

# ベイジアンネットワーク

岐阜大学 工学研究科  
博士後期課程 2年 志賀元紀

平成 17 年 1 月 27 日

## 1 はじめに

ベイジアンネットワークは、確率的な因果関係のモデル化や確率的な推論を目的として、様々に用いられている。ベイジアンネットワークを利用する際の問題点としては、大きく分けて 2 つある。一つ目は、標本が与えられたときに、どれだけ良い精度で、確率的な構造を推定できるか、つまり、ネットワーク構築 (または、学習) の問題である。2 つ目は、確率的な構造が既知のとき、推論のための周辺確率の計算効率に関する問題である。この文章では、後者の問題のみ述べる。ネットワーク構築の問題<sup>1</sup>を知りたい方は、他の文献を参考して頂きたい。

## 2 ベイジアンネットワーク

### 2.1 ベイジアンネットワークとは

ベイジアンネットワークとは、確率的な因果関係をモデル化する手法であるグラフィカルモデリングの一つである。ベイジアンネットワークは、因果関係の有無を有向グラフで表し、また、因果関係の強さを条件付確率で表す。ベイジアンネットワークは、非循環有向グラフ (directed acyclic graph, しばしば、DAG と略される)、つまり、矢印をたどったときに元のノードに戻ってこないグラフのみを扱う<sup>2</sup>。グラフィカルモデルの利点は、因果関係の有無を概観でき、詳細な確率分布とは切り離して考えることができることにありと考えられる (もちろん、それだけではない)。

ベイジアンネットワークの具体例として、Burglar Alarm<sup>3</sup>を図 1 に示す。泥棒 (Burglar) が侵入すると、警報装置 (Alarm) が作動すること、警報装置は、地震 (Earthquake) によっても作動してしまうこと、また、地震が起きるとラジオ (Radio) で伝えられることをモデル化したものである。

グラフの構造より、全ての確率変数の結合確率は、

$$P(B, A, E, R) = P(B)P(A|B, E)P(E)P(R|E) \quad (1)$$

と表現できる。ベイジアンネットワークを利用した推論は、その結合確率を周辺化した周辺事後確率により行われる。具体例として、まず、警報が作動したとき、泥棒が侵入した確率を推論する場合を考え

<sup>1</sup>現状としては、人間の主観的な感覚によって、ネットワークを構築することが多いように思われる。

<sup>2</sup>循環グラフにおいて、確率分布を定義できるのだろうか？

<sup>3</sup>ベイジアンネットワークの解説書を見ると、必ずと言って良いほど載っている超有名な例である。

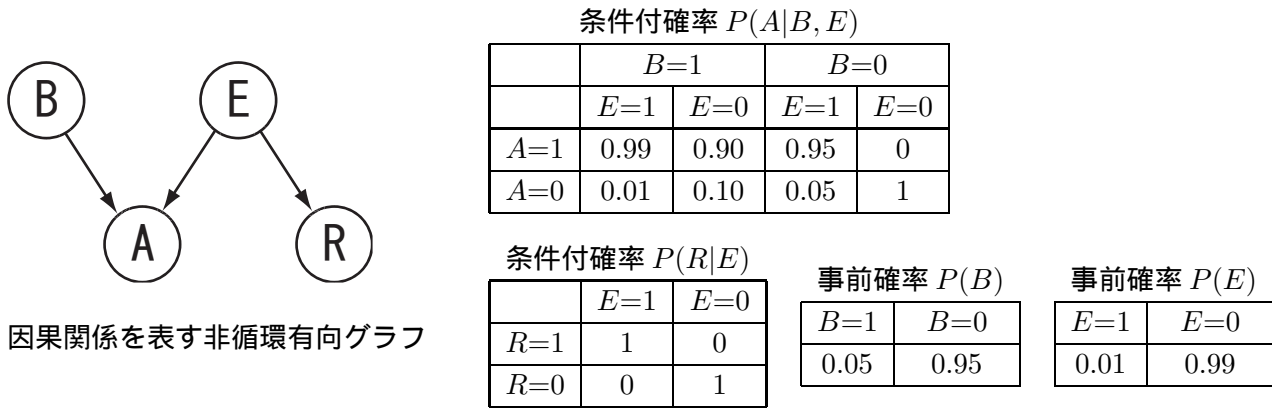


図 1: ベイジアンネットワークの例 (Burglar Alarm)

