

統計的推論と AI の推論

Statistical Inference and Inference in AI

麻生 英樹*
Hideki Asoh

赤穂 昭太郎*
Shotaro Akaho

本村 陽一*
Yoichi Motomura

* 電子技術総合研究所
Electrotechnical Laboratory.

1996年11月25日 受理

Keywords: statistical inference, Bayesian analysis, Bayesian network, finite mixture.

1. はじめに

「AIの推論」で、普通に思い浮かべるのは論理的(演繹的)推論であろう。「演繹的推論」の働きは「現在の世界の状態についての部分的な情報と世界についての知識とを組み合わせ、現在の世界の状態、あるいは未来の世界の状態についてより多くのことを推測すること」である。また、「過去の観測データと良く一致する知識を獲得すること」も「帰納推論」「仮説推論」などの名前で呼ばれている。

これらの「AIの推論」の多くは、決定論的な記号論理をベースとして研究が行なわれてきたが、近年いろいろな意味で、人工知能システムが実世界(Real-World)に出てくるのが求められていることの一つの側面として、人工知能システムに、あらかじめ人間によって抽象化された概念だけでなく、さまざまなセンサやネットワークを通じて得られる大量の生データをも扱うことが求められるようになってきている。こうした不確実性やノイズを含む大量のデータの処理、データからの汎用的知識の抽出・獲得は、統計学、あるいはデータ解析学の研究領域であり、これらの領域で蓄積された「統計的推論」の技法は、人工知能システムにおいても有効なはずである。また、人工知能システム固有の問題意識が、統計学やデータ解析学に新たな問題をもたらし、新たな展開を促す可能性もあるだろう。ここに、統計学研究と人工知能研究との接点がある。

このような接点の認識自体は古くからあった。たとえば、人工知能の分野の中でも、センサからの不確実かつ多様なデータを直接扱う必要のあったパターン認

識の分野では統計的手法の利用が古くから主流であるし[Fukunaga 72]、論理的推論や意思決定の分野における情報の不確実さへの対処に関する国際会議であるUncertainty in AI(UAI)も今年で12回目を数えている。しかし、人工知能の実世界化の要請が高まるにつれて、この接点の重要性も高まっていると言えるだろう。このような状況を背景として、本稿では、統計学あるいはデータ解析研究と人工知能研究との接点にある研究領域の最近の動向について述べたい。言うまでもなく、「推論」と呼ばれる研究領域は広大であり、以下で述べられるものはそのごく小さな一部である。

2. ベイズ決定理論

2.1 理論的枠組み

不確実な状況の下での意思決定のための確率的推論の問題は、特定の条件の下で、仮説的命題の確からしさを評価し、そこから適切な行動制御戦略を導くことである。行動選択に関係する各命題の真偽が完全に決定されていれば、通常の最適化問題であるが、そうではないために、確率的な扱いが必要になる。この種の問題に対する一般的な解決の一つはベイズ決定理論(Bayesian Decision Theory, Bayesian Decision Analysis)である(たとえば[Berger 85])。その理論は、対象となっている世界の記述、つまり Universe of Discourse、その上の各命題の同時確率分布の値、および各行動の効用関数(utility function)あるいは損失関数(loss function)がすべて既知であるという条件の下で、期待効用を最大化する、あるいは期待損失を最小化する(ベイズ最適(Bayes Optimal)な)行動戦略(policy)の選

択を行なうことに関する理論的な枠組みを与えている。

ここで必要となる統計的推論は、演繹的推論にあたる。世界の確率分布と現在の状況についての情報とを用いた各命題の条件付きの確率計算および確率値に基づく期待効用最大化、そして、帰納的推論にあたる、過去のデータからの確率分布の推定である。

ベイズ決定理論の概念を理解するための最も簡単な例として移動ロボットの経路選択を考える（もとの例は Dean らの教科書 [Dean 91] にあったもの）。ロボットが地点 A から地点 B へ移動するのに、二つの経路が考えられるとする。一つはドアを通過していくものであり、もう一つは廊下をまわっていくものである。ドアを通る経路は距離が短い、一定の確率でドアが閉まっている可能性がある。ドアが閉まっていた場合にはロボットは廊下まで戻って迂回しなくてはならない。

この場合、ロボットの行動に関係する世界の状態はドアが開いているか、閉まっているかの二つであり、過去の観測データなどから、それぞれの確率は 0.5 ずつであることがわかっているとす。ロボットにとっての損失関数は走行時間であるとし、ドアが開いていた場合が 6 分、廊下を回った場合が 10 分、ドアが閉まっていた迂回した場合が 12 分であるとする。可能な行動戦略はドアを通ることを狙うかはじめから安全な廊下を通ることを狙うかの二つである。この場合、ドアを狙った場合の期待損失は 9 分、廊下を回った場合の期待損失は 10 分であるので、ドアを狙う戦略のほうが期待損失が少ないことになる。

