

# ニューラルネットにより条件付き確率を学習するベイジアンネットワーク

## Bayesian Network that Learns Conditional Probabilities by Neural Networks

本村 陽一\*

Yoichi Motomura

**Abstract:** In this paper, Bayesian network that learns conditional probabilities by neural networks is introduced. Bayesian network is a probabilistic model for probabilistic reasoning in the AI context. In order to apply this model to a real world problem which includes non-linear, multi-dimensional and continuous random variable, we use neural networks as flexible representations of conditional probabilities of Bayesian network. The model is constructed and learnt from samples collected in the real environment. Learning conditional probability is realized by the back propagation algorithm. However, effective constructing graph structure of Bayesian network is still open problem. For experimental evaluation of graph learning, we develop Bayesian network software, "BAYONET". This simulator can connect to major database software, so practical large database can be tractable. The latter half of this article explains this software for structure learning of Bayesian network.

## 1 まえがき

実環境におけるロボットなどの知的情報処理システムに望ましい動作を実行させるために、不確定要素を含んだ各種変数の値を推定しそれに基づいて期待される効用を最大とするような出力を選択する必要がある。とくに複雑なタスクを処理するためには多数の要因からなる比較的複雑な関連性や多様な状況依存性を扱うことが必要であり、一様な分布モデルだけでは対象を記述しきれない場合がある。こうした場合にモデルを複雑で非一様なものに拡張してその上での確率的な演算によって情報処理を行なうことが考えられる。有向グラフ構造を持つ確率モデル, Bayesian network もその一つである。

本稿では Bayesian network を実環境における柔軟な情報処理システムに適用することを目的として、確率変数の連続値化, 条件付き確率の非線形化を実現するためにニューラルネットを用いて Bayesian network を構成するモデルを紹介する。また様々な実用的な問題領域に応用するために実際に観測されたサンプルのデータベースから、推論に用いる Bayesian network を学

習する枠組が重要である。Bayesian network の学習についての課題と、そのために開発した研究用ソフトウェア BAYONET についても述べる。

## 2 Bayesian network

各確率変数をノードで表し、その間の確率的依存関係を矢印で結んでいくとある領域についての関連知識を非循環性の有向グラフ (Directed Acyclic Graph: DAG) として図示することができる。こうして全ノードの依存関係を示す DAG と各子ノードでの局所的な依存関係を定量的に表した条件付き確率の 2 つで定義される確率モデルが Bayesian network である [Pearl 88, CGH 97, RW 96, 鈴木 94]。

確率変数を  $X, Y$  のように書きそれぞれの実現値を  $x, y$  とする。この  $X$  と  $Y$  の間に何らかの因果関係があり、 $Y$  が  $X$  に依存している時、例えば「If  $X = x$ , Then  $Y = y$ 」というように記述することができる。実環境で生じている様々な事象を考えると依存関係は複雑にこみ入っており、「If  $X_1 = x_1, \dots, X_i = x_i, \dots$ , Then  $Y = y$ 」のように逐一完全に記述しようとすると記述量が膨大になる。そこで主たる依存関係だけに注目し、さらに不確実性を吸収するために「もし  $X = x$  の

\*電子技術総合研究所, 〒 3058568 茨城県つくば市梅園 1-1-4 tel. 0298-55-3211, e-mail motomura@etl.go.jp, Electrotechnical Laboratories, AIST, MITI, 1-1-4 Umezono, Tsukuba, Ibaraki 3058568, Japan

時、 $Y = y$ となる確率は $P(Y = y|X = x)$ のような確率化を考える。この確率化は $X = x$ となる状況のもとで、繰り返し観測できる事象 $Y = y$ の発生頻度、あるいは $Y$ が $y$ であると思われる主観的な確信度 (Degree of belief) などとしても解釈することができる。この時、 $X \rightarrow Y$ と表し、 $X$ を親ノード、 $Y$ を子ノードと呼ぶ。このような確率的依存関係を持つノード間をアークで結びネットワークを構成したものが Bayesian network になる。

