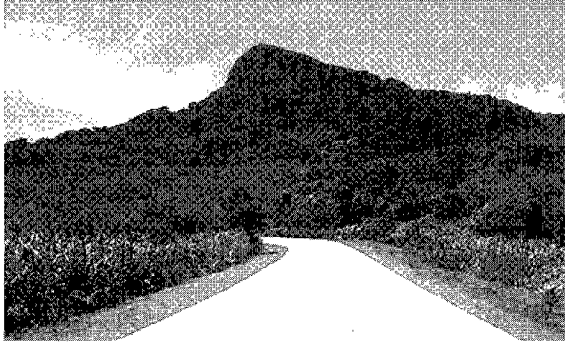


# 道しるべ： 統計的学習研究探訪



上田 修功

NTT コミュニケーション科学基礎研究所

学習とは過去の経験を活かして将来の予測を行うことといえる。近年、計算機に高度な学習機能を持たせたいという理由から、学習研究の発展に寄せる期待はますます高まりつつある。そこで、本稿では数値情報を取り扱い、広範囲な工学的応用を持つ統計的学習に焦点をあて、この分野のこれまでの研究成果、そして現状での中心研究テーマ、さらには今後の研究課題を筆者なりに概観し、初学者にも統計的学習研究の輪郭が理解できるよう導入的な説明を行う。

## 統計的学習研究の目指すもの

学習は、ほとんどの人に馴染みのあるキーワードであるが、統計的学習研究となると数理的色彩の濃い論文が多いせいか、何を目指し何を研究しているのか？という素朴な疑問を持つ読者も少なくないと思われる。そこで、まず統計的学習の研究内容の紹介に入る前に、学習研究分野において統計的学習研究はどの位置づけられるのかについて概説する。

学習とは、一言でいうと過去の経験を活かして将来の予測を行うことである。学習は統計的学習以外に、機械学習、認知科学、脳（神経）科学の分野でも研究されている。やや（かなり？）雑駁ではあるが、これらをあえて分類するならば、表-1のようになるであろう。

過去の経験を活かし将来の予測をする際、過去の経験を記号（言語情報）として蓄え、将来の予測も記号レベルで推論するような学習の形態と、数値（正確には信号レベル）を媒体とする学習の形態に大別される。人工知

能の研究分野で主に研究されている機械学習や認知科学は前者の形態に属し、脳科学、統計的学習は後者の形態に属す。脳科学では、脳磁界や脳波といった信号から人間の記憶や学習が脳のどの部位で、どう作用して実現されるかを探求している。

また、脳科学や認知科学では、人間の記憶や学習の仕組みの解明を目指すという立場で研究がなされているのに対し、機械学習や統計的学習は、真の仕組みの解明を探求するのではなく、むしろ、その仕組みの工学的実現

	数値	記号
解明	脳科学	認知科学
実現	統計的学習	機械学習

表-1 統計的学習研究の位置づけ

を目指している。極論すれば、人間がどうやっていようが、その機能を実現すればよいという立場である。

以上から明らかなように、統計的学習とは数値情報に基づいて学習の仕組みを工学的に実現する研究分野と位置づけられるが、より詳しくは、その実現法において、観測データの背後に未知の確率モデルを仮定し、観測データからそのモデルを推定するというアプローチをとる。それゆえ、数理統計学をベースとして理論展開される。統計的という修飾語はここに起因する。統計的学習は、主に非線形モデルを統一的に扱うニューロコンピューティングにおいて精力的な研究がなされている。

統計的学習研究は、画像、音声、言語といった特定のメディア情報を直接取り扱うわけではなく、より一般的な視点で学習の問題を考察する基礎研究である。それゆえ、成果が地味であるが、特筆すべきは、統計的学習の研究分野では、その時代に整合した“おはなし”だけが流行とともに現れては消え、消えては現れというのではなく、理論に支えられた技術がきちんと体系化されている点である。その意味で統計的学習研究は学問として健全な発展を遂げているといえる。

たとえば、多層ニューラルネットワークは約15年前、逆誤差伝搬法という学習アルゴリズムの誕生により大ブームを引き起こしたが、当時の技術は非常に粗削りで実用にはほど遠いものであった。それが統計的学習における長年の研究により、今日では多くの実領域で使用されるまでに至っている。