る．この確率は，式 (1) を利用することにより，

$$\begin{aligned}
 P(B = 1|A = 1) &= \frac{P(A = 1, B = 1)}{P(A = 1)} \\
 &= \frac{\sum_E \sum_R P(B = 1)P(A = 1|B = 1, E)P(E)P(R|E)}{\sum_B \sum_E \sum_R P(B)P(A = 1|B, E)P(E)P(R|E)} \\
 &= \frac{P(B = 1) \sum_E P(A = 1|B = 1, E)P(E)}{\sum_B P(B) \sum_E P(A = 1|B, E)P(E)} \\
 &=
 \end{aligned}$$

と計算できる．さらに，地震のニュースをラジオで聞いた，つまり，新たに確率変数が観測された<sup>4</sup>とき，泥棒が侵入した確率は，

$$\begin{aligned}
 P(B = 1|A = 1, R = 1) &= \frac{P(B = 1, A = 1, R = 1)}{P(A = 1, R = 1)} \\
 &= \frac{\sum_E P(B = 1)P(A = 1|B = 1, E)P(E)P(R = 1|E)}{\sum_B \sum_E P(B)P(A = 1|B, E)P(E)P(R = 1|E)} \\
 &= \frac{P(B = 1) \sum_E P(A = 1|B = 1, E)P(E)P(R = 1|E)}{\sum_B P(B) \sum_E P(A = 1|B, E)P(E)P(R = 1|E)} \\
 &=
 \end{aligned}$$

となる．その結果，泥棒が侵入した周辺事後確率が激減する．一見すると，奇妙に思われるかもしれないが，地震の発生が，泥棒の侵入に対する推論に大きく影響を与えていることがわかる．つまり，推論において，泥棒の侵入と地震の発生は，独立関係でないことがわかる．

## 2.2 条件付独立，d-separation

ベイジアンネットワークの構造を考えるとときに，条件付独立という概念がしばしば現れる．確率変数  $A$  が観測されたとき，確率変数  $B$  と  $C$  が条件付独立とは，

$$P(B, C|A = a) = P(B|A = a)P(C|A = a)$$

が成り立つことである．この定義式が成り立たないときは，推論において，2 変数が独立でないと考えられる．

<sup>4</sup>しばしば，evidence と呼ばれる

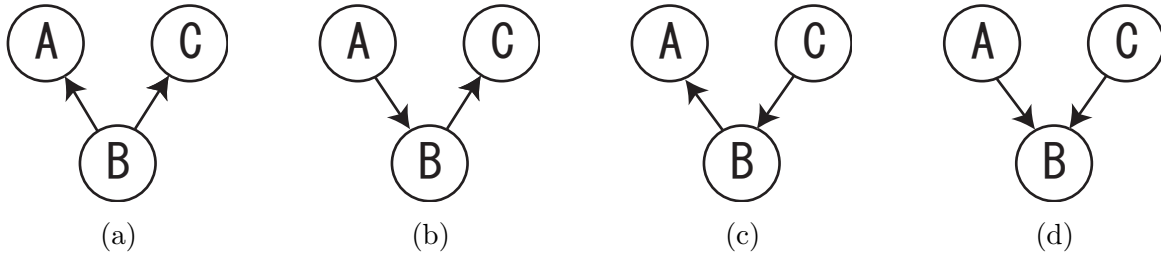


図 2: 条件付独立となる例はどれか?

