Research Memorandum No.595 Feb. 1996

# 統合技術のための統計数理的方法

改訂版 (2.5 版)

伊庭幸人

統計数理研究所
〒 106 東京都 港区 南麻布 4-6-7
Email: iba@ism.ac.jp

“ポスト記号論理” の人工知能の基盤としてのベイズ統計の利点と欠点について検討し、現在利用可能な技術をモデルとアルゴリズムの両面から列挙した。また、将来の情報処理技術のなかで統計の枠組に必ずしも収まらない部分についても考察した。調査報告書のための草稿 (筆者分担分) である。

KEYWORDS
ベイズ統計，シンボル，モデルとアルゴリズムの分離，ニューラルネットワーク，再帰的な方法，マルコフ連鎖モンテカルロ法，平均場近似，非平衡，因果関係

## 1 要旨

1. ベイズ統計の枠組の利点と欠点の検討

   ベイズ統計の枠組は従来のパターン処理、シンボル処理の手法を統一的に理解するために有益である。処理アルゴリズムの部分とモデルの部分を明確に分けて理解するのが、現代的な統計学の考え方の特徴である。これは、計算機言語における宣言型言語の考え方にも通じる。この見方によって、従来の方法を統一的な立場から見直すことができ、より柔軟あるいは高速な処理が可能になることが期待される。

   ベイズ統計の枠組はパラメータとデータをともに確率変数と考える点で一般の統計学の枠組とことなる。これは不完全なデータを扱う場合に柔軟な扱いを可能にする。また、階層的な構造を考えることで、学習・汎化を行なうことができる (階層ベイズ法あるいは ABIC 法)。

   ここでは、ベイズ統計の考え方の利点と欠点について論じる。記号論理や確率以外の “確信度” を利用する枠組との比較、ベイズ統計以外の統計の枠組との比較を試みた。また、将来の情報処理技術の基礎として考えた場合の問題点を検討した。

2. シンボル / パターン統合のために役立つ統計モデルの調査

   ここでは、アルゴリズムとモデルを分けたうちの、モデルの方について扱う。とくに、シンボル / パターン統合のために役立つモデルについて調査した。また、シンボルに関する統計モデルについても調査した。

3. 確率の計算のためのアルゴリズムの整理

   ここでは、アルゴリズムとモデルを分けたうちの、アルゴリズムの方について扱う。周辺確率の計算法として有力な手法を 3 つにまとめて、統一的にとらえることを試みた。また、周辺確率の計算法とエネルギー (距離) を最小化する手法の関係、情報処理の手法と統計物理における手法の関係をあきらかにした。

4. 自律システムの数理的基礎の探索

   ここでは、上記の枠組におさまり切らない手法の中から、将来の情報処理技術の基礎となるものについて論じる。特に非線形科学の応用に重点をおいた。

## 2 本文

1. ベイズ統計の枠組の利点と欠点の検討

   (a) 記号論理との比較・ “確信度” を用いる枠組との比較

   記号論理と比較した場合に、確率による記述を用いることの利点は、多少の内部矛盾があっても処理が無効になったり停止したりしないことである。

   上記の点は確率のかわりに何らかの数字で表される “確信度” のようなものを用いたシステムでも同じである。 “確信度” 一般に対しての確率の重要な特徴は足し算と掛け算の両方が意味を持つことにある。これは、それ自身とその対数の両方について足し算が意味を持つといっても同じである。たとえば、壊れたファイルを復元した場合に可能な直し方が複数あるとする。可能な復元結果のうちで多数のものに $A$ という部分が含まれているほど $A$ が本物である可能性が高いと感じられるだろう。このときには確率の和を考えていることになる。一方、 2 つの独立に復元できると思われる部分があったときに、ひとつが $A$ であると同時に他方が $B$ である確率は、それぞれの確率の積になる。一般に和と積の両者が意味を持つおかげで、エントロピー (場合の数) の効果を正面から扱うことができる。熱平衡の統計物理学やシャノン流の情報理論が確率論の上に構築されている根拠もそこにある。

