第7章ベイジアンネットワーク(1)

10NM706F　江口晃

7.1ベイジアンネットワークとは？

- 事象間の影響関係をグラフィカルなモデルとしてわかりやすく記述する目的がある。
- このとき作成されたモデルをベイジアンモデルという。

7.1.1いろいろな分析

ベイジアンネットワークによる分析

⇒変数間の相互関係を知りたいときに利用する

①見かけ上の関係などの要因を全て取り除く。

②両変数の影響関係を方向つきで分析する。

③想定する状況に関連する全ての変数に関して分析する。

④得られたデータから探索的にモデルを構築する。

⑤連続変数と離散変数の両方をモデルの中に組み込むこともできる。

7.1.2ベイジアンネットワークの具体例

「科学」「数学」「社会」「読解」「作文」の5つの能力と「性別」を利用してモデルの構築を試みた。

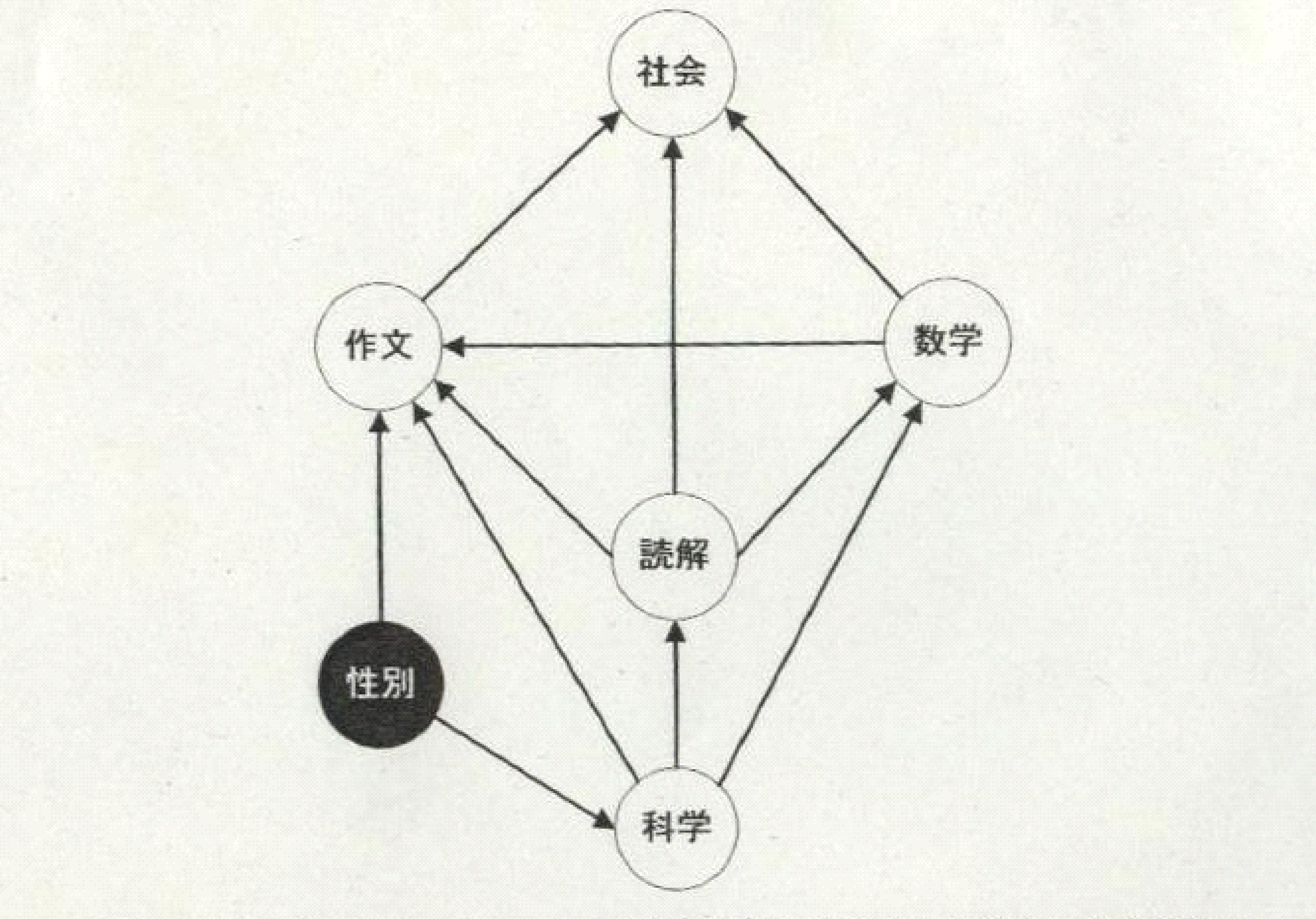

図 7.1　ベイジアンネットワークにより構築された学生の 5 能力に関する構造

7.2ベイズ統計学

- ベイジアンネットワークでは変数間の影響の強さを不確かさの度合いで率的に表現する。不確かさの度合いを「ベイズ的」という。

7.2.2ベイズの定理(1)

「Xであり、かつYである確率」

確率の乗法定理 1 $$p(X,Y)=p(Y \mid X)p(X)$$

