関係を明示的にはつけないようなグラフィカルモデルもある．そのような場合，確率モデルは無向グラフ上に定義され，確率変数間のリンクは明示的な親子の因果ではなく，双方向の依存性を表すこととなる．このような無向グラフ的グラフィカルモデルは BN に対して**マルコフネットワーク**(Markov network) もしくは**マルコフ確率場**(Markov random field) と呼ばれ，実用上非常に重要な確率モデルのクラスを成す．以下では簡単な例を通してマルコフ確率場がどのような確率モデルなのかを見ていく．

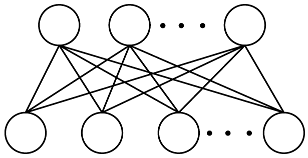

図 4　完全 2 部グラフ
上層・下層間はすべてのノード間にリンクがあり，層内のリンクはない．

3.1　マルコフ確率場の例（2 部グラフ型マルコフ確率場）

例として図 4 のような完全 **2** 部グラフ(bipartite graph) 型のマルコフネットワークを考える．ノードは BN のときと同様で確率変数を表す．しかしながら，今回の場合リンクは矢印ではなく単なる線（無向線分）である．リンクは BN のときのような明確な親子関係を表すものではなく，確率変数間の双方向の依存関係を表現する．上層のノードには確率変数 $\boldsymbol{X} = \{X_i \mid i = 1, 2, \ldots, n_x\}$ を割り当て，下層のノードには確率変数 $\boldsymbol{Y} = \{Y_j \mid j = 1, 2, \ldots, n_y\}$ を割り当てる．ここで n_x と n_y はそれぞれ上層・下層のノード数を表す．

Hammersley–Clifford の定理より，図 4 のマルコフネットワーク上での $\boldsymbol{X}$，$\boldsymbol{Y}$ の結合確率は一般に

$$P(\boldsymbol{X}, \boldsymbol{Y}) := \frac{1}{\mathcal{Z}} \prod_{i=1}^{n_x} \prod_{j=1}^{n_y} \Phi_{i,j}(X_i, Y_j) \qquad (14)$$

と表すことができる．ここで $\Phi_{i,j}(X_i, Y_j) > 0$ は上層ノード i と下層ノード j 間のリンクのポテンシャル関数と呼ばれる一種のコストであり，$\mathcal{Z}$ は規格化定数である．

このモデルを例えば医療診断などの診断システムのモデルとして利用することを考えてみよう．例えば，下層に対応するノードを問診データ（日々の生活習慣）に対応させ，上層のノードをかかる生活習慣病とすると，BN のときと同様に条件付き確率を計算することで問診データからとある生活習慣病にかかる確率などを算出できる（BN のときと同様に一般的なマルコフ確率場上での条件付き確率の計算は近似計算を必要とし容易でないが，2 部グラフ型のこのモデルの場合は条件付き独立性[2]より簡単に計算できる）．

BN との違いは，こちらは明確な親子の因果関係をモデルとして定めていないという点にある．普通は生活習慣が原因で生活習慣病になるので，一見明らかな因果関係があるようにも思えるが，場合によっては病気であることが原因で生活習慣に影響を及ぼすこともあるかもしれない（本当にそんな場合があるかどうかは定かではないが）．明確な親子関係を課していないため，そのような微妙な（因果関係の判断が難しい）場合も考慮に入れてモデル化できるというのがマルコフ確率場の一つの利点であり，そのため，現在のさまざまなパターン認識システム（画像理解システムなど）の設計の中核をマルコフ確率場に基づくモデルが担っている．

ここでは例として 2 部グラフ型のマルコフ確率場を紹介してきたが，一般のネットワーク上で定義することが可能であり，問題に合わせてさまざまな形のマルコフ確率場が存在する．マルコフ確率場に関する基礎数理と最近の動向については文献 [7, 8] などが非常に詳しい．

3.2　マルコフ確率場に対する統計的機械学習

確率モデル (14) の中のポテンシャル関数の値は BN の場合と同様に，観測データから統計的機械学習の方法を用いて決定される．そこでも最尤法が主力となる．しかしながら，マルコフ確率場に対する最尤推定は観測データが完全データであっても計算量爆発の問題を孕んでおり一般には困難である．BN の場合は単なる標本条件付き確率を観測データから構成するだけですんでいたが，マルコフ確率場の場合にはそう簡単にはいかない．そのため，マルコフ確率場に対する近似学習法の研究は古来から現在にわたるまで盛んに行われ続けており，今なお発展してきている分野である．

4.　おわりに

本稿では BN への導入のための基礎的な概念と数理について解説し，そして関連する確率モデルとしてマルコフ確率場を取り上げた．BN についての数理と応用についてさらに学びたい読者は文献 [2, 4, 5] などを参考にするとよいだろう．また，BN の数学的基礎付けに対しては文献 [6] が詳しい．BN 上での機械学習を含む，統計的機械学習全般に興味のある読者には参考書として文献 [7] をおすすめする．

謝辞　本稿執筆にあたり，東北大学大学院情報科学研究科の和泉勇治准教授ならびに浅利岳氏には多くの有益なコメントいただいたこと深く感謝する．

[2] 下層のノードの値を固定すると，上層のノードはそれぞれ独立な確率変数として扱える．

参考文献

[1] J. Pearl, Probabilistic Reasoning in Intelligent Systems, *Networks of Plausible Inference* (2nd ed.), San Francisco, CA: Morgan Kaufmann, 1988.
[2] 田中和之，ベイジアンネットワークの統計的推論の数理，コロナ社，2009.
[3] 田中和之，確率モデルによる画像処理技術入門，森北出版，2006.
[4] 繁桝算男，植野真臣，本村陽一，ベイジアンネットワーク概説，培風館，2006.
[5] 本村陽一，岩崎弘利，ベイジアンネットワーク技術 ユーザ・顧客のモデル化と不確実性推論，東京電機大学出版局，2006.
[6] 鈴木譲，ベイジアンネットワーク入門，培風館，2009.
[7] C. M. Bishop, Pattern Recognition and Machine Learning, Springer, 2006.
[8] M. J. Wainwright and M. I. Jordan, Graphical Models, Exponential Families, and Variational Inference, *Foundations and Trends in Machine Learning*, **1**, 1–2, 1–305, 2008.