Research Memorandum No.632 Feb. 1997

柔らかな情報処理と統計学

ver.1.961

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, URL http://www.ism.ac.jp/~iba/

統計学が最近 (あるいは近い将来) の情報処理の研究においてどのような役割を果たしているか (果たしうるか) について、一般論およびいくつかの例を通じて考察した。委員会報告書のための草稿で、昨年度の

統合技術のための統計数理的方法 (2.5 版)
ISM Reseach Memorandum No.595, February, 1996.

の続編 (応用編) に相当するものである。

KEYWORDS

統計モデル, 記述と構成, 情報統合, 自己組織化, generalization, 遺伝子情報処理, タンパク質の設計, 神経生理学データの解析

もくじ

1 柔らかな情報処理と統計学の考え方 **3**

1.1 統計モデルの有用性と特徴 3

1.2 記述的なアプローチと構成的なアプローチ 4

2 情報統合と統計学 **4**

2.1 情報統合と統計学 – “自己組織化” の基盤として 4

2.2 情報統合と統計学 – 部分を張り合わせる “糊” として 6

2.3 情報統合と統計学 – 問題点 6

3 統計的情報処理の新しい応用分野 **7**

3.1 遺伝子情報処理 7

3.1.1 テンプレートの生成 8

3.1.2 遺伝子情報処理における情報統合 8

3.2 タンパク質の設計 9

3.2.1 タンパク質設計の計算理論 9

3.2.2 ミクロな物体の設計 10

3.2.3 平衡? 非平衡? 10

3.3 神経生理学データの解析 11

3.3.1 位相同期の解析 11

3.3.2 機能のマッピング 12

1　柔らかな情報処理と統計学の考え方

ここでは、“モデル” がなぜ必要か、“統計” モデルの特徴は何か、という点を簡単に述べる。すでにいろいろな形で説明してきたので、詳しい解説はそれらを参照されたい。統計モデルとしてどのようなものが利用できるか、また、それらを扱うためにどういう技法があるかについては昨年の報告の筆者担当分 [伊庭 1996a] で枚挙した。また、階層的な統計モデルについての具体的説明は [伊庭 1996b] に、確率を用いる具体的利点については [伊庭 1993] にある。情報処理の基礎として役立つ統計学の文献の紹介は [伊庭 1997] を、(統計) モデルと人工知能についての一般的な議論は [伊庭 1996c] を参照されたい。

1.1　統計モデルの有用性と特徴

まず、基本的な前提として、高度な情報処理を実現するためには、断片的な知識を集積するだけでは不十分である、ということがある。かりにいままで起こったことのすべてを記憶しているとしても、新しく起こることは必ず過去の経験とどこか違っている。従って、過去の経験した状況のどの部分に着目してどの部分を切り出したらよいか、過去の経験のうち何が雑音であって何が実体であるか、ということは経験だけから知ることはできず、何か内的な規則・構造のようなものが必要とされることになる。もちろん、このような内的な規則もまた経験によって修正・獲得することができるが、そのためにはまたより上位の規則が必要になる。

このような階層的な構造をうまく表現するには、統計モデルによる方法 (確率ベースの考え方) が有効である。統計モデルにおける確率構造は、単に個々の事象の確からしさを表わすものというより、推論において考慮される事象の範囲を明示的に表現するものとして考えるべきである。一般に、言語を共有することは世界の範囲とその可能性についての思考を共有することにほかならず、世界のモデルを作るとは、世界の範囲とその可能性の全体を定めることにほかならない。 そして、 “いま考慮されている可能性の全体” を表現するもっとも手近な数学的手段が確率構造なのである。

一般にある言明の確からしさを考えるには、その言明で言及されていない部分を無視する (粗視化する) 操作が必要である。確率的枠組では、これはある条件をみたす状態の確率の和をとる操作 (周辺化) に相当する。従って、この枠組では、ある条件を満たす場合の数がどの程度あるかが、言明の確からしさに影響することになる。 階層的なシステムにおいては、この効果は特に重要である。これは、統計物理とのアナロジーでいえば、マクロな量の分布を考えるには、原子などのミクロな要素の配置ごとに定まる “エネルギー” だけでなく、ひとつのマクロな状態に含まれるミクロな状態の場合の数、すなわち “エントロピー” を考慮する必要があることに対応している。

