

【特集】「AIとデータデータに基づく意思決定と社会イノベーション創出」

データ関連の数理技術の変遷： 深層学習は春秋戦国時代の秦となるのか？

Transition of Mathematical Technique in Data Analytics:
Does Deep Learning Become Qin in the Warring States Period in China?

樋口 知之
Tomoyuki Higuchi

情報・システム研究機構統計数理研究所
Research Organization of Information and Systems, The Institute of Statistical Mathematics.
higuchi@ism.ac.jp, <http://www.ism.ac.jp/~higuchi/>

Keywords: deep learning, sparse modeling, generative model, emulation, experimental design.

1. はじめに

1.1 秦の始皇帝による国家統一

著者は中国の歴史に関する解説書や小説を読むのが趣味の一つである。日本人の例に漏れず私も、三国時代(『三国志』で有名な、魏・呉・蜀の三国が並立した時代)に関する書物を数多く読んでいるが、壮大な中国の歴史の中で個人的に特に興味深いのは、三国時代の後に続く魏晋南北朝時代である。この時代は、周辺の異民族の流入による紛争の多発もあり、統一国家が長らく現れず(約300年間)多数の国家が勃興・没落した、中国の歴史の中でも珍しい期間である。そのはるか昔に、この広大な中国を初めて統一したのが秦の始皇帝である。

秦の始皇帝がなぜ統一国家を樹立できたのか、その理由にはさまざまな説があるが、著者が注目するのは、立国および統一までの政治的なゲーム論の側面ではなく、統一後の大胆な制度改革である。特に、度量衡・文字・貨幣の統一といった情報統制が、統一国家の維持に大きな効果をあげた点である*1。それまでの中国は、長さや重さを人間の手や足で測量していたため、その正確な数値を知ることも伝えることもできなかった。始皇帝は、長さを計る基準を「度」、体積は「量」、重さは「衡」と定め、数値を測るための単位、いわゆる「度量衡」を制定した。これ以降は、農作物の生産高や商品の流通量が正確に把握および予測できるようになり、国家の収入は安定した。また、地域ごとにバラバラであった文字や貨幣、さらには道路の幅の統一も実施し、当時の情報統制を初めて実現した。誰もが知る「焚書坑儒」も、この情

報統制を確実に成し遂げるための“リセット用劇薬”であったとまでいわれるほどである。

1.2 深層学習の勃興

ここまで書くと、勘所の良い読者の中には、著者が何をどう語りたのか、ほぼ察した方も多であろう。深層学習の強みは、ビッグデータの登場により、人工知能技術の「度量衡」的性質をもっているためにある、と著者は考える。深層学習は、画像・動画処理、音声処理(認識)、自然言語処理の3分野では、それまでの機械学習の諸手法を、圧倒的性能により凌駕し、その計算プラットフォームもさまざまな企業から無償で提供されている。有力な統計的機械学習技術が競合していた、それまでのある種“長閑な”時代が一変してしまった。まるで、秦の国家統一前の春秋戦国時代の終焉のようである。この勢いで、深層学習はすべての統計的機械学習の手法を無価値に追いやるのであろうか？

データを有効活用するための推論技術の背後にあるのは帰納法であり、帰納法の基盤となるのは統計学およびその周辺数理である。統計の研究者からすれば、深層学習は非線形関数の一つに過ぎず、内在するパラメータ推定(学習)にまつわる帰納法由来のさまざまな問題を避けて通ることは決してできない。ではなぜ、深層学習はここまでの成果をあげ、そして多くの若者を虜にしたのであろうか？ たった15年しか続かなかった秦の国家統一と同じように、深層学習はまたまた“第三次人工知能ブーム”として終わってしまうのであろうか？ 本稿では、比較的規模の大きいデータを学生時代から取り扱ってきた著者の長年の経験に基づき、データに関わる諸数理技術の変遷を現在から過去へ遡るとともに、今の俯瞰的描像を示してみたい。

*1 始皇帝が本当に行ったかについては諸説あるようであるが、その後の漢が400年もの長期王朝を築けた理由は、秦の国家統一による諸々の制度改革にあることでは一致している。

2. 深層学習隆盛の側面

2.1 統計的機械学習の急所

統計的機械学習の実体は、タスク、数理モデル、学習アルゴリズムの三つのトリプレットで定まる。タスクとは、データを用いて何を行いたいのかという課題設定のような広義のものから、予測や判別といった統計的情報処理の目的のような狭義のものを意味する。数理モデルはそのタスク実現に利用する数理的構造体であり、深層学習では、深層ニューラルネットワークがそれに対応する。学習アルゴリズムとは、数理モデルに含まれるパラメータをデータから定める手続き（方法）である。深層学習の場合、バックプロパゲーション（逆誤差伝搬法）がそれになる。ここでは、深層ネットワーク構造などのハイパーパラメータや、学習効率などの実際の計算上必要となるパラメータの決定法も、学習アルゴリズムに含めることとする。

統計的機械学習は、データセットがどんなにビッグであろうと、所詮、有限であるデータ集合からの推論であることには変わりないので、工夫なしにはその根源的弱点を回避できない。具体的には、1) データサンプル数がパラメータ数よりも少ないケースで必ず生ずる過学習、2) サンプリングバイアス（データ集合の偏りから来る推定誤差）、3) 内挿法の限界、これら三つが顕著な弱点である [樋口 16a]。深層学習も統計的機械学習の一手法であるため、程度の違いはあれ、それらの弱点は生来のものである。本誌の読者は、過学習およびサンプリングバイアスの問題についてよくご存じであろうから、ここでは内挿法の限界について少し説明を加える。

