ベイジアンネットワークのモデル構築手法の検討

鈴木 康之† 中村 誠弘†† 木村 昌臣††
芝浦工業大学大学院工学研究科† 芝浦工業大学工学部情報工学科††

1 はじめに

不確かな事象を確率的に推論する手法としてベイジアンネットワークと呼ばれる予測手法がある．ベイジアンネットワークは確率変数をノードで表し，ノード間の因果関係をアークで結んだ非循環有向グラフで構成されるネットワークモデルを持ち，親ノードと子ノードの各変数がとる具体的な値ごとに割り当てられた条件付き確率分布を条件付き確率表として与え事象の予測を行う．ベイジアンネットワークを利用して確率推論を行うためには事象間の因果関係を反映した適切なネットワーク構造を持ったモデルを構築する必要がある．しかし，確率変数の選択や因果関係があると思われる変数同士をアークで結びモデルを構築する作業はベイジアンネットワークを利用するユーザ自身がアドホックに行うことが多く，かつ明確な構築ルールがない．そのため，必ずしも良い推論結果が得られるモデルが構築されるとは限らない．そこで，本研究では適切なネットワーク構造を持ったベイジアンネットワークを構築するため，学習データから変数間の因果関係を得てモデルを構築する手法の提案を行う．因果関係を抽出する指標として条件付きエントロピーを利用する．

2 提案手法

条件付エントロピーとは事象B（前件部）が生じているという条件下における事象A（後件部）の条件付き確率 $P(A\mid B)$ の情報量 $-\log P(A\mid B)$ の平均値であり，以下の式で表わされる．

$$H(A\mid B)=-\sum_{A,B}P(A,B)\log P(A\mid B) \quad (1)$$

H(A|B)は事象Bについての知識を得た後で事象Aにまだ残っている不確定さを表す式であるので，事象Aと事象Bの間に事象Bが確定したときに事象Aが一意に決まる確定的な関係に近い状況が成り立っているときには条件付きエントロピーの値は0に近づき不確定になるほど増大する，独立性・従属性の指標として扱われる．この条件付きエントロピーを用いて学習データから変数間の因果関係を抽出する．

Examination of the model construction method of Bayesian Networks
† Yasuyuki Suzuki, †† Seihiro Nakamura, ††Masaomi Kimura
†graduate school of Shibaura Institute of Technology
††Shibaura Institute of Technology

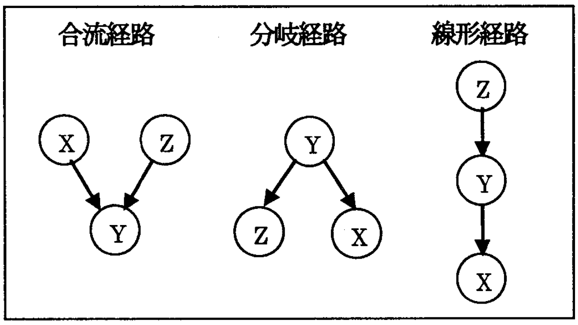

図 1. ネットワークモデルの基本構造

複雑なネットワーク構造を一度に決定することは困難であるため，ネットワークの基本構造であると考えられる合流経路・分岐経路・線形経路についてまず考える．ノードが 3 つの場合の例を図 1 に示す．これら 3 つの経路が組み合わされることによってネットワークが構成されていると考え，本研究では 3 つの基本経路の条件付きエントロピーの値を調べる．分岐経路と線形経路は条件付き独立性の観点から変数 Y を与えた時に X と Z が条件付き独立であると考えることができるので，両経路には類似している部分があると言える．一方で YX 間のアークの向きが等しく，ZY 間のアークの向きが異なるため，分岐経路と線形経路の違いが ZY 間のアークの向きの違いによるものか調べるため両経路の条件付きエントロピーを比較する．

式(2), (3)は X と Z の条件付き独立性を基に H(Z|X) と H(Z|X,Y)を変形したものである．

$$H(Z\mid X)=\sum_{X,Z}P(X)P(Z)\log P(Z) \quad (2)$$

$$H(Z\mid X,Y)=\sum_{X,Y,Z}P(X\mid Y)P(Z\mid Y)P(Y)\log P(Z\mid Y) \quad (3)$$

前提条件として両経路の Y の出現確率を同一にし，変数間の条件付き確率も同一にすると P(Y)と P(X|Y)は線形経路でも分岐経路でも共通であるから両経路の違いは(2)では P(Z), (3)では P(Z|Y)である．よって，(2)では P(Z)が両経路で出方が違うことによりエントロピーH(Z|X)に違いが生じ，(3)では P(Z|Y)が両経路で出方が違うため H(Z|X,Y)に違いがでる．また，H(Z|Y)についても同様のことが言える．さらに P(Z|Y)が両経路で異なる値をとることを式を用いて示す．前提条件より線形経路の条件付き確率 P(Z|Y) と分岐経路の P(Y|Z)を $\begin{pmatrix} p & 1-p \\ 1-p & p \end{pmatrix}$ とおき，Y の出現確率を P(Y=0)=r , P(Y=1)=1−r とおく．また，線形経

路の Z の出現確率は未知であるため P(Z=0)=s，P(Z=1)=1－s と定義する．このときの両経路の P(Z|Y)を下記に示す．

$$P(Z\,|\,Y)=\begin{pmatrix} p & 1-p \\ 1-p & p \end{pmatrix} \qquad \text{(線形経路)}$$

$$P(Z\,|\,Y)=\begin{pmatrix} \dfrac{ps}{r} & \dfrac{(1-p)(1-s)}{1-r} \\ \dfrac{(1-p)s}{r} & \dfrac{p(1-s)}{1-r} \end{pmatrix} \qquad \text{(分岐経路)}$$