この問題は最も簡単な場合であるが、これから出発しても、ドアの状態が間接的な観測によって得られる場合（たとえば途中で出あった人に尋ねる、など）、ドアの開閉確率自体をも推定しながら同じ移動を何度か繰り返す場合、など多くのバリエーションが考えられ、多くの理論的考察がなされるとともに、識別・診断問題（たとえば脳脈探査エキスパートシステム PROSPECTOR でのベイズ分析 (Bayesian Analysis) を使った帰納推論 [Duda 76, Duda 78] が有名）、計画問題（たとえば移動ロボットの地図学習や経路計画 [Dean 91, Nourbakhsh 95, Thrun 96]）などに応用されてきた。

ベイズ決定理論は概念的、理論的には明快な枠組みだが、実際の問題に適用しようとする、いろいろな問題が生じる。確率的推論の研究の多くは、そうした問題への対応であるとも言えるだろう。それらの問題を大きく以下の三つに分けてみる。

- Universe of Discourse の決定に関する問題
- 効用関数の決定に関する問題
- データ量、計算量、記憶量に関する問題

以下では、まず上の二つの問題に簡単に触れ、さらに、三番目の問題に関連する最近の研究動向について詳しく述べる。

移動ロボットの例では、ロボットの行動戦略決定に関与するのはドアの開閉だけであることがあらかじめ知られていることを前提としていた。しかし、現実にはある意思決定に関与することがあらかじめすべてわかっていることは少ない。たとえば、移動ロボットは、今まで知らなかった通路やドアに遭遇することがありうる。これが Universe of Discourse の不確定の問題であり、フレーム問題とも関連する深刻な問題である。こうした事態に対処するには確率推論システム自らが、新たなカテゴリや仮説命題を逐次的かつ自己組織的に導入できる仕組みを持つことが必要であり、そのためにクラスタリング技法を使う試みなどが行なわれている。

実世界的な意思決定においては、異なる選択を比較するための価値尺度自体がよくわからないことも考えられる。この効用関数の不確定の問題については、とりあえずの評価を用いて意思決定を行ない、結果が望ましくなければ、評価自身も修正する、というような探索的な技法が試みられているようである。また、たとえば、一連の行動の最終結果に対して評価が与えられるが、個別の要素行動についての評価は与えられないような場合に、個別要素行動の期待効用を推定する問題は、強化学習 (Reinforcement Learning) の問題の一部として研究されている [Sutton 88, Watkins 92]。

2.2 組合せ構造を持つ確率分布モデルの導入

確率分布の推定は不良設定問題であり、安定に推定するには多数の偏りのないデータが必要である。しかし、実際にはそうしたデータが得られないことはごく普通である。この問題への対処法は大きく二つに分けることができる。一つは、Dempster-Shafer 理論 [Dempster 67, Dempster 68, Ishizuka 82, 石塚 83, Shafer 76] に代表される、確率分布の不確定性をそのまま扱えるように確率論を拡張するもので、理論的にも美しく、応用例も多いが、確率値の不確定性を扱うために、通常確率計算よりもさらに計算量が増えるという問題がある。従って、対象の規模拡大への対処が一般には難しい。

もう一つの方法は、分布推定問題を良設定にすることである。具体的には、正則化条件や確率分布モデルによって記述される構造的制約を明示的に導入することである。古典的な統計においては、正規分布を中心とした分布モデル族が扱われてきたが、主に計算

機の発達と有効な算法の発明によって、現在では、使用可能なモデルの種類は大きく広がってきている。特に、近年、Bayesian network や隠れマルコフモデル (Hidden Markov Model) [中川 88], deformable template [Amit 91, Yulie 90], Markov Random Field [Chellapa 93, Geman 84], 確率木文法などに代表されるような組合せ構造を持つ確率分布モデルの活用が盛んになりつつある。また、ニューラルネットワークやファジー推論もまた、構造的な確率モデルによる確率的知識獲得と推論という枠組みの中で眺めることができる (たとえば [Mackay 92])。

対象と用途にあったモデルおよびそのモデルを用いた知識獲得と利用のアルゴリズムを発明/発見することは、確率推論における最も重要な問題の一つである。

以下の 2 章では、これらの構造を持つ確率分布モデルの例として、近年アルゴリズム的な側面および応用可能性の側面から注目されている Bayesian network と混合分布の二つについてより詳しく解説する。