統計的機械学習が実現する諸操作は、カーネル法 [赤穂 08, 福水 10] のイメージでざっくり言えば、もともとのデータ集合は超高次元空間に写像されており、そこで形成されるデータ近傍に含まれる点集合を引数とする線形写像とみなせる。それゆえに、写像関数が極めて特異的性質を保持しない限り、諸操作の結果は、超高次元空間でのデータ近傍内の要素を用いた内挿となっている [樋口 16c]。したがって、極端現象のような、一度も発生したことの無い事象を手元にあるデータ点のみでもって表現すること、つまり外挿は、統計的機械学習の最も不得手なところである。そのようなデータを一度でもよいから計測・観測することの重要性は、この説明からも明らかであり、だからこそ、データ量が少なくなるときに強化学習やシミュレーションが必須となるのである。

2.2 人工知能技術の度量衡

深層学習に含まれる膨大な数のパラメータは、ビッグデータなしにはけっして定まらない。データが足りなくなるときに強化学習によって増やさなければならない。深層学習にとってビッグデータは必要不可欠の栄養であ

り、栄養なしにマシンは育たないのである。では、ビッグデータの恩恵は最大限活用するとして、なぜ深層学習が他の手法と比較して圧倒的存在感を示しつつあるのか？ その理由として、著者はパラメータ推定法、計算プラットフォーム（計算機環境を含む）、そして成果発表の場の三つが、人工知能技術の「度量衡」となりつつある点をあげたい。

§1 統一化されたパラメータ推定法

統計の研究者から見て深層学習の一番の強みは、なんといっても、パラメータ推定がバックプロパゲーションと確率的勾配降下法（Stochastic Gradient Decent：以後SGD）で統一されている点である。深層学習同様に膨大なパラメータ数を持ち、高い表現能力をもつ数理モデルとしてノンパラメトリックベイズモデル [持橋 10] があるが、そのパラメータ推定法は、問題ごとに解析的な導出計算が必要となり、利用者にとっては面倒である。また、その妥当性の理解に比較的高度な数理的素養を必要とする一方、バックプロパゲーションの理解は容易である。この恩恵は、深層ニューラルネットの層別構造により微分の連鎖率が適用できることと、活性化関数の引数が重みパラメータと前段のニューロンの出力値の線形結合であることの2点から来る。

SGD自体は特出したアイデアと思えないように、原始的な形はさまざまな分野で古くから提案されていた [甘利 17]。深層学習に実装されたその驚異的実効性能も、ビッグデータがもたらしたものである。表現能力の高い統計的機械学習では、ビッグデータをフル活用する際に、パラメータ学習計算がサンプル数のべき乗で増えてしまうという大きな障害があった。カーネル法がその代表例である [赤穂 08, 福水 10]。SGDの採用により、計算効率は飛躍的に向上するとともに、パラメータ学習法が統一された点は最大の収穫である。

§2 “真に汎用化”された計算プラットフォーム

パラメータ推定が、深層ニューラルネットの構造によらず、また、多くの場合タスクにもよらず統一化されたため、応用開発を行う現場にも大きな変化が生じた。深層学習の計算プラットフォームの登場である [大野 17]。基本ソースのオープン化に先鞭をつけたのは、2015年終わりにライブラリ TensorFlow（テンソルフロー）を出した Google である。本誌の読者はよくご存じであろうが、その他のライブラリとして、Caffe（C++をベース、UC Berkeleyが開発）、Keras（TheanoやTensorflowをベース）、Chainer（Preferred Networks：PFN）が有名である。また、General-purpose computing on graphics processing units：GPGPUの利用形態として、それまでの数値シミュレーションの高速化から深層学習の大規模化にいち早く商機を見いだしたNVIDIAの大躍進により、GPUは前述のライブラリ群とともに計算プラットフォームの両輪となっている。なお、本稿で数値シミュレーションとは、第一原理（支配方程式）を数値的に解く計算、

およびエージェントシミュレーションのような微視的数理モデルを大規模に計算するタイプの両方を指す。

§3 学術成果の瞬時共有の場 arXiv の集中的活用

arXiv (アーカイブ) はコーネル大学が運営するプレプリントサーバで、数学、物理学の、主に理論を中心とする論文が PDF の形式で保存および公開されている Web サイトである。実はこの arXiv が、今、世界で最も注目される学術サイトに成長している。2015 年時において、ダウンロード数は年間 1 億 4 000 万件程度 (毎日 40 万件近くに相当)、収録する総論文数は 110 万本に上る [Ginsparg 11]。

2005 年あたりからは、前述の分野に加えてコンピュータサイエンスや統計学の論文のシェアが急増し、人工知能の研究者にとって、もはや研究開発に欠かせない情報基盤となっている。現在は投稿論文の 20% 以上がコンピュータサイエンス分野であり、2015 年終わりに TensorFlow が公開されてからは、それまでは週当たり 500 程度の投稿数だったのが、その 1 年後には 2 000 近くにまで増加している。つまり、毎日、250 ~ 300 件のコンピュータサイエンス分野の論文が投稿されているのである。人工知能、特に機械学習の理論的研究分野では、投稿されたある論文が、翌日には、全く別の研究グループによる投稿論文で引用されることも珍しくない [樋口 17b, 北本 17]。また、機械学習の研究成果を公表する国際会議の論文採択率およびスコープによる階層化がこの 10 年の間に急速に進み、arXiv とともにハイレベルの国際会議が効率的な情報収集の場となっている。

