

ベイジアンネットワーク構築ソフトウェア: BAYONET-PRO

本村 陽一*

産業技術総合研究所 情報処理研究部門

Abstract: ベイジアンネットワークを各種の問題に適用するためには、適切なベイジアンネットワークモデルを構築する必要がある。これまでの多くのベイジアンネットワークソフトウェアでは主に確率推論部分の実装に注意が向けられており、利用者がベイジアンネットワークを構築することはそれほど容易であるとは言えなかった。そこで、ベイジアンネットワークを構築するために重要である、変数の決定、グラフ構造の選択、条件付き確率の獲得、のそれぞれの処理を支援するベイジアンネットワーク構築システムの開発を進めてきた。さらに Wizard 形式の対話的 GUI によって、誰もが容易にデータを解析し、ベイジアンネットワークを構築できるような拡張を加え、本格的な実用化のために提供できる体制を整えたので今回その概要を紹介する。本ソフトウェアの利用希望者は (株) 数理システムから入手することが可能である。

1 はじめに

ベイジアンネットワークにより確率推論を行うためには、問題構造を良く反映した適切な変数、グラフ構造、条件付き確率を持ったモデルを構築する必要があるが、多くのベイジアンネットワークシステム (例えば Hugin) ではこのモデルの設計をほとんど利用者に委ねているために、確率推論を行う以前のモデル化の段階で困難が生じることが多い。例えば、実際にベイジアンネットワークを構築しようとすると、注目する確率変数をどのように定義し、また変数間の主要な依存関係をどのように選ぶか、また条件付き確率を頻度データから求める時にデータ数が十分でない場合の問題をどう解消するか、などの問題がある。またユーザが構築したモデルが不適切なために、確率推論の精度が低下することもしばしば起こる。そこでベイジアンネットワークを実際に応用したいというニーズが増えるに当たって、適切なモデル、確率推論の精度を高めるようなモデルを構築する作業を支援するシステムの必要性も高まっている。

本稿では適切なベイジアンネットワークモデルを構築するために、確率変数の決定、グラフ構造の決定、条件付き確率値の決定、の3つを支援するために開発した BAYONET の最新版、BAYONET-PRO について紹介する。

2 ベイジアンネットワーク

ベイジアンネットワークとは (1) 確率変数と (2) 確率変数間の条件付依存関係、(3) その条件付き確率の3つによって定

義されるネットワーク状の確率モデルである。(1) はノード、(2) はノード間に張った有向リンクで表され、リンクの先に来るノードを子ノード、リンクの元にあるノードを親ノードと呼ぶ。(3) は親ノードがある値をとった時に、子ノードがある値を取る条件付き確率のことである。ある子ノード X_j に注目した時、親ノードが複数あるときの親ノードの集合を $\pi(X_j) = \{X_1, \dots, X_i\}$ と書くことにする。この時の変数 X_j の値が親ノードの変数の値によって影響をうけるが、それが非決定的、つまり親ノードの値だけではよらない不確実性がある時、この関係を子ノードの変数 X_j について親ノードの値を条件とする条件付き確率を、

$$P(X_j | \pi(X_j)) = p_k \quad (1)$$

のような、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列挙した表 (条件付き確率表) の形で表す。

このベイジアンネットワークを使うことで、(1) 一部の確率変数の値が観測できた時、未観測の変数についての確率分布を求めたり、(2) 確率が最も大きい値を変数の予測値として得ることができる。この計算を確率推論と呼ぶ。

車の故障診断の例を考えよう。エンジンが始動するためには点火系、燃料系、それにセルモータが勢いよく回ることが必要である。もしエンジンがかからない時、故障の原因を探るために我々はどのように考えるであろうか。真の原因は部品の不具合を直接観測するまではわからないとしても、バッテリーの古さや、燃料計の値、セルモータの回転する勢い (音) などから事前におおよその見当がつけられることもあるのではないだろうか。こ

* 〒 305-8568 茨城県つくば市梅園 1-1-1 つくば中央第二, tel: 0298-61-5415, e-mail: y.motomura@aist.go.jp

