

ベイジアンネットワーク：入門からヒューマンモデリングへの応用まで

本村 陽一

産業技術総合研究所

デジタルヒューマン研究センター

Abstract: ベイジアンネットワークはグラフ構造を持つ確率モデル (グラフィカルモデル) の一つである。本講演では、このベイジアンネットワークについて、モデルの定義、モデル上で実行される確率推論アルゴリズム (確率推論)、モデル構築のアルゴリズム (統計的学習)、ソフトウェアなどの話題を中心に概説する。またその応用として、障害診断、また人間のモデル化とそのモデルに基づく予測技術としてのヒューマンモデリング研究に活用するアプローチなどについて、いくつかの応用事例の紹介もあわせて行う。

1. まえがき

ベイジアンネットは不確実性を含む事象の予測や合理的な意志決定、観測結果から原因を探る障害診断などに利用することのできる確率モデル (グラフィカルモデル) の一種である。最近、このモデル上での確率推論アルゴリズムの進歩や、不確実性を含む様々な問題への応用事例、ソフトウェアの普及などにより、知的情報処理システムへの利用が現実的になってきている。ベイジアンネットは条件付確率分布群によって対象をモデル化する一般的な枠組みであると言える。条件付確率分布として表現できる対象は非常に多く、また条件付確率分布を表によって表すために自由度が高く、従来他のモデルで扱われている問題の多くもベイジアンネットの枠組みによっても表現できる。

さらにこうしたベイジアンネットによってモデル化を行うことにより可能となる有効な情報処理がある。一つは観測した変数群から未観測の対象の確率分布を計算する確率推論であり、もう一つはそのためのモデルを統計データから構築する統計的学習である。

本稿では、まずベイジアンネットモデルについて解説し、さらに確率推論、統計的学習の仕組みなどの解説を行う。また実際にベイジアンネットを利用するために筆者が開発したソフトウェアについて紹介する。最後に不確実性を対象とする知的情報システムへの適用例として障害診断と人間についてのモデリング、ヒューマンモデリングへの応用についても述べる。

2. ベイジアンネットワーク：モデル

ベイジアンネットワーク (Bayesian network, Bayesnet, belief network) とは複数の確率変数の間の定性的な依存関係をグラフ構造によって表し、個々の変数の間の定量的な関係を条件付確率で表した確率モデルである。英語の成書としては [1, 2, 3, 4, 5] が、また日本語で読める文献としては [6, 7, 8, 9, 10] などがあるので、さらに興味のある方は適宜参照してほしい。

確率モデルとしては、確率変数、その間の依存関係を表すグラフ構造、条件付確率、の集合によって定義される。これを用いた確率計算によって不確実性を含む事象の予測やシステムの制御、障害診断などの知的情報処理に利用することができる大きな特徴である。

確率変数としては例えば「会議に関係するメールが来る」というような事象について定義し、その可能性に関して 0 から 1 の間の確率値をとる。「次の会議が行われる曜日」(月曜から日曜日) のように複数の状態と、それぞれの確率をとる場合もある。「前回の会議の曜日」のように観測の結果、状態を確定できる場合にはその状態の確率値は 1、それ以外の状態については 0 となる。

変数はノードとして、変数間の依存関係は向きを持つ有向リンクで図示する。例えば、確率変数 X_i, X_j の間の条件付依存性をベイジアンネットワークでは $X_i \rightarrow X_j$ と表す。リンクの先に来るノード (この場合は X_j) を子ノード、リンクの元にあるノード (この場合は X_i) を親ノードと呼ぶ。

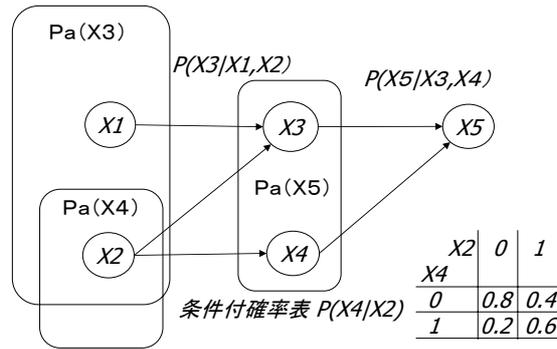


図 1: ベイジアンネットワーク
Bayesian network

親ノードが複数あるとき子ノード X_j の親ノードの集合を $Pa(X_j)$ と書くことにする。 X_j と $Pa(X_j)$ の間の依存関係は次の条件付確率によって定量的に表される。(ただし $Pa(X_j)$ が空集合の時は事前確率分布。)

$$P(X_j | Pa(X_j)) \tag{1}$$

さらに n 個の確率変数 X_1, \dots, X_n のそれぞれを子ノードとして同様に考えると、全ての確率変数の同時確率分布は式 (2) のように表せる。

$$P(X_1, \dots, X_n) = \prod_{j=1}^n P(X_j | Pa(X_j)). \tag{2}$$

こうして各子ノードとその親ノードの間にリンクを張って構成したベイジアンネットワーク(図1)によって、これらの変数の間の確率的な依存関係がモデル化できる。先の例で言えば「前回の会議の曜日」と「会議に関係するメールが来る」という二つの事象と「次の会議のある曜日」の間に依存関係があるならば¹、この3つのノードを使ったベイジアンネットワークによって、その確率的な関係がモデル化でき、それを使ってシステムが今日のスケジュールを予測して事前に入力しておいたり、会議を忘れないようにユーザに通知することなどが可能である。またこのような確率的関係が成立していれば、それを何段にもつないだネットワークと後で述べる確率推論アルゴリズムによって、より複雑な推論も行うことができる。

	$Pa(X_j) = \mathbf{x}_1$	\dots	$Pa(X_j) = \mathbf{x}_m$
$X_j = y_1$	$p(y_1 Pa(X_j) = \mathbf{x}_1)$	\dots	$p(y_1 Pa(X_j) = \mathbf{x}_m)$
\vdots	\vdots	\ddots	\vdots
$X_j = y_n$	$p(y_n Pa(X_j) = \mathbf{x}_1)$	\dots	$p(y_n Pa(X_j) = \mathbf{x}_m)$

表 1: 条件付確率表 (CPT)
Conditional probability table(CPT)

変数が離散の場合、式 (2) の右辺にある条件付確率は例えば $P(\text{Next} = \text{Mon} | \text{Mail} = \text{true}, \text{Prev} = \text{Mon}) = 0.8$ のような形で列挙することができ、子ノードと親ノードがとる全ての状態のそれぞれにおける確率値を定めた表(条件付確率表:CPT)を使うことで完全に表現できる。

一般的に書くと、親ノードがある状態 $Pa(X_j) = \mathbf{x}$ (\mathbf{x} は親ノード群の各値で構成したベクトル)のもとでの n 通りの離散状態 (y_1, \dots, y_n) を持つ変数 X_j の条件付確率分布は $p(X_j = y_1 | \mathbf{x}), \dots, p(X_j =$