   確率の別な特徴として、和が 1 に規格化されなければならないということがある。これはしばしば計算を面倒にする原因になるが、規格化定数の部分に含まれる情報が重要な役割を果たすこと

も多いので、必ずしも欠点とはいえない。後出の finite mixture model を用いた推論はその例である。

(b) 通常の統計学の枠組との比較

通常の統計学 (頻度主義) の立場では、はじめにどれが確率変数かを決めてしまう。これに対して、ベイズ統計の枠組ではパラメータも確率変数と見なす。このため、ベイズ統計の枠組では、データの一部が欠けていればそれらをそのまま推定すべきパラメータとみなすことができる。これは、確率モデルにそれぞれ確率を割り当てるような階層的な構成を扱う際にも好都合である。

ベイズの枠組では、情報を積極的に採り入れた事前分布 (informative prior) を考えることで、従来の枠組では推定手続きの一部であった "罰金項" をモデルの一部として解釈することが可能である。これは単なる解釈の問題ではなく、データから罰金項の大きさを決める手段を与えるという意味でも重要である (たとえば、 Good、赤池弘次、 MacKay)。

しかし、罰金項をモデルの一部と解釈することには問題もある。罰金項であれば、外部と独立に意味を持つ必要はないが、事前分布はデータがない場合でも独立に意味を持たなくてはならない。これはシステムが事前分布の形で外界の (ある意味での) コピーを所有することを意味する。このような考え方は、外部と密接に相互作用した状態でのみ機能を持つようなシステムの記述に際しては適切でないかも知れない。

(c) 将来の情報処理技術の基礎として考えた場合の問題点

現代的な統計学の枠組に共通することであるが、ベイズ統計の枠組においてもモデルとアルゴリズムははっきりと分離して記述される。これは宣言型言語において、充足すべき論理式とそれを充足する手続きを分離して考えることに対応している。また、統計物理において平衡状態を規定するギブス分布とこれを実現し維持するダイナミクスを分離して考える立場とも共通点がある。このような方式には利点も多いが、将来にわたってこの考え方で良いのかどうかはわからない。統計物理においては、研究の中心はこのような分離が意味を持たない現象の探求に移行しつつある。これとの比較で考えると、ベイズ統計は熱平衡統計物理に対応する記述形式であって、その先には、手続きの分離が意味を持たないような本質的に非平衡の世界があるのかもしれない。

外界の環境と active に相互作用しながら動作をするシステムを設計するためにベイズの枠組が適切かどうかは難しい問題である。ベイズの枠組ではある時点までのデータで計算された事後確率の値を利用して次の行動を制御することができる。この意味では、ベイズは環境と active に相互作用するシステムに向いているといえる。しかしながら、この場合に、外界に active に働きかけてデータを採取する部分とデータから事後確率を計算して予測を行なう部分はやや異質であり、枠組としての一体感に乏しいようにも感じられる。

2. シンボル / パターン統合のために役立つ統計モデルの調査

下記のようなモデルが興味ある研究対象として指摘される。以下では、通常の統計モデルとベイズ統計特有のモデルとを区別せずに並べた。通常の統計モデルはベイズ統計の枠組でも解釈できるからである。

(a) シンボルと画像の統合に関して

初期視覚レベルでのシンボルと画像の統合に関しては Geman 等により導入された階層マルコフ場モデルの重要性が従来から指摘されている。階層マルコフ場モデルにおけるハイパーパラメータの学習 (環境適応) の問題は重要であるが、完全には解決されていない。より大域的で高次の画像の構造を表すモデルとしては **active contour model** や **deformable template model** (たとえば、 Grenander) などが提案されている。これらは興味深いモデルであるが、人間に近い視覚をもったシステムを実時間で動作させる目的にはいまひとつものたりない。高次視覚の認知科学的、脳科学的な理解に基づいたモデル化が必要なように思われる。