「Yであり、かつXである確率」と言い換える

確率の乗法定理 2 $$p(X,Y)=p(Y,X)=p(X \mid Y)p(Y)$$

ベイズの定理

$$p(Y \mid X)p(X)=p(X \mid Y)p(Y)$$

ベイズの定理 1 $$p(Y \mid X)=\frac{p(X \mid Y)p(Y)}{p(X)}$$

7.2.2ベイズの定理(2)

ベイズの定理

$$p\,(原因\mid 結果) = \frac{p\,(結果\mid 原因)\,p\,(原因)}{p\,(結果)}$$

ベイズの定理とは「原因から結果へ」を計算することで、ある結果が得られたときにその原因となったのは何かという確率を求めることができる手法である。

7.2.3独立と条件付き独立

- 独立

  XとYを確率変数としたとき、同時確率が$p(X,Y)=p(X)p(Y)$のように表せた場合、XとYは互いに独立であるという。

- 条件付き独立

  X,Y,Zの3つの確率変数を仮定したとき、$p(X,Y|Z)=p(X|Z)p(Y|Z)$のようにZが所与の下で分解可能である場合、XとYはZを与えた下で条件付き独立であるという。

7.3.1グラフ理論

- グラフ理論の「グラフ」は複数の「点」とそれらを結ぶ「線」のみで構成された概念を意味している。
- ベイジアンネットワークではこのグラフの形でモデルを表現する。