¹もちろん定例的な会議とそれをメールで通知する習慣があり統計的にこうした傾向があるならば、である。

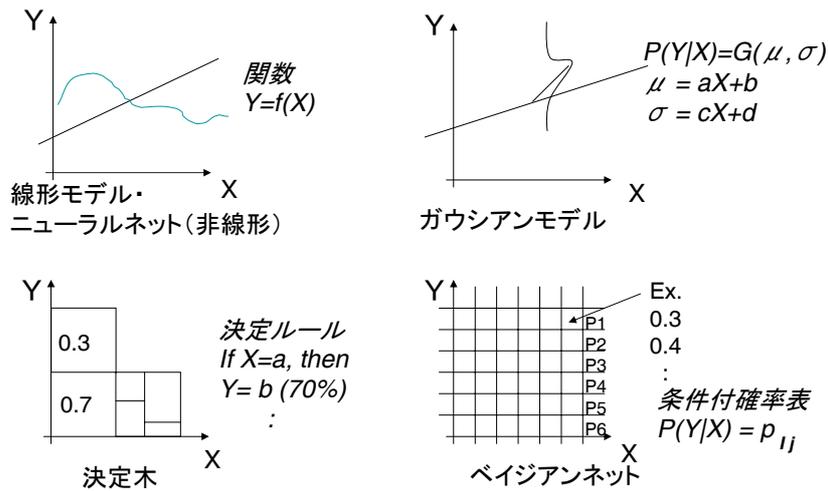


図 2: Bayesian network and other probabilistic models

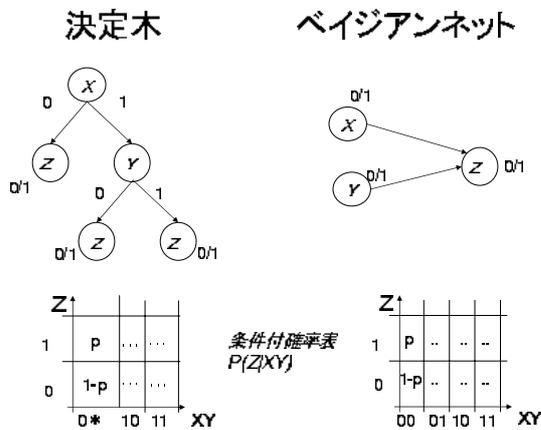


図 3: Bayesian network and Decision tree

$y_n|x)$ となる (ただし $\sum_{i=1}^n p(y_i|x) = 1.0$)。これを各行として、親ノードがとりうる全ての状態 $Pa(X_j) = x_1, \dots, x_m$ のそれぞれについて列を構成した表 1 の各項目に確率値を定めたものが X_j にとっての条件付確率表 (CPT) である。

ベイジアンネットは $X - Y$ 空間を条件付確率表にしたがって離散化し、個々の確率値を割り当てた不連続な確率分布によるモデル化である。その自由度は比較的高く、線形から非線型な依存関係まで柔軟に近似することができる (図 2)。また各項目毎に十分な数の統計データがあれば、変数の各状態についての頻度を正規化して、各項目の確率値を求めることが容易にできる。

決定木もベイジアンネット同様に変数空間を分割するようにモデル化するが、分割の仕方がやや異なっている (図 3)。

さらに、いわゆるパターン認識におけるベイズ識別との関連で言えば、親ノードにパターンクラス、子ノードに特徴量を与え、パターンクラスの事後確率を計算し、これを最大化するようなパターン認識を行うことができ、この場合は Bayesian classifier や naive Bayes と呼ばれるものと等価である。(ベイジアンネットとして拡張したベイズ識別器を使うことで識別精度が向上したという報告 [11, 12] もある。) また音声認識やバイオインフォマティクスなど、時系列データの認識に使われる確率モデル、Hidden Markov Model と等価なモデルをベイジアンネットとして作成することもできる (図 4)。このように状態変数を導入したモデルは Dynamic Bayesian Net とも呼ばれてい

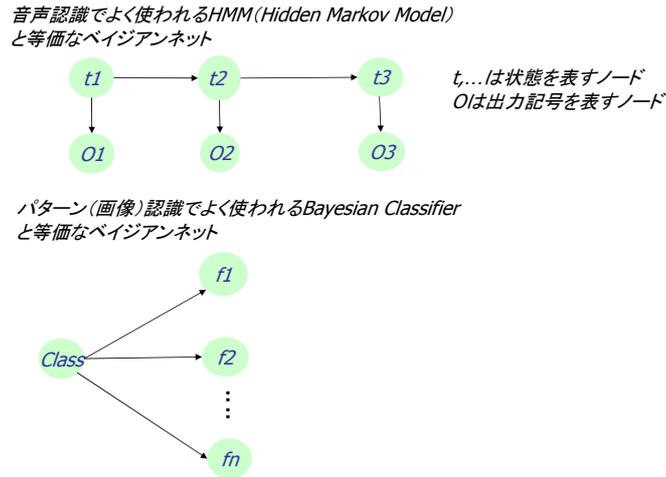


図 4: Bayesian network, Bayesian classifier and HMM

る．このようにベイジアンネットワークはその表現力の高さから，これまで有用とされてきた多くの確率モデルを包含し，統一的に理解できるものとみることができる．

3. ベイジアンネットワークの確率推論

ベイジアンネットワークを使うことで，一部の変数を観測した時のその他の変数についての確率分布を求めたり，確率値が最も大きい状態をその変数の予測結果として得ることができる．観測された変数の情報 (e) から，求めたい確率変数 (X) の確率値，すなわち事後確率 $P(X|e)$ を求め，それにより X の期待値や事後確率最大の値 (MAP 値)，ある仮説の確信度 (いくつかの変数が特定の値の組をとる同時確率) などを評価するわけである．先の例では「次の会議のある曜日」を予測したり，異常を観測したセンサーの状態からシステムの障害原因を推定するような計算処理である．こうした確率計算に基づく推論が確率推論と呼ばれている．

3.1 確率推論アルゴリズム

ベイジアンネットワークによる確率的推論は，i) 観測された変数の値 e をノードにセットする，ii) 親ノードも観測値も持たないノードに事前確率分布を与える，iii) 知りたい対象の変数 X の事後確率 $P(X|e)$ を得る，という手順で行なわれる．iii) における事後確率を求めるために，観測された情報からの確率伝搬 (変数間の局所計算) によって各変数の確率分布を更新していく確率伝搬法 (belief propagation) と呼ばれる計算法がある．ここでは簡単に，図 5 の構造のもとでの計算の実行例を示す．

$X_1 \rightarrow X_2, X_2 \rightarrow X_3$ の間に依存性があり，条件付確率が与えられているとする．今，計算しようとしているノードを X_2 として，上流にある親ノードに与えられる観測情報を e^+ ，下流の子ノードに与えられる観測情報を e^- と書く．計算したい事後確率 $P(X_2|e)$ は， e を e^+ と e^- に分け， X_2 と e^- に注目してベイズの定理を使うと次のようになる．

$$\begin{aligned} P(X_2|e) &= P(X_2|e^+, e^-) \\ &= \frac{P(e^-|X_2, e^+)P(X_2|e^+)}{P(e^-|e^+)} \end{aligned}$$

また e^+ と e^- は X_2 を固定した時には条件付き独立になるので， $\alpha = \frac{1}{P(e^-|e^+)}$ を X_2 の値によらない正規化定数とすれば，次のように変型できる．

$$P(X_2|e) = \alpha P(e^-|X_2)P(X_2|e^+). \quad (3)$$

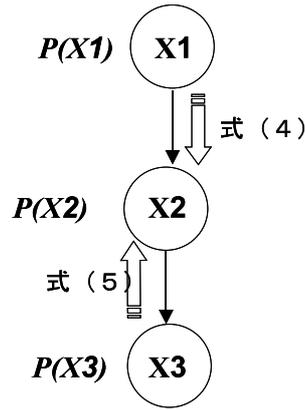


図 5: 簡単な構造での確率伝搬
Belief propagation on simple network

このうち e^+ による X_2 への寄与分, つまり親ノードから伝搬する確率を $P(X_2|e^+) = \pi(X_2)$ と書く. これは, $P(X_1|e^+)$ と X_2 の CPT を使って, X_1 についての周辺化, 次式 (4) によって求めることができる.

$$\pi(X_2) = \sum_{X_1} P(X_2|X_1)P(X_1|e^+). \quad (4)$$

$P(X_1|e^+) = \pi(X_1)$ は観測値が与えられているならば, その値は決定できる. 観測値がなく, さらに親ノードを持たない最上流のノードであるならば, 事前確率を与える. その上流に親ノードを持つ場合には式 (4) を再帰的に適用していけば最終的にはもっとも上流にあるノードによって, その値が求まる.