3. 非線形および非ガウス型情報処理

3.1 深層学習と生成モデルの融合

前述した「度量衡」により、深層学習は、少なくとも画像・動画、音声、自然言語の三つの情報処理分野においては常套手段となり、適用にあたり深い数理的考察を不要とするまでのコモディティとなっている。さらには、近年、深層学習の利用形態が、識別モデルから生成モデルの構築へ移行しつつある [得居 16, Uber 17]。生成モデルを手にするメリットは、データの欠損・異常値処理、リスク解析、ベイズの定理を用いた逆解析など広範囲にわたる。よって今後、深層学習研究のターゲットは、生成モデルの自動構築に向かうものと著者は予想する。

深層ニューラルネットは、これまで識別モデルとして利用されることが多かったが、生成モデルに利用する研究もこの 2 ~ 3 年とみに盛んである [岡野原 17]。その理由は、結果の可読性とスモールデータへの対応といった深層学習における「弁慶の泣き所」の克服にある。結果の可読性とは、パラメータ推定やネットワークの構造学習 (モデルの選択) などのデータからの学習結果を、“人間” が理解しやすいかどうか、を意味する。具体的には、深層学習はブラックボックスであるため何がどう効いて

いるのかよくわからないが、線形回帰モデルのようなシンプルな統計モデルは変数間の関係が明示化されているため“からくり”が見える、と、よくいわれる話である。もう一つのスモールデータの取扱いに関する課題は、これもよく知られた、深層学習が膨大なデータを必要とする点である。その場合、データを模擬的に大量発生することが有効でシンプルな解決策である。本章では以下に複数の技術を取りあげながら、深層学習適用の今後の有望分野について私見を述べる。

§1 変分 Auto Encoder

本誌の読者には説明不要と思うが、Auto Encoder (以後 AE) とは、入力値を模擬する値を出力値とする深層ネットワークである。ただし、その構造的な特徴として、最中間層のノード数がデータの次元と比して著しく小さい点があげられる。統計学の主成分分析の非線形版とイメージいただければ良い。原始的 AE の出力は確定的、つまり入力値に対して出力値は一つに決まるが、入力値にわざわざホワイトノイズを入れる Denoising AE (DAE) が 2008 年に [Vincent 08]、さらには、ノイズを入れた効果をペナルティ項としてより一般的な形で定式化した Generalized DAE (GDAE) が 2013 年に提案され [Bengio 13]、AE を直接的に確率モデル化し深層ネットワークを生成モデルとして利用する契機となった。

確率モデル化はさらに巧妙化、高度化し、変分 AE (以後 VAE) に発展していく [Kingma 14]。DAE や GDAE で導入された擾乱項はデータ空間で直接的に付加されているが、一方、VAE では、後段の Decoder への入力となるベクトルを生成する際にホワイトノイズが加えられる。つまり VAE の場合は、中間層部分の生成モデル化、および擾乱の付与、といったベイズモデルでいう潜在変数ベクトルの確率的挙動をモデル化しているため、生成構造が VAE と GDAE は本質的に全く異なる。また、VAE は、潜在変数ベクトルの存在により、学習の結果の可読性が高い点も優れた点である。生成モデルの自動構築法としての汎用性から、この VAE と敵対的生成ネットワーク (Generative Adversarial Network : GAN) [Goodfellow 14, Radford 16] に今、大きな期待が寄せられている [樋口 17a, 岡野原 17]。

§2 統計的ダウンスケーリング

深層学習をベースとする生成モデル研究に関連して、数値シミュレーションとの融合も残された魅力的な分野である。数値シミュレーションは通常、時間・空間解像度をあげると、そのべき乗で計算時間がかかるため、計算リソースの限界から超高分解能の数値シミュレーションは簡単には実現できない。そこで、第一原理計算を経由せずに、粗い解像度の数値シミュレーションの結果から、細かい解像度の数値シミュレーションの計算を推定するのが、統計的ダウンスケーリングと呼ばれる計算である [樋口 17a, 稲津 10, 高藪 10]。すでに、深層学習を用いて統計的ダウンスケーリングを試みる研究も現れ、

流体や構造物のような連続体力学計算には著者は有望であるとみている。例えば、気象予報や構造体振動計算などに深層学習を適用して、超解像度の計算を高速で実現することも可能である。ただし、これらはあくまで近似計算であり、力学的保存則やさまざまな拘束条件を満たす保証がないことは肝に銘ずるべきである。