3. Bayesian network

3.1 グラフ構造による確率分布モデルの表現

確率変数をノードで表し、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 ($X \rightarrow Y$ 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 (Belief network, Causality network, あるいは総称して Probabilistic network と呼ばれることもある) である (教科書として [Neapolitan 90, Pearl 88], 他にも解説としては [Charniak 91, Pearl 91, Spiegelhalter 93, 鈴木 94] がある。また最近では WWW によるガイド [AFIT 96] なども公開されている)。具体的な Bayesian network として古典的な医療診断の例題 [Lauritzen 88] を図 1 に示す。他にも著名なネットワークの例が [Norsys 96] にまとめられている。

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が複雑に込み入っているため、「もし $X_1 = x_1, \dots, X_i = x_j, \dots$ ならば $Y = y$ 」のように完全に記述しようとする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

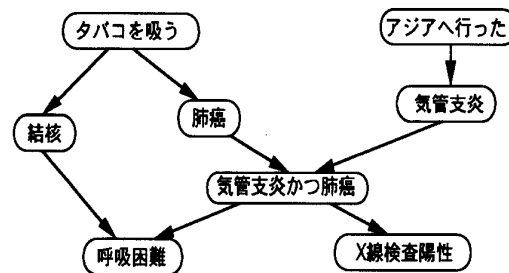


図 1 Bayesian network の例

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変数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, \dots, Z の全変数の中で因果関係が X, Y の間にだけ存在する場合には全ての変数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。

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

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 (変数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。

3.2 Bayesian network による推論

Bayesian network による確率的推論は、

- (i) 未観測の事象についての確率値を得る
- (ii) その条件のもとでのもっとも確信度の高い実現値の組合せを推論結果として出力する

の2段階がある。それぞれその状況のもとでの仮説の確率分布の推定と観測結果との同時確率最大の仮説の選択に対応している。(i)は逐次的にノード間の局所的な確率伝播として実行される。ここでは $X \rightarrow Y$ についての計算を示す。

- (1) 観測結果 $X = e$ が与えられた時の $P(Y)$ は各条件付き確率 $P(Y = y|X = e)$ から構成する。
- (2) 子ノードに観測結果 $Y = e$ が与えられた時の $P(X|Y = e)$ は次式 (ベイズの定理) で求める。

$$P(X|Y = e) = \alpha P(Y = e|X)P(X) \quad (1)$$

(正規化係数 $\alpha = 1 / \int_{x \in X} P(e|x)dP(x)$.)

- (3) 親ノードの $P(X)$ が更新された時, $P(Y)$ は次式によって更新する。

$$P(Y) = \int_{x \in X} P(Y|x) dP(x) \quad (2)$$

- (4) 逆に子ノードの $P(Y)$ が更新された場合の親ノードの $P(X)$ は次式で更新する。

$$\begin{aligned} P(X) &= \alpha \int_{y \in Y} P(X|y) dP(y) \\ &= \alpha' P(X) \int_{y \in Y} P(y|X) dP(y) \end{aligned} \quad (3)$$

ただし α, α' は正規化係数。

- (5) 観測値が与えられない最上位のノード (root node) については事前分布を与える。

以上の各計算ステップが形式的に X, Y 間で受渡しされる局所的な確率情報であり, これがネットワーク上で大域的に伝播することで確率的推論が実行される。このとき計算がグラフの上で逐次的に行なわれるために, 任意の二つのノード間を結ぶ経路が一つ以下のネットワークの場合 (singly-connected network) は各観測ノードから発生した伝播のみで計算は終了し計算量もノードに対する線形のオーダーで済む。しかし任意の二つのノード間を結ぶ経路が二つ以上ある場合を含む一般的なグラフ構造 (multiply-connected network) の場合には, 計算順序によっては計算結果の一致性が保証されない可能性があり最悪計算量は NP-hard になる [Cooper 90, Dagum 93]。このためこうした問題を解決するために様々な効率的な伝播方法やグラフを前処理する方法, モンテカルロ計算による近似的な計算方法などが 80 年代の主要な研究課題であった [AFIT 96, Henrion 90]。どの方法もそのパフォーマンスが対象とするグラフ構造や確率値に依存しており, 完全な解決は見られていないが, 最近では問題対象に応じてグラフ構造を限定したりするなどしてこの問題を回避し, どこまで実際の課題に適用できるかに関心が

移っている。例えば音声対話モデル [Akiba 94], 音源分離 [Kashino 95], 自律移動ロボットの地図学習 [本村 96a], 自律走行車 [Forbes 95], アクティブビジョン [Rimey 94], タンパク質構造解析 [Klingler 94] などの研究例や特集記事 [Heckerman 95b] がある。近似的手法を実装した汎用パッケージも多く開発され [Andersen 89, Washington Univ. 96], Microsoft 社の商用ソフトウェアや GE, Intel 社, NASA などにおいて実際のフィールドシステムの障害診断部などに応用され, 実問題における有効性が示されている。