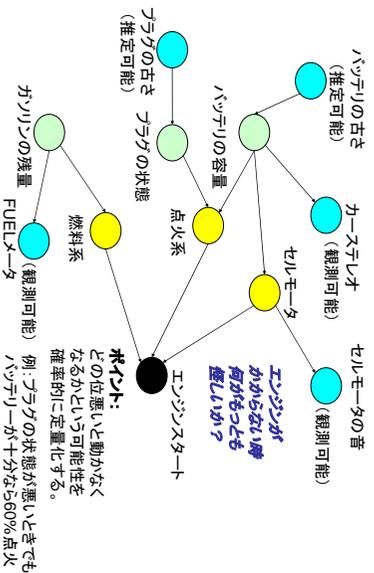


図 1: ベイジアンネットワーク: 車の故障診断の例

れをベイジアンネットワークでモデル化したものが図 1 である。バッテリーの容量が十分であればカーステリオの音がきちんと聞こえている、などの定性的な依存関係をグラフ構造で表し、さらにその上で、プラグがどの位古くなるか状態がどの程度劣化する可能性があるかといった定量的な依存関係を各条件付確率分布で表す。このベイジアンネットワークに運転席から観測したり推定できた情報、(例えば燃料計は Full であるとか、ラジオは聞こえているなど) を代入して、それぞれの要因に障害が発生している確率を計算する。その結果、もっとも不具合のある可能性が高い要因を中心に故障を診断することで最小のコスト (診断回数, 時間) で問題解決をはかることが可能となる。

3 ベイジアンネットワークの構築

さて、このような障害診断の例のように不確実性を含む重要な問題領域は多い。そのような様々な問題にたいして、対象についてベイジアンネットワークを構築し、その上で確率計算を実行する、という統一的な方法論によって全て同じ様に取り扱うことができるのがベイジアンネットワークのメリットであろう。例えばバイオインフラストラクチャ分野においても、DNA の解析データから遺伝子間の依存関係を表し、重要遺伝子を特定することで創薬へ応用する研究や、大量の顧客情報からユーザの要望や特性などを予測するユーザモデリングへの応用なども最近注目されている。

ところで、実際にベイジアンネットワークを応用しようとした時に最初にぶつかる壁は問題に応じて、どのようなベイジアンネットワークを構築すれば良いかという問題である。そこで興味の対象となる変数をもっとも良く予測し、本質的な問題構造を的確に表すベイジアンネットワークを構築する手法が重要である。また幅広い様々な分野において、ベイジアンネットワークのモデル構築手法を、誰もが、容易に、すぐに利用できるように、ソフトウェアの形で整備する

ことはこれまでの基礎的な研究成果を実用的な応用に結び付けるためにも非常に意義の深いことである。

4 BAYONET

筆者はこれまでデータベースや変数間の規則を与えること、それにもっとも良く適合するベイジアンネットワークを構築するソフトウェア、BAYONET の研究開発を進めてきた。BAYONET とは JAVA による実装としては世界でも最初期のベイジアンネットワークソフトウェアである 1996 年に作成されたバージェン [本村 96] を始めとし、以後通産省の RWC プロジェクト [麻生 02] などいくつかの研究プロジェクトの中で機能拡張が進められてきたソフトウェアシリアーズの総称である。本来は、筆者の行う知能システム・機械学習研究の中で必要とする機能を実現するためのプロトタイプとして開発が行われたものである。当時まだベイジアンネットワークのモデル構築機能を持つソフトウェアが一般的でなかったことから、SQL データベースに格納された大量のデータとのインタフェースを始め、ベイジアンネットワークを構築するためのモデル構築機能をほぼ揃えており、海外のベイジアンネットワーク構築ソフトウェアと比べても遜色はない。

ベイジアンネットワークを実際に構築する手順は以下の通りになる。

- モデルで使用する確率変数, X_j を決定しノードを作成する。
- 変数間の依存関係にしたがって、親ノードから子ノードにリンクを張っていく ($\pi(X_j)$ の決定)。
- 変数間の依存関係を定量的に表す条件付き確率表, $CPT: P(X_j | \pi(X_j))$ を決定する。