§3 エミュレーション

統計的ダウンスケーリングは空間的内挿法であるが、ほかにも、深層学習による数値シミュレーションの代替は考えられる。特に、エミュレーションと呼ばれる、数値シミュレーションの結果をビッグデータとして取り扱い、統計的機械学習により、ある興味ある物理量の予測や判別のための統計モデルを得る手法の高度化には大いに役立つであろう [樋口 17a, 中野 14]。すでにマテリアルズインフォマティクスの分野では、密度汎関数理論 (density functional theory : DFT) に基づいて、さまざまな物性値を数値シミュレーションの結果からエミュレーションで予測し、新物質の探索をスマートに実現する試みがなされている [McMahon 16, 瀧川 15]。また、気象海洋分野において、大規模な数値シミュレーションの結果から異常気象 (現象) につながる因子を発見し、異常気象を予知する試みも始まっている [松岡 17, Prabhat 17a, Prabhat 17b]。このような中、米国の IT メガ企業も深層学習と数値シミュレーションの融合に最近、とみに力を入れている [Stoecklein 17, Tompson 17]。

3.2 スパース性と非ガウス性

深層学習を用いた生成モデルの研究が進めば、変数 (要素) 間の関係が非線形で、パラメータ推定に用いる最適化関数が L_2 (尤度関数がガウス分布) である非線形・ガウス型の統計的モデルは、深層学習によって自動的に獲得可能となろう。一方、最適化関数に L_1 (パラメータの事前分布が非ガウス分布) を用いた線形・非ガウス型の統計的モデリングは、スパースモデリング [Sparse 13] によりほぼ自動化された。これは、統計的モデリングのこの 20 年間における最大成果である。なお、独立成分分析 (Independent Component Analysis : ICA) [村田 04]、圧縮センシング (Compressed Sensing : CS) [田中 10]、非負値行列分解 (Non-negative Matrix Factorization : NMF) [亀岡 14] などの行列分解、さらにはさまざまなテンソル分解 [鹿島 15] などの技術も、線形表現、次元圧縮、スパース性の仮定に基づく信号分解なので、スパースモデリングの一種である。

§1 特徴ベクトルの自動選択とデータ解析法の“野蛮化”

スパースモデリングの典型的利用法は、線形回帰モデルの説明変数の自動選択である。説明変数を記述子、また説明変数のセットを特徴ベクトルと呼ぶことも多い。これまで特徴ベクトルの構成法は、ほぼ人間の知的作業

に委ねられており、実は、この構成法が、予測・判別性能といったパフォーマンスをほとんど決めていたといっても過言ではなかった。つまり、特徴ベクトル構成法は、機械学習の『匠の技』といえ、機械学習にもかかわらず人間の判断が最も性能向上に大切という、羊頭狗肉の面もあった。深層学習では、“生” データを直接入力することで、この特徴ベクトルの構成作業を省略しながらも大きな成果を上げている。深層学習を使えば、ユーザは特徴ベクトルの選択・決定に悩まずに最高の予測・判別性能を得ることができる。まさに、「深層学習は『匠の技』の習得からユーザを解放した」と喧伝される所以である。

この方針をさらに進め、前処理などをすべて省いた生データから、最終的な目的まで直接、深層ネットワークでつないだ End-to-End という研究も加速している [篠田 17]。さらには、この力を逆手に取り、生データを深考なしにとりあえず可視化および画像化し、合わせて、データ処理に関わる諸作業の様子を撮像するなどにより、大量の画像や動画を作成すれば、特徴ベクトルの選択問題を回避しつつ目的を達成することも原理的には可能である。これまでのデータ解析においては、生データのもつ特性と解析目的に沿った解析法の慎重かつ適切な選択が成功の肝であった。しかしながらこれからは、上述したようなすべて画像化・音声化するといった、データ解析の目的からすると遠回りに見える、“野蛮” なデータ解析が普通感覚になっていくかもしれない。

§2 可読性と機序の解読

特徴ベクトルの選択という、最も応用ドメインや計測現場の知識・経験を必要とするプロセスが自動化されるメリットは大きい。前述のとおり、深層学習には、結果の可読性に関して決定的な弱点がある。一方、スパースモデリングでは、変数選択の操作が明示的であるため、可読性は高い。スパースモデリングは、予測や判別性能の向上といった面よりも、どの変数がどのように結果につながるのかといった機序 (メカニズム) や因果関係の同定に主たる興味がある場合、つまり、現象の解明に対して有効である。事実、地球科学や環境科学で頻出する逆問題の解法において、パラメータや潜在変数に対して L_1 正則化を加えることで、空間や時間上の不連続点の発見など、非ガウス型の情報処理の一般化に大きく寄与した [Nakata 17]。

線形モデルに限定すれば、計算機性能の向上を直接的に生かして、組合せ爆発するモデル群の中からも、赤池情報量規準 (Akaike Information Criterion : AIC) などの情報量規準を参照して、厳密に最適モデルを全探索する道もある [五十嵐 17]。ここまで見てくると、ビッグデータ時代においては、情報量規準によるモデル選択作業を経由した変数選択は非現実的で、データの形式やデータ解析の目的によって、深層学習あるいはスパースモデリングのどちらかを利用することになろう。

4. スモールデータと現場主義

4.1 ビッグデータ時代の AIC

スパースモデリングの登場前は、情報量規準によるモデル比較を通じて、説明変数の選択がなされていた。統計数理研究所第8代所長の故赤池弘次先生は、統計モデルを比較するための情報量規準 AIC を提案され、統計学そのものに大きなインパクトを与えると同時に、モデルの永続的な改良を通じて知識獲得を目指す研究スタイルを広範な研究分野において確立することに成功した。IEEE-AC に発表された 1974 年の論文 [Akaike 74] は、AIC の思想とその有効性が明確に示されたものとして人口に膾炙しており、その年ごとの引用数は発表から 40 年以上にわたって増え続けている。Web of Science (WoS) および Scopus の 2 大論文情報データベースの情報をもとに、その様子を図 1 に示した。注視されたいのは、縦軸は累積数でなく、毎年の引用数である。図からも見て取れるように、統計数理の研究成果が評価されるには 20 ~ 30 年ほどの長い時間が必要であること、またその影響が長期間にわたって継続することがわかる。