ここで領域内の全ての組合せ的な関係を記述するのではなく、明示的な因果関係のない事象間は独立であると仮定し、主要な因果関係だけに限定してグラフ構造を構成することにより計算量、記述量の爆発が避けられる。つまりこれは対象を効率的に記述し近似する確率モデルになっていて、例えば  $A, \dots, Z$  の全状態の中で因果関係が  $X, Y$  の間にだけ存在する場合には全ての事象についての同時確率分布が次式右辺のように簡単化されることになる ( $P(X|Y)$  についての条件付き確率だけを考慮し、他の変数は独立なものとして確率計算が簡単になる)。

$$P(A, \dots, X, Y, Z) \\ = P(A) \dots P(X|Y)P(Y)P(Z)$$

この様に Bayesian network による確率分布の表現は

- 直観的に捉えられた依存関係の間の構造 ( 事象間の条件つき独立性 ) をグラフィカルに表現する。
- 確率的知識データの局所的な表現の組み合わせから成り、他の領域、文脈へ転用する際にメンテナンスが容易である。
- ルールベースで記述された古典的なエキスパートシステムに比べて直観的でわかりやすく、実環境での対話性に優れる。また局所的な計算のため並列分散化しやすい。

といった特長がある。最近では実環境における情報処理についての応用も盛んであり [HMW 95] 例えば、ロボティクス [Dean 91, Forbes 95], アクティブビジョン [Rimey 94] などの応用例がある。

### 3 Bayesian network の学習

Bayesian network の学習はおおまかに以下の2つのステップに分けられる。

- 変数間の依存関係を抽出して、グラフを構築する、
- 変数間の依存関係の特徴づける各条件付き確率を統計データなどから学習する、

### 3.1 ニューラルネットによる条件付き確率の学習

後者の条件付き確率の学習は例からの確率分布の推定として考えることができる。従来エキスパートシステムなどへの応用として良く使われる Bayesian network は特に確率変数は離散変数に限定し、条件付き確率を表現で与えることが多い。また連続変数の場合は線形の依存関係のみを考えた多次元正規分布として考えることが多い [GH 94]。しかし、実環境で動作するロボットが対象とするような確率変数を考えると、これらは非線形な依存関係を持つことも少なくなく、多次元ベクトル、{ 離散, 連続 } { 線形, 非線形 } の区別なく統一的に取扱えるような条件付き確率表現であることが望ましい。

そこで筆者らは確率変数の次元に応じた入出力素子数を用意した一層の隠れ層を持つ次のような3層ニューラルネット  $f(x)$  を用いて条件付き確率を表現するモデル (図1) を考え種々のサンプルデータについての学習などを行なっている [本村 97][Motoura 97]。

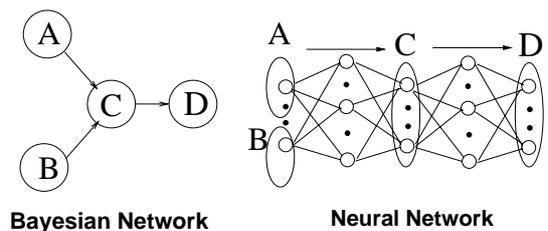


図1: ニューラルネット上に表現される Bayesian network

$Y$  が離散多値の確率変数  $Y = (y_1, y_2, \dots, y_k)$  の場合、 $k$  個のニューロン (クラスター) により  $Y$  の確率ベクトル  $P(Y = y_1), P(Y = y_2), \dots, P(Y = y_k)$  を表す。各確率変数に対応したニューロンクラスターを入力層、出力層として1つの隠れ層を持つ3層ニューラルネットを用いて  $X = x$  のもとでの条件付き確率  $P(Y|X = x)$  を次のように表現する。

$$f_k(x) = g \left( \sum_j v_{jk} g \left( \sum_i w_{ij} x_i + b_j \right) + b_k \right) \\ g(x) = \frac{1}{1 + \exp(-x)} \\ P(y_k|x) = f_k(x) / \sum_k f_k(x)$$

$Y$  が連続値の確率変数の場合、これの確率分布を  $\mu, \sigma$  でパラメタライズしたガウス分布と考えて次のように表す。

$$f_\mu(x) = \sum_j v_j g \left( \sum_i w_{ij} x_i + b_j \right) + b_\mu, (f_\sigma \text{ も同様})$$

$$P(y|x) = \frac{1}{\sqrt{2\pi}f_\sigma(x)} \exp\left(-\frac{(y - f_\mu(x))^2}{2f_\sigma(x)^2}\right)$$