図 2(a) ~ (d) に示した例が条件付独立であるかどうかを調べてみよう．まず，図 2(a) について考える．その結合確率  $P(A, B, C)$  は，

$$P(A, B, C) = P(B)P(A|B)P(C|B)$$

となることより，確率変数  $B$  が観測されたとき，確率変数  $A$  と  $C$  が独立になることは，自明である．したがって，図 2(a) の例は，条件付独立である．次に，図 2(b) について考える．その結合確率は，ベイズの定理を用いることにより，

$$P(A, B, C) = P(A)P(B|A)P(C|B) = P(A, B)P(C|B) = P(B)P(A|B)P(C|B)$$

となり，図 2(a) の結合確率と同じになり，条件付独立になる．また，図 2(c) の結合確率は，

$$P(A, B, C) = P(C)P(B|C)P(A|B) = P(A|B)P(B, C) = P(B)P(A|B)P(C|B)$$

となり，この例も図 2(c) の結合確率と同じになり，条件付独立になる．最後に，図 2(d) について考える．その結合確率  $P(A, B, C)$  は，

$$P(A, B, C) = P(A)P(B|A)P(C)P(B|C) \neq P(B)P(A|B)P(C|B)$$

となり，条件付独立とはならない．したがって，これらの例の中で，条件付独立となるのは，(a) ~ (c) であり，さらに，これらは，実は，確率的な構造を持っていることもいえる．

条件付独立であるか条件付独立でないかを判断するには，観測された確率変数に 2 つとも矢印が向けられているかを調べればよい．観測された確率変数が，2 つとも矢印が向けられているとき，この確率変数は，collider<sup>5</sup>であるという．さらに，条件付独立に関して，d-separation<sup>6</sup>という概念がある．d-separation は，d-connection でないときをいう．d-connection とは，次のように定義される．グラフ  $G$  が，確率変数の集合  $X_s, Y_s, Z_s$  からなるとしよう． $Z$  によって， $X_s$  と  $Y_s$  が d-connection であるとは， $X_s$  と  $Y_s$  の間の経路  $U$  を考えたとき， $U$  上の確率変数が全て collider であるとき，または， $Z_s$  またはその子が collider であるとき，または，経路  $U$  上の確率変数で collider であるものが  $Z_s$  に含まれているときをいう．

<sup>5</sup>和訳すると，粒子加速器．

<sup>6</sup>d は，dependency の略と考えられる．

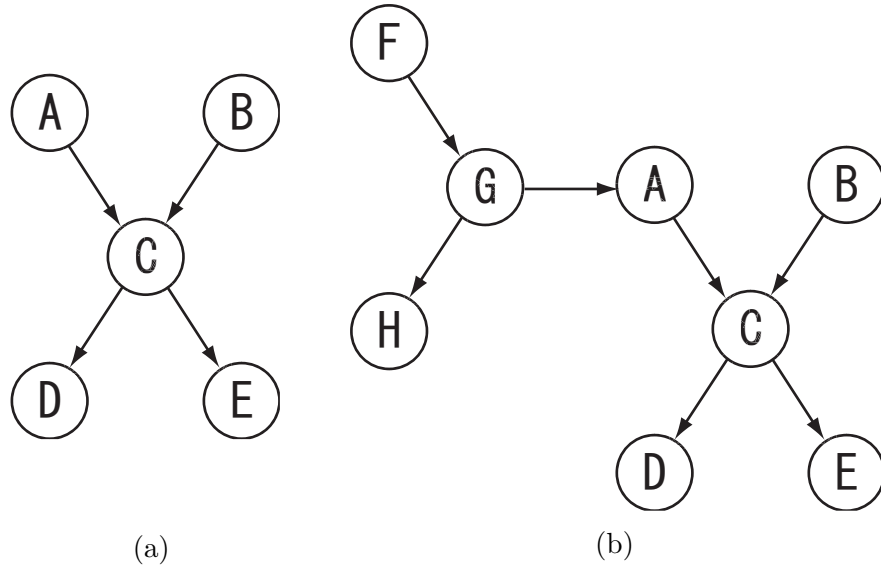


図 3: 非循環有向グラフ (ループなし)

### 2.3 周辺事後確率の計算例

周辺事後確率を，結合確率をそのまま総和を計算すると<sup>7</sup>，確率変数（ノード）の数の増加とともに，計算コストは，指数オーダーで増加する．そこで，構造を考慮する<sup>8</sup>ことにより，できるだけ計算を効率化することにより，多項式のオーダーまで抑えることを考える．この計算コストの効率化は，後に述べる Belief Propagation の考え方につながるので，例を具体的に示す．