一方, 式 (3) の子ノード側の e^- の寄与分, つまり子ノードから伝搬する確率を $P(e^-|X_2) = \lambda(X_2)$ とすると, これを計算するためには, すでに定義されている条件付き確率 $P(X_3|X_2)$ を使って X_3 の全状態について周辺化する次式を用いればよい.

$$\lambda(X_2) = \sum_{X_3} P(e^-|X_2, X_3)P(X_3|X_2).$$

観測から得られる情報 e^- は X_2 の値によらず独立であることを利用すると, これは次式のように書き直せる.

$$\lambda(X_2) = \sum_{X_3} P(e^-|X_3)P(X_3|X_2). \quad (5)$$

ここで, $P(X_3|X_2)$ は条件付確率表として与えられている. $P(e^-|X_3) = \lambda(X_3)$ は観測情報が与えられているならば値が確定できる. また観測値がなく, その下流に子ノードを持たない下端のノードの場合には, 無情報であるから一様確率分布であるとして X_3 の全ての状態について等しい値とする. また一般の構造のネットワークの場合, さらに下流に子ノードを持つならば, 式 (5) を再帰的に適用していけば, 最終的にはもっとも下流にあるノードによって値が求まるので, やはり $\lambda(X)$ を計算することが可能である.

したがって, 以上式 (4),(5) を, 式 (3) に代入すればノード X_2 の事後確率が求まる. 同様に次式によって任意のノードの事後確率も局所的に計算できる.

$$P(X_j|e) = \alpha\lambda(X_j)\pi(X_j).$$

ベイジアンネットのリンクの向きを考慮しないグラフ構造内の全てのパスがループを持たない時, そのベイジアンネットは singly connected なネットワークと呼ばれる. この場合には, 親ノード, 子

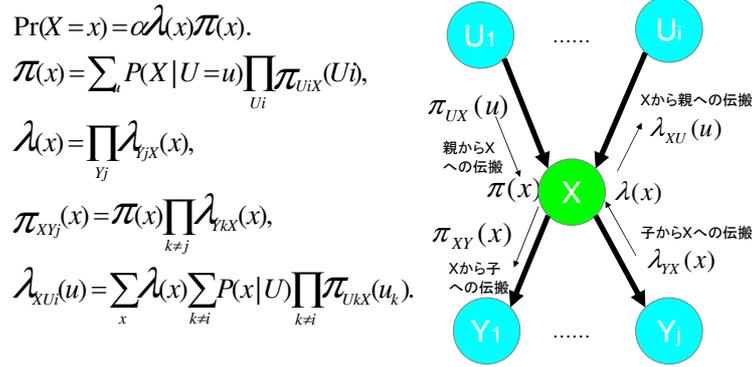


図 6: 確率伝搬アルゴリズム
Belief propagation algorithm

ノードが複数存在するような構造のネットワークでも，条件付独立性の性質を使うことで，各ノードについて上流からの伝搬，下流からの伝搬，上流への伝搬，下流への伝搬の4種について先の確率伝搬計算を図6のように行なうことで計算は完了する．この計算量はネットワークのサイズにたいして線形オーダーである．多くの場合メモリサイズの制約から子ノードに接続される親ノードの数が制限されるので，その場合にはノード数に対しても線形オーダーとなる．

リンクの向きを考慮しないでネットワークを見たときに，どこか一つでもパスがループしている部分がある時，このベイジアンネットは multiply connected と呼ばれる．この場合には単純にリンクに沿って確率を伝搬していただくだけでは，その計算の収束性が保証できない．しかし1990年代のはじめに junction tree アルゴリズムと呼ばれるグラフ構造を事前に変換してから確率計算を実行する手法が開発されたことで，様々な構造にたいするベイジアンネットの有用性が高まった．この Junction tree アルゴリズムを実装したソフトウェアとして Hugin (<http://www.hugin.com>) があり，これを使った実用化も進んでいる．

一方で multiplyconnected なグラフを変換せずにそのまま確率計算を行う近似アルゴリズムがあり，決定的なサンプリング法や loopy belief propagation アルゴリズムなどがある．ここではこの3つの異なるアルゴリズムについての紹介を行う．

3.1.1 Junction Tree アルゴリズム

junction tree はまず，適切な親ノードを併合する操作を繰り返してノードのクリークをクラスタとして生成し，元のベイジアンネットのノードをクリークとして結合した singly connected な木構造からなる junction tree (または moral graph) と呼ばれる無向グラフに変換する．次に，こうしてできた singly connected な木構造にしたがってクリーク毎に確率伝搬を行うことでやはり確率値が計算できる．複雑なネットワークの場合には，グラフ変換にかかる計算コストが大きくなるが，一度グラフ構造の変換に成功した後何度も確率計算を行うような場合には非常に効率の良い確率伝搬を実行することができる．

ただし junction tree アルゴリズムは，ノード数が増え，グラフ構造が複雑になるにつれ，変換操作自体の計算コストが無視できない問題となる．例えばネットワーク構造が常に変わらない場合には，一度だけ変換を行えば良いが，状況によってネットワーク構造が変化する場合にはその都度グラフ構造を変換しなければならず，変換のための計算コストは深刻である．またグラフ構造の性質によっては効率のよい junction Tree に変換できず，結果として巨大なクラスターが生じることも起こりえる．その場合にはクラスター内の確率計算のために多数の確率変数の全状態の組み合わせについての計算が必要なために計算量とメモリ消費が増大する．

3.1.2 サンプリングメソッド

グラフ構造を変換しない確率推論アルゴリズムのうち、比較的厳密計算に近い近似アルゴリズムがストカスティックサンプリングによる確率推論である。これはベイジアンネットによる全ての確率変数がとりえる全状態ごとに、その結合確率を事前に求めておき、その確率に基づいて確率変数の具現値の集合をあるサンプル数だけストカスティックに生成する。さらにそのサンプル群のうち与えられたエビデンスと合致するものだけを対象に、知りたい変数の具現値の頻度を数えあげ正規化することで、対象とする確率変数の事後確率 $P(X|e)$ を求める。サンプル数が多ければ解は厳密解にいくらかでも近づくが、ノード数、ノードの状態数が増加すると必要なサンプル数が指数的に増大する。しかし、解の精度が低くてよいならばサンプリング数を少なくすることで計算時間を減少させることができ、利用する状況に応じて計算時間と解の精度を制御することができるメリットがある。