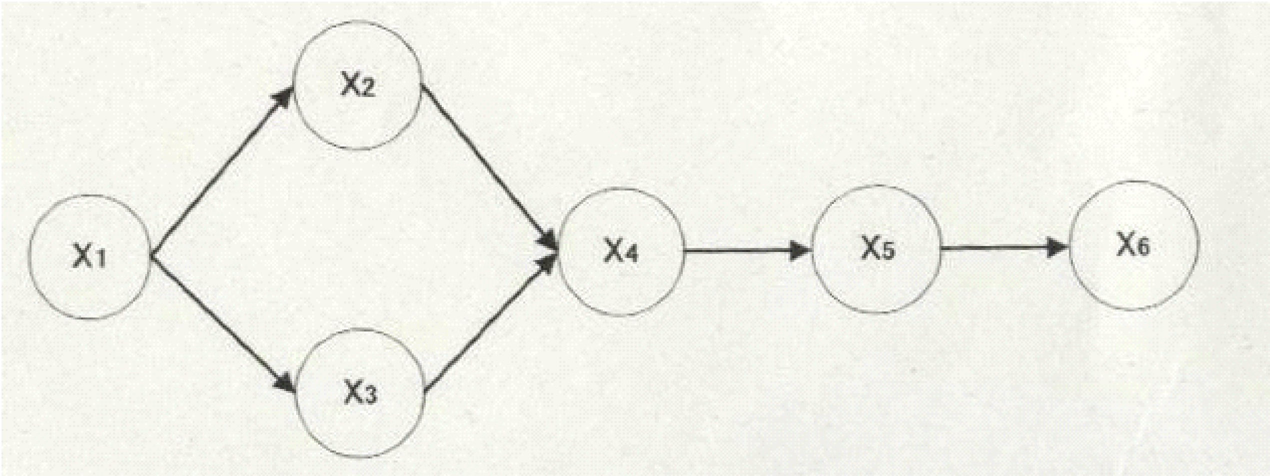

図 7.2 ベイジアンネットワークにおけるグラフ構造の例

7.3.2グラフの因数分解性(1)

- ベイジアネットワークモデルにおいて、条件付き確率は極めて重要である。

X,Y,Zの3つの確率変数を仮定する。

$$p(X, Y, Z) = p(Z \mid X, Y)p(X, Y)$$

7.3.2グラフの因数分解性(2)

p(X,Y)に対して乗法定理を用いると以下のように表現できる。

$$p(X,Y,Z) = p(Z \mid X,Y)p(Y \mid X)p(X)$$

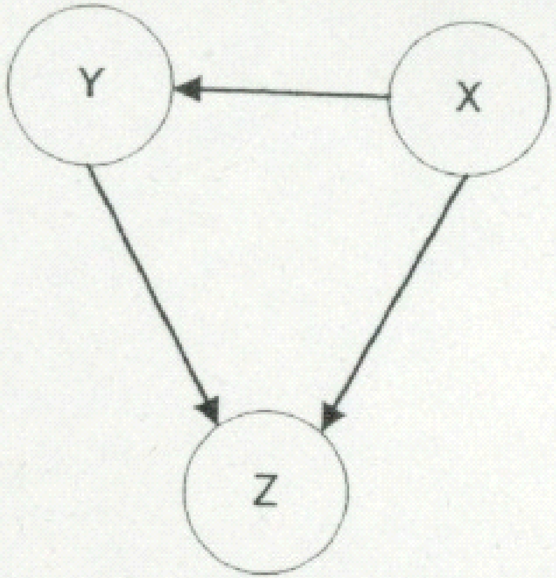

図 7.3　条件付き確率のグラフによる表現（完全連結の場合）

7.3.2グラフの因数分解性(3)

ここでノードXとノードYが独立であるとすると以下のようになる。

$$p(X,Y,Z) = p(Z \mid X,Y)p(Y)p(X)$$

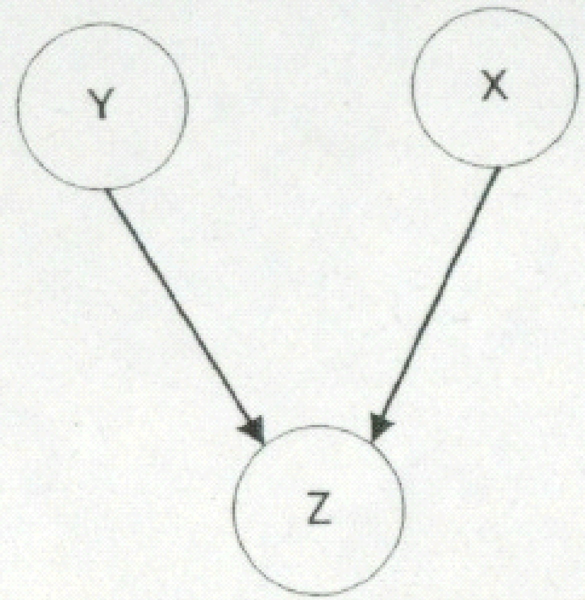

図 7.4　条件付き確率のグラフによる表現（X と Y が独立の場合）

7.3.2グラフの因数分解性(4)