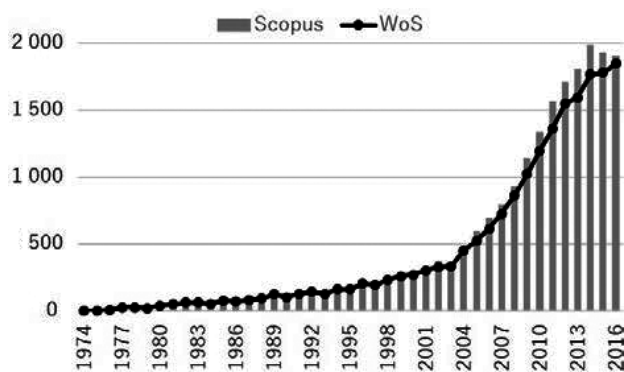


図 1 AIC 論文引用数の時系列。棒グラフが Scopus, 折れ線グラフが Web of Science (WoS) での引用数

実はこの引用数が 2015 年以降、ほぼ高止まり状態にある。これは AIC の典型的な利用法の一つとして(説明)変数選択があるが、前述したように、それが深層学習とスパースモデリングにより代替されはじめたため、引用数増加が見られなくなってきたのであろうか? AIC の誕生からほぼ半世紀、統計数理にも新しい時代が到来した予感がある。

今後 AIC の存在価値は、深層学習やスパースモデリングの普及により、緩やかに減少していくのであろうか? それを明らかに否定する事実がある。1 日限定ではあるが、2017 年 11 月 5 日の Google の検索サイトトップのロゴとして世界 15 か国で赤池先生の似顔絵が掲示された。その日は赤池先生の生誕 90 周年の記念日であった。掲示された似顔絵および赤池先生の貢献に関する文章が Google Doodle として今も閲覧可能である [Google

17]。深層学習を含めて先進的な機械学習を総動員し研究開発を進めている人工知能業界のトップ企業が、ビッグデータ時代のこの時期に赤池先生の業績に深い敬意を表した点に、Google のすごさを改めて実感した。

4.2 クロスバリデーションと汎化能力

人間は日常生活において、常に先を予測し、意思決定し、行動している。そのため頭の中には、過去の経験に基づいて構成されたモデルが存在する。研究開発上は、そのモデルを未来永劫改善し続けることが大切であり、AIC は“真”のモデルを完全に同定するために考えられたわけではない [Akaike 95]。繰返しになるが、AIC は単に、変数選択や、少数個のモデルの中からの“真”のモデルの決定のためだけのものではない。モデル群の海原を航海するための羅針盤が AIC なのである [Akaike 07]。また、原理としても汎用的であったからこそ、その variant がこれまで多数提案され、現代においても情報量規準に関する研究が綿々と続いている。その原理とは、一言で言えば、汎化能力の評価である。

今のビッグデータの時代、データサンプル数に比してパラメータ数が少ないモデルを利用する状況では、モデルの比較に何の疑いもなくクロスバリデーションを採用する。一方、赤池先生の時代はスモールデータしかなかったため、モデル構築(パラメータ推定)にはすべてのデータを用いるべき、というのが赤池先生の持論であった。とはいえ、表現能力の高い複雑なモデルを採用すると過学習の問題は必ず生ずる。このように、スモールデータの時代には汎化能力の評価は難しかったのである。赤池先生は分布の視点を導入し、スモールデータの時代にかかわらず、汎化能力の具体的な評価導出法の確立に成功した。今はクロスバリデーションが汎化能力を見積もる常套手段となっており、細かい技術的工夫に類する研究もなされてはいるが、スモールデータ時代の先人の賢察を学ぶと、もっと面白い発想が生まれるのではないかと思う。

4.3 ネオ実験計画と強化学習

1980 年代にスーパーコンピュータの利用が広く普及すると、第三の科学ともいわれる計算機科学が確立し、特にこの近年、流体解析や構造解析に利用されるソルバーの汎用化とともに、設計段階における計算機シミュレーションの相対的地位はかなり高まっている。ただし、前述したように、通常、計算機シミュレーションには膨大な時間がかかるため、エミュレータなどの統計的機械学習を用いて計算負荷を大幅に軽減する研究が盛んである。また近年は、試作品製作をごくわずかにして、大規模なシミュレーション計算でもって製品開発を行うことも普通になってきているが、設計パラメータ空間の次元が増えると、組合せ爆発する計算時間の増大が課題である。そのため、ガウス過程回帰 [Rasmussen 06] や、その古典版ともいえるクリギング [村上 11]、また次元削

減を目的としたスパース回帰などを総合的に用いて、設計パラメータ空間内の探索をスマート化する研究も盛んである [Seko 15]. つまり、従来の実験計画法から、コモディティ化したシミュレーション実験や高精度センサ情報を高度利用する実験計画法のリノベーションが、焦眉の急となっている。著者はそれを「ネオ実験計画法」と呼んでいる [樋口 17a, 中野 14, Risk 15].