(b) シンボルに関して

シンボルに関する統計的な手法としては、まず、シンボルの間に直接距離を定義してそれを用いる方法がある。具体的には、

i. シンボルを仮想的な高次元空間に配置する。不要な次元は落す。
例: 数量化

ii. シンボルを仮想的な木の上に配置する。意味のない枝先は落す。
例: 階層的クラスター化

などがある。これらの手法は形式的な統計モデルの枠組の外で論じられることも多い。

別のアプローチとして、シンボルの組に対して頻度を与えるモデル (多項分布モデル) のパラメータをさらにパラメトライズするタイプの方法がある。この方法のポイントは、頻度を直接用いるかわりに適当な同一視もしくは再パラメータ化を行なうことによって、データ数が少ない場合にも安定した学習を可能にするところにある。

i. 無視できる要因を除去して予測に必要なもののみを残す。
例: 分割表モデル、決定木モデル

ii. 確率の対数を要因の 1 体、 2 体、.. の交互作用に分解して高次を落す。
例: 対数線形モデル

(i) は **CATDAP**(坂元慶行) や決定木 (たとえば、 Quinlan) によるデータ解析ツールとして実用性が示されている。いわゆる連想記憶や Hinton らのボルツマンマシン (ここではアルゴリズムではなくモデルを指す) では、 (ii) に近いモデルを考えているが、高次の交互作用のかわりに隠れユニット (潜在変数) を多数使用している。最近流行の **Bayesian network** も (ii) の一種とみなせる。

シンボルの (時) 系列に関しては、マルコフ連鎖をベースにしたモデルが良く利用される (例: 隠れマルコフ連鎖 (**hidden Markov model**))。また、いわゆる形式言語の特定のクラスに確率的規則を導入して統計モデルとしたものがある (例:(確率的) 有限状態機械 (**finite state machine**), 確率的文脈自由文法 (**probabilstic context free grammer**))。

シンボル系列のモデルはテキスト圧縮の見地からも研究されている (例: **FSMX** モデル)。データの確率構造を良く表現できるモデルがあれば、情報を高度に圧縮して貯えることができるからである (シャノンの定理)。

この分野で利用できるモデルは現在のところ限定されている。今後の発展のためには、研究者が分野や分野による術語の違いに因われずに経験を共有し、既存の知識の最大限の活用をはかることが望ましい。また、モデル選択やあてはめに時間のかかるようなモデルも genetic algorithm やマルコフ連鎖モンテカルロ法などの手法を導入することで有効に使えるようになるかもしれない。これは一種の自己組織化技法とも考えられる。

(c) シンボルの生成に関して

対象のクラス分けを教師なしで学習させることを考える。生成された各クラスをシンボルに対応させることにすれば、これはシンボル生成の最も簡単な場合とみなせる。この問題は統計モデルにもとづいたクラスター分析の問題として定式化することができ、従来から知られている **finite mixture model** に帰着する。このとき、確率分布が正しく規格化されていなくてはならないという条件はある種の競合学習に対応する。

finite mixture model は理論の立場からも興味深い。特に重要なのは最適なクラスター数 (シンボルの個数) をデータから決定する問題であって、通常のカイ 2 乗検定や AIC などが適用できないことが知られている。ひとつのクラスターが消滅するときに、 (i) その重みが零になる、 (ii) 他のクラスターと完全に重なってしまう、の両者があることが問題を難しくしている。

この他、すぐ後でのべるニューラルネットワークのトポロジーの学習なども、シンボル生成の一種としてとらえられる。社会的な状況、たとえば群れの中でのコミュニケーションを通じてのシンボル生成の問題 (たとえば、橋本敬・池上高志) も重要であるが、統計モデルの範囲には収まりにくい。

今後の展望についていえば、単にシンボルが生成するというだけでなく、生成されたシンボルがうまく使われるような本格的な例が必要である。また、上述のシンボル観では、シンボルの組合せ的な特質が十分現れていないという問題がある (谷淳)。

後者に関しては、何を内部状態として選ぶかを含めて finite state machine の学習を考えるのがひとつの方向である (たとえば、 Mataric のロボット)。