大規模データからのネットワーク構築, 条件付き確率の学習も最近の重要な研究課題である。ネットワーク構築を自動的に獲得する問題は統計学におけるモデル選択の問題であり, MDL による方法 [Suzuki 93], ベイズ的な方法 [Heckerman 95a] や GA の適用 [Larrañaga 96], 因果関係を抽出していく方法 [Hofmann 96, 植野 96] など様々な試みがなされている。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, パラメー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 [Pearl 88], Sigmoid 関数 [Neal 92, Saul 96] 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 [Geiger 94] や階層型ニューラルネット [本村 96b] の適用などがある。

4. 混合分布と EM アルゴリズム

4.1 混合分布

複数のモジュール (あるいはエージェント) が分散・協調する「エージェント指向」アーキテクチャが人工知能の実現のために有効であると期待されているが, 統計学においても, 近年, 複数のモジュール構造を持つ確率分布の推定やそれを用いた統計的推論の研究が進められている。

最も簡単な例として, 二つの正規分布 $\mathcal{N}(\mu_1, 1)$, $\mathcal{N}(\mu_2, 1)$ の混合を考える。 μ_1, μ_2 は平均であり, 簡単のため分散は 1 に固定した。この二つの正規分布を二つのモジュールだと思ふことにする。正規分布はあまりにも単純な形をしており, とっても知的なエージェントには見えないが, 複雑なエージェントの場合も話はそれほど変わらない。

観測データは, 1 次元の直線上に与えられた n 個の点である。二つの正規分布を組み合わせてこのデータ

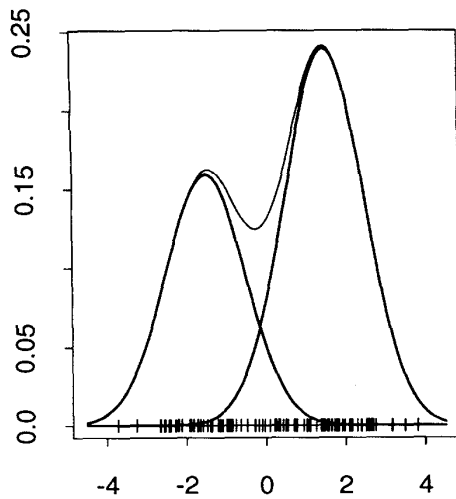


図2 二つの正規分布の混合分布

にできるだけよくフィットさせるようにすることが分布推定の目的である。

式で書けば、このモデルは

$$p(x | \theta, \mu_1, \mu_2) = \theta \phi(x | \mu_1) + (1 - \theta) \phi(x | \mu_2),$$

と表される。ただし、

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

このように、複数の確率分布を重みつきで足しあわせた分布を一般に混合分布 (Finite Mixture) と呼ぶ。

サンプル x_1, \dots, x_n があつたときに、このサンプルが出現する確率

$$\prod_{i=1}^n p(x_i | \theta, \mu_1, \mu_2) \quad (4)$$

をパラメータ θ, μ_1, μ_2 の関数とみなした式 (4) を尤度と呼び、確率分布の推定では尤度を最大化するパラメータを求める (最尤推定) ことがよく行なわれる。しかし、式 (4) は指数関数の和の積の形になっていて、そのまま最大化することは非常に難しい。もし、 $\theta = 1$ とおいて、一つだけの正規分布でフィットさせるとすると、非常に簡単に解ける (ちなみに $\mu_1 = \sum_{i=1}^n x_i / n$ となる)。

一つずつの正規分布のフィットは易しいのだから、すぐに考え付くのはあらかじめサンプルを粗く二つのグループに分け、片方のグループのサンプルで一方の正規分布を学習し、もう一方のグループのサンプルでもう一方の正規分布を学習する方法である。もちろん、最初の分け方は荒っぽいから、学習した正規分布を用いてそれぞれのサンプルがどちらに属するかを改めて決

め直す。そしてまた正規分布を学習するということを繰り返していけばなんとなくそれらしい解に収束しそうである。

実はこの粗いアイデアのアルゴリズムでは最尤法で得られる解には収束しないが、混合分布の最尤推定のアルゴリズムとしてよく知られている EM アルゴリズムは、基本的にこのような精神で考え出されたもので、より柔軟なデータのグループ分けをすることにより、尤度の単調な増加を保証している [赤穂 96, Amari 94, Dempster 77, 宮川 87]。上の問題の場合の具体的な推定過程は下記のようになる。