まず，図 3(a) の有向グラフを考える．図 3(a) において，その全ての確率変数の結合確率は，

$$P(A, B, C, D, E) = P(A)P(B)P(C|A, B)P(D|C)P(E|C)$$

となる．まず，周辺事後確率  $P(C = c)$  を計算すると，

$$\begin{aligned} P(C = c) &= \sum_A \sum_B \sum_D \sum_E P(A)P(B)P(C|A, B)P(D|C)P(E|C) \\ &= \left\{ \sum_A \sum_B \underbrace{P(A)}_{\rho_{A,C}(a)} \underbrace{P(B)}_{\rho_{B,C}(b)} P(C = c|A, B) \right\} \underbrace{\left\{ \sum_D P(D|C = c) \right\}}_{\lambda_{D,C}(c)} \underbrace{\left\{ \sum_E P(E|C = c) \right\}}_{\lambda_{E,C}(c)} \end{aligned}$$

となる．ここで， $\rho_{A,C}(a)$  を親  $A$  から子  $C$  へのメッセージ， $\lambda_{E,C}(c)$  を子  $E$  から親  $C$  へのメッセージと考えれば，周辺事後確率  $P(C = c)$  の計算は，確率変数  $C$  に入ってくる全てのメッセージの積により計算できる．ただし，子からのメッセージが観測された確率変数かに関するものでないとき，そのメッセージは，推論に影響を及ぼさない<sup>9</sup>．観測された確率変数は，推論に影響を及ぼすので，evidence（証拠）と呼ばれる．

<sup>7</sup>例えば， $\sum_{X_1} \cdots \sum_{X_M} P(X_1, \dots, X_M)$  のように計算する

<sup>8</sup>括弧をできるだけ，内側にいれるだけ．

<sup>9</sup>総和したとき，1 になってしまう．

また，確率変数が  $E = e$  と観測され，evidence がある場合の周辺事後確率  $P(C = c|E = e)$  を計算すると，

$$\begin{aligned}
P(C = c|E = e) &= \frac{P(C = c, E = e)}{P(E = e)} \\
&= \frac{\{\sum_A \sum_B P(A)P(B)P(C = c|A, B)\} \{\sum_D P(D|C = c)\} P(E = e|C = c)}{\sum_C \{\sum_A \sum_B P(A)P(B)P(C|A, B)\} \{\sum_D P(D|C)\} P(E = e|C)} \\
&= \left\{ \sum_A \sum_B \underbrace{P(A)}_{\rho_{A,C}(a)} \underbrace{P(B)}_{\rho_{B,C}(b)} P(C = c|A, B) \right\} \underbrace{\left\{ \sum_D P(D|C = c) \right\}}_{\lambda_{D,C}(c)} \underbrace{\frac{P(E = e|C = c)}{Z}}_{\lambda_{E,C}(c)}
\end{aligned}$$

となる．ただし， $Z$  は， $\sum_c P(C = c|E = e) = 1$  となるための規格化定数である．この例のように，evidence がある場合には，規格化定数が含まれ，計算が複雑になる．しかしながら，周辺事後確率の計算は，確率変数  $C$  に入ってくる全てのメッセージの積により計算できることには変わりがない．

次に，図 3(b) のように，図 3(a) を拡張した有向グラフを考える．周辺確率  $P(C = c)$  を計算すると，

$$\begin{aligned}
P(C = c) &= \sum_{A,B,D,E,F,G,H} P(A, B, C = c, D, E, F, G, H) \\
&= \left\{ \sum_A \sum_B \underbrace{\left\{ \sum_G \underbrace{\left\{ \sum_F \sum_H \underbrace{P(F)}_{\rho_{F,G}(f)} \underbrace{P(H)}_{\rho_{H,G}(h)} P(G|F, H) \right\}}_{\rho_{G,A}(g)} \right\}}_{\rho_{A,C}(a)} P(A|G) \right\} \underbrace{P(B)}_{\rho_{B,C}(b)} P(C = c|A, B) \right\} \\
&\quad \cdot \underbrace{\left\{ \sum_D P(D|C = c) \right\}}_{\lambda_{D,C}(c)} \underbrace{\left\{ \sum_E P(E|C = c) \right\}}_{\lambda_{E,C}(c)}
\end{aligned}$$

