ベイジアンネットとは

産業技術総合研究所 脳神経情報研究部門 一杉裕志

2009-06-12

１．ベイジアンネット

ベイジアンネット(Bayesian network)は、複数の**確率変数**(random variable)の間の確率的な因果関係を計算機のメモリ上に記憶するためのデータ構造である。ベイジアンネットを用いて複雑な知識を効率的に表現することができ、その知識に基づいて事後確率計算やＭＰＥ計算など様々な確率的推論を行うことができる。音声認識などでよくつかわれる隠れマルコフモデル(HMM, Hidden Markov Model)もベイジアンネットの一種である。

ベイジアンネットは、確率変数を表す**ノード**と、確率変数間の因果関係を表す**エッジ**（リンク、アークとも言う）で構成される。さらに、各ノードごとに**条件付確率表**(conditional probability table)と呼ばれるものを保持する。条件付確率表は、あるノードの親ノードの集合がある値の組み合わせを取ったときにそのノードがある値を取る**条件付確率**(conditional probability)を表にしたものである。

図１は４つの確率変数を表すノード S, R, W, C から構成される簡単なベイジアンネットの例である。確率変数 S は「スプリンクラーが動いたかどうか」、確率変数 R は「雨が降ったかどうか」、確率変数 W は「芝生が濡れているかどうか」、確率変数 C は「雲が出ているかどうか」を表しているとしよう。

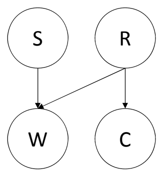

図１　４つのノードから成るベイジアンネットの例。

P(S=no)	P(S=yes)
0.8	0.2

P(R=no)	P(R=yes)
0.98	0.02

S	R	P(W=no|S,R)	P(W=yes|S,R)
no	no	0.88	0.12
no	yes	0.2	0.8
yes	no	0.1	0.9
yes	yes	0.02	0.98

R	P(C=no|R)	P(C=yes|R)
no	0.7	0.3
yes	0.005	0.995

図２条件付確率表の例。

４つの確率変数の間の因果関係は、条件付確率表として与える。図２は、図１のベイジアンネットの各ノードに付随する条件付確率表の例である。

条件付確率表は、確率変数の間の因果関係の強さの知識を記憶するデータである。例えば P(W=no|S=no,R=no)=0.88 は、スプリンクラーも動かず雨も降っていないときに芝生が濡れていない確率は 0.88 であるという知識を表している。また P(R=yes)=0.02 は、単に雨が降る確率（**事前確率**）が 0.02 であるという知識を表している。

2. ＭＰＥ

ＭＰＥ（most probable explanation）とは、ベイジアンネットにおいて、与えられた観測データを最もよく説明する変数の値の組のことである。与えられた観測データを表す確率変数とその値の組の集合を ｉ 、**隠れ変数**（観測データ以外の確率変数）とその値の組の集合を h とすると、 MPE となる値の組 m は次の式で与えられる。

$$\mathbf{m} = \arg\max_{\mathbf{h}} P(\mathbf{h},\mathbf{i})$$

ただし $P(\mathbf{h},\mathbf{i})$ は h と ｉ という値の組み合わせが起きる**同時確率**（joint probability）で、以下の式で表せる。

$$P(\mathbf{h},\mathbf{i}) = \prod_{x \in \mathbf{h} \cup \mathbf{i}} P(x \mid parents(x))$$

ここで *parents*(x) はノード X の親ノードの値の組である。

例えば図１のベイジアンネットにおいて、観測値 W=yes が与えられたとする。求めるＭＰＥは、観測値との同時確率がもっとも高い隠れ変数 S, R, C の値の組 {s, r, c}で、以下の式で表される。

$$\mathbf{m} = \arg\max_{\{s,r,c\}} P(S=s, R=r, C=c, W=yes)$$

$$P(S=s,R=r,C=c,W=yes) = P(W=yes \mid S=s,R=r)P(C=c \mid R=r)P(S=s)P(R=r)$$

以下に、具体的なＭＰＥの計算手順の一例を示す。まず、 S=no, R=no, C=no という値の組の、観測値 W=yes との同時確率は、図２の条件付確率表の値を用いて以下のように計算される。

$$\begin{aligned}
&P(S=no,R=no,C=no,W=yes)\\
&= P(W=yes \mid S=no,R=no)P(C=no \mid R=no)P(S=no)P(R=no)\\
&= 0.12 \times 0.7 \times 0.8 \times 0.98\\
&= 0.065856
\end{aligned}$$