まず最初にパラメータを適当な値の初期値におき、 $\theta = \theta^{(0)}$, $\mu_1 = \mu_1^{(0)}$, $\mu_2 = \mu_2^{(0)}$ とする。次に、各サンプルについて、それぞれの正規分布に属する確率を計算する。つまり、

$$q_{i,1} = \frac{\theta^{(0)} \phi(x | \mu_1^{(0)})}{\theta^{(0)} \phi(x | \mu_1^{(0)}) + (1 - \theta^{(0)}) \phi(x | \mu_2^{(0)})}$$

および $q_{i,2} = 1 - q_{i,1}$ を計算する。続いて、その確率を重みとしてパラメータを計算する。つまり、

$$\theta^{\text{new}} = \frac{1}{n} \sum_{i=1}^n q_{i,1}, \quad (5)$$

$$\mu_1^{\text{new}} = \frac{1}{n\theta^{\text{new}}} \sum_{i=1}^n q_{i,1} x_i, \quad (6)$$

$$\mu_2^{\text{new}} = \frac{1}{n(1 - \theta^{\text{new}})} \sum_{i=1}^n q_{i,2} x_i, \quad (7)$$

とする。更新されたパラメータを用いて、同様の手順を繰り返して収束するまで続ける。このアルゴリズムの収束先は元の最尤推定問題の局所最適解になっている。

Jordan らは、ニューラルネットワークを要素モジュールとするようなより複雑なモデルを提案し、学習制御などに応用して良好な結果を示している。EM アルゴリズムの特徴は、モジュール毎に処理を分散できることである。すなわち、一旦データからの情報をモジュール毎に分配した後では、各モジュールが独立に学習を行なうことができるような構造になっている。

4・2 EM アルゴリズム

EM アルゴリズムは、観測できない変数をもつような最尤法の局所最適解を求めるための反復アルゴリズムである。

前節で述べた二つの正規分布の場合には、各サンプルがどちらの正規分布に属しているかというラベルが観測できない変数になっている。もし、このラベルが与えられていればそれぞれ一つずつの正規分布を学習

すればよいのだから問題は非常に簡単になる。一般に EM アルゴリズムは、このように観測できない変数を含めることによって単純になるような問題に適している。

ここで、観測できる変数を y 、観測できない変数を z とし、合わせて $x = (y, z)$ という変数を考える (x はより一般に、 y への多対一の写像が存在するような任意の空間の元で構わないが、ここではこのように表される場合のみを考える)。

y の分布を $g(y | \theta)$ とし、 $x = (y, z)$ の分布を $f(y, z | \theta)$ とする。すると、与えられたデータに対する z の条件つき分布はベイズの定理を用いて、

$$h(z | y, \theta) = \frac{f(y, z | \theta)}{g(y | \theta)} \quad (8)$$

と表すことができる。

EM アルゴリズムは次の 2 つのステップの計算を繰り返して行なうアルゴリズムである。パラメータの初期値を $\theta = \theta^{(0)}$ とおく。

- (1) (E(xpectation) step) x の分布の対数の (データ y とパラメータ $\theta^{(0)}$ に対する) 条件つき期待値を計算する、つまり、

$$\begin{aligned} Q(\theta | \theta^{(0)}) \\ = \int h(z | \theta^{(0)}) \log f(y, z | \theta) dz \end{aligned} \quad (9)$$

を計算する。

- (2) (M(aximization) step) $Q(\theta | \theta^{(0)})$ を最大にする θ を θ^{new} とおく。

前節の問題に対してこの計算を適用してやれば、式 (5)~(7) の逐次更新式が得られる。一応の対応関係を述べると、 $q_{i,1}$ の計算が条件つき確率 h の計算に対応しており、 Q の最大化は前節の問題の場合には陽に解けてしまう。

4・3 EM アルゴリズムの問題点

EM アルゴリズムは局所最適解に収束するアルゴリズムであり、大域的な収束は保証されていない。しかし、多くの実例では他の手法に比べて良い解に収束することが知られており、しかも繰り返しの初期には Newton 法と同程度に速い。ただし解の周りでは収束が遅くなるので加速が必要となる。

また、一般には E と M のそれぞれのステップが単純になるとは限らないので、計算量を減らすためのいろいろなテクニックが研究されている [Tanner 93]。例えば、E ステップは積分 (離散の場合は総和) 計算なので計算が大変な場合には Gibbs サンプリングなどのモンテカルロ法で近似計算することがよく行なわれる。

一方、M ステップは基本的に最適化問題なので解が陽に書き下せない場合には Newton 法などの反復法がよく用いられる (もとの問題を直接 Newton 法で解くよりは単純な最適化になっていることが多い)。さらに、そもそも計算が単純になるような確率モデルを選ぶことが重要であり、情報幾何学的な観点からの研究が行なわれている [Amari 94]。