## どのような問題を取り扱うのか

統計的学習の具体的方法論に入る前に、統計的学習では主にどのような問題を取り扱っているのかについて説明しておく。

学習は、一般に、

- 教師無し学習 (Unsupervised Learning)
- 教師有り学習 (Supervised Learning)

に分類される。教師無し学習では、観測データは入力情報のみからなる。したがってそこでの学習のタスクは、観測データを何らかの形で表現 (representation) することである。表現法として、クラスタリング (clustering) のように、単に、データを複数のクラスタ (塊) として表現する方法や、さらには、データの背後にある確率分布として表現する方法がある。前者の代表例として、画像や音声のような歪みを許すデータ圧縮法であるベクトル量子化 (Vector Quantization) があり、後者の代表例として混合正

規分布推定 (観測データに  $m$  個の正規混合分布をあてはめる問題) がある。

一方、教師有り学習では、観測データは入力情報と出力情報からなる。つまり、入力に対しその理想出力値も教師信号として与え、ある定められた尺度で予測出力値をできるだけその理想出力値に近づけるように学習させる。ここでの代表的タスクとして

- 回帰問題 (Regression)
- 識別問題 (Classification)

がある。回帰問題とは、 $N$  組の入出力からなる観測データ  $D = \{(x_i, y_i), i = 1, \dots, N\}$  に対し、その入出力の写像関係をパラメータ  $\theta$  で特徴づけられた関数  $y = f(x; \theta)$  の形で学習させる。実験で得られたデータに最小自乗法で直線 (1次関数) をあてはめる問題はこの典型例である。

一方、識別問題とは入力をあらかじめ定めた  $C$  個のカテゴリに分類する問題である。したがって、教師信号  $y$  はカテゴリインデックス  $\{1, \dots, C\}$  のいずれか1つをとる。この代表例として、郵便番号認識 (0から9の10種類の数字の認識) などのパターン認識問題がある。

また、報酬と罰則という大ざっぱな教師信号を用いる強化学習 (Reinforcement Learning) と呼ばれる教師有り学習もある。

以上が、学習の大まかな分類であるが、以下では説明を容易にするため、工学的な応用が豊富な“教師有り学習”の枠組みで統計的学習の研究内容を紹介する。ただし、本稿はあくまで導入的なものなので、キーワードとその直観的な説明にとどめる。なお、表には記載しなかったが、数値情報を取り扱う学習研究として数学、情報理論から派生した計算論的学習や情報論的学習という研究分野もある。

## 一を聞いて十を知る

以下では読者に馴染みのある“試験勉強”とのアナロジーで統計的学習の研究内容を概観してみよう。

試験勉強 (学習) では、その分野の参考書もしくは問題集により知識を固め、来るべく試験に望むわけであるが、最も重要なことは試験でいかに合格点をとれるかということに尽きる。いくら問題集で知識を詰め込んだとしても、同じ問題が出ないと点がとれないようでは本来の学習とはいえない。一を聞いて十を知る学習がベストな学習といえよう。

統計的学習においても同様で、学習データに対して予測がうまくいったからといって、未知 (未学習) データ

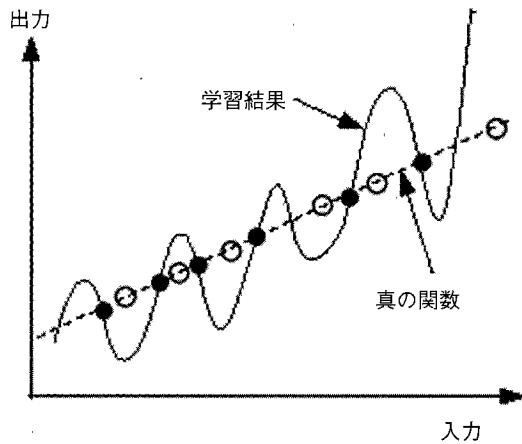


図-1 汎化能力の低い学習結果

に対して予測精度が悪ければ意味がない。統計的学習では、未知データに対する予測性能を汎化能力と呼び学習法の評価基準としている。

試験勉強では、やみくもに問題数をこなすのではなく、適切な教材（問題集）、（可能であれば？）教師を選び、重要事項を十分理解し、自身の弱点を繰り返し補強するという学習がより効率的であろう。もちろん2、3問で応用力がつくわけではないので、こなすべき最小限の問題数がある分野（試験勉強の科目）の難易度に応じて決まるはずである。また、“過去問”を利用して出題傾向を分析しておくことも重要であろう。