となる．この計算式は，ノード  $G$  に入ってくるメッセージの積を計算し，ノード  $A$  にメッセージを送る部分が，先ほどの例より余分に加わっていることがわかる．これらの例より，局所的に，周辺事後確率を計算し，それをメッセージとして伝播することにより，知りたい周辺事後確率を計算できることがわかる．

## 2.4 Belief Propagation

前節で述べた周辺事後確率の計算を，一般化した計算アルゴリズムである Belief Propagation (以下，BP) について述べる．なお，BP は，ループがないベイジアンネットワーク<sup>10</sup>にのみ，周辺事後確率を厳密に計算できる．ループがあるベイジアンネットワーク<sup>11</sup>に対し，BP を用いることについては，次節で述べる．

BP のアルゴリズムは，以下で述べるように，各時刻 (または，ステップ)  $t$  で，全てのノード間のメッセージを更新することになる．まず，時刻  $t$  における，ノード  $X$  からその子  $D_1$  へのメッセージ  $\rho_{XD_1}^{(t)}(x)$  の更新方法は，次の手順になる．

<sup>10</sup>singly connected graph と呼ばれる．

<sup>11</sup>multiply connected graph と呼ばれる．

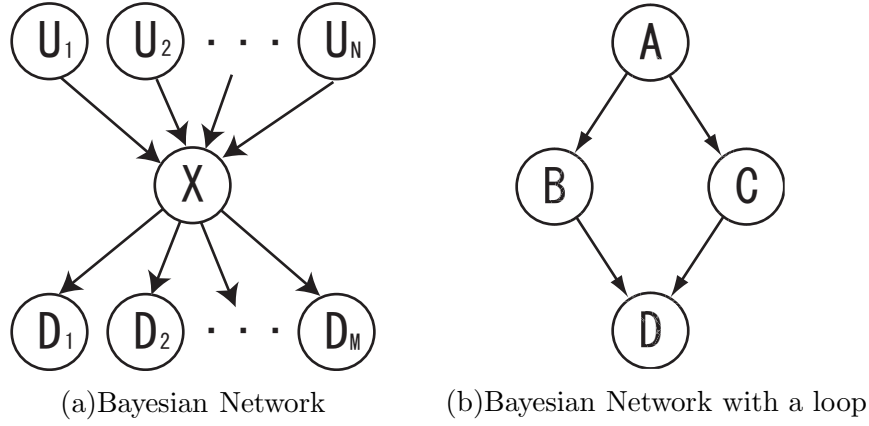


図 4: Belief propagation for Bayesian network

1. 親からのメッセージ  $\rho_{U_i X}^{(t)}(u_i), i = 1, \dots, N$  と条件付確率の積の全ての状態に関する総和

$$p_{in}^{(t)}(x) = \sum_{U_1, \dots, U_N} P(X = x | U_1 = u_1, \dots, U_N = u_N) \prod_{i=1}^N \rho_{U_i X}^{(t-1)}(u_i)$$

を計算する .

2.  $D_j$  以外の子からのメッセージ  $\lambda_{X D_i}^{(t-1)}(x), i = 1, \dots, j-1, j+1, \dots, M$  の積

$$p_{out}^{(t)}(x) = \prod_{i \neq j}^M \lambda_{D_i X}^{(t-1)}(x)$$

を計算する

3. これらの積  $\rho_{X D_j}^{(t)}(x) = p_{in}^{(t)}(x) p_{out}^{(t)}(x)$  により, ノード  $X$  から子  $D_1$  へのメッセージ  $\rho_{X D_1}^{(t)}(x)$  が計算できる .

また, 時刻  $t$  における, ノード  $X$  からその親  $U_j$  へのメッセージ  $\lambda_{X U_j}^{(t)}(x)$  の更新は, 次の手順になる .