この時、サンプルの生成方法にいくつかバリエーションがあり、古典的なものとしてはランダムサンプリング、また最近では Markov Chain Monte Carlo 法 (MCMC) の適用も考えられている。また決定論的なサンプリング法としては systematic sampling があげられる [3]。

3.1.3 Loopy belief propagation

確率伝搬法を強引に multiply connected なベイジアンネットに適用する方法が loopy belief propagation と呼ばれている。Multiply connected なベイジアンネットに対して強引に局所的な確率伝搬法を繰り返し適用してみると、経験的には良好な性質、たとえば多くの場合事後確率最大の状態に収束していること、また収束しないような場合の多くはノードの値が振動するので、比較的容易に判別できること、などが実験的に示されている [13]。

また loopy belief propagation はベイジアンネットを含むグラフィカルモデル一般について、アルゴリズムの改良やその収束性の解析などに関する理論研究 [15, 16, 17, 18] が最近盛んであり、この中で統計力学や情報幾何学が重要な役割を果たしている。また日本でも若手研究者による研究が進められ、重要な国際会議の一つである Neural Information Processing Systems (NIPS) において日本から提案したワークショップが開催されるなど活発な活動が行われている。

4. ベイジアンネットの統計的学習

ベイジアンネットを実際に使う時にはまず適切なベイジアンネットモデルを構築しなければならないのであるが、これが人手ではなかなか容易でないという問題がある。変数、条件付確率、グラフ構造は対象領域をよく知ったエキスパートの経験や知識により適切に決めなければならない。つまり適切なモデルを構築する手法が本格的な実用化の大きな鍵になっている。

モデルの構築は大量の統計データと変数の定義を与え、それらを最もよく説明するようにグラフ構造と条件付確率を決定することで行われる。グラフ構造を仮定できれば、条件付確率表だけを求めれば良い。ここではベイジアンネットの学習の中心となる、条件付確率の学習と、グラフ構造の学習のそれぞれについて述べる。

4.1 条件付き確率の学習

条件付確率表において、 X, Y のとりうる値の全ての組み合わせについてデータが存在する場合を完全データと呼び、そうでないものを不完全データと呼ぶ。完全データの場合にはデータの頻度により CPT の全ての項を埋めることができる。例えば簡単のため確率変数が真偽二値とすると、親ノード群 $Pa(X_j)$ がある値をとる時の X_j が真であった事例数を n_j^t 、偽であった事例数を n_j^f とする。また、 $Pa(X_j)$ を与えた時に X_j が真となる条件付確率が $P(X_j = 1 | Pa(X_j)) = \theta_j^*$ であるとすると、条件付確率の学習では、この θ_j^* をデータの頻度 n_j^t, n_j^f から推定することになる。データの数

が多く、 n_j^t, n_j^f が十分大きい場合には、最尤推定により $\tilde{\theta}_j = n_j^t / (n_j^t + n_j^f)$ とすることができる。

各項目内の頻度が少数となる場合には、 θ の推定量として $\tilde{\theta} = n/N$ の信頼性が低くなる。その場合には、事前確率 $\theta_0 = 1/Z$ を仮定して $\tilde{\theta} = (n+1)/(N+Z)$ と考えることもある。また、条件付確率の点推定値ではなく、条件付確率値の確率分布を考えることもできる。つまりこのような観測結果 (n_j^t, n_j^f) が得られる確率は二(多)項分布 ${}_{n_j^t+n_j^f}C_{n_j^t} \theta_j^{*n_j^t} (1-\theta_j^*)^{n_j^f}$ で表せるから、逆に条件付確率の分布としては頻度 n をパラメータとする Dirichlet 分布として表現することができる(ただし、そのままでは確率推論の実行が困難なため、実用的には条件付確率表として表すことが多い)。

全ての起こり得る組合せのデータを持たない場合を不完全データと呼ぶ。また実際の問題に適用するにあたっては確率変数の状態数が増加する傾向があり、これにともなって条件付確率表(CPT)のサイズが増大することがあり、いくつかの項についてはデータが欠損したり、相対的に各条件付確率を推定するのに十分なデータ数が確保できないことになる。データから条件付確率を決定するためには、各項に十分な数のデータが必要になるので、このような場合には、最尤法により全ての条件付確率の推定値を得ることは難しい。

したがって不完全データの場合には条件付確率について事前分布を仮定したり、未観測データについての確率分布を推定し、さらにその分布によって期待値計算を行なうことが避けられない。またそのためにEMアルゴリズムを適用することも考えられる。ただし、実際にはEM法での繰り返し計算に時間がかかることや実装上の困難から、こうした場合についてはまだ実用化が進んでいないと言える。

そこで、実際の問題領域においてはより現実的な解決策が望まれているのが現状である。そのための現実的なアプローチの一つとしては、Bound and Collapse 法やエントロピー最大化に基づく方法が提案されている [21, 22]。また筆者は不完全データからの学習において、連続確率分布を表すニューラルネットをはじめにデータから学習し、それを用いて条件付確率表の欠損値を補完する手法とそのためのソフトウェアを開発している [23]。

4.2 グラフ構造の学習

以上で述べた条件付き確率の学習はある特定のグラフ構造のもとでのパラメータ推定である。一方、このグラフ構造もデータから決定したいという要求がある。データから構造を評価するときはその構造のもとでの最適な条件付確率パラメータを確定する必要があるため、グラフ構造の学習はパラメータの探索を含むものになる。最適なパラメータが確定できず準最適なパラメータで代用する場合には、構造の評価は正確でないことに注意が必要である。

グラフ構造を決定するにはデータをもっとも良く説明するようにグラフ構造を探索する。グラフ構造の探索空間はグラフのノード数を k とすると、 ${}_3C_k$ となり、 k が大きくなると、探索空間が爆発的に増大し、全解探索が難しくなる。そのためグラフ構造の学習は現在も未解決の多くの問題を含んでおり、今後もなお重要な研究課題である。また親ノードの数が増えるとCPTのサイズが増加し、学習に必要な事例数が増えたり、欠損が発生しやすくなるため、尤度に加え、できるだけ少ない親ノードをとるようなペナルティ項を考慮する必要がある。

現在、良く知られているベイジアンネットワークの構造学習アルゴリズムとしては現実的な時間でグラフ構造を探索するためのヒューリスティクスを用いた K-2 アルゴリズムがある [19]。以降ではこれを例にとりグラフ構造の学習方法を説明する。

子ノード1つを根、これに接続する親ノード群を葉とした木に注目すると、ベイジアンネットワークはこの木が複数組み合わせられたものになっている。そして条件付確率分布はこの局所的な木のそれぞれについて一つ定義される。そこでグラフ構造の決定は各子ノード毎に最適な局所木を探索する Greedy アルゴリズムとして実現できる。つまり、(1) 各ノードについて親ノードになりえる候補を限定しておく、(2) ある子ノードを一つ選び、候補となる親ノードを一つづつ加えてグラフを作る、

(3) そのグラフのもとで条件付確率を学習し、情報量基準を評価する、(4) 評価が高くなった時だけ親ノードとして採用し、(5) 親ノードとして加える候補がなくなるか、加えても評価が高くならなくなったら他の子ノードへ移る、(6) 全ての子ノードについて(1)-(5)を繰り返す、という Greedy サーチアルゴリズムである。また(3)でのモデルの評価基準においては MDL の適用 [20] が一般的であるが、どのような場合にどのモデル選択基準が適切であるかという問題は完全にはわかっていない問題でもある。ベイジアンネットの構造学習の研究でも他にも様々な情報量基準を使った研究が報告されている。