実は、興味深いことに、統計的学習における汎化能力を高めるための学習法はまさに上記と類似した方法論が展開されている。これについて以下順に説明する。

(1) データ数と汎化能力の関係：一般に、汎化能力は、学習データ数の増加とともに向上する。ただし、その向上率は学習すべき問題の複雑さに応じて異なる。この問題に対する基礎理論としてVC次元（Vapnik-Chervonenkis dimension）がある。

(2) 能動学習（Active Learning）：通常の学習では、個々の学習データを公平に取り扱うが、能動学習では、ある特定のデータをより積極的に繰り返し使用することにより汎化性能を向上させる。主要な成果として、回帰問題の場合、“予測出力値の分散の大きなデータをより積極的に使用する”，また識別問題の場合、“識別誤りが生じているデータを積極的に使用する”ことにより汎化性能が向上するという結果が示されている。分散が大きい、あるいは識別誤りを起こしているということは、直観的には信頼性が低いことと等価なので、信頼性が低い結果を生むデータは何度も学習させよというまさに試験勉強に

おける“弱点補強”に通じる。

(3) モデル構造選択：試験勉強の例で適切な教材を選ぶことが重要だと述べたが、統計的学習においても、データにあてはめるべきモデルの構造を適切に選ぶ必要がある。モデルの構造とは、より正確にはモデルの複雑さのことをいう。

たとえば、データに多項式曲線をあてはめる場合、多項式の次数がモデルの構造（複雑さ）に相当する。このモデル構造選択は、汎化性能の高い学習機械を構成するうえできわめて重要である。たとえば、本来、図-1の点線の直線上にあるデータに対し、黒丸で示した学習データから、自由度の高い（柔軟性の高い）高次の多項式をあてはめると、図-1に示すように得られた曲線はすべて学習データを通るが、図-1の白丸に示す未学習データに対してはうまく予測できていない。学習データに過度に学習するあまり、未知データでの性能劣化を引き起こす現象を過学習（over learning）と呼ぶ。

線形モデルに対しては、AIC（Akaike's Information Criterion）、MDL（Minimum Description Length）などの情報量基準を用いて最適モデル選択が実現できる。しかしながら、非線形モデルに対しては、理由は専門的になるので省略するが、一般に情報量基準はうまく作用しない。これについては、最近、ベイズ学習の枠組みでいくつかの解決方法が提案されている。

(4) ベイズ学習：試験勉強では、“過去問”から出題傾向をあらかじめ把握することにより効率の良い学習ができる。これは事前知識の利用である。統計的学習でも、ベイズ学習と呼ばれる事前分布（事前知識の確率分布）を利用する学習の枠組みがある<sup>1)</sup>。ベイズ学習は古くからあるが、困難な期待値計算（多次元積分計算）を必要とするため、実用面ではこれまで敬遠されていた。ところが、近年、マルコフ連鎖モンテカルロ法、Variational近似法などの各種近似法の提案によりベイズ法が見直されている。現在、ベイズ法は統計的学習研究におけるブームとなりつつある。

(5) 局所解の回避：非線形モデルの場合、学習是最急降下（もしくは上昇）法に代表される非線形最適化アルゴリズムが用いられる。この種のアルゴリズムでは初期解が真の解に十分近い場合、真の解に収束するが、そうでない場合は、初期解の近傍に収束するという局所最適性の問題に悩まされる。この問題も統計的学習研究の重要研究課題である。これは試験勉強の例では、問題を解く際、本質的でない個所にこだわり過ぎて正解にたどり着

けないという状況に相当するであろう(若干, アナロジーに無理があるが...).

