

# Bayesian network による大規模データのモデル化について

## Modeling with learge scale data by Bayesian networks

本村 陽一 Yoichi Motomura

(motomura@etl.go.jp, <http://www.etl.go.jp/~motomura/>)

電子技術総合研究所 情報科学部

Information science division, Electrotechnical Laboratory

### 1 はじめに

大規模な統計データをモデル化する上でデータに含まれる観測ノイズ、例外などの不確実性を扱うためには確率的な枠組が有効である。特に各確率変数の間に何らかの因果関係が存在すると考えられる場合、その因果関係を矢印で結びんだ有向グラフによって構成し、各変数の依存性を条件付き確率として表した Bayesian network [1] を用いてモデル化することができる。これによって部分的にしか得られない観測から、未知の変数の値を予測したり、新たな観測結果をその他の観測から検定するといったことが可能である。本稿では、大規模な統計データを Bayesian network によってモデリングするための2つの異なるアプローチ(グラフ構造中心の学習と条件付き確率中心の学習)と、それにまつわる研究例、問題点や課題について述べる。最後に筆者らの行なった、条件付き確率をニューラルネットで学習する方法の簡単な応用例として、気象庁が公開している年間の気象データを Bayesian network によりモデル化し、降雨予測を行った結果を紹介する。

### 2 Bayesian network

#### 2.1 グラフ構造を持つ確率モデル

各事象をノードで表し、その間の因果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)として表すことができる。(X → Y において、X を親ノード、Y を子ノードと呼ぶ)この DAG と条件付き確率により構成される確率モデルが Bayesian network である(教科書として[1, 2]などがある)。リアルワールドデータについての応用も盛んであり例えば、ロボットナビゲーション[3]、アクティブビジョン[4]などが有名である。

さて、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る場合、例えば「もし X =

x ならば Y = y」というように記述することができる。しかし実世界の様々な事象を考えると因果関係はとて複雑に込み入っているため複雑な問題に対して「もし  $X_1 = x_1, \dots, X_i = x_i, \dots$  ならば Y = y」のように完全に記述しようとするとやがて破綻をきたす。そこで主たる因果関係だけに注目し、さらに不確実性を吸収するために「もし X = x の時、Y = y となる確率は  $P(y|x)$ 」という確率化を考える。この確率は 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)として解釈することができる。

X と Y の全状態についての因果関係は条件付き確率分布  $P(Y|X)$  によって記述することができる。また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を抽出したグラフ構造を与えることにより計算量、記述量の爆発を避けている。これはある種の近似的なモデルであり、例えば A, ..., 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。

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

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因果関係の間の構造(事象間の条件つき独立性)を容易に表現することができる。
- 確率的知識データの部分的な表現に適しており、他の領域、文脈へ転用する際にもメンテナンスしやすい。
- 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。逆

に Neural network と比べた場合、学習によって獲得した知識は明示的に表現されるので取扱い易い。

といった特長がある。

## 2.2 Bayesian network の学習

推論に用いる Bayesian network が事前に存在することはまれであり、実際には観測により得られたデータに基づいて構成していく必要があることが多い。これが Bayesian network の学習である。Bayesian network の学習には次の2つの要素を決定しなければならない。

- グラフ構造
- 各確率変数についての条件付き確率

グラフ構造を獲得するアルゴリズムについては K2- アルゴリズム [7] B- アルゴリズム [8], ベイズ的学習 [10] や GA の適用 [12], 因果関係を抽出していく方法 [15] など様々な試みがなされている。またネットワーク構造を決定する問題は統計学におけるモデル選択の問題でもあり, AIC や MDL [13] により簡単なグラフを選択する方法も適用できる。

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。特に確率変数が離散二値の場合などでは条件付き確率は一般的には表で与えられるが、確率変数が連続値である場合などでは分布形を仮定し、パラメタ推定を行なうことでよりコンパクトな表現にすることができる。線形和 [1] Sigmoid 関数 [5] などの単純なモデルによる解析や、より一般的な条件付き確率を学習、近似する方法として Gaussian Mixture や階層型ニューラルネット [17, 18, 19] の適用、などがある。

## 3 グラフ構造中心の学習

グラフ構造  $g$  についてデータ  $D$  に対するフィッティングの良さ  $p(D|g)$  を Closed form で解き、ベイズ的な学習を行なうためにはパラメータの事後確率を事前確率と同族の確率分布とするために共役事前分布でなくてはならず、事前分布には Dirichlet 分布を用いることが行なわれている [6]。さらに Heckerman, Geiger らのグループは主にネットワーク構造をデータによって学習するのに有効な条件についての検討を行ない、以下のような条件をまとめている。 [9, 10, 11]