BAYONET では、データベースに格納された統計データを検索し、モデルとの適合性を確認しながら変数、ネットワーク構造、条件付き確率のそれぞれを求めていくことになる。とくに最新版の BAYONET-PRO では Wizard 形式の GUI により、必要事項を選択していくだけで、データベースや変数規則から最適なベイジアンネットワークを構築できるようになった。一方、アニュアルモードでは、ユーザが適切に変数選択を行えるように、結果として構築したモデルの良さや、条件付き確率表を可視化してモデルを洗練させることができる。またネットワーク構造はモデル選択基準として尤度、情報量基準 MDL, AIC などを用いたアルゴリズムや、 $C4.5$, $K2$ といったアルゴリズムによって自動的に行われる。これらのモデル選択アルゴリズムは利用者が自由に選んだり、追加することも可能である。条件付き確率についてはデータベース

ニューラルネットの汎化能力を利用した 欠損データの補完 (疎データへの対応)

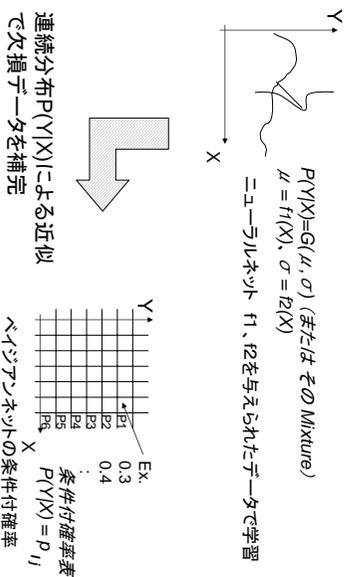


図 3: ニューラルネットによる条件付き確率の学習

に格納された頻度から求められるが、とくに SQL データベースと連携し、メモリ上に読み込むことの困難な大量のデータからでも計算できるのが BAYONET の特長である¹。

その他にも、BAYONET 独自の特徴として、ニューラルネットにより条件付き確率を学習、補完する方法を導入している [Motomura 97, Motomura 00b]。この手法は実際のデータの中には、しばしばデータに偏りや欠損値があるという問題や、連続値や多次元ベクトルの扱いが従来のベイジアンネットソフトウェアでは十分でないという問題を解決するために開発された。これは条件付き確率表を展開した連続空間上で階層型ニューラルネットが張る連続的な確率分布を考え、与えられたデータから学習したニューラルネットの汎化 (近似) 能力によって、データにない欠損している項目についての条件付き確率を推定するものである (図 3)。

また JAVA のリフレクション機能を利用して利用者が階層型ニューラルネットの代わりに独自の学習モジュールを追加したり、構築したベイジアンネットを使った確率推論を他のアプリケーションから TCP/IP 接続を経由して利用するための API を備えるなど、拡張性にも優れている。構築したモデルは先に紹介した確率推論用のソフトウェア、Hugin と互換性のあるファイナルフォーストにより、これらの推論ソフトウェアでそのまま利用することができる。これを活かして BAYONET と Hugin を組み合わせた次世代コールセンター用のアプリケーションを開発するプロジェクトも始動している [佐藤 02]。

現在、最新版である BAYONET-PRO は製品化が決定し、一般への供与、商用利用が可能になった。またもちろん研究用途のために共同研究や技術研修の中で無償で利用することや、学術・教育用に限定した安価なライセンス

