ベイジアンネットワーク概説

1.4 ノードの確率の計算

1.4.1 確率変数の分布の表記

1.4.2 トポロジー順序

発表者　加藤 友宏
発表日　2006/10/20

1.4 ノードの確率の計算

- 各ノードの不確定性を確率によって評価することが目的
  - 条件付き確率によって評価
  - どこからも影響されないノード（ルート）は確率を直接評価
  - 実際には、いくつかのノードは不確定性を失い、確定した値を持っている。

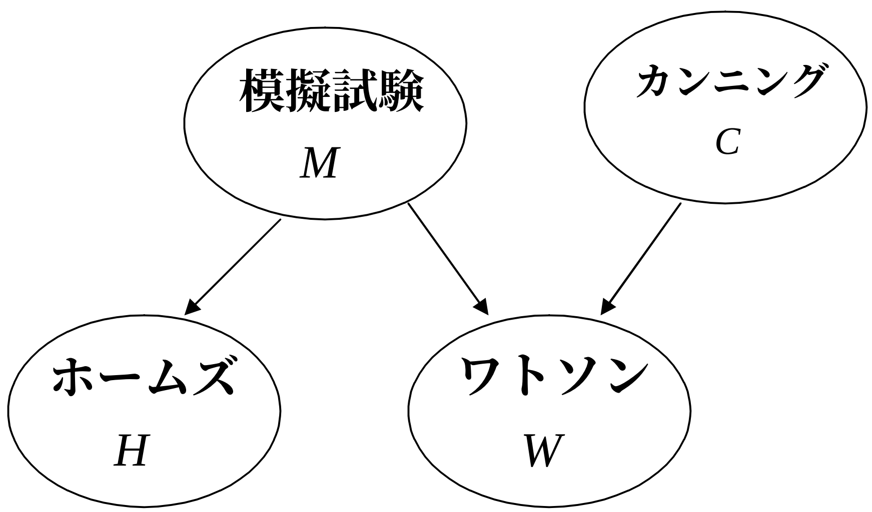
1.4 ノードの確率の計算

- 例題
  - ホームズ君とワトソン君が非常に難しい難問を解いています。このように難しい問題でも、ホームズ君は、5割解けます。また、類似問題を模擬試験ですでに解いている場合は、ホームズ君は、9割解けます。ワトソン君は、それほど数学が得意ではなく、通常解ける割合が1割、模擬試験で既出でも4割の正答率です。ワトソン君は、かつてカンニングをしたことがあります。

    ここで問題です。ワトソン君が問題を解きました。ワトソン君がカンニングをした可能性はどの程度でしょうか？

1.4 ノードの確率の計算

- 例題の不確実な状況をグラフにすると...



M : 模擬試験を受ける
C : カンニングをする
H : ホームズが正解
W : ワトソンが正解

それぞれを否定

$\overline{M}$: 模擬試験を受けない
$\overline{C}$: カンニングをしない
$\overline{H}$: ホームズが不正解
$\overline{W}$: ワトソンが不正解

各ノードは、その事象が起きたか起きなかったかの2値とする

1.4 ノードの確率の計算

- 他のノードに影響されないノードの事前確率を評価
  - （1）ホームズとワトソンが模擬試験を受ける確率P(M)=0.5（ホームズとワトソンが共に受けるか否か）
  - （2）カンニングをする確率をP(C)=0.1
- 各ノードは2値としたので
  - （3）模擬試験を受けない確率 $P(\overline{M})$は0.5
  - （4）カンニングをしない確率 $P(\overline{C})$ は0.9

1.4 ノードの確率の計算

- MとCの各状況におけるワトソンの正答確率の評価
  - （5）「模擬試験を受ける、かつ、カンニングをする」場合、ホームズの解答を用いると仮定すると、ワトソンが正解する確率は、P(W|M,C)=0.9
  - （6）「模擬試験を受けない、かつ、カンニングをする」場合、ホームズの解答を用いるとすると、ワトソンが正解する確率は、$P(W \mid \overline{M}, C) = 0.5$
  - （7）「模擬試験を受ける、かつ、カンニングをしない」場合、ワトソンが正解する確率は、$P(W \mid M, \overline{C}) = 0.4$
  - （8）「模擬試験を受けない、かつ、カンニングをしない」場合、ワトソンが正解する確率は、$P(W \mid \overline{M}, \overline{C}) = 0.1$