統計的に獲得した  $x, y$  のデータセットから定まる条件付き確率を教師信号として与えることで、 $x$  を入力した時に  $P(Y|x)$  が得られるようにニューラルネット、 $f(x)$  を BP 法により学習する。

### 3.2 グラフ構造の学習

グラフ構造の学習についてはグラフを探索するアルゴリズム [CH 92, 植野 96] や生成したグラフを比較する際の MDL の適用 [Suzuki 93] などの研究がある。しかし Bayesian network の場合、探索空間が  $3^{n(n-1)/2}$  と比較的大きくなること、実際には対象となる問題について真のモデルを仮定することが容易でないことからその評価方法も含めてなお研究課題は多く残されている。大局的な探索は膨大であることから、局所的な構造選択による自己組織的なグラフ構造の獲得手法の研究が重要である。こうした手法の実験評価を効率良く行うために、局所的な構造を仮説として複数保持し、これらの自動選択が可能な仕組みを導入した Bayesian network シミュレータ (BAYONET)[本村 98] が開発されており、これについて次章で紹介する。

## 4 BAYONET

問題に応じて適切なモデル化を行うために、データから忠実にベイジアンネットを構成することが重要である。そこで実際的な大規模データベースに接続し、適切なグラフ構造を求め、条件付き確率を学習するためのシステムとして BAYONET を開発した (図 2.)。

本システムは、外部のプログラムとの接続性や拡張性を利用して確率理論に基づくモデル選択アルゴリズムの評価実験や電総研で開発中の事情通口ポット [松井 97] の柔らかな知識表現などの研究用途に用いられている。Bayesian network を操作するオペレーションは言語としても提供され、ファイルや TCP/IP を通じて実行することで他のシステムのための推論エンジンとしても機能する。

本システムは既に実際に利用されているような大規模なデータベースからの学習を容易にするため、多くの一般のデータベースソフトへの接続を可能にしている。また GUI による学習の条件の変更や近似誤差のモニター機能などによって学習結果の解析を支援する。これにより GUI による対話的な操作と学習の繰り返しにより最適な構造の探索を行い、データベースからの知識発見を支援することができる。

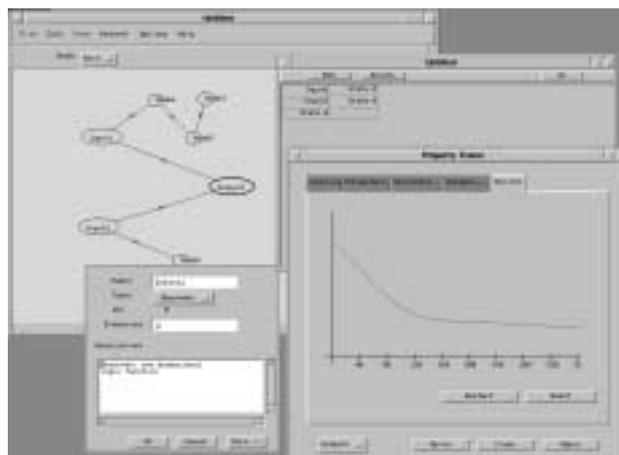


図 2: BAYONET

本システムの最新の情報やプログラムは <http://www.etl.go.jp/~motomura/BN/> でも公開されている。

### 4.1 グラフ構造の探索と動的な変更

条件付き確率の表現能力の拡大 (非線形化) によりデータへのモデルフィッティングが容易になり適用範囲の拡大とともに、とりえるグラフ構造の解空間が広がる。しかしベイジアンネットの構造学習については、現在も様々な手法が提案されつつあり研究が活発であるが、探索空間が広大であるために最適な構造を完全に自動的に得ることはなお容易ではない。

そこでユーザーが興味ある確率変数を整理することでとりえる局所構造を制約し、それらを多様な基準で比較、選択できるような仕組みが必要である。例えばニューラルネットによる条件付き確率の学習の収束 (成功) は与えた構造の妥当性の判定基準になるので、対話的なグラフ構造の修正、条件付き確率の学習、の繰り返しによってより適切なグラフ構造を探索する。またデータへのフィッティングだけでなくパラメータ数を考慮した様々な情報量基準の適用やさらに他の制約条件を導入した選択基準によりグラフ構造を決定することが考えられる。