(d) ニューラルネットワークに関して

ニューラルネットワークの研究の流れとしては、実際の生物あるいは脳に即した方向と工学技術としての確立を目指す方向がある。

実際の生物の神経系のダイナミクスは、マルコフ連鎖モンテカルロ法のダイナミクスのような熱平衡状態のまわりの揺らぎではなく、もっと非平衡的なものである。これはシナプス結合の一方向性 (ミクロレベルでの双方向性の欠如) からも明らかである。もちろん、ダイナミクス (アルゴリズム) のレベルで非平衡的であっても、モジュール全体としてはベイズ的な計算を行なっているという主張は可能であるが、現在の脳研究の流れはミクロレベルの機構を尊重する方向にあると思われる。

一方、工学技術として使われている “ニューラルネット” のほとんどは統計モデルとして解釈できる。ただし、各種のニューラルネットが統計モデルとしてひとつのカテゴリーに属するわけではない。統計の立場からみると、工学者はパラメータが多数あるモデルのいくつかをなんとなくニューラルネットと呼んでいるように思われる。また、モデルとアルゴリズムの分離が可能であるにも関わらず、両者が混同されていることが多い。たとえば、 “バックプロパゲーション” はアルゴリズムの名称で “多層パーセプトロン”はモデルの名前であるが、前者がひとり歩きしているような観がある。 “ボルツマンマシン”にいたっては、アルゴリズムの名称かモデルの名称か全くわからない。

統計モデルとしてのニューラルネットの研究で興味深いのは、汎化の問題である。通常の統計処理においても、たとえば 100 個のデータ点を 100 変数の多項式で当てはめれば、当てはめは完全であるが、情報処理としては全くナンセンスである。ニューラルネットもデータ数に対していたずらにサイズを大きくすれば同じ結果を招くことは明らかである。種々の方策が提案されているが、シナプス強度に罰金を課する方法がもっとも合理的であるように思われる (Rumerhurt)。この方法は階層的なベイズモデルの文脈で解釈することもできる (MacKay)。特に興味深いのは罰金を非線形にした場合の振舞いであって、ネットワークのトポロジー (結合の仕方) を学習するのに相当する効果が生じる (石川真澄、石黒真木夫、 Williams)。対応する事前分布を石黒は準ガウス分布と呼んでいる。 Williams の **Laplace prior** はその特殊な場合に相当する。汎化のための手法は、簡単な例での実験を終って、より実用的なシステムで有効性を検証すべき段階にさしかかっていると思われる。

3. 確率の計算のためのアルゴリズムの整理

少し大規模な問題では、事後分布による統計量の期待値や周辺分布を求めようとすると、超高次元の積分や組合せ爆発する和の計算が必要になる。このことはベイズ統計を実際の問題に適用する場合の大きな障害となって来たが、近年ではいろいろな対処法が開発されている。

熱平衡系の統計物理においても高次元・多変数のギブス分布を扱う上で類似の困難があり、それを克服する手法の研究が発達しているので、それらを利用するのがひとつの考え方である。実際には、(数値的) 転送行列法と非線形フィルタリング・ Baum-Welch 法のように両方の分野で並行に開発された技法もある一方で、マルコフ連鎖モンテカルロ法のように、統計物理からの技術移転が成果をあげた例もある。

連続変数の場合にもっとも簡単な方法は、何らかの最適化手法を利用して事後分布が最大になる点 (mode,MAP 推定値) を求めてそのまわりでガウス近似 (線形近似) を行なう方法である。それ以外の主な方法としては、 (a) 再帰的な方法、 (b) マルコフ連鎖モンテカルロ法、 (c)self-consistent な近似法、がある。

(a)(b)(c) のそれぞれの基本的な特徴は次の通りである。

(a) 再帰的な方法 (フィルタリング・ **Baum-Welch** 法・転送行列法)

- 丸め誤差・離散化誤差を除けば厳密 (更に近似を加えることも可能)。
- 適用範囲は限定される。
- メモリが多量に必要。
- 乱数は通常使わない (例外あり)。
- 基本的には予め決まったステップ数で終了する。

(b) マルコフ連鎖モンテカルロ法

- 無限時間かければ厳密。
- 適用範囲は広い。
- CPU が多く必要。
- 乱数が不可欠。
- iteration により無限に続く列を生成する。

(c) **self-consistent** な近似法 (平均場近似など)