1.2 記述的なアプローチと構成的なアプローチ

上記のような考え方は、複雑なシステムを理解したり作ったりするための方法としては“記述的”な方法といえる。“記述的”といっても“表面的で本質に届かない”とか、“工学的”に対して“科学的”という意味ではなくて、「システム全体の目的やそのおかれている環境を分析して、できるだけ明示的・組織的に考えたり設計したりする」という意味である。統計モデルに基づく方法は良い意味でも悪い意味でも記述的な方法の特徴を強く持っている。D.Marr([マー 1987])の意味での“計算理論的方法”も記述的なアプローチのひとつの典型である。

これに対して、最近“複雑系”などに関連して、話題になっているのは“構成的”方法である。“構成的方法”のちゃんとした定義は難しいが、筆者の個人的な考えでは、「そいつが何だかはっきりとは掴めないままに、とにかくそれをやってみる」(“Doing something without knowing exactly what it is.”)というのがその精神をよく表現しているのではないかと思う(もともとの主張については[金子・津田 1996]を参照)。科学や技術が進歩する際には、とにかく何かをやってみた結果、典型的な特殊な例がつくられ、そこから一般論が構成されることが多い。人工知能や情報処理の研究だからといって、はじめから組織的にやることを要求するのはおかしいかもしれない。数値実験によるカオス研究は構成的なアプローチが成功した例であるし、Brooks流のロボット研究に注目が集まっているのもこの観点からであろう。

実際には、このふたつの観点は相補的であって、そのときのその分野の状態によって一方が優勢になったり、他方が優勢になったりする。また、両者の有効性もその時の状況に依存する。たとえば、アドホックな方法が多数繁茂して何が本質だかわからなくなった場合、あるいは、部分的な目標にこだわったために本来の目標が見失われている場合には、記述的な観点を強調することで新しい展望がひらけることが多い。これに対して、そもそも何が問題なのか判断する材料すらないような場合は、無根拠でも、とにかく何かやってみることが有効だろう。一方の観点を教義のように考えるのではなく現状を打開するための発想法としてとらえる柔軟さが必要である。

2 情報統合と統計学

ここでは、‘情報統合”の目的からみて統計的な技法がどう使えるかということを2つの側面(“自己組織化”、“糊”)にわけて考察する。また、その問題点についても論じる。

2.1 情報統合と統計学 – “自己組織化”の基盤として

認知科学やロボットの研究において、初期視覚や運動制御の理解のためには、“計算理論”のような記述的な方法がかなり有効であった。また、これらを統計学的な枠組で理解

しようとする試みも盛んに行なわれた ([Marroquin et al. 1987], [川人 1996], 批判的な議論は [Aloimonos 1994])。これに対して、 RWCP など最近の研究計画では、高次の視覚や行動、一般に意味的・シンボル的な要素を含む問題に挑戦したいという動きがある。この問題は、しばしば “シンボル・パターンの統合” あるいは “縦の情報統合” として理解されている。この見方自体は、かなり一面的な見方である (この点の詳しい議論は [伊庭 1996c] を参照) が、これが重要な問題であることは間違いない。

このような問題に対するひとつのアプローチは、単純で一般的なモデルから出発して、システムに大量の経験を積ませることで自己組織的に意味的・シンボル的な要素を習得させるという手法である。これを自己組織化手法というのは古いイメージがあるが、たとえば、進化的アプローチといってみても、実態としては同じようなものである (この種の手法に対する楽天的な展望を面白おかしく描いた SF に [ラッカー 1996] がある)。

このような手法が、たとえば自律ロボットの製作という課題にたいしてどの程度有効かはまだわからない。困難の第一は、どの程度の “generalization” が可能かという点にある。たとえば、家の中を歩いて移動する、という課題に対して進化的アプローチ (遺伝的アルゴリズム) を使うことにする。ランダムにロボットのプログラムを変更したものを沢山つくって、玄関から台所まで歩かせ、うまく移動できたものを選ぶことにすれば、いずれ、玄関から台所まで効率的に歩いて移動できるものが残るだろうと予測される。しかし、そうなったとしても、そのロボットが家の中での歩行というタスク一般について能力を獲得したとはいえない。玄関から台所までの床の凹凸や細かい障害物をを単に丸暗記しているものが生き残っている可能性もある。その丸暗記の仕方も、配置のわずかな差に対してロバストとは限らないので、通常の意味では、その家の構造を記憶したとさえいえないかもしれない。よりシンボル的な行動の例として、パズルを解かせることと、パズルを解く能力を発達させることの違いを考えてみよう。前者と後者は全く違うレベルの問題である。前者が簡単に最適化問題として定式化できる場合でも、後者で何を最適化すればいいのかは必ずしも自明ではない。