4・4 他の確率モデルへの適用

EM アルゴリズムは混合分布の他にも数多くの確率分布モデルに適用できる。音声認識や自然言語処理でよく用いられている隠れマルコフモデルはその状態遷移が観測されない変数となる確率モデルとなっている。確率的に発火する 2 値素子から成る確率的なニューラルネットワークであるボルツマンマシン [Hinton 84] は、隠れ素子を観測できない変数とした確率モデルになっている [Amari 94]。多層のニューラルネットもまた、隠れ層の発火パターンを観測できない変数とした確率モデルと考えることができる。このようなモデルに対しても EM アルゴリズムの有効性が検討されている。

5. おわりに

統計的推論と AI の推論の接点について、特に Bayes 決定理論を中心とする観点から述べた。近年、組合せ構造を持つ確率モデルを利用することによって、より複雑で柔軟な統計的推論の可能性が開かれ始めている。また、統計的推論における最大の問題は、確率分布の推定や確率値計算に伴う計算およびデータ採取のコストであり、この問題を避けるためにさまざまな技法が提案されているが、古典的な確率・統計に準拠することで、しばしば ad hoc になりがちな推論のためのヒューリスティクスを事前分布などの形で明示的に定式化できるようにする効果も大きい。

しかしながら、正確な確率分布のモデル化や分布の推定、それをういた意思決定などの計算コストは依然として高い。従って、確率推論の応用に関して実際の観点から重要な問題は、あまり正確ではないかもしれない近似的な確率計算をシステム全体の中で具体的にどう利用していくか、である。このためには、問題 (すなわち Universe of Discourse) の適切な分割や適切な部分構造モデルの構成とともに、システムが自ら能動的に情報を収集する能力、自らの状態を常に評価し、誤った判断からも復元できる能力を持つこと、などが重要であると思われる。