- 近似的 (系統誤差あり)。
- 適用範囲は比較的広い。
- メモリも CPU も比較的少なくて済む (近似精度をあげると急速に増える)。
- 乱数は通常使わない。
- iteration により収束する列を生成する。

情報処理と統計物理、さらに情報処理の異なる分野において同じアルゴリズムに全く違う名称が用いられている場合が多く、相互理解を深める上で問題である。以下では用語を列挙するとともに、それぞれの方法の簡単な解説を記す。

(a) 再帰的な方法

これは分割統治法の一種である。システムが $A$ と $B$ の 2 つの部分に分けられ、 $A$ のうち $B$ と直接相互作用している部分 $\hat{A}$ と $B$ のうち $A$ と直接相互作用している部分 $\hat{B}$ がそれぞれ少数の要素からなるとする。このとき、 $\hat{A}$ のあらゆる組合せについて $A$ の部分に関する和を計算し、 $\hat{B}$ のあらゆる組合せについて $B$ の部分に関する和を計算しておけば、それを組み合わせてシステム全体 $A \cup B$ についての和を求めることができる。 $A$ 及び $B$ を更に分割してゆくことによって、再帰的に全体の和を計算できる。

上の説明では、システムの “つながり具合” がネットワーク的になっている場合が頭に浮かぶかもしれない。このような応用は実際に存在している。家系図上での遺伝子の伝搬を追跡する問題で **peeling** と呼ばれている方法がその例である。しかし、圧倒的に多くの応用は、システムが 1 次元の鎖状あるいは帯状をしている場合に対してである。このような問題は時系列から生じることが多いが、画像を帯状の領域に分けて近似的に取り扱うことによっても生じる。この場合、長い方向に直交するような分割を考えることになる。また、再帰的な分割を行なうというより、前向きと後向きの計算をそれぞれ行なうという形になるのが普通である。代表的な例としては、時系列解析におけるカルマンフィルタ、その一般化としての非線形フィルタ (北川源四郎)、隠れマルコフモデルに対する **Baum-Welch** 法 (の和の計算の部分) などがある。また、確率的文脈自由文法で使われる **inside-outside** 法も基本的には同じ原理に基いている。

モデル族が特別な性質を持っている場合は、モデルに関する和を再帰的な方法でとることにより、混合モデルを用いた情報処理を行なうことができる。 FSMX モデルに対する **context tree weighting** の技法はその例である。

非線形フィルタを近似的に乱数を用いて実行するのが、モンテカルロフィルタ (北川源四郎) である。乱数を用いる手法には **sequential imputation** と呼ばれるものもある。

統計物理では再帰的な計算法は転送行列法・転送積分法として古くから知られている。また、量子力学におけるシュレディンガー方程式はファインマン経路積分に対する再帰的な計算法とみなせる。

(b) マルコフ連鎖モンテカルロ法

統計物理では、熱平衡状態での確率分布 (ギブス分布) は、システムのミクロなダイナミクスによって生成されると考えられている。このダイナミクスは本来はニュートンの運動方程式を解くことによって得られるものであるが、もう少し粗視化して適当なマルコフ連鎖やランジュバン方程式を考える場合もある。この関係を逆転させれば、与えられた分布を定常分布として持つようなマルコフ連鎖やランジュバン方程式を利用することによって、その分布からのサンプリングを行なうことができる。これが、マルコフ連鎖モンテカルロ法 ($(\mathbf{MC})^2$) の基盤となる考え方である。

与えられた分布 $P(X)$ を定常分布として持つようなマルコフ連鎖を作るための簡単な方法は、遷移確率 $q(X \to Y)$ を詳細釣合の式

$$P(X)q(X \to Y) = P(Y)q(Y \to X) \tag{1}$$

を満たすように定めることである。 $q(X \to Y)$ はごく近い $X$、 $Y$ を除いては 0 として良い。但し、遷移を複数回繰り返すことで任意の状態の対の間を移れる必要がある。