本稿のはじめに述べた “いままで起こったことのすべてを記憶しているとしても、新しく起こることは必ず過去の経験とどこか違っている。過去の経験した状況のどの部分に着目してどの部分を切り出したらよいか、過去の経験のうち何が雑音であって何が実体であるかは自明でない” というのはまさにこうした問題を指している。自己組織化といっても何らかの形で高次の規則をあたえなければならないのである。

統計学においては、こうした問題は早くから注目されていて、情報量規準や階層モデルなどの道具が開発されてきた。したがって、この困難の本質をもっとも明確に理解することができるのは統計学の立場からであると思われる。また、逆に、統計モデルを使うという前提でいえば、意味的・シンボル的な要素を扱うために利用可能なモデルは比較的少数の汎用的なものに限られている ([伊庭 1996a] 参照) ため、必然的に大量のデータ、大量の

計算を前提にした “自己組織化的” 方法をとらざるおえないという面もある。いずれにしても、両者の間には密接な関係があるといってよい。

自己組織化アプローチには他にも問題がある。このうち、よく知られているのは、 “最初の一歩の困難” である。 2 足歩行を学習させる問題では、文字通りこれは “最初の一歩の困難” であって、いつもすぐにころんでしまうなら、得点はつねに零であって、学習も進化もしようがない。この問題はかなり本質的であるが、極めて特別な場合を除いて、統計学的な枠組によって一般的に処理するのは難しそうである。

2.2 情報統合と統計学 – 部分を張り合わせる “糊” として

複数のモジュール (エージェント) からなる情報処理システムにおいて、確率的な枠組を用いて各モジュールの統合をはかることが考えられる。確率の値は情報の種類によらない普遍的な意味を持つので統合には好都合である。また、モジュールの間のやりとりのためには、状況全体 (たとえば、画像全体) に対する確からしさの代わりに部分 (たとえば、特定の場所が赤いかどうか) についての確からしさをうまく利用した方が効率がよい。確率的な枠組では周辺確率 (着目している部分以外を無視して、それについての和をとった確率) と条件つき確率を使いわけることで、部分と全体の関係をうまく処理することができるので便利である。

これに関係した興味ある考え方として統計物理における平均場近似 (自己無矛盾近似) がある ([Peterson and Hartman 1989], [伊庭 1996d], [伊庭 1996e])。この手法の基本は、ここでの文脈に即した用語で説明すると、 “モジュールの間のやりとりに周辺確率を利用し、 iterative な計算によってそれらの値を辻褄が合うように決めることでモジュールの統合をはかる” というふうに表現できる。

確率あるいは情報量が普遍的な意味を持つように見えるのは、統計モデルを与えることで、考えている “可能性の全体” を明示的に与えた結果であることに留意する必要がある。確率的な枠組では情報の “意味” の部分は統計モデルによって規定されているのである。確率の値が便利なことは上に述べた通りであるが、より重要なのはその背後にひかえているモデルである。このことを忘れて、確率的枠組それ自体の力を過大評価してはならない。 “情報理論は情報の ‘意味’ を考えていないので不完全である” といった議論の多くも、この点についての誤解にもとづいている。

2.3 情報統合と統計学 – 問題点

RWCP の有力な研究テーマである “自律ロボットにおける情報統合” や “マルチモード統合” においては、各部分がそれぞれ良く理解されているものをより良くするために統合するというわけでは必ずしもない。むしろ、探索的な研究を行なう際の発想のヒントとして “情報統合”、すなわち、貧弱なセンサーや不十分な個別処理からの情報をうまく利用す

るようなシステムで何か面白いものはできないか、ということがあるのだと思う。従って、統合によって何パーセント良くなるとかいうことよりも、統合の過程でどういうアイディアが生まれるかが重要である。そこでは、先の分類でいう “構成的なアプローチ” が重要な地位を占めることが予想される。 “特定の目的を持った行動” の単位で情報を再分割して考えるという方針 ([Aloimonos 1994]) は、 “情報統合” の有力な方向と思われるが、この場合も、どのような行動を単位として捉えるかが最も重要であって、情報統合のための一般的手法はどちらかといえば実装のための道具である。

このような状況で、部分を張り合わせるための一般的手法としての確率的な定式化に頼りすぎると、もっぱら “糊” や “つめもの” の開発を行なうことになってしまい、本質を見失う可能性もないとはいえない。先に Marr の “計算理論” を、記述的アプローチの例として、統計モデルの考えに近いと分類したが、 Marr 自身は確率論的な枠組にはむしろ批判的であった。確率論的な枠組の便利さにかえって危惧を感じたためであろう。今まで述べてきたような文脈での “情報統合” の研究においては、一般論の研究と同時に、自由な発想にもとづいた個別の研究を奨励することが重要と思われる。