## 学問に王道なし

以上, 統計的学習研究のこれまでの研究内容を概観した. 誌面の都合上, 表層的な説明に終始し, とても読者に内容を理解していただいたとは思っていないが, “試験勉強とのアナロジー” からおよその内容は想像できたのではないかと思っている. 以下では, 上記内容に興味を抱き, 統計的学習にこれから参入しようと考えている読者に有用な参考書等を紹介する.

“学問に王道なし” といわれるが, 特に統計的学習の場合, 確率統計の基礎を知らずしてこの分野の研究をしようと思うのはちょっと厚かましい. 確率統計に関して, 大学院レベルの知識は持ち合わせて欲しいものである<sup>☆1</sup>. 確率統計に馴染みのない読者に, 海外でも工学系の授業でポピュラーな Papoulis の本<sup>2)</sup> をお勧めする. 本書では最初に確率論の公理からきちんと説明されているが, 本質を理解するうえで必要でない厳密な議論は省略され, その分, 理解を促す計算例が豊富で記述も分かりやすい.

また, 前述した, VC次元, 能動学習, さらに, 最適化アルゴリズムも含め, 統計的学習法全般についてきちんと整理された入門書として, Bishop の本<sup>3)</sup> がある. この業界において最多引用数を誇る世界的な名著である. ただし, 9, 10章でベイズ学習が説明されているが, これはラプラス近似という古典的な手法であり, 最新のベイズ学習の成果は記載されていない(出版が1995年なので当然ではあるが). ベイズ学習の最新動向については, 手前みそであるが, 筆者の最近の研究会資料<sup>4), 5)</sup> を参考にしたい. さらに, 識別問題(パターン認識)に関しては, Bishopの本はやや手薄で, これも手前みそで恐縮だが, 文献6) が入門者からあるレベルの専門家までを想定してより詳しい内容を分かりやすく説明している.

統計的学習研究の最新動向を日本で手っ取り早く知る方法は, NIPS (Neural Information Processing Systems)<sup>7)</sup> の国際会議録を見ることであろう. NIPSはニューロコンピューティングの分野で最高峰の国際会議で, インパクト性の高い論文のみが採録される. 言い換えれば, 会議録は最新テーマの宝庫といっても過言ではない.

<sup>☆1</sup> “確率変数 (random variable) とは何か” を説明できなければ, 残念ながら確率統計に関して大学院レベルの知識がないと自覚すべきであろう.

## 展望

これまでの統計的学習研究内容は, まさに我々人間の学習を工学的に実現するための方法論の研究であったといえる. ということは統計的学習の今後の有望な研究テーマを発掘するには, 我々人間の学習に目を向けることが一番の近道かもしれない.

表-1に示したように, これまでの統計的学習研究は, 取り扱う情報の種類で, 機械学習と独立して研究がなされてきた. しかしながら, たとえば, 人間が数字を記憶する際, 単純に数値を記憶するのは困難でござる合わせのように言語(記号)情報を利用する方がはるかに記憶しやすい. ところが, 現在の統計的学習では, 多様な情報を直接取り扱うことができない. これに対する1つの解は, 統計的学習と機械学習とを融合した新たな学習の枠組みの構築である. 筆者自身挑戦したいテーマの1つと考えている.

### 参考文献

- 1) たとえば, Bernardo, J. M. and Smith, A. F. M.: Bayesian Theory, John Wiley & Sons (1994).
- 2) Papoulis, A.: Probability, Random Variables and Stochastic Processes, Third Edition, McGraw-Hill, Inc. (1991).
- 3) Bishop, C. M.: Neural Networks for Pattern Recognition, Oxford Press (1995).
- 4) 上田修功: 併合分割操作付き Variational Bayes 学習, 統計数理研究所共同研究会報告 (Sep. 1999).
- 5) 上田修功: 局所解の回避と最適モデル探索を同時実現する Variational Bayes 学習, 信学技報, NC99-132, pp.113-120 (2000).
- 6) 石井健一郎, 上田修功, 前田英作, 村瀬洋: わかりやすいパターン認識, オーム社(1998).
- 7) Neural Information Processing Systems: <http://www.cs.cmu.edu/Group/NIPS> (平成12年4月28日受付)