具体的な $q(X \to Y)$ の決め方にはいろいろな種類があり、それに応じて、メトロポリス法、**Gibbs sampler**(統計物理では **heat bath** 法という) などのアルゴリズムがある。マルコフ連鎖モンテカルロ法はそれらの総称である。

ボルツマンマシンという用語はしばしばマルコフ連鎖モンテカルロ法と同義に使われる。しかし、前述したようにこの用語はアルゴリズムでなく特定のモデルの意味に使われることもあり、また、特定の学習則 (ボルツマンマシンの学習方程式) を示唆することもある。さらに、後述の simulated annealing 法を指す場合もあるようである。

マルコフ連鎖モンテカルロ法の収束を改善するための工夫として、与えられた分布に手を加えたものにマルコフ連鎖モンテカルロ法を適用する方法、あるいは、複数の分布を同時にシミュレートする方法が開発されている (**simulated tempering** 法 (Marinari and Parisi), マルチカノニカル法 (Berg 他), **Metropolis-coupled MCMC** 法 **(MCMCMC)** あるいは 温度並列シミュレーテッド・アニーリング法 (Geyer、木村・瀧))。また、マルコフ連鎖モンテカルロ法と密接に関係した計算法としてランジュバン方程式を利用する方法、およびその発展としてのハイブリッドモンテカルロ法がある。

マルコフ連鎖モンテカルロ法は統計物理に起源を持つが、統計物理ではマルコフ連鎖モンテカルロ法という名称は使われない。統計物理では固定した分布を用いて importance sampling を行なう方法はあまり使われないので、単にモンテカルロ法といえばマルコフ連鎖モンテカルロ法を指すことが多い。特に区別の必要のあるときは動的モンテカルロ法と呼ぶ。

(c) **self-consistent** な近似法

ベイズ統計では考えている世界全体についての (同時) 確率が基本になる。たとえば、画像を考える場合には、画像全体 $X$ のとりうる状態のすべてについて事後確率 $P(X)$ がまず定義される。その上で、 $X$ が複数の部分 $A, B, C...$ からなっている場合の各部分に関する確率は、

$$P(A) = \sum_B \sum_C ..P(A, B, C..) \tag{2}$$

のように他の部分について和をとる (周辺化する) ことによって定義される。われわれにとって考えやすく、実際に欲しいのは、 $P(A)$ などの部分に関する確率である。再帰的な方法やマルコフ連鎖モンテカルロ法の場合も計算の目標はこれらの周辺化された確率 (周辺分布) であった。

このように考えてみると、なんらかの方法で周辺確率 $P(A)$ の満たすべき式を求めて、その方程式を iterative に解くことで無駄なくこれらを求めることができないかという考えに導かれる。残念ながらこれを厳密に行なうことはできないが、高次の相関 (依存関係) を無視することで、$P(A), P(B)..$ の間に近似的な閉じた関係式が得られる。このようにして導かれる計算法 (のひとつ) が平均場近似 **(mean field approximation)** である。最も簡単なのは $A, B..$ が最小の要

素 (画素など) である場合であって、単に平均場近似というとこれを指すことが多い。連続変数を離散化して平均場近似を行なおうとすると計算量が増え過ぎることがあるが、この場合も、同じ精神に基づいて期待値などに関する閉じた関係式 (self-consistent な関係式) を立てることが行なわれる。これも平均場近似と呼ばれる。

部分に関する確率 (確信度)$P(A), P(B)..$ に対して iterative な計算を行なう方法としては、画像解釈における **relaxation labeling** 法が良く知られている。平均場近似との最も大きな違いは、通常の relaxation labeling 法ではデータが初期値に入るだけなのに対し、平均場近似ではデータが iteration の式に含まれていることである (たとえば、 Titterington)。 relaxation labeling 法ではモデルとアルゴリズムは明確に分離されていない。

平均場近似の名は統計物理に由来する。また、 self-consistent 近似の考え方は物理全体で使われている。量子力学における **Hatree** 近似はその例である。