3 統計的情報処理の新しい応用分野

ここでは、広い意味での柔らかな機械・柔らかな情報処理の分野で統計学の考え方がどう役立つかを検討する。具体的には

- 遺伝子情報処理
- タンパク質の設計
- 神経生理学データの解析

について述べる。生物関係が主となったが、もちろん生物関係以外でも興味あるテーマはある。たとえば、

- 大規模数値計算への統計学的アプローチ
- 計算機ハードウエアの誤りの検出

についても触れる予定であったが、〆切の関係で割愛せざるおえなかった。これらについては別の機会に述べたい。

3.1 遺伝子情報処理

いわゆる “遺伝子情報処理” は現在流行の分野であり、統計学的な手法も重要な寄与をしている。以下では、シンボルの生成や階層モデルによる情報統合など柔らかな情報処理に共通の観点から、この分野を論じる。より広範囲の展望や日本の研究者の寄与については、[星田 ほか 1994] 及び [長谷川・岸野 1996] を参照されたい。

3.1.1 テンプレートの生成

DNA などの配列に対して、与えられた特徴的な部分配列を高速で検出するという問題は、データベースの検索などに必要であり、熱心に研究されてきた。比較的最近になって、隠れマルコフモデルなどの階層的な統計モデルに基づく方法で、特徴的な部分配列のテンプレートを自動生成することが試みられている ([Lawrence et al. 1993], [Krogh et al. 1994], [Liu et al. 1995])。これらの方法では、テンプレートの生成と各配列上での位置の推定が同時に行なわれる。技術的には Gibbs Sampler や Baum-Welch 法 (非ガウスフィルタ) などが使われている。

この例は、階層的な統計モデルを用いて、シンボル (あるいは、シンボル間の関係) を自己組織的に獲得する例として興味深い。関連した理論的な問題としては、検出された特徴が有意であるかどうかを判定する問題がある。この種の問題では、通常のカイ 2 乗検定や AIC がうまく適用できるとは限らないことに注意する必要がある。

3.1.2 遺伝子情報処理における情報統合

この分野で、複数の見方・レベルを同時に処理することでどういう利益があるかという点を考えてみたい。こうした試みは、自律ロボットやマルチモード統合のようなテーマでの情報統合とはだいぶ異なっているが、広い意味では情報統合の一種といえるだろう。

例として、系統樹推定における alignment と分岐年代推定の統合を考えよう ([Throne et al. 1991], [Thorne and Churchill 1995])。複数の生物の DNA やタンパク質を比較して系統関係を推定する ([長谷川・岸野 1996]) ためには、 DNA やタンパク質の配列のどの部分とどの部分を比較するかを決める作業が必要である。この作業を (global) alignment という。進化の過程で配列にはしばしば欠失や挿入が起こるので、この作業は自明ではない (立体視の場合の左右の眼からの情報の位置合わせ ([マー 1987]) に似た作業になる)。一般的には、まず最初に alignment を行ない、次に配列の比較によって分岐年代・進化速度を推定するという 2 段階の作業を行なうことになるが、 Throne たちの仕事では配列の欠失・挿入と置換 (点突然変異) の両者を含む統計モデルを設定することで両者を統合的に行なうことが試みられている。

Throne らの研究では、点突然変異の量を推定する際に、最適な alignment だけではなく、あらゆる可能な欠失・挿入について和をとった周辺尤度を利用している。これは、統計物理とのアナロジーでは、 “基底状態エネルギー” のかわりに “自由エネルギー” を (あるいは、最適経路のかわりに経路積分を) 考えることに相当する。彼らは、最適な alignment のみで点突然変異の量を推定した場合には、突然変異の量の推定値にバイアスを生じることを指摘している。最適解のまわりの揺らぎの効果 (“エントロピー” の効果) が効く例として、理論的にはたいへん面白い。

別の例として、タンパク質の高次構造推定と系統樹推定の統合がある。もともと、分子

時計による系統樹推定は、 DNA の塩基配列 (あるいはタンパク質のアミノ酸配列) が一定の確率で置換されるという仮説 (進化速度の一定性の仮説) から出発したが、実際には進化速度は、種によって、また、配列の部位によって異なっている。このうち、種による進化速度の相違については、統計モデルを利用した方法によってかなり詳しく測定できるようになったが、配列の異なる部分における進化速度の違いの研究は困難とされてきた。