1.  $U_j$  以外の親からのメッセージ  $\rho_{U_i X}^{(t-1)}(u_i), i = 1, \dots, j-1, j+1, \dots, N$  と条件付確率の積の全ての状態に関する総和

$$p_{in}^{(t)}(x, u_j) = \sum_{U_i, i \neq j} P(X = x | U_1 = u_1, \dots, U_N = u_N) \prod_{i \neq j}^N \rho_{U_i X}^{(t-1)}(u_i)$$

を計算する .

2. 全ての子からのメッセージ  $\lambda_{X D_i}^{(t-1)}(x), i = 1, \dots, M$  の積

$$p_{out}^{(t)}(x) = \prod_{i=1}^M \lambda_{X D_i}^{(t-1)}(x)$$

を計算する

3. これらの積をとり，さらにノード  $X$  の状態による総和

$$\lambda_{XD_j}^{(t)}(u_j) = \sum_X p_{out}^{(t)}(x) p_{in}^{(t)}(x, u_j)$$

により，ノード  $X$  からその親  $U_j$  へのメッセージ  $\lambda_{XD_j}^{(t)}(x)$  が計算できる．

この更新手順は，初期段階において，ノードに入ってくるメッセージが計算できない．そこで，適当に初期値を設定する必要がある．このようにしても，結局，真の周辺事後確率に収束する<sup>12</sup>ことを示すことができる．

最後に，ノード  $X$  における事後確率の計算手順を述べる．

1. 親からのメッセージ  $\rho_{U_i X}^{(t)}(u_i), i = 1, \dots, N$  と条件付確率の積の総和

$$p_{in}^{(t)}(x, u_j) = \sum_{U_1, \dots, U_N} P(X = x | U_1 = u_1, \dots, U_N = u_N) \prod_{i=1}^N \rho_{U_i X}^{(t)}(u_i)$$

を計算する．

2. 全ての子からのメッセージ  $\lambda_{XD_i}^{(t)}(x), i = 1, \dots, N$  の積

$$p_{out}^{(t)}(x) = \prod_{i=1}^M \lambda_{XD_i}^{(t)}(x)$$

を計算する

3. これらの積  $P^{(t)}(X = x) = p_{out}^{(t)}(x) p_{in}^{(t)}(x)$  により，ノード  $X$  における周辺事後確率が計算できる．時刻  $t$  が大きくなれば， $P^{(t)}(X = x)$  は真の周辺事後分布に収束する．繰り返し述べるが，本節で述べた BP は，ループがないベイジアンネットワークにのみ，周辺事後確率を厳密に計算できる．

## 2.5 Loopy Belief Propagation

図 4(b) のようにループがあるベイジアンネットワークを考える．このとき，ある時刻  $t$  におけるメッセージ  $\rho_{B,D}^{(t)}(b)$  の更新は，

$$\begin{aligned} \rho_{B,D}^{(t)}(b) &= \sum_A \rho_{A,B}^{(t-1)}(a) P(B|A) \\ &= \sum_A \left\{ P(A) \lambda_{C,A}^{(t-2)}(a) \right\} P(B|A) \\ &= \sum_A \left\{ P(A) \left\{ \sum_C P(C|A) \lambda_{D,C}^{(t-3)}(c) \right\} \right\} P(B|A) \\ &= \sum_A \left\{ P(A) \left\{ \sum_C P(C|A) \left\{ \sum_D \rho_{B,D}^{(t-4)}(b) P(D|B=b) \right\} \right\} \right\} P(B|A) \end{aligned}$$

となり，過去に同一のノードで送られたメッセージに依存する．そのため，周辺事後確率の計算において，同じ計算が重複することになり，これが理由で，周辺事後確率が正確に計算できない．しかしながら，一応，周辺事後確率の近似値を与えることができる．そこで，適当な回数だけ BP を実行して，そのときに，得られた周辺確率を，周辺確率の近似値とする計算アルゴリズムを Loopy Belief Propagation (以下，Loopy BP) という．