ベイジアンネットワークモデルとは、すべてのノードの同時確率を条件付き確率と周辺分布の積の形に分解して構造化されたグラフである。

完全連結のグラフから条件付き独立の関係にあるノード間の有向性を切断して式を簡略化していく。

7.3.3ベイジアンネットワークモデルにおける確率計算(1)

- モデル構築

あるシステムSにはバルブB1とバルブB2がついている。この2つのバルブを監視するために1つのセンサーを取り付けた。

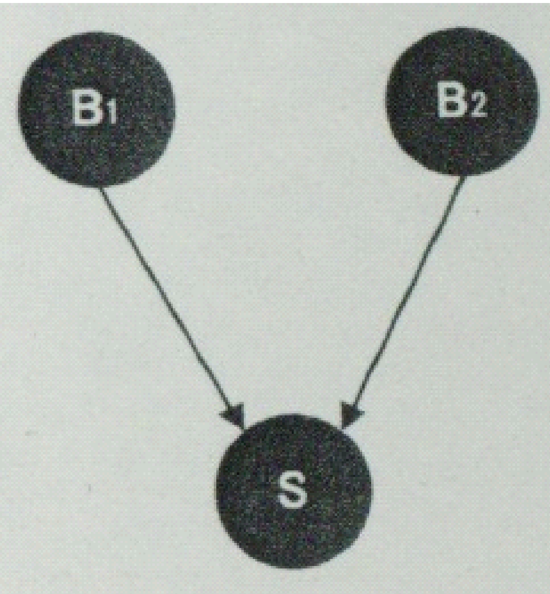

図 7.5　故障診断におけるベイジアンネットワークモデル

7.3.3ベイジアンネットワークモデルにおける確率計算(2)

バルブの故障する確率は以下の通り。

表 7.1　バルブに関する事前確率表

	0（正常）	1（故障）
$p(B_1)$	0.999	0.001
$p(B_2)$	0.995	0.005

7.3.3ベイジアンネットワークモデルにおける確率計算(3)

センサーの条件付き確率は以下の通り。

表 7.2　センサーに関する条件付き確率表

B_1	B_2	$p(S \mid B_1, B_2)$	
		0（反応なし）	1（反応あり）
0（正常）	0（正常）	0.999	0.001
0（正常）	1（故障）	0.750	0.250
1（故障）	0（正常）	0.100	0.900
1（故障）	1（故障）	0.020	0.980

7.3.3ベイジアンネットワークモデルにおける確率計算(4)

- センサーに反応が見られた場合にバルブB1に以上がある確率を求める。

ベイズの公式より

$$p(B_1 = 1 \mid S = 1) = \frac{p(S = 1 \mid B_1 = 1)p(B_1 = 1)}{p(S = 1)}$$

加法定理・乗法定理より値を代入して

$$p(B_1 = 1 \mid S = 1) = \frac{0.9004 \times 0.001}{0.003143155} = 0.2864638$$

7.4.1低出生体重児に関する研究

以下のデータを使用する。

表 7.3　低出生体重児に関する医療データ

	体重	喫煙	過敏	高血圧	人種
1	正常	喫煙なし	過敏あり	高血圧なし	黒人
2	正常	喫煙なし	過敏なし	高血圧なし	その他
3	正常	喫煙あり	過敏なし	高血圧なし	白人
⋮	⋮	⋮	⋮	⋮	⋮
188	低体重	喫煙なし	過敏なし	高血圧あり	黒人
189	低体重	喫煙あり	過敏なし	高血圧あり	白人

7.4.2モデルの評価

- モデルの評価

  複数のモデルの中から最良と思われる作業を決定する作業。

  ネットワークスコア

$$S(G) = p(G, d) = p(d \mid G)p(G)$$