上で説明した期待値・周辺分布の計算法は、最適化の諸手法と密接に関係している。これを明示するためには、 "温度"$T$ を導入してもとの分布 $P(X)$ を

$$P_T(X) = \frac{P(X)^{1/T}}{\sum_X P(X)^{1/T}} \tag{3}$$

と変形すれば良い。 "温度"$T \to 0$ の極限で、変形された分布 $P_T(X)$ はもとの分布が最大となる点 (mode,MAP 推定値) $X = X_{mode}$ の上の $\delta$ 分布に収縮する。従って、任意の量 $A(X)$ の期待値は $T \to 0$ で $A(X_{mode})$ に一致する (分布が最大となる点が唯一つの場合)。このことから、十分小さい $T$ において $P_T(X)$ に対して期待値・周辺分布の計算法を適用することは、分布の最大値を求めるための最適化を行なうのと本質的に同じであるといえる。

上では、考えの便宜のために "温度"$T$ を導入した。これに対して、十分大きい $T$ で分布 $P_T(X)$ に対して期待値・周辺分布の計算法を適用し、その結果生じた状態を少し小さい $T$ についての計算の初期値として与えることを繰り返す最適化の方法がある。十分ゆっくりと $T \to 0$ とすることによって、局所的な極値にとらえられることを避け、大域的な最適解に近いものを求めるのが狙いである。**simulated annealing** 法や **Hopfield-Tank** 法はこのような方法の例である。

以下では、上の対応づけによって、各手法がどのような最適化手法に対応するかを示す。

(a) 再帰的な方法

再帰的な方法は最適化手法としての動的プログラミングに対応する。大雑把にいえば期待値・周辺分布の計算法において確率の和をとる部分で、動的プログラミングではエネルギー (距離) の最小値をとっていることになる。確率の積とエネルギーの和はそのまま対応する。再帰的な方法の場合には、ゆっくり $T \to 0$ とすることには通常意味がない。

(b) マルコフ連鎖モンテカルロ法

$T \to 0$ ではマルコフ連鎖モンテカルロ法は山登り法もしくは最急上昇法 (最急降下法) にほぼ対応する。有限の温度 $T$ でマルコフ連鎖モンテカルロ法を適用し、ゆっくり $T \to 0$ とするのが、**simulated annealing** 法である。

なお、 simulated annealing の考えはどの最適化問題にも適用可能であるが、温度 $T = 1$ のときの分布に意味を与えることができるのはベイズ統計 (または統計物理) の問題に適用した場合のみである。たとえば、巡回セールスマン問題における温度 1 での期待値には工学的な意味はないと考えるのが普通である。

(c) **self-consistent** な近似法

$T \to 0$ では self-consistent な近似法は山登り法もしくは最急上昇法にほぼ対応する。有限の温度 $T$ で近似的な方法でも、 $T \to 0$ では厳密な局所的極大を与える場合がある。有限の温度 $T$ で平均場近似の計算を行なって、ゆっくり $T \to 0$ とする (もしくは小さいが 0 でない $T$ で計算を行なう) のが、 **Hopfield-Tank** 法に対応する。分布に意味を与えられるのがベイズ統計や統計物理の問題に適用した場合のみであるのは、 simulated annealing の場合と同じである。

4. 自律システムの数理的基礎の探索

以前から、非平衡・非線形システムの科学が情報処理技術やロボット技術に寄与することが期待されてきた。最近の話題としてはカオスの応用がある。またリミット・サイクルの応用も以前から議論されている。これらの応用の中には興味あるものもあるが、全体としては十分に満足できる内容とはいい難い。たとえば、カオス・アニーリングにおいてはカオスが外部から加えた乱数以上の役割を果たしているかどうかという疑問が避け得ないし、カオスの制御においては結局カオスは制御によって排除される存在でしかないのではないかという疑問が生じる。生物の示す周期的な変化や運動をリミット・サイクルに同定してみても、単なる “見立て” 以上のものがあるかどうかは必ずしもはっきりしない。非線形、非平衡の系の興味ある現象を工学的に利用する場合、その現象の本質的な部分がシステムの機能にとって不可欠の形で生かされていなくては研究の本来の意図が達成されたとはいえないであろう。