謝 辞

本研究の一部は、Real-World Computing 計画の下で行なわれました。

◇ 参 考 文 献 ◇

- [AFIT 96] <http://www.afit.af.mil:80/Schools/EN/ENG/LABS/AI/BayesianNetworks/> (1996).
- [赤穂 96] 赤穂昭太郎: EM アルゴリズムの幾何学, 情報処理, Vol.37, pp.43-51 (1996).
- [Akiba 94] Akiba, T. and Tanaka, H.: A Bayesian approach for user modelling in dialogue systems, *Proc. of the International Conference on Computational Linguistics*, pp.1212-1218 (1994).
- [Amari 92] Amari, S., Kurata, K. and Nagaoka, H.: Information Geometry of Boltzmann Machines, *IEEE Trans. Neural Networks*, Vol.3, pp.260-271 (1992).
- [Amari 94] Amari, S.: Information Geometry of the EM and em Algorithms for Neural Networks, *Neural Networks*, Vol.8, pp.1379-1408 (1994).
- [Amit 91] Amit, Y., Grenander, U. and Piccioni, M.: Structural image restruction through deformable templates, *Journal of the American Statistical Association*, Vol.86, pp.376-387 (1991).
- [Andersen 89] Andersen, S.K., Olesen, K.G., Jensen, F.V. and Jensen, F.: Hugin—a Shell for Building Bayesian belief Universes for Expert Systems, *Proc. of IJCAI'89*, pp.1080-1085 (1989).
- [Baum 70] Baum, L., Petrie, T., Soules, G. and Weiss, N.: A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains, *Annals of Mathematical Statistics*, Vol.41, pp.164-171 (1970).
- [Berger 85] Berger, J.O.: *Statistical Decision Theory and Bayesian Analysis* (2nd. Ed.), Springer-Verlag (1985).
- [Charniak 91] Charniak, E.: Bayesian Networks without Tears, *AI Magazine*, Vol.12, pp.50-63 (1991).
- [Chellapa 93] Chellapa, R. and Jain, A. (eds.): *Markov Random Fields, Theory and Application*, Academic Press (1993).
- [Cooper 90] Cooper, G.: The Computational Complexity of Probabilistic Inference using Bayesian Belief Networks, *Artificial Intelligence*, Vol.42, pp.393-405 (1990).
- [Dagum 93] Dagum, P. and Luby, M.: Approximating probabilistic inference in Bayesian Belief Networks is NP-hards, *Artificial Intelligence*, Vol.60, pp.141-153 (1993).
- [Dean 91] Dean, T.L. and Wellman, M.P.: *Planning and Control*, Morgan Kaufmann (1991).
- [Dempster 67] Dempster, A.P.: Upper and lower probabilities induced by a multivaluated mapping, *Annals of Mathematical Statistics*, Vol.38, pp.325-339 (1967).
- [Dempster 68] Dempster, A.P.: A generalization of Bayesian inference, *The Journal of the Royal Statistical Society B*, Vol.30, pp.205-247 (1968).
- [Dempster 77] Dempster, A.P., Laird, N.M. and Rubin, D.B.: Maximum likelihood from incomplete data via the EM algorithm, *The Journal of the Royal Statistical Society B*, Vol.39, pp.1-38 (1977).
- [Duda 76] Duda, R.O., Hart, P.E. and Nilsson, N.J.: Subjective Bayesian methods for rule-based inference systems, *Proc. 1976 National Computer Conference*, Vol.45, pp.1075-1082 (1976).
- [Duda 78] Duda, R.O., et al.: Development of the PROSPECTOR Consultant System for Mineral Exploration, Final Report for SRI Projects 5821 and 6915, Artificial Intelligence Center, SRI International (1978).
- [Forbes95] Forbes, J., Huang, T., Kanazawa, K. and Russell, S.: The BATmobile: towards a Bayesian automated taxi, *Proc. of IJCAI'95*, pp.1878-1885 (1995).
- [Fukunaga 72] Fukunaga, K.: *Introduction to Statistical Pattern Recognition*, Academic Press (1972).
- [Geiger 94] Geiger, D. and Heckerman, D.: Learning Gaussian networks, *Uncertainty in AI*, Vol.10, pp.235-243 (1994).
- [Geiger 95] Geiger, D. and Heckerman, D.: A characterization of the dirichlet distribution with application to learning Bayesian networks, *Uncertainty in AI*, Vol.11, pp.196-207 (1995).
- [Geman 84] Geman, S. and Geman, D.: Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, *IEEE Trans. on PAMI*, Vol.6, pp.721-741 (1984).
- [Heckerman 95a] Heckerman, D., Geiger, D. and Chickering, D.M.: Learning Bayesian networks: the combination of knowledge and statistical data, *Machine Learning*, Vol.20, pp.197-243 (1995).
- [Heckerman 95b] Heckerman, D., Mamdani, A. and Wellman, M.: Real world applications of Bayesian networks, *Communications of the ACM* (1995).
- [Henrion 90] Henrion, M.: An introduction to Algorithms for Inference in Belief Nets, *Uncertainty in AI*, Vol.5, pp.129-138, Elsevier Science Publishers B.V. (1990).
- [Hinton 84] Hinton, G.E., Sejnowski, T.J. and Ackley, D.H.: Boltzmann machines: constraint satisfaction networks that learn, Technical Report, CMU-CS-84-119, Carnegie-Mellon University, Department of Computer Science (1984).
- [Hofmann 96] Hofmann, R. and Tresp, V.: Discovering structure in continuous variable using Bayesian networks, *Advances in Neural Information Processing System*, Vol.8, pp.500-506, MIT Press, Cambridge MA (1996).
- [Ishizuka 82] Ishizuka, M.: An extention of Dempster & Shafer's theory to fuzzy set for constructing expert systems, 生産研究, Vol.34, pp.312-315 (1982).
- [石塚 83] 石塚 満: Dempster & Shafer の確率理論, 信学誌, Vol.66, pp.900-903 (1983).
- [Jacobs 91] Jacobs, R.A., Jordan, M.I., Nowlan, S.J. and Hinton, G.E.: Adaptive mixtures of local experts, *Neural Computation*, Vol.3, pp.79-87 (1991).
- [Kashino 95] Kashino, K., Nakadai, K., Kinoshita, T. and Tanaka, H.: Organization of hierachical perceptual sounds, *Proc. of IJCAI'95*, pp.158-164 (1995).
- [Klingler 94] Klingler, T.M., Brutlag, D.L. and Douglas, L.: Discovering structural correlations in alpha-helices, *Protein Science*, Vol.3, pp.1847-1857 (1994).
- [Larrañaga 96] Larrañaga, P., Poza, M., Yurramendi, Y., Murfa, R.H. and Kujipers, C.M.H.: Structure learning of Bayesian networks by genetic algorithms: a performance analysis of control parameters, *IEEE Trans. on PAMI*, Vol.18, pp.912-926 (1996).
- [Lauritzen 88] Lauritzen, S.L. and Spiegelhalter, D.J.: Local computations with probabilities on graphical structures and their application to expert systems, *Journal of the Royal Statistical Society, Series B*, Vol.50, pp.157-224 (1988).