[定義 1] independence equivalent

確率変数間の因果関係が等価。同一の無向グラフ構造と同一の V-structure を持つベイジ

アネットワークは independence equivalent であり、学習時のグラフ選択の評価基準はこのようなネットワークにたいしては同値でなければならない。

[定義 2] v-structure

$x, y$  からのリンクが  $z$  に収束し、かつ  $x$  と  $y$  の間にリンクがない時 ( $\{x \rightarrow z \leftarrow y\}$ )、これを v-structure と言う。

[定義 3] Global independence

あるノードに関連する条件付き確率のパラメータは他のノードに関連するパラメータとは独立。

[定義 4] Local independence

そのノードの親ノードの具現値 ( $x$ ) にたいする条件付き確率  $p(Y|x)$  は、別の具現値にたいする条件付き確率  $p(Y|x')$  と独立。

客観的な評価基準によってグラフ構造を決定するために、表現する因果関係が等価 (independence equivalent) でありながら、異なる DAG については同値となるような基準となることを要請し、さらに便宜的な学習、確率伝播を容易にするために、条件付き確率が定義 3,4 を満たすことを要請すると、そのような分布族は離散の場合は Dirichlet, 連続の場合には normal-Wishert 分布に限られることが示される [11]。ここではこのように条件付き確率についての制約を設けて、グラフ構造を優先して探索するアプローチをグラフ構造中心の学習と呼ぶことにする。

ここで挙げたような条件に従うためには例えば連続変数をモデリングする場合には、線形な dependency を持つネットワークしか許されないなど実問題に適用する上では不便な制約が生じてしまうことが問題になる。

## 4 条件付き確率中心の学習

Bayesian network によるモデル化は不確実性を吸収するため、その表現を一意に決定することはできない (あるデータに対して、それを良く表すモデルには様々な表現が取り得る)。ここでは先のグラフ構造中心の学習に対して、条件付き確率をより柔軟なものにしていくことによりデータをモデル化するアプローチを条件付き確率中心の学習と呼ぶことにする。

### 4.1 ニューラルネットによる条件付き確率の学習

三層の階層型ニューラルネットは任意の非線形写像を近似することが知られ、また BP 法による学習は様々な応用領域に対して成果を挙げ

ている。このニューラルネットを各確率変数の条件付き確率として用いることができる。

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, BNNN: Bayesian network on neural networks (ニューラルネットの上に実現されたベイジアンネット) を提案している [17](図 1)。これにより確率変数が離散多値, 連続値の場合についても統一的に取扱え、学習も容易になるメリットがある。

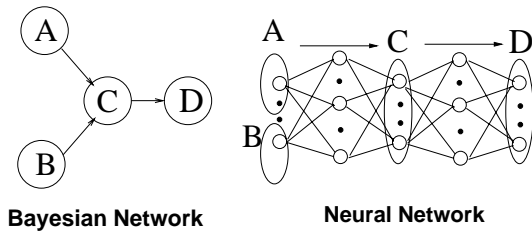


図 1: Bayesian network on neural networks

$Y$  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。

## 4.2 BAYONET

筆者らはニューラルネットを用いた Bayesian network 学習システム、BAYONETを開発している。これは事情通口ポット [14] の知識獲得、推論機能のためにも使用される。主な特長として JAVA で書かれていることにより、1. マルチスレッドによる並列計算、2. JDBC(Java DataBase Connectivity) による主要な商用データベースソフトと接続することによる大規模データベースの利用、などがある。

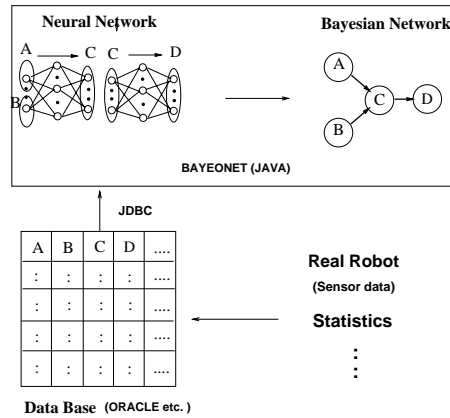


図 2: BAYONET

## 5 気象観測データベースからの学習

気象庁監修の平成7年気象庁年報 CD-ROM に格納されている気象観測データ、一年間の各観測地点における毎時刻の降水量に基づいて Bayesian net を学習して、次の予測タスク [19] を実行する例を示す。