シミュレーションに高性能センサ情報を取り込む「新製品エミュレータ」があれば、製品・サービス設計と検証（テスト）作業を一体化でき、結果としてプロセスの超高速化と開発予算の大幅な削減が実現できる。また、数値シミュレーション結果の信頼性を具体的に確立するための方法論の研究が欧米において急速に熱を帯びてきており、Uncertainty Quantification（以後、UQ）と呼ばれる学問分野にまで成長している。さらには NASA は、NASA UQ challenge というデータコンペを 2014 年から開始し [NASA 14], 米国統計学会は Society for Industrial and Applied Mathematics : SIAM と共同で Journal on UQ を同年に刊行しだした [SIAM 14]. その雑誌の取り扱う主たる分野として sensitivity analysis, model validation, model calibration, data assimilation の四つがあげられている。

エミュレーション、ネオ実験計画、および UQ などの発想は、ロボットや自動運転など、リアル空間への人工知能技術の埋込みににおいて参考になる点も相当多いと感じている [樋口 16c]. 現在の人工知能では、Google DeepMind の AlphaGo [Alpha 15] および、ごく最近では AlphaGo Zero [Alpha 17] のボードゲームでの圧倒的性能により、強化学習の旋風が吹いている。これからは、強化学習策の改善に関する理論研究に加えて、リアル空間との橋渡しに必須の UQ、学習をより効果的に行うネオ実験計画、さまざまなタイプのエミュレータ構築法の研究が大切になっていくであろう。言い換えれば、現場（リアル空間）に学ぶ重要性が顕在化し、現場を知る人間をマシンが代替していく時代になっていく。今から 50 年前、赤池先生は現場に通い、自ら実験を工夫することで、実験計画を体現していた。前述した Google Doodle は、“For Dr. Akaike, experience was core to creativity.” と述べている [Google 17]. シンプルではあるが、非常に力強く素朴で、心に響くメッセージである。ぜひ読者とともに心に刻みたい。

5. おわりに

本稿では、データ解析法の変遷を、あえて現在から昔へ遡る形で解説した。著者にとっても初めての試みである。機械学習の教科書的存在である書物「パターン認識と機械学習」を人工知能学会の若手・中堅研究者ら（今

や、日本の AI コミュニティを支える実力者ばかり）と一緒に翻訳して [Bishop 07], 2017 年末でちょうど 10 年になった。本来ならその書籍のスコープに則り、著者の専門であるベイジアンモデリングと周辺分野を解説するのが適切であっただろう。その線上で、ノンパラメトリックベイズ [持橋 10, 佐藤 16] やグラフィカルモデル [鈴木 16, 渡辺 16] などの確率モデル、また、さまざまな計算技法（マルコフ連鎖モンテカルロ法 [伊庭 05, 久保 12] や粒子フィルタ [樋口 11] など）について触れるべきであったかもしれない。さらには、機械学習の基盤数理として重要な最適化 [金森 16] の変遷について触れるべきであった。しかしながら、紙面の制約もあったが、今は、深層学習を中心に置かずデータ解析技術の句を語ることができないため、このような構成とした。よって、読みづらい点はすべて著者の責に帰するものである。データ解析に関わる数理技術の俯瞰的解説については、著者の過去の解説文を参考にされたい [樋口 16b].

本稿では、深層学習が人工知能の度量衡になってきている点を強調した。度量衡や情報統制により、秦は統一国家を中国国土に初めて打ち立てることができたが、その滅亡は非常にあっけなかった。15 年で滅びた理由として、万里の長城や阿房宮（皇帝の宮殿）の建設などによる国費の浪費と人民の酷使、また、不老不死を求めた珍妙な行動による人心の離反が有名である。これら以外にも諸説あるが*2、過度な中央統制が項羽と劉邦に代表される偉人の反感を招いたともいわれている。では、本稿の主演を務めた深層学習は、今後も人工知能技術においてずっと「度量衡」となるのであろうか？ 人はワンパターンに飽きる動物であり、他人とは同じ方向にはあえて進まない変わり者は必ず出てくる。深層学習を超える発想をもった、現代の項羽と劉邦は現れるのであろうか？ 目が離せない激動の時代はまだまだ続く。

著者の頭の回転はかなり悪くなってきたが、専門とする研究領域のめざましい進展と劇的な変化を同時体験できることの楽しさ、素晴らしさ、そして有り難さをかみしめつつ、筆をおろすことにする。

◇ 参考文献 ◇

- [赤穂 08] 赤穂昭太郎：カーネル多変量解析—非線形データ解析の新しい展開（シリーズ確率と情報の科学），岩波書店（2008）
- [Akaike 74] Akaike, H.: A new look at the statistical model identification, *IEEE Trans. on Automatic Control*, Vol.19, No.6, pp. 716-723 (1974)
- [Akaike 95] 赤池弘次：時系列解析の心構え、時系列解析の実際（2），朝倉書店（1995）
- [Akaike 07] 赤池弘次ほか：特集 赤池統計学の世界，総研大ジャーナル，Vol. 12, <https://www.soken.ac.jp/disclosure/pr/publicity/journal/no12/> (2007)
- [Alpha 15] <https://deepmind.com/research/alphago/> (2017)
- [Alpha 17] <https://deepmind.com/blog/alphago-zero-learning-scratch/> (2017)
- [甘利 17] 甘利俊一：もうちょっとだよなー，ディープリニング，人工知能，Vol. 32, No. 6, pp. 827-835 (2017)