分岐・線形経路の P(Z|Y)の値は上式より s=r のとき等しくなるが，それ以外のとき両経路の P(Z|Y)は異なる．以上により両経路のエントロピーの違いが P(Z|Y)からくるものだと分かる．

次に両経路の条件付きエントロピーの違いを式(1)に具体的な値を当てはめて実験を行い分岐経路・線形経路の特徴を導出する．また，合流経路についても同様に実験を行う．

3 実験

合流経路・分岐経路・線形経路の条件付きエントロピーを求めるために 3 経路の特徴を持つ学習データを人工的に生成し，それら各経路のエントロピーを算出し各経路についてどのような特徴が得られるか実験を行った．今回生成するデータは図 1 の構造を基本とし，ノードの状態が 0 と 1 の 2 値であるデータとした．また，前提条件より分岐経路と線形経路の Y の出現確率を同一にし，変数間の条件付き確率をすべて一定とした学習データを生成する．各経路について生成した学習データの詳細について表 1 に記載する．実験では経路上にあるすべてのノードについて各経路の条件付きエントロピーを求め，条件付きエントロピー表として提示する．

表 1. 実験で用いる学習データ

<table>
<tr><th></th><th>合流経路</th><th>分岐経路</th><th>線形経路</th></tr>
<tr><td>レコード数</td><td>10,000 件</td><td>10,000 件</td><td>10,000 件</td></tr>
<tr><td>先頭ノードの出現確率</td><td>P(X=0)=0.50
P(Z=0)=0.50</td><td>P(Y=0)
=0.7047</td><td>P(Z=0)
=0.246</td></tr>
<tr><td>ノード間の条件付き確率</td><td>$P(A=0|B=0,C=0)=0.9$
$P(A=0|B=1,C=0)=0.1$
$P(A=0|B=0,C=1)=0.1$
$P(A=0|B=1,C=1)=0.9$
$P(A=1|B=0,C=0)=0.1$
$P(A=1|B=1,C=0)=0.9$
$P(A=1|B=0,C=1)=0.9$
$P(A=1|B=1,C=1)=0.1$</td><td colspan="2">$P(A=0\,|\,B=0)=0.9$
$P(A=1\,|\,B=0)=0.1$
$P(A=0\,|\,B=1)=0.1$
$P(A=1\,|\,B=1)=0.9$</td></tr>
</table>

4 結果

表 2，表 3，表 4 はそれぞれ分岐経路・線形経路・合流経路の条件付きエントロピー表である．

はじめに表 2 と表 3 で示す分岐経路と線形経路の条件付きエントロピーの特徴について説明する．分岐経路のエントロピーでは H(X|Y)≒H(Z|Y)となり同程度の値を示している．これは X と Z はどちらも Y を親に持つ変数であり，かつ条件付き確率が XY 間，ZY 間で同定度であるためエントロピーが同程度になる．一方，線形経路ではこのような特徴が表れていない．よって，分岐と線形経路を区別する特徴は X と Z を入れ替えてもエントロピーが変わらないという対称性の表れである．また，両者のエントロピーを比較して大きく異なっている部分では H の右側のパラメータに Z を持つ H(Z|X,Y), H(Z|X), H(Z|Y)であり，式の変形で求めた結果と一致する．表 4 の合流経路では前件部のパラメータが 2 つの場合と 1 つの場合のエントロピーには大きな差があり，パラメータが 2 つの場合のエントロピーの方が小さい．これは，「親(前件部)が 1 つ決まると子(後件部)が決まる」よりも「親が 2 つ決まると子が決まる」不確定さの方が低いことを意味し，合流経路の特徴を示すものである．

表 2. 分岐経路の条件付きエントロピー表

H(X|Y,Z)	H(X|Y)	H(X|Z)
0.3151733	0.315234	0.467505
H(Y|X,Z)	**H(Y|X)**	**H(Y|Z)**
0.17120191	0.315409	0.323533
H(Z|X,Y)	**H(Z|X)**	**H(Z|Y)**
0.32348183	0.467689	0.323542

表 3. 線形経路の条件付きエントロピー表

H(X|Y,Z)	H(X|Y)	H(X|Z)
0.33274516	0.332915	0.475457
H(Y|X,Z)	**H(Y|X)**	**H(Y|Z)**
0.17727968	0.31253	0.319991
H(Z|X,Y)	**H(Z|X)**	**H(Z|Y)**
0.28902048	0.424271	0.289191

表 4. 合流経路の条件付きエントロピー表

H(X|Y,Z)	H(X|Y)	H(X|Z)
0.319159427	0.692890781	0.692890811
H(Y|X,Z)	**H(Y|X)**	**H(Y|Z)**
0.319167488	0.693102002	0.692898872
H(Z|X,Y)	**H(Z|X)**	**H(Z|Y)**
0.319211619	0.693146133	0.692942973

5 おわりに

本研究では分岐経路と線形経路の条件付きエントロピーの式の変形により 2 経路間のエントロピーを比較した．また，人工的に生成した合流経路・分岐経路・線形経路の学習データから条件付きエントロピーを求め各経路の特徴の導出を行った．これにより，学習データから変数間の因果関係を抽出することができ，適切なネットワーク構造を持ったモデルが構築できると考えられる．今後の展望として，より一般的な学習データに対しても 3 経路を判別するルールを明確にし，学習データから自動でモデルを構築するシステムの開発を目的とする．

参考文献

[1] 北研一，確率的言語モデル，pp.9-10，東京大学出版 (2001)．

[2] 本村陽一・原功，確率ネットワークによるユーザモデル構築システム，第 1 回 IPA Technology EXPO (ITX2001)．