このような適切な構造の探索を支援するため、子ノード毎に局所的な木構造 (各子ノードと (複数の) 親ノード間の木構造) と条件付き確率を複数保持し、容易に切り替える機構が本システムは導入されている。これにより、様々なベイジアンネットの構造学習アルゴリズムの性能評価に利用できる。

また対象とする問題において、変数間の依存性が観測結果や状況に応じて変化する場合には静的な単一の構造を持つ確率ネットワークでは対応しきれないことがあ

るが、このような場合にも動的なベイジアンネットの構造の変更 [GH 93] を実現するために確率推論時の親ノードの値によって局所構造を切り替えるアルゴリズムを組み込むので、従来の静的で単一な構造では表現できない複雑な問題も扱うことができる。

## 5 まとめ

Bayesian network は複雑な依存関係を持つ多数の確率変数から構成されるシステムの近似モデルであり、これによって柔軟な知的システムへの応用が期待できる。本稿では Bayesian network を連続、非線形な領域に適用するためのニューラルネットワークにより条件付き確率を表現するモデルについて述べた。また実世界に存在する多種多様なデータを格納した大規模なデータベースからデータ間の依存構造や条件付き確率を獲得して予測や推論を行うためのシステムを紹介した。とくに適切なグラフ構造を構築するための局所構造の選択のための仕組みが実装されており、これによる情報論的な観点からのモデル選択基準の提案と実験評価が今後の課題である。

## 参考文献

- [CGH 97] Castillo,E., Gutierrez,J. and Hadi,A. (1997). Expert Systems and Probabilistic Network Models, Springer.
- [CH 92] Cooper,G. and Herskovits,E. (1992). A Bayesian Method for Induction of Probabilistic Network from Data, Machine Learning, 9,309-347.
- [Dean 91] Dean,T.L. and Wellman,M.P. (1991) .Planning and Control, Morgan Kaufmann.
- [Forbes 95] Forbes,J., Huang,T., Kanazawa,K., Russell,S. (1995). The BATmobile: Towards a Bayesian Automated Taxi, Proc. of IJCAI'95, 1878-1885.
- [GH 93] D.Geiger, D.Heckerman (1993). Inference Algorithms for Similarity Networks, Uncertainty in AI,9, 326-334.
- [GH 94] Geiger,D. and Heckerman,D. (1994). Learning Gaussian Networks, Uncertainty AI,10,235-243.
- [HMW 95] Heckerman,D., Mamdani,A., and Wellman,M. (1995). Real World Applications of Bayesian Networks, Communications of the ACM.
- [Motoura 97] Motomura,Y., Hara,I., Asoh,H. and Matsui,T.(1997). Bayesian Network that Learns Conditional Probabilities by Neural Networks. Proc. of ICONIP'97 Vol. 1. pp.584-587.
- [Pearl 88] Pearl,J. (1988). Probabilistic Reasoning in Intelligent Systems, Morgan Kaufmann Publishers.
- [Rimey 94] Rimey,L.D. and Brown,C.M. (1994). Control of selective perception using Bayes nets and decision theory, *International Journal of Computer Vision*, 12, 173-207.
- [RW 96] Russel,S and Norvig,P. (1995). Artificial Intelligence – A Modern Approach –, Prentice Hall.
- [Suzuki 93] Suzuki,J.(1993).A Construction of Bayesian Networks from Databases Based on an MDL Principle, Uncertainty in AI,9,266-273.
- [麻生 97] 麻生, 赤穂, 本村 (1997). 統計的推論と AI の推論, 人工知能学会論文誌, Vol.12,2, pp.196-203.
- [植野 96] 植野真臣 (1996). 意志決定アプローチによる Bayesian Network の因果モデル構築, 人工知能学会誌, 11, 725-734.
- [鈴木 94] 鈴木讓 (1994). 確率ネットワーク, AI 白書, 113-117, コンピュータエージ社.
- [松井 97] 松井俊浩 (1997). おせっかいロボットとも呼ばれる事情通ロボットの計画, bit,Vol.29, 12, pp.4-11.
- [本村 97] 本村 (1997). Bayesian network による大規模データのモデル化について, 人工知能学会 第 4 回情報統合研究会 Technical Report SIG-CII-9707-04.
- [本村 98] 本村, 原 (1998). データベースからの学習機能を持つ確率推論システム: BAYONET, 人工知能学会 第 12 回 全国大会.