もし、単に、 DNA の塩基座位 (あるいは、タンパク質レベルでアミノ酸の位置) ごとに進化速度が異なるとすれば、データの数に対してパラメータ数が多くなりすぎるため意味のある結果は得られない。このような場合に有効な手法として隠れマルコフモデル (HMM) のような階層的な統計モデルの利用が考えられる。最も簡単な方法は、 “隣接部位の進化速度の違いが少ない” という平滑化的な条件を考えることだろう ([Felsenstein and Churchill 1996])。より面白い考え方は (タンパク質 / アミノ酸レベルで問題を考えたときに) 隠れ状態としてタンパクの 2 次構造を考えるというものである ([Goldman et al. 1996])。 2 次構造の予測問題では従来から隠れマルコフモデルが使われている ([星田 ほか 1994]) から、それに系統比較・進化の要素を付け加えたと考えることもできる。いずれにしても、この “情報統合” によって、系統比較、構造の予測の両方の面に利益が生じることが期待される。

どちらの例でも、階層的な統計モデルをうまく利用することによって “情報統合” がなされている。複数のレベルの処理を簡単に一体化・双方向化できるという統計モデルの利点がうまく利用された例といえるだろう。

3.2 タンパク質の設計

タンパク質設計の問題では、揺らぎを考慮すること、すなわち、確率分布として設計することの重要性が指摘されている。この立場では、設計問題は、最尤推定 (あるいはボルツマンマシン学習) と形式的に類似した問題となる。ただし、有限個のデータからの推定ではなく、ターゲットとなる分布が正確に与えられているという点では、厳密な意味での統計的推定の問題ではない (AIC や MDL などはとりあえず意味を持たない)。

3.2.1 タンパク質設計の計算理論

タンパク質、あるいは、一般にひも状の物体を考える。ひもを構成するユニットであるアミノ酸 (モノマー) の配列を与えてどのような構造が形成されるかを予測せよというのが、高次構造の予測問題である。これに対して、逆方向の問題、すなわち、ターゲットとなる形状の方を与えてその形になるような配列を求める問題が考えられる。これが、ここでいうタンパク質 (ヘテロ・ポリマー) の設計問題である。高次構造の予測には、物理法則に近いところからシミュレーションによって形状を予測する立場と、 (前節で触れたような) 隠れマルコフモデルなどを利用した統計的アプローチがある。どちらの立場からも、 “逆問題” としての設計問題を考えることができるが、ここでは、前者の立場で考える。

設計に際して、どのような量を最適化するアミノ酸配列を選んだらよいであろうか？ ちょっと考えると、ターゲットとなる形状のエネルギーを低くする配列を選ぶのが良いように思われるが、これでは駄目である。他の可能な形状のエネルギーも同時に低くなってしまったのでは意味がない。そこで、種々の付加条件を加えた最適化 ([Shakhnovich and Gutin 1993], [Sasai 1995]) が考えられているが、ひとつの筋道として、“適当な温度での **Gibbs** 分布が、ターゲットとなる形状のところにできるだけ大きい重みを持つように配列を決める”という方針が考えられる ([Kurosky and Deutsch 1995], [Seno et al. 1996])。これによって、ターゲット形状のエネルギーと他の形状のエネルギーの差だけが効くようにするとともに、ターゲット形状と近いエネルギーを持つ形状の場合の数の因子 (エントロピーの因子) をとり入れて設計することができる。熱揺らぎを考慮して設計することは、物理的にみても、ある程度は妥当であろう。

この考え方を採用すると、設計問題は与えられた確率分布を確率モデルでできるだけ良く近似するようにパラメータを定めるという問題に帰着される。この問題は、“パラメータ” が “アミノ酸の配列” という離散的なものである点を除いて、通常の最尤推定の問題、あるいは (隠れユニットなしの) ボルツマンマシン学習の問題と形式的に類似している。

3.2.2 ミクロな物体の設計

タンパク質の設計問題は、製薬などへの応用が期待されているが、あまり性急な実用の期待は持たない方が無難であろう。ここで論じているのは、とりあえず “形状” の設計についてであるが、“機能” の設計はまた別問題かもしれない。タンパク質の酵素活性と形状、アミノ酸配列の関係について、実験的な理解がもっと必要と思われる ([Trakulnaleamsai et al. 1995])。

現段階では、タンパク質の設計を考えることは、マイクロマシンなどを含むミクロな物体の設計一般を考える上での例として意味があるのかもしれない。上の方法 (熱平衡的設計、確率モデルによる設計) では、 Gibbs 分布の分母の規格化定数 (分配関数) を通じて、対象のとりうる形状全体についての情報が設計に取り込まれている点に特徴がある。

3.2.3 平衡? 非平衡?