[タスク]

西方の地点の降雨開始時刻から東方の地点の降雨開始時刻を以下の式のように予測する。

$$P(T_X|T_{B_1}, T_{B_2}, T_{B_3}) = \int_{A_1, A_2} P(T_X|T_{A_1}, T_{A_2})$$

$$\prod_{i=1}^2 P(T_{A_i}|T_{B_i}, T_{B_{i+1}}) dP(T_{A_1}) dP(T_{A_2}).$$

( $T_X$  観測地点  $X$  における降雨開始時刻)

学習条件、手順は以下の通り。

[学習条件]

- 1995年日本各地気象観測地点の毎時降水量 (8760/地点)
- 降雨開始時刻: 降水量 1.0mm 未満から 1.0mm 以上への変化

- 隣接地点の24時間以内の降雨開始時刻のみ有効

#### [学習手順]

- 全気象データ (CD-ROM) を Oracle データベースに格納。
- 実験システム (JAVA プログラム) から JDBC でアクセス。
- 該当データをニューラルネットの教師信号として抽出。
- 隠れ素子 10 個の NN で学習回数 10,000 回の BP 学習を行なう。

ニューラルネットの学習が収束すれば変数間の関係のモデル化が成功したと判定することができる。

#### 降雨開始時刻予測ベイジアンネット

ある観測地点の降雨開始時刻はそれより西方の地点の降雨開始時刻と依存関係があると仮定して図 (5) のようなグラフ構造を考える。



図 3: Bayesian network for rainfall prediction

- 福井, 岐阜, 名古屋 ( $B_1, B_2, B_3$ )  
→ 松本, 飯田 ( $A_1, A_2$ )
- 松本, 飯田 ( $A_1, A_2$ ) → 東京 ( $X$ )

上記の各観測地点を選んで、これらに関する観測データをデータベースから取り出し条件付き確率をニューラルネットにより学習した。変数は降雨開始時刻 ( $0, \dots, 23[h]$ ) で、ニューラルネットは連続変数、ガウス型の条件付き確率分布として表現した。

個々の条件付き確率の学習は平均二乗誤差 0.02 以下でいずれも収束した。このことから先のグラフ構造のもとで BNNN によるモデル化が成功したことがわかる。

次に予測タスクの性能評価を単一の三層ニューラルネットと比較した。

福井, 岐阜, 名古屋の3地点の観測を得た時の東京における降雨開始時刻の予測を前記の気象観測データを基に行なった所、BNNN の場合の平均二乗誤差 (m.s.e) は 0.013, 三層ニューラルネットの場合の平均二乗誤差は 0.023 となり、BNNN を用いたモデル化により単純なニューラルネットよりも予測精度は向上した。

#### 6 議論

Bayesian net では柔軟なモデル化が可能である一方、それゆえ一意にモデル表現を決定することは困難である。それゆえグラフ構造を主にデータへのフィッティングを行なうグラフ構造中心の学習と、条件付き確率を非線形化するなどする条件付き確率中心の学習という2つの異なるアプローチがある。しかし、前者の中でグラフ構造を決定するために客観的な評価基準を導入することには様々な理論的な制限が伴い、現在のところ実問題に適用するためには不都合な点も多い。これとは別に計算効率やタスクの遂行に都合の良い主観的な観点によりグラフ構造を設定し、条件付き確率を柔軟に学習させることによりモデル化を行なう条件付き確率中心の学習のアプローチでは、実世界の非線形性を含む関係を比較的良く近似することができる。

これは条件付き確率に制約を置いたグラフ構造中心のモデル化は比較的「硬い」シンボル寄りモデル化であるのに対し、条件付き確率中心の学習によるモデル化の方がより「柔軟な」パターン寄りのモデル化になっているためと考えることができる。一方 後者のアプローチである BNNN により気象データの学習、予測を行なった例では単純なニューラルネットによるモデル化と比較して予測精度はやや上回っていた。これは単純な入力と出力のパターンのみによるモデル化よりもある程度の明確なグラフ構造を与えたことにより変数間の関係がより精密にモデル化できたためと解釈できる。