*2 急激な制度改革および法家思想による統治。

- [Bengio 13] Bengio, Y., Yao, L., Alain, G. and Vincent, P.: Generalized denoising auto-encoders as generative models, *Advances in Neural Information Processing Systems* (2013)
- [Bishop 07] Bishop, C. M. 著, 元田 浩, 栗田多喜夫, 樋口知之, 松本裕治, 村田 昇 訳: パターン認識と機械学習, シュプリンガー・ジャパン (現在丸善) (2007)
- [福水 10] 福水健次: カーネル法入門: 正定値カーネルによるデータ解析, 朝倉書店 (2010)
- [Ginsparg 11] Ginsparg, P.: ArXiv 創設 20 年, 情報管理, Vol. 54, No. 7, pp. 415-420 (2011)
- [Goodfellow 14] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozai, S., Courville, A. and Bengio, Y. G.: Generative adversarial nets, *Advances in Neural Information Processing Systems*, Vol. 27, pp. 2672-2680 (2014)
- [Google 17] <https://www.google.com/doodles/hirotugu-akaikes-90th-birthday> (2017)
- [樋口 11] 樋口知之: 予測にいかす統計モデリングの基本—ベイズ統計入門から応用まで, 講談社 (2011)
- [樋口 16a] 樋口知之: スモールデータ, ビッグデータ, そしてスマートデータ—人工知能ブームの中での統計学—, 統計, Vol. 67, No. 1, pp. 9-14 (2016)
- [樋口 16b] 樋口知之: ビッグデータが変える日常と非日常の境目, 信学誌, Vol. 99, No. 1, pp. 30-35 (2016)
- [樋口 16c] 樋口知之: 人工知能はみよみまねマシンの究極形, 情報管理, Vol. 59, No. 5, pp. 331-335 (2016)
- [樋口 17a] 樋口知之: シミュレーション, データ同化, そしてエミュレーション, 岩波データサイエンス, Vol. 6, pp. 98-105 (2017)
- [樋口 17b] 樋口知之: 爆速化する研究スピード, 数学セミナー, No. 12, p. 1 (2017)
- [伊庭 05] 伊庭幸人, 種村正美: 計算統計 2 マルコフ連鎖モンテカルロ法とその周辺 (統計科学のフロンティア 12), 岩波書店 (2005)
- [五十嵐 17] 五十嵐康彦, 竹中 光, 中西 (大野) 義典, 植村 誠, 池田 思朗, 岡田真人: 全状態探索による線形回帰のスパース変数選択, 第 31 回人工知能学会全国大会予稿集 (2017)
- [稲津 10] 稲津 将, 佐藤友徳: 大は小を兼ねるのか: ダウンスケーリング, 天気, Vol. 57, No. 4, pp. 195-199 (2010)
- [亀岡 14] 亀岡弘和: 非負値行列因子分解とその音響信号処理への応用, 日本統計学会誌, Vol. 44, No. 2, pp. 383-407 (2014)
- [金森 01] 金森敬文, 鈴木大慈, 竹内一郎, 佐藤一誠: 機械学習のための連続最適化, 講談社サイエンティフィク (2016)
- [鹿島 15] 鹿島久嗣: 関係データの機械学習—行列・テンソル分解によるアプローチ—, 日本ロボット学会誌, Vol. 33, No. 2, pp. 72-76 (2015)
- [Kingma 14] Kingma, D. P. and Welling, M.: Auto-encoding variational Bayes, *Int. Conf. on Machine Learning* (2014)
- [北本 17] 北本朝展: ディープラーニングとオープンサイエンス: 研究の爆速化が引き起こす摩擦なき情報流通へのシフト, https://www.nii.ac.jp/sparc/event/2016/pdf/20170214_3.pdf (2017)
- [久保 12] 久保拓也: データ解析のための統計モデリング入門—一般化線形モデル・階層ベイズモデル・MCMC (確率と情報の科学), 岩波書店 (2012)
- [松岡 17] 松岡大祐: 気象ビッグデータからの極端現象発生予測—台風タマゴ発見から豪雨予測まで—, https://www.jst.go.jp/kisoken/presto/project/1112077/1112077_17.html (2017)
- [McMahon 16] McMahon, J. M.: Density functionals from deep learning, <https://arxiv.org/abs/1608.00316> (2016)
- [持橋 10] 持橋大地: 最近のベイズ理論の進展と応用 (III)—ノンパラメトリックベイズ—, 信学誌, Vol. 93, No. 1, pp. 73-79 (2010)
- [村上 11] 村上大輔, 堤 盛人: Kriging を用いた実用的な面補間法, GIS—理論と応用—, Vol. 19, No. 2, pp. 59-69 (2011)
- [村田 04] 村田 昇: 入門 独立成分分析, 東京電機大学出版局 (2004)
- [中野 14] 中野慎也, 樋口知之: 地球科学におけるシミュレーションとビッグデータ—データ同化とエミュレーション—, 信学誌, Vol. 97, No. 10, pp. 869-875 (2014)
- [Nakata 17] Nakata, R., Hino, H., Kuwatani, T., Yoshioka, S., Okada, M. and Hori, T.: Discontinuous boundaries of slow slip events beneath the Bungo Channel, southwest Japan, *Scientific Reports*, doi:10.1038/s41598-017-06185-0 (2017)
- [NASA 14] NASA LaRC UQ Challenge 2014, <https://uqtools.larc.nasa.gov/nda-uq-challenge-problem-2014/> (2014)
- [岡野原 17] 岡野原大輔: ニューラルネットの逆襲から5年後, <https://research.preferred.jp/2017/11/deeplearning-5years-later/> (2017)