タンパク質の折り畳み過程そのものは、動的・非平衡的な過程であって、 Gibbs 分布によっては本来記述できないものである (さらに過激な立場としては、折り畳まれた後で酵素として機能している状態がすでに熱平衡状態とはみなせないという考えも可能かもしれない)。従って、少なくとも、折り畳み過程を明示的に考慮して “ランダムな初期状態からはじめて、ターゲット形状になめらかに折り畳まれること”を要求して設計するのが本当であるという意見もありうる。実際、 [Sasai 1995] の条件 IV や V は折り畳み過程に対する要求を反映させたものとも解釈できる。

動的・非平衡的なものが重要なのは明らかだとして、はじめからそのような要素を採り入れた設計をすべきだという立場もある。しかし、系統的扱いが可能な Gibbs 分布に基づく設計を試した上で、非平衡の面がどの程度効いてくるかを考えるのが、より組織的な方針であるように思われる。

3.3 神経生理学データの解析

今後の情報処理の研究において、人間や動物の脳から学ぶことは多いと思われるが、逆に、人間や動物の脳の働きを実験データから理解するには高度な情報処理が必要である。

複数の電極による同時測定や光学的方法による測定などの実験方法の高度化に伴って、また同期コーディング (相関コーディング) など新しい仮説の導入に伴って、脳研究におけるデータ処理の重要性はますます増大すると思われる。このうち、ここでは、位相同期の解析という非線形データの解析一般にかかわる問題と、機能のマッピングという脳特有の問題を考えてみる。

3.3.1 位相同期の解析

脳のなかで情報がどのような形でコードされているかという問題は非常に重要である。古典的には、パルスの頻度 (発火頻度) の短時間平均が情報の担い手であるとされてきた。これに対して、最近話題になっている説 (同期コーディング、相関コーディング) では、個々のパルスの相関あるいは位相の同期が情報の一部を担っているのではないかという提案が行なわれている ([Gerstein et al. 1989], [櫻井 1996], [藤井 1996])。

このような問題を実験的に解析するためには、さまざまな困難が予想される。そもそも “パルスの頻度の短時間平均” という場合に、どの程度の時間の平均がもっとも多くの情報を担っているのかは、データ解析の過程で見い出されるべきことである。また、平均をとる時間を極めて小さくとれば、頻度の短時間平均に着目する立場は個々のパルスに着目する立場に連続的に移行するから、上の 2 つの立場を区別するのは見た目ほど簡単ではないかもしれない。さらに、両者のコーディングが併用されているケースも考えられる。こうした複雑な状況で、系統的なデータ解析を行なうためには、階層的な統計モデルに基づく手法が効果的なのではないかと予想される。

より一般に、非線形力学系の解析という立場に立つと、位相同期・非線形引き込みの統計学的な扱いの重要性が指摘される。線形システムと非線形システムの大きな違いは、位相が情報を担う仕方の違いにある。 AR モデルのような線形モデルに基づく方法でも時系列同士の相関を扱うすることはできるが、それらで記述できるのは周波数表現で同じ周波数に対応する部分の情報の流れであって、 1:2 や 3:1 などの引き込みに対応する情報の流れは記述することができない。また、 1 つの時系列のピリオドグラムに山が 2 つあるとして、その間の位相の関係がどの程度揺らいでいるかを調べることもできない。位相同期を

記述できる柔軟なモデルを作ることは時系列解析にとって重要な課題であり、神経生理データの解析にも有効な可能性がある。

3.3.2 機能のマッピング

脳の神経細胞が外界のどのような刺激に反応しているのかは、古くから調べられている ([川人 ほか (1994)])。しかしながら、高次の領域、視覚でいえば物の形状などに反応する部分では、どういう刺激に神経細胞が反応しているのかを分析するのは容易ではない。複数の神経細胞によってひとつのことが表現されている可能性を考慮すると、困難はさらに増加する。