<sup>12</sup>このことを不思議に思った人は，Mackay の教科書を参考にするとよい．

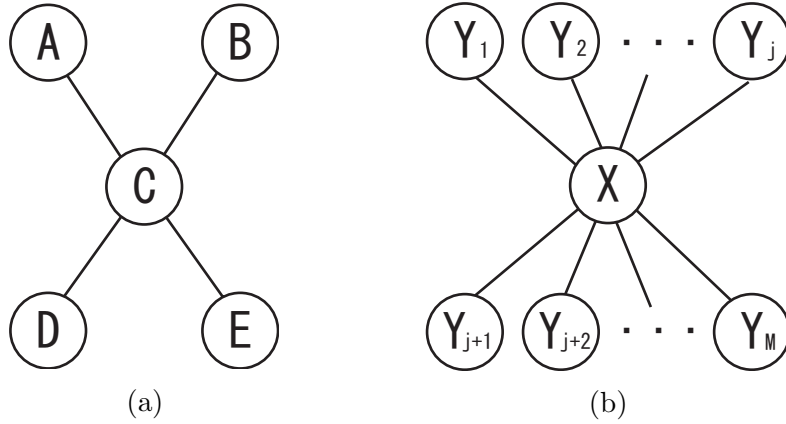


図 5: Markov network

### 3 マルコフネットワーク

#### 3.1 マルコフネットワークとは

確率的な構造を無向グラフ (undirected graph) で表現するマルコフネットワークについて述べる．無向グラフを用いたグラフィカルモデルには，様々な呼び方があるようだが，ここでは，論文 [2] に従う呼び方を使用する．マルコフネットワークは，図 5(a) のような無向グラフで表現され，グラフ中のノード  $A, B$  の結合の強さは，非負の値のポテンシャル関数  $\Psi(A, B)$  で表される．ポテンシャル関数は，2 変数関数であり，変数の順序に依存しない，つまり， $\Psi(A, B) = \Psi(B, A)$  である．図 5(a) のマルコフネットワークにおける全ての確率変数の結合確率は，

$$P(A, B, C, D, E) = \frac{1}{Z} \Psi(A, C) \Psi(B, C) \Psi(C, D) \Psi(C, E)$$

と表すことができる．ただし， $Z$  は規格化定数である．変数  $D$  の周辺確率の計算は，

$$P(C = c | E = e) = \frac{1}{Z} \underbrace{\sum_A \Psi(A, C = c)}_{\pi_{AC}(c)} \underbrace{\sum_B \Psi(B, C = c)}_{\pi_{BC}(c)} \underbrace{\sum_D \Psi(C, D)}_{\pi_{C=c, D}(c)} \underbrace{\Psi(C = c, E = e)}_{\pi_{EC}(c)}$$

となる．ただし， $Z$  は規格化定数である．マルコフネットワークにおいても，ノード間のメッセージ (例えば， $\pi_{AC}(c)$ ) の伝播により，知りたい周辺事後確率が計算できる．マルコフネットワークの場合，親子関係がないため，ベイジアンネットワークよりも計算が単純になる．

#### 3.2 マルコフネットワークにおける Belief Propagation

図 5(b) のようなマルコフネットワークにおける Belief Propagation について述べる．マルコフネットワークには，ノード間の結合に向きがないため，その BP は，ベイジアンネットワークにおける BP より簡単なアルゴリズムになる．まず，時刻  $t$  における，ノード  $X$  からその子  $D_j$  へのメッセージ  $\pi_{XD_j}(x)$  の更新方法は，次の手順になる．



1. ノード  $Y_j$  以外の全てのノード  $Y_i, i \neq j$  からノード  $X$  へのメッセージ  $\lambda_{Y_i X}^{(t-1)}(x)$  の積

$$p_{in}^{(t)}(x) = \prod_{i \neq j} \pi_{Y_i X}^{(t-1)}(x)$$

を計算する .

2. これとノード  $Y_j$  と  $X$  のポテンシャル関数の積の , ノード  $X$  の状態による総和 , つまり ,

$$\pi_{XY_j}^{(t)}(y_j) = \sum_X \Psi(Y_j = y_j, X = x) p_{in}^{(t)}(x)$$

により , メッセージを更新できる .