1.4 ノードの確率の計算

- ワトソンが正答したときカンニングをした確率

$$P(C|W) = P(C,M|W) + P(C,\overline{M}|W)$$

- ここでCとWは互いに独立であると仮定

$$P(C,M|W)$$

$$= \frac{P(W|M,C)P(M)P(C)}{\begin{Bmatrix} P(W|M,C)P(M)P(C) + P(W|\overline{M},C)P(\overline{M})P(C) \\ + P(W|M,\overline{C})P(M)P(\overline{C}) + P(W|\overline{M},\overline{C})P(\overline{M})P(\overline{C}) \end{Bmatrix}}$$

$$= \frac{0.9 \times 0.5 \times 0.1}{\begin{Bmatrix} (0.9 \times 0.5 \times 0.1) + (0.5 \times 0.5 \times 0.1) \\ + (0.4 \times 0.5 \times 0.9) + (0.1 \times 0.5 \times 0.9) \end{Bmatrix}}$$

$$= \frac{0.045}{0.295} = \frac{5}{59}$$

1.4 ノードの確率の計算

同様にして、 $P(C,\overline{M} \mid W) = \dfrac{0.025}{0.295} = \dfrac{5}{59}$

$$P(C \mid W) = \frac{9}{59} + \frac{5}{59} = \frac{14}{59}$$

ワトソンが正答したとき、

カンニングをしたた確率は $\dfrac{14}{59}(\quad 0.237)$

1.4.1 確率変数の分布の表記

- P (X) は、確率変数Xの分布を示すものと定義

Xがx_1, x_2, x_3の値をとるとき、

$P(X)$は、$P(x_1), P(x_2), P(x_3)$を成分とするベクトルを示す

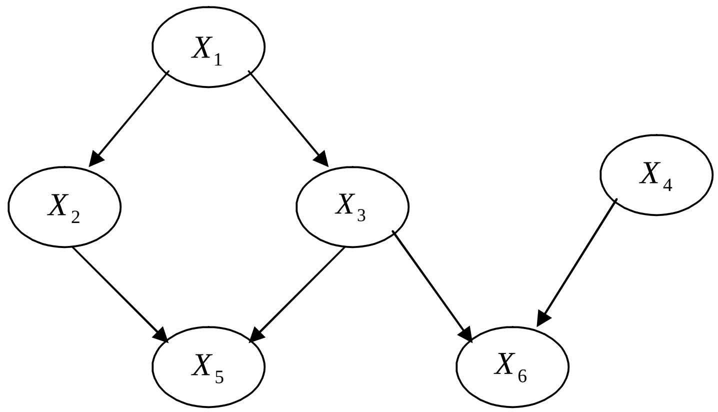
Xがx_1, x_2, x_3の値をとり、Yがy_1, y_2の値をとるとき、

$$P(X \mid Y) = \begin{bmatrix} P(y_1 \mid x_1) & P(y_2 \mid x_1) \\ P(y_1 \mid x_2) & P(y_2 \mid x_2) \\ P(y_1 \mid x_3) & P(y_2 \mid x_3) \end{bmatrix}$$

1.4.2 トポロジー順序

- ベイジアンネットワークモデルの目的

  → 各ノードの数値の確率分布を求めること



目的は、$P(X_1, X_2, X_3, X_4, X_5, X_6)$を評価すること

同時分布から、

周辺分布$P(X_1), P(X_2), P(X_3), P(X_4), P(X_5), P(X_6)$が得られる

1.4.2 トポロジー順序

- トポロジー順序
  - 子孫が、その先祖の前にくることがないようなノードの順序
  - 1通りではなくいくつかの可能な順序が存在

順序$X_1, X_2, X_3, X_4, X_5, X_6$が一つのトポロジー順序

$$
\begin{aligned}
&P(X_1, X_2, X_3, X_4, X_5, X_6) \\
&= P(X_1)P(X_1 \mid X_2)P(X_3 \mid X_1, X_2)P(X_4 \mid X_1, X_2, X_3) \\
&\quad \times P(X_5 \mid X_1, X_2, X_3, X_4)P(X_6 \mid X_1, X_2, X_3, X_4, X_5)
\end{aligned}
$$

条件付き独立性を考えると、

$$
= P(X_1)P(X_2 \mid X_1)P(X_3 \mid X_1)P(X_4)P(X_5 \mid X_2, X_3)P(X_6 \mid X_3, X_4)
$$

1.4.2 トポロジー順序

- 一般的に、ノードの同時確率は、

ノードの全体を$U=[X_1, X_2, \ldots, X_n]$

X_j**の親ノードを**$pa(X_j)$

$$P(u)=\prod_j P(X_j \mid pa(X_j))$$