このようにモデルの構築は条件付確率の最尤推定と情報量基準に対して最適なグラフ構造を探索する手続きの繰り返しとして実現される。全てのグラフ構造を探索することはノード数が増えると計算量の爆発を起こすため、事前に親ノードになれる候補を制限するなどのヒューリスティクスや欲張り探索とすることが実用上の工夫である。また情報量基準や親ノードの探索戦略の違いによって他にも様々なバリエーションがあり、問題に応じて適宜選択することが重要である。

データに基づいて学習したグラフ構造は選択時に用いた情報量基準について最適化したものであるが、依存関係の向きがベイジアンネットとして最適であるという保証はない。つまり、統計データから相関は読み取れるが、依存関係の向きを確定するための情報としては十分ではない。そこで質の高いベイジアンネットモデルを構築するためには実際にはデータから学習したグラフ構造をさらに洗練することが必要である。

ベイジアンネットの一つの子ノードについてのグラフ構造(親ノード)の決定としてみなせば、これは決定木(decision tree)の学習アルゴリズムやクロス集計表の独立性検定と関連が深い。また得られた構造がデータの背後にある有用な知識を表していると考えられることからベイジアンネットの学習はパス分析やデータマイニングの一つの形としても考えることもできるだろう。しかしベイジアンネット独自の特徴としては多段の依存関係をネットワーク化し、その上で確率推論が実行可能なことがある。この特徴によって、後で述べるようにデータから学習して動作する自律的な知的システムが実現可能になるのである。

5. ベイジアンネットソフトウェア: BayoNet

先に述べたように、広範な問題に対して容易にベイジアンネットを利用できるようにするためには、データから適切なモデルを構築する技術が必要である。BayoNet[23, 24, 25]はこうした問題意識によりベイジアンネットモデル構築ソフトウェアとして開発された。また大規模なモデルでも高速に確率推論を実行するため、近似的な確率推論アルゴリズムの実装も行っている。ここではこの BayoNet の概要を述べ、確率推論アルゴリズムの実験評価結果を示す。

5.1 BayoNet の特徴

筆者が本ソフトウェアの研究・開発を始めた当初はベイジアンネットのモデル構築機能を持つソフトウェアは他にはあまりなかった。そこで本格的な SQL データベースからの統計的学習、ベイジアンネットを構築するためのグラフ構造の自動学習、事前知識(ルール)の併用、質問に答えることで対話的にモデルを構築できる Wizard スタイルの GUI など、一般のユーザがベイジアンネットを用いる際に必要となる機能についても実装を進めながら、一方で研究用途のための高度な機能も積極的に導入していることが BayoNet の大きな特徴である²。

また BayoNet は特に機能拡張や他のプログラムと連携を行うことが考慮されており、次のような特色がある。(i) JDBC により、主要な各種データベースシステムとの連携、(ii) JAVA のリフレクションにより、各種アルゴリズムの追加が容易、(iii) 他のベイジアンネットソフトウェアとのモデ

²現在 BayoNet は産総研イノベーションズを通じて企業へのソースコードライセンスや商業実施権の供与や、研究用途のために共同研究や技術研修の中で利用することも可能である。また無償の評価版のダウンロードや学術・教育用の安価なライセンスの販売、商業利用ライセンスの販売なども行っている。

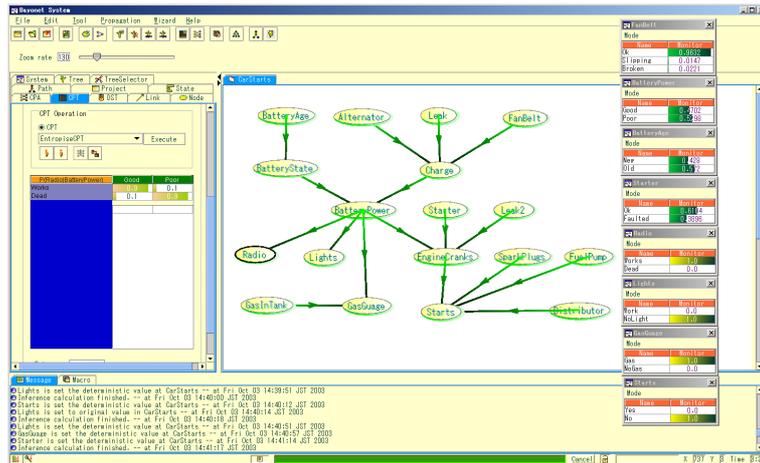


図 7: BayoNet

ルの共有, (iv) ネットワーク (TCP/IP) 経由で外部プログラムと接続し, 統計的学習・確率推論機能の利用が可能. データベースとの連携においては, データをメモリに読み込むことなく, SQL コマンドを利用して条件付確率をデータベース側で計算するため, 全てのデータを転送する必要がなく, 大規模データの取り扱いが容易である. また GUI 上で操作するだけでなく, TCP/IP 経由でサーバとして動作し, クライアントとなる他のソフトウェアから各機能を利用することもできる. これによりユーザの操作履歴やフィードバックからモデルを学習し, それによって予測や推論を行う知的支援システムを開発するような応用が可能である.

当初は Junction tree アルゴリズムによる推論機能を持つソフトウェア Hugin に, BayoNet で構築したモデルを互換性のあるファイルとして渡すことで確率推論を実行していた. 最近, BayoNet に Systematic sampling と Loopy Belief Propagation の 2 つの確率推論アルゴリズムの実装が行われ, BayoNet だけで確率推論を実行することも可能になった.

したがって, 現在 BayoNet で使用できる確率推論アルゴリズムは System sampling と Loopy BP の 2 つの近似アルゴリズムと, 厳密計算である Hugin による Junction tree アルゴリズムの 3 つである. この 3 つは計算速度, 解や収束性などについて異なる特性を持ち, 確率推論を行うベイジアンネットの構造によって適切な推論アルゴリズム判断する必要がある. そこで各アルゴリズムの傾向を把握するための実験評価を行ったので紹介する.

5.2 確率推論アルゴリズムの評価実験

BayeoNet で利用できる確率推論アルゴリズムについて性能評価を行った実験評価を示す. Loopy BP は最近その性質が注目されている近似アルゴリズムであり, メモリ消費が少なく, 実行速度が速いという特長がある. しかしその一方で解の精度や収束性には依然不明な点が多い. そこで次の各項を実験により評価する.

- loopy BP の収束性
- loopy BP により得られた解と Junction tree により得られた厳密解との比較
- Systematic sampling, loopy BP, Junction tree の計算時間の比較

用いるベイジアンネットとしては, ノードの数を変えながらランダムに生成したグラフ構造の中で適当なものを選び, さらに証拠状態の割合・CPT のエントロピーなどを次のように条件を変えて実験を行った.

- 厳密解がわかっている multiply connected なグラフ構造に限定して下記のパラメータについて, 組み合わせ的にモデルを生成し評価を行う.