同様にして、他の値の組み合わせの同時確率も計算し、２の３乗個あるすべての組み合わせの各同時確率をまとめると下のようになる。

$$\begin{aligned}
P(S=no,R=no,C=no,W=yes) &= 0.065856\\
P(S=no,R=no,C=yes,W=yes) &= 0.028224\\
P(S=no,R=yes,C=no,W=yes) &= 0.000064\\
P(S=no,R=yes,C=yes,W=yes) &= 0.012736\\
P(S=yes,R=no,C=no,W=yes) &= 0.123480\\
P(S=yes,R=no,C=yes,W=yes) &= 0.052920\\
P(S=yes,R=yes,C=no,W=yes) &= 0.000020\\
P(S=yes,R=yes,C=yes,W=yes) &= 0.003900
\end{aligned}$$

この中では、 S=yes, R=no, C=no がもっとも同時確率の高い値の組み合わせになるので、これがＭＰＥである。したがって、図１および図２の形式で記憶されている知識に基づいて、もし芝生が濡れているならば、「スプリンクラーは動いたが雨は降らず雲も出ていない」という組み合わせがもっとも可能性が高いと推論されたことになる。

一般にはＭＰＥの計算には、ノード数 n に対して $O(2^n)$ の計算時間が必要である。しかし、ネットワークの構造に制約がある場合は計算量を減らせる場合がある。また、近似解でよければより少ない計算量で計算できる場合がある。

3.条件付確率表の学習

ベイジアンネットのネットワーク構造が与えられていて、各確率変数の値の組についての大量の観測データがあれば､それをもとに条件付確率表の要素の値を決めることができる。これを**条件付確率表の学習**と呼ぶ。

例えば、 1000 個の観測データのうち R=no であるものが 980 個であれば、 P(R=no) は 980/1000 となる。また、その中でさらに C=no であるものが 686 個であれば、P(C=no|R=no) は 686/980 となる。

隠れ変数（観測データが与えられない変数）がある場合は、**ＥＭアルゴリズム**という手法用いて、隠れ変数の推定値に基づいて条件付確率の値を決定する。

条件付確率表の学習は､通常大量のデータを一度に処理することで行われる｡しかし､時々刻々と新しい観測データが与えられるたびに逐次的に条件付確率表を更新することも可能である｡そのような学習アルゴリズムは､**オンライン学習アルゴリズム**と呼ばれる。

例えば過去に得られた条件付確率の値が P(Y=yes|X=yes)=3/10, P(Y=no|X=yes)=7/10 だったとする。新しく観測データとして得られた確率変数 X, Y の値がそれぞれ X=yes, Y=yes であったなら、条件付確率の値を P(Y=yes|X=yes)=(3+1)/(10+1)=4/11, P(Y=no|X=yes)=7/(1+10)=7/11 に更新する。なお、 P(Y=yes|X=no), P(Y=no|X=no) の値は更新する必要はない。

なお、観測データの数が少ない場合は、上で述べたような素朴な統計量は真の条件付確率の値からかけ離れており、実用にならない。実はこの状況は機械学習の分野で言う**過適合**(overfitting)に相当する。この問題は、条件付確率の値を学習パラメタとみなして事前分布を与えることで改善できる。

4.構造学習

ネットワーク構造が不明な場合でも、観測データに基づいて最適なネットワーク構造を推定することができる。これを**構造学習**(structure learning)と呼ぶ。構造学習も一般にはノード数 n に対して $O(2^n)$ の計算量を必要とする。しかし、ネットワーク構造に制約を入れたり近似解にすることで計算量を減らせる可能性がある。

隠れ変数があり得る場合は、解の探索空間が増えるので、一般には問題はさらに難しくなる。

5.ベイジアンネットと脳

筆者が提案する**ＢＥＳＯＭモデル**と呼ぶ大脳皮質の神経回路モデルによれば、ヒトの大脳皮質は約２０万個のノードからなる巨大なベイジアンネットである。大脳皮質の主な機能は、感覚器を通じて与えられる観測データにもとづいて外界の状況を認識・学習することである。大脳皮質の認識はベイジアンネットのＭＰＥ計算、大脳皮質の学習はベイジアンネットの条件付確率表の学習と構造学習に対応する。神経細胞の間をつなぐシナプスが、条件付確率をオンライン学習する。

大脳皮質は認識と学習に必要な計算量を巧妙な方法で O(n) にまで減らしていると考えており、その具体的アルゴリズムが明らかになりつつある。

6.参考文献

ベイジアンネットの参考文献については、下記ページをご覧ください。

「脳を理解するための情報源メモ」ベイジアンネット
http://staff.aist.go.jp/y-ichisugi/besom/brain-memo.html#Bayesian-network

ＢＥＳＯＭモデルについては下記テクニカルレポートをご覧ください。

一杉裕志、「脳の情報処理原理の解明状況」
産業技術総合研究所テクニカルレポート AIST07-J00012, Mar 2008.
http://staff.aist.go.jp/y-ichisugi/besom/AIST07-J00012.pdf

間違いの指摘、コメントなど歓迎いたします。

y-ichisugi@aist.go.jp