ひとつの方向として、社会調査の分析とのアナロジーを考えることができる。与える刺激をアンケート調査における設問に、それに対する細胞の反応を設問の答えに対応させて考えるわけである。神経細胞の反応の分析に多次元尺度法を利用することはすでに試みられている ([Young and Yamane 1992])。分割表の自動選択 ([坂元 1985]) も使えるかもしれない。これらの方法には心理学で開発されたものも多いので、刺激に対する反応の分析に利用できるのは当然かもしれないが、ミクロなレベルへの適用は新しい方向といえる。

高次の領域、たとえば高次視覚に関与する IT 野においても、似た反応をする細胞がかたまって存在するといわれている ([Wang et al. 1996])。これを解析するためには空間的側面を考慮したモデルが必要かもしれない。さらに、同期コーディングのような時間的側面まで考えると問題は極めて複雑になる。こうしたデータの解析は統計学にとって魅力的な題材である。

ここで考えたような問題では、脳研究の循環的側面があらわになっている。つまり、既存の情報処理手法によって、脳で用いられている表現やモデルを明らかにすることができれば、その理解を情報処理手法の改善に使うことができる。こうした良循環が続けば、脳の理解もわれわれの使える情報処理手法も進歩していくだろう。逆にこれが悪循環になれば、脳の理解も進まないし、脳研究の情報処理手法への寄与もないということになるかもしれない。これは、津田のいう解釈学的循環 ([金子・津田 1996]) のひとつの側面とも考えられる。

文献一覧

[Aloimonos 1994]

Aloimonos,Y. (1994), What I have learned (A reply to Tarr and Black's paper), CVGIP: Image Understanding, Vol.60, pp.74-85.

[Felsenstein and Churchill 1996]

Felsenstein,J. and Churchill,G.A. (1996), A hidden Markov model approach to variation among sites in rate of evolution, Mol. Biol. Evol., 13, pp.93-104.

[藤井 1996]

藤井宏 (1996), 脳の符号化機構への新しいシナリオ、科学, Vol.66, pp.880-888, (1996 年 12 月号), 岩波.

[Gerstein et al. 1989]

Gerstein,G.L., Bedenbaugh,P. and Aertsen,M.H.J., Neuronal assemblies, IEEE Transaction on Biomedical Engneering, 36, pp.4-14.

[Goldman et al. 1996]

Goldman,N., Thorne,J.L and Jones,D.T. (1996), Using evolutionary trees in protein secondary structure prediction and other comparative sequence analyses, Journal of Molecular Biology, 263, pp.196-208.

[長谷川・岸野 1996]

長谷川政美, 岸野洋久 (1996), 分子系統学, 岩波書店.

[星田 ほか 1994]

遺伝子情報処理への挑戦, 共立出版.

[伊庭 1993]

伊庭幸人 (1993), ベイズ統計と統計物理, — 有限温度での情報処理 —, 「物性研究」 60-6 (1993 年 9 月号), pp.677-699.

伊庭幸人 (1996), ベイズ統計と統計物理 (物性研究 1993 年 9 月号) への訂正と追加, 「物性研究」 65-5 (1996 年 2 月号), pp.678-685.

[伊庭 1996a]

伊庭幸人 (1996), 統合技術のための統計数理的方法 (2.5 版), The Institute of Statistical Mathematics, Research Memorandum No.595.

[伊庭 1996b]

伊庭幸人 (1996), 学習と階層, — ベイズ統計の立場から —, RWC 情報統合ワークショップ '95 予稿集, pp.189-198. (「物性研究」 65-5 (1996 年 2 月号), pp.657-677 にも掲載).

[伊庭 1996c]

伊庭幸人 (1996), 基礎的問題から見た情報統合, 人工知能学会誌, Vol.11, No.2, (1996 年 3 月号), pp.193-200.

[伊庭 1996d]
伊庭幸人 (1996), 統計学者・数理工学者のための統計物理入門 — 暫定版 (2.1 版), — 格子スピン模型とマルコフ連鎖モンテカルロ法を中心にして —, The Institute of Statistical Mathematics, Reseach Memorandum No.592, January, 1996. (http://www.ism.ac.jp/~iba/ で図を除いて入手可能.)

[伊庭 1996e]
伊庭幸人 (1996), 回帰分析の変数選択におけるベイズ的方法, – 平均場近似に基づくアプローチ –, The Institute of Statistical Mathematics, Reseach Memorandum No.624, October, 1996. SITA 96(Dec. 1996, 箱根) 予稿. (http://www.ism.ac.jp/~iba/ で入手可能.)

[伊庭 1997]
伊庭幸人 (1997), 統計科学 文献案内, ISM Reseach Memorandum No.630, January, 1997. (http://www.ism.ac.jp/~iba/ で入手可能.)