- ノード数: 20 ~ 300
- 証拠状態の含有率 (Ev): 10 ~ 90%まで変化
- ネットワークの複雑性 (Nw): 一つの子ノードに結合する親ノードの数で制御
- CPTのエントロピー (En): 各ノードの CPT をエントロピーの平均値が 0.1 から 0.9 まで一定の値になるようにした上でランダムに生成する .

LoopyBP アルゴリズムの解は、用いるグラフが持つループ構造と証拠状態の割合により、結果は以下の3つに分類される .

- 厳密解に高い精度で収束する (準厳密解)
- 厳密解とは一致しないがある解に収束 (非厳密解)
- 一定周期で振動し一定値に収束しない (非収束)

また表 2 のような傾向が見られた .

	En 小	En 大	Ev 多	Ev 少	Nw 小	Nw 大
準厳密解	-	-	少	多	多	少
非厳密解	-	-	多	少	少	多
非収束	多	少	多	少	少	多

表 2: 様々なネットワークにおける LoopyBP の傾向

条件付確率のエントロピーが小さいと非収束となることが多いが、解の精度にはあまり影響が見られない . また証拠状態が少ない場合や、ネットワークの複雑度が増大すると準厳密解が減少する傾向がある . これは multiply connected なネットワークでは変数間の相互作用が強いと解の収束性や精度に悪影響を与えるためと考えられる .

各アルゴリズムの計算速度の比較は表 3 のようになった . ノード数が多くなると LoopyBP は圧倒的に高速である . Junction tree はノード数 300 では消費メモリが増大して計算が実行不可能となった .

ノード数	LoopyBP	Junction tree	Sys. Sampling
20	119ms	112ms	445ms
50	314ms	997ms	1845ms
100	2.283sec	10.820sec	4.197sec
300	4.765sec	実行不可能	20.367sec

表 3: 実行速度比較 : Pentium III 975MHz,512MBmem

以上の実験から Loopy BP は大規模なネットワークに対しても非常に高速であり、メモリ消費も少ないというメリットがあるが、なお解の精度や収束性に問題のある場合があることがわかる . Junction tree アルゴリズムではグラフ変換のための計算時間が膨大になり、ノード数が多いと実行が不可能な場合もある . Systematic sampling では小規模なネットワークの場合ではサンプル数を十分とることにより高い精度で解が得られるが、大規模なネットワークでは junction tree と同様、メモリ消費計算時間が膨大になり推論の実行は難しい . 現在、実験評価により得られた傾向を分析し、LoopyBP の特長を生かしつつ、解の精度と収束に関する問題を解決するための新規アルゴリズムの検討を行っている .

6. ベイジアンネットの応用

これまで見たように、ベイジアンネットの特徴的な情報処理は次の二つである . 一つは観測した変数群から未観測の対象の確率分布を計算する確率推論であり、もう一つはそのためのモデルを統計データから構築する統計的学習である . この両者が組み合わさることで、従来の固定的な計算を

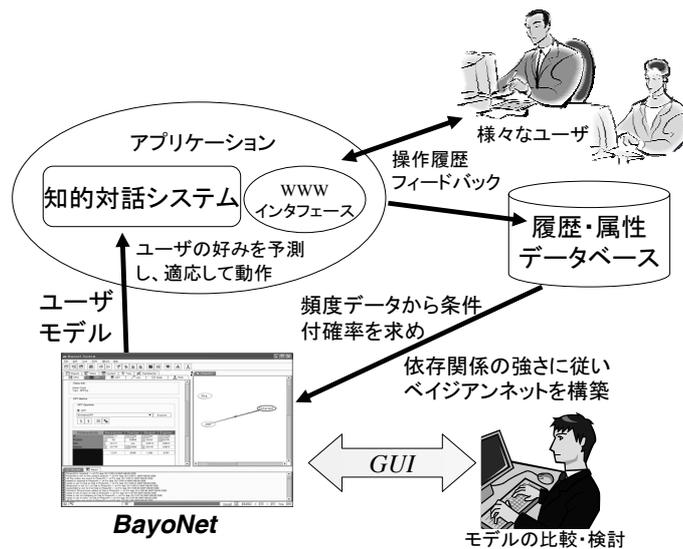


図 8: ユーザの嗜好性などを自律的に学習する応用システムのイメージ

超えた自律的な情報システムが実現できると期待される。これは、まず実環境で得られるデータから計算モデルを自動的に構築し、さらにこれを利用し予測や推論、最適な制御を行うシステムである。さらにその動作結果とあらたな観測をあわせてデータとして集積することでモデルをさらに洗練させて、システムの動作を改善する、という一連の手続きのループが実現できることが、ユーザの利用状況の変化などに対応できる自律性のポイントである (図 8)。

またこの時、確率として不確実な対象が扱えることも従来の古典的な計算機にはない重要なポイントである。つまり大量の統計データの中に存在する統計的にはある程度の頻度で成立しているような漠然とした知識を抽出し、不確実な対象に対して何%の確率で起こり得るといった予測や平均的に最適になるような制御を行うといったロバストな情報処理を実現することが、これからの知的情報処理の大きな課題である。

こうしたベイジアンネットの特徴を活用する上でも、計算機が高速になり、実用的なベイジアンネットソフトウェアが利用できるようになった意義は大きい。これまで個々に実装が必要だった高度な確率推論やモデル構築機能を、外部から呼び出せる API によって簡単に利用可能になったことで、ベイジアンネットを応用したアプリケーションシステムの開発効率が飛躍的に向上するのである。ここではそのようなベイジアンネットを応用した例をいくつか紹介する。

これまでベイジアンネットの応用としてもっとも成功しているのは複雑なシステムの障害診断がある。また最近ではユーザに適應するソフトウェアエージェントのためのユーザモデリングに利用する例なども増えてきている [24]。

そこで、故障診断の例としてはすでに実務レベルとなっている Hugin を用いた Hewlett-Packard 社のプリンタ障害診断と Dynasty 社の汎用トラブルシューティング用ソフトウェアパッケージを、ユーザモデリングの例としてマイクロソフト社で研究されているプロトタイプソフトウェアを例にあげて解説する。また最後に現在我々が進めているデジタルヒューマン研究についても紹介する。

6.1 障害診断への応用

ベイジアンネットのソフトウェアを開発している Hugin Expert 社と Hewlett-Packard 社のカスタマーサポート R&D は Systems for Automated Customer Support Operations (SACSO) プロジェクトという共同開発により、HP 社のプリンタに関する障害診断・発見システムを開発した [26]。他にも障害診断システムへのベイジアンネットの応用は NASA や Intel, Nokia 社をはじめ多くの例が

ある [30] が、この SACSO プロジェクトの場合は、プリンタのような民生品で、エンドユーザがアクセスするカスタマーサポートにおけるシステムとして実用化した点が目を引く。

また Hugin の API を利用した応用例としては Dynasty 社の WWW サービスシステム構築用のアプリケーションソフトウェア群がある。これを用いて実際に作成されたアプリケーションとしては対話型の医療系の診断システムがある。これは救急医療の窓口を訪れた患者の症状から正しく状況を判断し、緊急医へ情報提示を行うもので、専門外の症例に対する応急対応、緊急性の有無の判断、最適な専門部科がどこであるか、などの緊急時の判断を支援するシステムである。このソフトウェア群を適用することで、推論エンジン部のベイジアンネットを問題領域に応じて適切に構築し、それに合わせた入出力インタフェースを組み合わせれば、短期間で本格的なアドバイス型障害発見・診断システムが構築できると期待される。