また , ノード  $X$  における事後確率の計算は , 全てのノード  $Y_i, i = 1, \dots, M$  からノード  $X$  へのメッセージ  $\lambda_{Y_i X}^{(t)}(x)$  の積を計算し , その規格化をする , つまり ,

$$P^{(t)}(X = x) = \frac{1}{Z} \prod_{i=1}^M \pi_{Y_i X}(x)$$

により , 求めることができる . マルコフネットワークの BP は , 親子関係がないため , ベイジアンネットワークよりも計算が単純になるだけでなく , メッセージの伝播を行列演算を用いて表現できる [2] ことでも優れている .

## 4 ベイジアンネットワークからマルコフネットワークへの変換

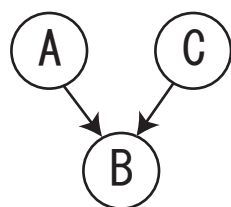
最後に , 任意のベイジアンネットワークは , マルコフネットワークに変換方法を示す . この変換により , 任意のベイジアンネットワークは , マルコフネットワークに変換できる .

次のページの図 6 の具体例を用いて説明する . collider (複数の親を持つノード)  $C$  に対して , 共通の親がクラスター化されたノード  $C'$  を新たに作る . ノード  $C'$  の状態は , 親ノード  $A, B$  の状態の直積  $C' = (A, B)$  で表す . そこで ,  $C'$  の状態を ,  $(A, B) = (0, 0), (1, 0), (0, 1), (1, 1)$  の順に 0, 1, 2, 3 と番号を付ける . そして , その子としてノード  $C$  を付け加える . そして , 他の部分のネットワークの変更せず , リンクの向きを無くす .

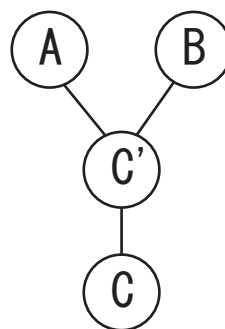
ノード  $C, C'$  に関するもの以外のポテンシャル関数は , 条件付確率と置き換える . ただし , 根となる部分は事前確率と条件付確率の積と置き換える . 合成ノード  $Z$  とその親ノードの間のポテンシャルは , 親の一致するとき , 1 として , それ以外ときは , ゼロとおく . これらの規則により , ポテンシャル関数は , 図 6 中の表のようになる .

## 参考文献

- [1] F. V. Jensen, "Bayesian Networks and Decision Graphs," Springer-Verlag, 2001.  
 [2] Y. Weiss, "Correctness of local probability propagation in graphical models with loops," Neural Computation Vol.12 pp.1-41, 2000.



(a) Bayesian Network



(b) 変換後の Markov Network

ポテンシャル関数  $\Psi(A, C')$

	$C'=0$	$C'=1$	$C'=2$	$C'=3$
$A=1$	0	1	0	1
$A=0$	1	0	1	0

ポテンシャル関数  $\Psi(B, C')$

	$C'=0$	$C'=1$	$C'=2$	$C'=3$
$B=1$	0	0	1	1
$B=0$	1	1	0	0

ポテンシャル関数  $\Psi(C', C)$

	$C'=0$	$C'=1$	$C'=2$	$C'=3$
$C=1$	$P(A=0)P(B=0) \times P(C=1 A=0, B=0)$	$P(A=1)P(B=0) \times P(C=1 A=1, B=0)$	$P(A=0)P(B=1) \times P(C=1 A=0, B=1)$	$P(A=1)P(B=1) \times P(C=1 A=1, B=1)$
$C=0$	$P(A=0)P(B=0) \times P(C=0 A=0, B=0)$	$P(A=1)P(B=0) \times P(C=0 A=1, B=0)$	$P(A=0)P(B=1) \times P(C=0 A=0, B=1)$	$P(A=1)P(B=1) \times P(C=0 A=1, B=1)$

図 6: Bayesian Network から Markov Network への変換