より有効なモデル化のポイントは適切なグラフ構造をどのように実現するかであり、例えばニューラルネットの学習の収束は与えた構造の妥当性の判定基準として用いることができる。仮説として与えるグラフ構造についてはタスクや環境に埋め込まれた主観的な観点から選択する必要があるため、ユーザーがこれを比較的容易に行なえるような対話的な学習支援の実現が本アプローチによる重要な課題である。

さらに学習・推論システムの開発課題としては以下を検討している。

- モデルの多様性を扱うためのマルチグラフ構造のサポート
- マルチグラフの場合の確率計算アルゴリズム

- 主観評価を円滑に行なう I/F (Data Minig)
- 大規模化のための近似計算の制御

## 謝辞

本研究は Real World Computing Program の一環として行われました。電総研大津知能情報部長、麻生学習統合基礎ラボリーダー、松井事情通口ポットラボリーダー、Bayonet の共同開発者である原研究員、ならびにプロジェクトメンバーの方々に感謝します。

## 参考文献

- [1] J.Pearl: Probabilistic Reasoning in Intelligent Systems, Morgan Kaufmann Publishers, (1988).
- [2] E.Castillo, J.Gutierrez, and A. Hadi: "Expert systems and Probabilistic network models", Springer-Verlag N.Y., (1997).
- [3] T.Dean and M.Wellman: "Planning under uncertainty, Planning and Control", Morgan Kaufmann, CA(1991).
- [4] L.Rimey and C.Brown: "Control of Selective Perception Using Bayes Nets and Decision Theory", International Journal of Computer Vision, 12:2/3, pp.173-207, (1994)
- [5] R.Neal: "Connectionist Learning of Belief Networks", Artificial Intelligence, 56, pp.71-113, (1992).
- [6] D.Spiegelhalter, L.Lauritzen, P.Dawid, and R.Cowell: "Bayesian analysis in Expert Systems", Statistical Science, 8(3), pp.219-247(1993).
- [7] G.Cooper: "The Computational Complexity of Probabilistic Inference using Bayesian Belief Networks", Artificial Intelligence, 42, pp.393-405(1990).
- [8] W.Buntine: "Theory Refinement on Bayesian networks", *Uncertainty in Artificial Intelligence '91*, pp.52-60, (1991).
- [9] Geiger, D. and Heckerman, D.: Learning Gaussian Networks, *Uncertainty in Artificial Intelligence '94*, pp.235-243, (1994).
- [10] Heckerman, D. and Geiger, D.: "Learning Bayesian Networks: A Unification for Discrete and Gaussian Domain", *Uncertainty in Artificial Intelligence '95*, pp.274-284, (1995).
- [11] Dan Geiger and David Heckerman: "A Characterization of the Dirichlet Distribution with Application to Learning Bayesian Network", Maximum Entropy and Bayesian Methods, pp.61-68, (1996).
- [12] P.Larrañaga, M.Poza, Y.Yurramendi, R.Murfa, and C.Kujipers: "Structure Learning of Bayesian Networks by Genetic Algorithms: A Performance Analysis of Control Parameters", IEEE Trans. on PAMI, vol.18, no9, pp.912-926(1996).
- [13] J.Suzuki: "A Construction of Bayesian Networks from Databases Based on an MDL Principle", *Uncertainty in Ai*, 9, pp.266-273(1993).
- [14] H.Asoh, Y.Motomura, I.Hara, S.Akaho, S.Hayamizu, and T.Matsui: "Combining probabilistic map and dialog for robust life-long office navigation, *Proc. of IROS '96*, (1996).
- [15] 植野真臣: "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, vol.11(5), pp.725-734(1996).
- [16] 麻生英樹, 赤穂昭太郎, 本村陽一: "統計的推論と AI の推論", 人工知能学会論文誌, Vol.12, No.2, pp.196-203, (1997).
- [17] 本村陽一, 多層パーセプトロンを用いたベイジアンネットによる確率的推論, 神経回路学会全国大会講演論文集 (1996).
- [18] 本村, 麻生 原, 赤穂, 松井: "事情通口ロボットにおけるニューラルベイジアンネットの学習", 人工知能学会第一回情報統合研究会, <http://www.etl.go.jp/~motomura/paper/> (1996).
- [19] Y.Motomura, I.hara, H.Asoh, T.Matsui: "Bayesian Network that Learns Conditional Probabilities by Neural Networks", International Conference of Neural Information Processing, (1997).