6.2 ユーザモデリングへの応用

ベイジアンネットを用いたユーザモデリングの実証的なプロジェクトとして先駆的なものはマイクロソフトの Lumière Project [27] である。ベイジアンネットはある機能がユーザにとって適切である確からしさを推定するために用いられる。同様のアプリケーションとして Windows 上の代表的なソフトウェア Outlook をもじった LookOut という、1998 年頃からマイクロソフトリサーチの社内ですべて試されているプロトタイプを紹介する。LookOut は Windows のメールソフトである Outlook を使うユーザの操作をモニターし、ユーザが新しいメールを開くとそれを検出して、そのメールの内容を読み取り、スケジュール調整を行うためにカレンダーを開くかどうか、またいつ、どのようにユーザ支援動作を行うのが最も良いかなどを判断する。

LookOut がメールの内容を読み取る時の確率推論は次のようになる。それぞれのメールが到着すると LookOut はユーザが Outlook のカレンダーとスケジュールサブシステムを使う可能性を確率として計算する。この確率はメールヘッダの内容（メールの差し出し人や時間など）とテキスト文中に含まれるパターンによる確率推論で計算できる。この確率と、サービスを提供した場合にユーザがどれだけありがたいと思うかという効用、動作のコストを考慮した期待効用が最適なアクション（場合によっては何もしないこと）などを決定する。

各アクションの期待効用の値によって、ユーザを支援する動作を行うか、カレンダーの表示とスケジュール入力を自動的に実行するかどうか、またはその必要がないので邪魔をしない、といったいくつかの候補の中から最良の動作を決定する。この推論の中で LookOut は人がミーティングなどを記述するとき用いる典型的な語句、“Friday afternoon”, “next week”, “lunch” などを検出する。それらの表現から想定される様々な場合の可能性を考慮して、スケジュールの決定に関連する可能性を確率として計算し、適切なユーザ支援動作を決定する。

もしも特定の日時と相手とのミーティングを行う確率が非常に高い場合には、その内容に基づいたサブジェクトと内容を入力済みの Outlook が起動し、この内容で良いか、あるいは修正するかをユーザに確認する。もし他のスケジュールとぶつかっていたら代替案として他の時間を探してからそれを表示する。一方、メールの情報が特定の日時を示している確率や、ミーティングを表す確率が低く、LookOut が起動するスケジュールリングサービスの期待効用が低い場合には、単にもっとも必要と思われる時期のカレンダーを表示するだけにとどめ、それ以上の余計な動作はしない。

ユーザが何をしたいと望んでいるかを確率と効用を通じてモデル化しているところ、そして観測だけからは確定できない事象に基づく不確実な意思決定を確率推論を用いることで実行しているところが、このシステムのポイントである。また、支援動作をキャンセルするなどのユーザからの適合性フィードバックを通じて、システムが学習できる点も非常に先進的である³。

³従来の Windows 製品にあるイルカが余計なお世話と感じていたユーザにとっても、必要性を判断して本当に必要な時だけ現われる支援機能と、適合性フィードバックにより自分に適応する機能であれば好ましいと感じられるだろう。

こうした自律的な判断の信頼性が向上すれば、複雑な機能を提供するシステムの場合でもユーザは比較的単純な操作を行うことでも所望のサービスを受けられるようになると期待され、運転中のドライバーに対するアシストなどではとくに重要な技術となるだろう。

6.3 デジタルヒューマン：ベイジアンネットによるヒューマンモデリングの応用

情報システムが以前より社会の中で人と密接に関わるようになり、人間がシステムに合わせるのではなく、システムが人間の方へと合わせる人間に優しい情報技術を実現することが益々重要な課題となっている。しかし情報技術の進歩の一方で、人間に対する理解が十分でないことによる困難が依然としてある。システムが人間のために動作するのであれば、システムがそのユーザである人間と無関係に一方的に動作するわけにはいかない。システム全体としてみると人間の行動や反応もそのシステムの一部として考えなければならないのであるが、人間がある状況ではどのように行動し、どのような状態にあるのか、ということについて十分理解されているとは言えない。例えばシステムが機能やサービスをどのように提供すれば良いのかを判断するためにはユーザのそれまでの行動からその意図や要求を汲み取ることが必要なのであるが、これをシステムが予測することはいかに計算機が高性能になった現在でもなお容易な問題ではない。

こうした問題への一つのアプローチとして我々はデジタルヒューマンという概念を提唱し、研究を行っている。人に関する情報を計算機内で取り扱うことを可能するために、人間の本質的な機能をシステムの一部として計算機上にモデル化する、このモデルがデジタルヒューマンである。このデジタルヒューマンを計算機上に実現することで、人にとっても最適な動作や制御を実現することが可能となる。

このデジタルヒューマンを構築するためのモデル化をここではヒューマンモデリングと呼んでおく。ヒューマンモデリングにおいては事前知識の取りこみと観測された大量の統計データからの学習が重要な役割を果たす。そこで我々はこうしたヒューマンモデリングのためにベイジアンネット技術の応用技術を独自のソフトウェアを用いて研究している。事前知識の利用、統計的学習の他にも、構築したモデルが情報システム稼動時にそのまま利用でき、計算機上で確率推論を実行できることからベイジアンネットによるヒューマンモデリングの有用性は非常に高い。

ここでとくに重要なことはベイジアンネットは非常に基礎的であるため幅広い問題に適用できる可能性を持つが、解決して意味のある重要な問題に適用しなければ、十分な社会的意義を果たしたとは言えない。こうした問題意識から慎重に応用対象となる問題を選び、現在次のようなプロジェクトを共同研究を軸として進めている。

- 1) 家庭内でのヒューマンモデリング
- 2) 自動車運転における運転手のモデル
- 3) 手術中の医師と患者の相互作用のモデル
- 4) 電子商取引やマーケティングにおける顧客モデル

1) は家庭内における子供の事故予防や、高齢者の介護支援、住みやすい住宅設計の支援などに貢献するものである。とくに子供の事故予防については現在セコム科学技術振興財団の助成を受け、山中医師（緑園こどもクリニック院長）らとの共同研究を進めている。2) は自動車運転中のドライバーを支援することを目的として高度な知的支援を行う自動車や、幅広い意味でのナビゲーションを行う次世代カーナビを目指して、企業との共同研究が行われている。3) については手術中の患者の心拍、血圧などの観測データから心理状態を推定するモデルの検討を専門医との共同研究により進めている [29]。また 4) については、最近電子商取引や POS システムの普及により、顧客の購買履歴の収集と、それを使ったデータマイニングが一般的になっている。このデータベースマーケティングの発展として顧客のモデル化にベイジアンネットを適用した例もすでにある [28]。ベイジアンネットの場合には顧客動向の解析だけでなく、その結果に基づく確率推論を自動推奨などの CRM や SCM システムに活用できることが特長である。目的変数として来店頻度（または確率）、購入頻

度(確率), 広告反応率などをとり, それらをもっとも良く予測できるような説明変数からなるベイジアンネットモデルを構築するのである. それにより顧客がとる行動, 店への来店, 物品の購入, 広告へ興味を示す, などの確率を予測したり, その行動をとる確率が大きくなる顧客層を見分けることができるようになる. これについては, 現在産総研ベンチャー開発戦略センターと共同して, ベンチャー創出モデルプロジェクトの一つとして計画が進められている.