¹この時、Linux など幅広く利用できるフリーのデータベース、PostgreSQL や商用の Oracle など多くのデータベースを使用できるのも特長である。

BAYONET の応用例

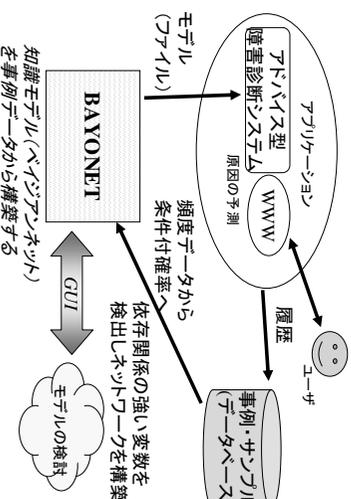


図 4: BAYONET の応用例

センサ、評価版なども準備しているので、利用を希望される方は筆者²、または (株) 数理システム³へ詳細を確認していただきたい。

5 まとめ

不確実性を含む問題領域において確率モデルの応用を進めるために、データベースと連携したベイジアンネットシステムを開発した。とくにこのシステム独自の特徴として、一般によく使われる SQL データベースと連携することで大量のデータに対して SQL 検索コマンドを用いた高度な操作を行い、適切な変数選択を支援する。また情報量基準にしたがって最適な局所モデルを自動的に選択することでグラフ構造を決定していく仕組みを導入している。さらにニューラルネットや回帰モデルなどを用いて学習することで欠損データがある場合やデータ数が十分でない場合でも条件付き確率値を近似することを可能にした。これらの特長によって、ユーザが問題構造に即した適切なベイジアンネットモデルを構築することが容易になる。本システムでは、条件付き確率の学習モデル、構造選択アルゴリズムなどを交換、拡張可能にしており、利用者が新たなモデル、アルゴリズムを追加することも容易である。これによって新たな理論的手法を実用的な規模の統計データにより短期間で評価することも可能である。このソフトウェアを多くのユーザが様々な問題に応用することで、ベイジアンネットの実用化に寄与できることを期待している。

参考文献

- [麻生 02] 麻生英樹: 特集「RWC-実世界知能」理論・アプリケーションサイズ基盤, 人工知能学会誌, Vol. 17, No. 2, pp. 124-129 (2002).

²<http://staff.aist.go.jp/y.motomura/bayonet/>
³<http://www.nlsi.co.jp>

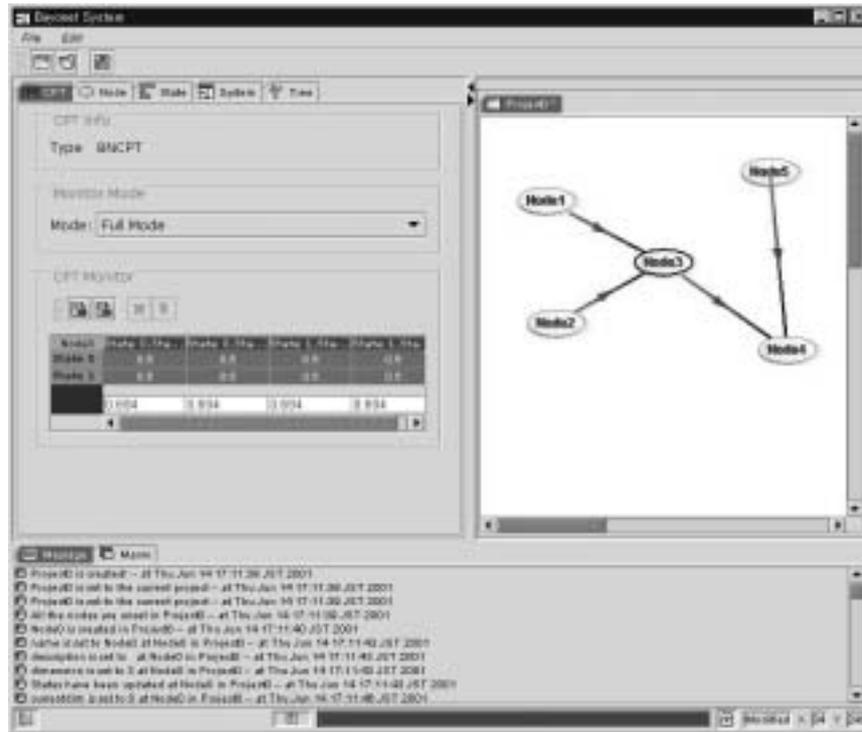


図 2: ベイジアンネット構築システム: BAYONET-PRO

[佐藤 02] 佐藤, 橋本, 本村: ベイジアンネットによるアドパイス型問題解決システム, ベイジアンネットセミナー BN2002, (2002).

[本村 96] 本村, 麻生, 原, 赤穂, 松井: 事情通口ボットにおけるニューラルベイジアンネットの学習, 人工知能学会第 4 回情報統合研究会 Technical Report SIG-CII-9601-04 (1996).

[Motomura 97] Motomura, Y. and et.al., : Bayesian Network that Learns Conditional Probabilities by Neural Networks, in *Proc. of the Int. Conf. on Neural Information Processing and Intelligent Information Systems* (1997).

[本村 00a] 本村, 佐藤: ベイジアンネットワーク-不確定性のモデリング技術-, 人工知能学会誌, Vol. 15, No. 4, pp. 575-582 (2000).

[Motomura 00b] Motomura, Y. and Hara, I.: Bayesian Network Learning System based on Neural Networks, in *Proceedings of International Symposium on Theory and Applications of Soft Computing* (2000).