- [大野 17] 大野健太: 深層学習フレームワークの設計と実装, 情処連続セミナー 第4回: ディープラーニングの活用と基盤, <https://www.ipsj.or.jp/event/seminar/2017/program04.html> (2017)
- [Prabhat 17a] Prabhat: A look at deep learning for science, <https://www.oreilly.com/ideas/a-look-at-deeplearning-for-science> (2017)
- [Prabhat 17b] Prabhat: Semi-supervised deep learning for climate @ Scale, https://qconnewyork.com/system/files/presentation-slides/prabhat_-_deep_learning_for_climate_v4.2.pdf (2017)
- [Radford 16] Radford, A., Metz, L. and Chintala, S.: Unsupervised representation learning with deep convolutional generative adversarial networks, <https://ishmaelbelghazi.github.io/ALI/> (2016)
- [Rasmussen 06] Rasmussen, C. E. and Williams, C. K. I.: *Gaussian Processes for Machine Learning*, MIT Press (2006)
- [Risk 15] 気候変動予測に確率のプロとして貢献, http://www.ism.ac.jp/ism_info_j/labo/project/130.html (2015)
- [佐藤 16] 佐藤一誠: ノンパラメトリックベイズ 点過程と統計的機械学習の数理 (機械学習プロフェッショナルシリーズ), 講談社 (2016)
- [Seko 15] Seko, A., Togo, A., Hayashi, H., Tsuda, K., Cha-put, L. and Tanaka, I.: Prediction of low-thermal-conductivity compounds with first-principles anharmonic lattice-dynamics calculations and Bayesian optimization, *Phys. Rev. Lett.*, Vol. 115, No. 20, pp. 1-5 (2015)
- [篠田 17] 篠田浩一: 音声言語処理における深層学習: 総説, 日本音響学会誌, Vol. 73, No. 1, pp. 25-30 (2017)
- [SIAM 14] *SIAM/ASA Journal on Uncertainty Quantification (JUQ)*, <https://www.siam.org/journals/juq.php> (2014)
- [Sparse 13] Sparse modeling, <http://sparse-modeling.jp/> (2013)
- [Stoecklein 17] Stoecklein, D., Lore, K.G., Davies, M., Sarkar, S. and Ganapathysubramanian, B.: Deep learning for flow sculpting: Insights into efficient learning using scientific simulation data, *Scientific Reports*, Vol. 7, <https://www.nature.com/articles/srep46368> (2017)
- [鈴木 16] 鈴木 讓, 植野真臣 著 編集, ほか: 確率的グラフィカルモデル, 共立出版 (2016)
- [高藪 10] 高藪 出, 金光正郎: 力学的ダウンスケーリングのレビュー, 天気, Vol. 57, No. 7, pp. 435-447 (2010)
- [瀧川 15] 瀧川一学: 大規模データに基づく電子物性予測のための深層学習技術の創出, <https://www.jst.go.jp/kisoken/presto/project/1112073/15665449.html> (2015)
- [田中 10] 田中利幸: 圧縮センシングの数理, 信学会 基礎・境界サイエティ, Vol. 4, No. 1, pp. 39-47 (2010)
- [得居 16] 得居誠也: 生成モデルの Deep learning, <https://www.slideshare.net/beam2d/learning-generator> (2016)
- [Tompson 17] Tompson, J., Schlachter, K., Sprechmann, P. and Perlin, K.: Accelerating Eulerian fluid simulation with convolutional networks, <https://arxiv.org/pdf/1607.03597.pdf> (2017)
- [Uber 17] Pyro Development Team: Uber AI Labs Open Sources Pyro, A deep probabilistic programming language, <https://eng.uber.com/pyro/> (2017)
- [Vincent 08] Vincent, P., Larochelle, H., Bengio, Y. and Manzagol, P.-A.: Extracting and composing robust features with denoising autoencoders, *Proc. 25th Int. Conf. on Machine Learning* (2008)
- [渡辺 16] 渡辺有祐: グラフィカルモデル (機械学習プロフェッショナルシリーズ), 講談社 (2016)

—— 著 者 紹 介 ——



樋口 知之 (正会員)

1989年東京大学大学院理学系研究科地球物理学専攻博士課程修了。理学博士。同年に文部省統計数理研究所に入所。2011年より、情報・システム研究機構理事および統計数理研究所長。専門はベイジアンモデリング。最近は、データ同化およびエミュレーション（シミュレーションの機械学習による簡便代替法）の研究に注力している。日本統計学会、応用統計学会、電子情報通信学会、日本応用数理学会、日本バイオインフォマティクス学会、日本マーケティング・サイエンス学会、International Statistical Institute, American Geophysical Union などの各学会の会員。一般社団法人データサイエンティスト協会顧問。また、日本学術会議の数値科学および情報学分野の連携会員でもある。