7. おわりに

計算機的高速化と, 大量データが現実利用可能となってきたことで, 情報システムが実際の問題領域における観測データから自律的に学習することも現実的になってきた. これまでの情報処理システムはプログラマーが設計した範囲でしか機能できなかったのであるが, 統計的学習と確率的な計算アルゴリズムを用いることによってシステムの稼動中に得られた統計データにしたがいその動作を最適化し, 様々な状況に適應しながら知的な処理が行える可能性が出てくるのである.

身近な例では, 送られて来たスパムメールを学習して自動分類を行うプログラムや, インターネット検索システムである Google の WWW ページのランク付けなどにもベイズ確率に基づく確率計算が利用されている. またインターネットの発達により, ユーザの年齢, 職業などの基本属性, 嗜好性, などと, これまでした買物や質問のような行動履歴などのデータなどが大量に集積され, これらの因果関係を積極的に活用したサービス提供や問題解決, 有用な知識の発見などを行うことも期待されている. これは従来から決定木などを用いたデータマイニングによる解析が行われている分野でもあるが, ベイジアンネットのように解析結果を知識モデルとして構築し, 推論まで自動的に実行できることはさらに大きなアドバンテージである.

今後はさらに新しいセンサ技術や, ネットワーク環境, 有機的に結合したデータベースなどの発展と連携した形で, ベイジアンネットが統計的学習に基づく有用な知識のモデル化の一つとして活用され, その上での確率的推論技術が計算機の新しい利用技術として確立されることが期待される.

参考文献

- [1] J. Pearl: “*Probabilistic reasoning in intelligent systems*”, Morgan Kaufmann (1988).
- [2] F.Jensen: “*An Introduction to Bayesian networks*”, University College London Press (1996).
- [3] E.Castillo, J. Gutierrez and A.Hadi: “*Expert systems and probabilistic network models*”, Springer-Verlag (1997).
- [4] R.G. Cowell, A.P.Dawid, S.L. Lauritzen and D.J. Spiegelhalter: “*Probabilistic Networks and Expert Systems*”, Springer-Verlag (1999).
- [5] M. Jordan: “*Learning in Graphical Models*”, The MIT Press (1999).
- [6] S. Russell and P. Norvig: “*Artificial intelligence, A modern approach*”, Prentice Hall 1995, 邦訳: 古川康一監訳, “*エージェントアプローチ人工知能*”, 共立出版 (1997).
- [7] 本村陽一, 佐藤泰介: “*ベイジアンネットワーク-不確定性のモデリング技術-*”, 人工知能誌, vol.15, no.4, pp.575-582 (2000).
- [8] 本村陽一: “*ベイジアンネットによる確率的推論技術*”, 計測と制御, vol.42, no.8, pp.649-654 (2003).
- [9] ベイジアンネットセミナー BN2002: “<http://staff.aist.go.jp/y.motomura/bn2002/>”, (2002).
- [10] ‘*ベイジアンネットセミナー BN2003*’, “<http://www.bn2003.org/>”, (2003).
- [11] Cheng, J. and R.Greiner, : Comparing Bayesian Network Classifiers, *proceedings of the fifteenth conference on uncertainty in artificial intelligence* (1999).
- [12] Cheng, J., Hatzis, C., Hayashi, H., Krogel, M., Morishita, S., Page, D., and Sese, J.: KDDD cup 2001 report, *ACM SIGKDD Explorations*, Vol. 3, No. 2 (2002).
- [13] K. P. Murphy, Y. Weiss, M. I. Jordan: Loopy Belief Propagation for Approximate Inference: An Empirical Study, *Proceedings of Uncertainty in AI* (1999).

- [14] 樺島祥介, 田中利幸: 統計力学と情報処理, 電子情報通信学会誌, 83-8, pp.630-636, (2000).
- [15] Y.Kabashima and D.Saad: Belief Propagation vs. TAP for decoding corrupted messages, Europhys. Letter, 44-5, pp.668-674 (1999).
- [16] T.Tanaka: A theory of Mean Field Approximation, Advances in Neural Information Processing Systems, 11, pp.351-357 (2000), MIT Press.
- [17] Shiro Ikeda, Toshiyuki Tanaka, and Shun-ichi Amari: Information geometrical framework for analyzing belief propagation decoder, , Advances in Neural Information Processing Systems, 14, pp.407-414, (2002).
- [18] 田中和之: 確率推論に対する統計力学的アプローチ-クラスター変分法と信念伝搬アルゴリズム-, 計測自動制御学会 システム・情報部門学術講演会 SSI-2002 予稿集 (2002).
- [19] G. Cooper and E. Herskovits: "A Bayesian method for the induction of probabilistic networks from Data", *Machine Learning*, vol.9 (1992), 309.
- [20] J. Suzuki: "A construction of Bayesian networks from databases based on an MDL principle", *Proc. of the 9th Conference on Uncertainty in Artificial Intelligence* (1993), 266.
- [21] M. Ramoni and P. Sebastiani: "Parameter estimation in Bayesian networks from incomplete data", *Intelligent Data Analysis Journal*, vol.2, no.1 (1998).
- [22] R. G. Cowell: "Parameter learning from incomplete data for Bayesian networks", 7th Int. workshop on Artificial Intelligence and Statistics (1999).
- [23] Y. Motomura, "BAYONET: Bayesian Network on Neural Network", Foundation of Real-World Intelligence, pp.28-37, CSLI California, (2002).
- [24] 本村陽一: ベイジアンネットワークソフトウェア, 人工知能学会誌, vol.17, no.5, pp.559-565 (2002).
- [25] 本村陽一: ベイジアンネットワークソフトウェア BayoNet, 計測と制御, vol.42, no.8, pp.693-694 (2003).
- [26] F. Jensen, U. Kjarulff, B. Kristiansen, H. Langseth, C. Skaanning, J. Vomlel and M. Vomlelova: "The SACSO methodology for troubleshooting complex systems", *Artificial Intelligence for Engineering Design, Analysis and Manufacturing (AIEDAM)*, vol.15, pp.321, (2001).
- [27] E. Horvitz, J. Breese, D. Heckerman, D. Hovel and D. Rommelse: "The Lumiere Project: Bayesian User Modeling for Inferring the Goals and Needs of Software Users", in *14th National Conference on Uncertainty in Artificial Intelligence* (1998).
- [28] Sebastiani, P., Ramoni, M. and Crea, A., Profiling Your Customers using Bayesian Networks: A tutorial exercise and the Bayesware submission to the KDD Cup 99 (1999).
- [29] 酒井, 持丸, 横山: 局所麻酔下副鼻腔手術における手術進行と患者反応モデル構築のための生理指標解析, 生体医工学, vol.41, no.4, pp.483-492 (2004).
- [30] <http://excalibur.brc.uconn.edu/~baynet/fieldedSystems.html>

著者 略歴



本村 陽一(正会員)

1993年電気通信大学大学院電子情報学専攻修士課程修了。同年通産省工業技術員電子技術総合研究所に入所。現在、独立行政法人 産業技術総合研究所デジタルヒューマン研究センター主任研究員。確率モデルの統計的学習・確率推論, 知的情報システムなどの研究に従事。日本行動計量学会, 電子情報通信学会, 人工知能学会, 日本神経回路学会, 認知科学会各会員。