[金子・津田 1996]
金子邦彦, 津田一郎 (1996), 複雑系のカオス的シナリオ, 複雑系双書 1, 朝倉書店.

[川人 1996]
川人光男 (1996), 脳の計算理論, 産業図書.

[川人 ほか 1994]
川人光男, 行場次朗, 藤田一郎, 乾敏郎, 力丸裕 (1994), 視覚と聴覚, 岩波講座 認知科学 3, 岩波. (特に、 2 章 (大脳視覚野の生理学), 4 章 (音響・聴覚系の生理学)).

[Krogh et al. 1994]
Krogh,J.M., Brown,M., Mian,I.S., Sjölander,K. and Haussler,D. (1994), Hidden Markov models in computational biology — application to protein modeling, Journal of Molecular Biology, 235, pp.1501-1531.

[Kurosky and Deutsch 1995]
Kurosky,T. and Deutsch,J.M. (1995), Design of copolymetric materials, Journal of Physics A, Mathematical and General, 27, L387-393.

[Lawrence et al. 1993]
Lawrence,C.E., Boguski,S.F., Liu,M.S., Neuwald,A. and Wooton,J.(1993), Detecting subtle sequence signals: A Gibbs sampling strategy for multiple alignment, Science, 262, pp.208-214.

[Liu et al. 1995]
Liu,J.S., Neuwald,A.F. and Lawrence,C.E. (1995), Bayesian models for multiple local sequence alingment and Gibbs sampling strategies, Journal of the American Statistical Association, 90, pp.1156-1170.

[マー 1987]
Marr,D. (1987), ビジョン, — 視覚の計算理論と脳内表現 —, 乾 敏郎, 安藤 広志 訳, 産業図書 (原著 (1982)).

[Marroquin et al. 1987]
Marroquin,J., Mitter,S. and Poggio,T. (1987), Probabilistic solution of ill-posed problems in computational vision, Journal of the American Statistical Association, 82, pp.76-89.

[Peterson and Hartman 1989]

Peterson,C. and Hartman,E. (1989), Explorations of the mean field thoery learning algorithm, Neural Networks, 2, pp.475-494.

[ラッカー 1996]

Rucker, R. (1996), ハッカーと蟻, ハヤカワ文庫 SF1161, 早川書房 (原著 1994).

(筋はとくに目新しくないが、もっともらしい詳細が面白い.)

[坂元 1985]

坂元慶行 (1985), カテゴリカルデータのモデル分析, 共立出版.

(英語版) Sakamoto,Y.(1991), Categorical Data Analysis by AIC, Kluwer.

[櫻井 1996]

櫻井芳雄 (1996), ニューロン集団の相関活動をみる, 科学, Vol.66, pp.784-792, (1996 年 11 月号), 岩波.

[Sasai 1995]

Sasai,M. (1995), Conformation, energy, and folding ability of selected amino acid sequences, Proc. Natl. Acad. Sci. USA, 92, pp.8438-8442.

[Seno et al. 1996]

Seno,F., Vendruscolo,M., Maritan,A. and Banavar,J.R. (1996), Optimal protein design procedure, Physical Review Letters, 77, pp.1901-1904.

[Shakhnovich and Gutin 1993]

Shakhnovich, E.I. and Gutin, A.M. (1993), Engineering of stable and fast-folding sequences of model proteins, Proc. Natl. Acad. Sci. USA (1993),90,pp.7195-7199.

[Thorne et al. 1991]

Thorne,J.L., Kishino,H. and Felsenstein,J. (1991), An evolutionary model for maximum likelihood alignment of DNA sequences, Journal of Molecular Evolution, 33, pp.114-124.

[Thorne and Churchill 1995]

Thorne,J.L. and Churchill,G.A. (1995), Estimation and reliability of molecular sequence alignments, Biometrics, 51, pp.100-113.

[Trakulnaleamsai et al. 1995]

Trakulnaleamsai,S., Yomo,T., Yoshikawa,M., Aihara,S. and Urabe,I. (1995), Experimental sketch of landscapes in protein sequence space, Journal of Fermentation and Bioengineering, 79, pp.107-118.

[Wang et al. 1996]

Wang,G., Tanaka,K. and Tanifuji,M. (1996), Optical imaging of functional organization in the monkey inferotemporal cortex, Science, 272, pp.1665-1668.

[Young and Yamane 1992]

Young, M.P. and Yamane, S. (1992), Sparse population coding of faces in the inferotemporal cortex, Science, 256, pp.1327-1331.