- [Mackay 92] Mackay, D.J.C.: A practical Bayesian framework for backpropagation networks, *Neural Computation*, Vol.4, pp.448-472 (1992)
- [宮川 87] 宮川雅巳: EM アルゴリズムとその周辺, 応用統計学, Vol.16, pp.1-19 (1987).
- [本村 96a] 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ワークによる確率的推論, 神経回路学会全国大会講演論文集 (1996).
- [本村 96b] 本村陽一, 麻生英樹, 原功, 赤穂昭太郎, 松井俊浩: 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ワークの学習, 人工知能学会情報統合研究会予稿 (1996). <http://www.etl.go.jp/People/motomura/BN/> (1996).
- [中川 88] 中川聖一: 確率モデルによる音声認識, 電子情報通信学会 (1988).
- [Neal 92] Neal, R.M.: Connectionist learning of belief networks, *Artificial Intelligence*, Vol.56, pp.71-113 (1992).
- [Neapolitan 90] Neapolitan, R.E.: *Probabilistic Reasoning in Expert Systems*, Wiley, NY (1990).
- [Norsys 96] <http://www.norsys.com> (1996).
- [Nourbakhsh 95] Nourbakhsh, I., Powers, R. and Birchfield, S.: DERVISH an office-navigating robot, *AI Magazine*, Summer, pp.53-60 (1995).
- [Pearl 86] Pearl, J.: Fusion, propagation and structuring in belief network, *Artificial Intelligence*, Vol.29, pp.241-288 (1986).
- [Pearl 88] Pearl, J.: *Probabilistic Reasoning in Intelligent Systems*, Morgan Kaufmann, CA (1988).
- [Pearl 91] Pearl, J. (南川訳): Bayes の決定法, 人工知能大辞典, pp.1026-1034, 丸善 (1991).
- [Peterson 87] Peterson, C. and Anderson, J.R.: A mean field theory learning algorithm for neural networks, *Complex Systems*, Vol.1, pp.995-1019 (1987).
- [Rimey 94] Rimey, L.D. and Brown, C.M.: Control of selective perception using Bayes nets and decision theory, *International Journal of Computer Vision*, Vol.12, pp.173-207 (1994).
- [Saul 96] Saul, L.K., Jaakkola, T. and Jordan, M.I.: Mean field theory for sigmoid belief networks, *Journal of Artificial Intelligence Research*, Vol.4, pp.61-76 (1996).
- [Shafer 76] Shafer, G.: *A Mathematical Theory of Evidence*, Princeton Univ. Press (1976).
- [Spiegelhalter 93] Spiegelhalter, D.J., Lauritzen, S.L., Dawid, P.A., and Cowell, R.G.: Bayesian analysis in expert systems, *Statistical Science*, Vol.8, pp.219-247 (1993).
- [Sutton 88] Sutton, R.S.: Learning to predict by the methods of temporal difference, *Machine Learning*, Vol.3, pp.9-44 (1988).
- [Suzuki 93] Suzuki, J.: A construction of Bayesian networks from databases based on an MDL principle, *Uncertainty in AI*, Vol.9, pp.266-273 (1993).
- [鈴木 94] 鈴木謙: 確率ネットワーク, AI 白書, pp.113-117, コンピュータエージ社 (1994).
- [Tanner 93] Tanner, M.A.: *Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions*, Springer-Verlag (1993).
- [Thrun 96] Thrun, S.: A Bayesian approach to landmark discovery and active perception in mobile robot navigation, Technical Report CMU-CS-96-122, Carnegie Mellon Univ. (1996).
- [植野 96] 植野真臣: 意思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, No.5, pp.725-734 (1996).
- [Washington Univ.96] <http://bayes.stat.washington.edu/almond/belief.html> (1996).
- [Watkins 92] Watkins, C.J.C.H.: Q-Learning, *Machine Learning*, Vol.8, pp.279-292 (1992).
- [Yuille 90] Yuille, A.L.: Generalized deformable models, statistical physics, and matching problems, *Neural Computation*, Vol.2, pp.1-24 (1990).

 著者紹介



麻生 英樹 (正会員)

1981年東京大学工学部計数工学科卒業。1983年東京大学大学院工学系研究科情報工学専攻修士課程修了。同年電子技術総合研究所に入所。現在情報科学部主任研究官。知的学習システムの実現に関する研究に従事。行動計量学会, 日本神経回路学会, 電子情報通信学会各会員。 <asoh@etl.go.jp>



赤穂 昭太郎

1988年東京大学工学部計数工学科卒業。1990年東京大学大学院工学系研究科計数工学専攻修士課程修了。同年電子技術総合研究所に入所。現在情報科学部研究官。知的学習システムの実現に関する研究に従事。日本神経回路学会, 電子情報通信学会各会員。 <akaho@etl.go.jp>



本村 陽一 (正会員)

1991年電気通信大学電気通信学部通信工学科卒業。1993年電気通信大学電気通信学研究科電子情報学専攻修士課程修了。同年電子技術総合研究所に入所。現在情報科学部研究官。知的学習システムの実現に関する研究に従事。日本神経回路学会, 電子情報通信学会, 日本認知科学会各会員。 <motomura@etl.go.jp>