ベイズ統計の枠組で行なわれたいくつかの研究には、この点で評価できるものがある。たとえば、階層的なマルコフ場モデルとマルコフ連鎖モンテカルロ法を組み合わせた情報処理はマルコフ場の自己組織的な特性、確率的な揺らぎなどが、情報処理能力に本質的に貢献している点で評価できる。ただし、このシステムは統計物理の立場からみたときには熱平衡系 (あるいは熱平衡に緩和していく系) であり、リミットサイクルやカオス、乱流などの非平衡系特有の多様な現象は示さない。強非平衡系の工学的利用が次の課題となるが、前述のようにこれは簡単ではない。

しかし、最近になって、興味深い研究がいくつかあらわれてきた。ストレンジアトラクタを内部表象としたロボット・ナビゲーション (谷淳) やリミットサイクル振動子による 2 足歩行モデル (多賀厳太郎) がその例である。これらの研究の特徴は、システムが自律的でダイナミックな内部を持っているのにもかかわらず、内部とは異質の外部 (環境・工学的目標) とうまく統合されている点にある。一般に工学者はシステムの変化の駆動力を外部に求める傾向が強く、逆に非線形科学の研究者は内部の自発的な変化の研究がすべてであると考える傾向にある。しかし、自律的な知的システムを開発するためには、その両者に目配りをすることが不可欠のように思われる。外部と内部の両者、あるいは、それらの interface について考察することは、異質な表現系の統合 (麻生英樹)、記述不安定性 (津田一郎) といった問題意識にも通じるかもしれない。

強非平衡系で重要になるかもしれない問題としては、予測のレベルと制御のレベルのギャップの問題がある。熱平衡系においては、相関関数と微小外力に対する応答のあいだには比例関係があることが古くから知られている (Einstein、久保亮五)。これに相当するものを利用した情報処理手法としてはボルツマンマシンの学習方程式がある。非平衡系ではこの関係は成立しないので、予測と制御の間にはギャップが生じる。われわれが日常生活で考えている因果関係は、予測よりむしろ制御に関係するという考え (たとえば、中島秀之・松原仁・大澤一郎) がある。もし、相関関係 (厳密には非独立性) を予測に、事象の因果関係を制御に対応させるなら、予測と制御の間のギャップは相関関係と因果関係のギャップに対応することになる。

わが国の非線形科学研究のレベルは伝統的に高い。現状では外部への適応とシステムの自律性の両方に目を向けた研究は少ないが、その潜在力が次世代の情報処理技術に生かされることを期待したい。

いくつかの内容、たとえばモデルとアルゴリズムの分離と宣言型言語の関係について述べた部分は、麻生英樹氏との議論に多くを負っています。また、いつも議論していただいている統計数理研究所のスタッフの方々、 RWC の Vision 検討会・ PKA ワークショップ・ info-integ メイルリストのメンバの皆さん、東京大学教養学部池上研究室をはじめとする複雑系・非線形科学の研究者の方々に感謝の意を表します。

REGISTRATION
of
Research Memorandum

(month) (day) (year)

Research Memorandum _____ Received on _____ _____, _____ by

Section of Research Information, the Institute of Statistical Mathematics

(*Do not write above this line* )

. . . . . .

(*Fill in the blanks below and return to Section of Research Information*)

| |
|---|
| Title<br>**統合技術のための統計数理的方法** |
| Author(s) and Affiliation(s)<br>YUKITO IBA<br>Institute of Statistical Mathematics 106 Minami Azabu Tokyo<br>Email: iba@ism.ac.jp |
| Key Words<br>**ベイズ統計, シンボル, モデルとアルゴリズムの分離, ニューラルネットワーク, 再帰的な方法, マルコフ連鎖モンテカルロ法, 平均場近似, 非平衡, 因果関係** |
| Abstract<br>**“ポスト記号論理” の人工知能の基盤としてのベイズ統計の利点と欠点について検討し、現在利用可能な技術をモデルとアルゴリズムの両面から列挙した。また、将来の情報処理技術のなかで統計の枠組に必ずしも収まらない部分についても考察した。** |