

モデリング変革の四半世紀を振り返って

25 years of a paradigm shift in a statistical modeling technology

前所長 樋口 知之 (Tomoyuki Higuchi)

キーワード：ベイズモデリング、粒子フィルタ、カーネル法、深層学習

私が研究所に入ったのは博士号取得の直後である平成元年4月、また所長を退任すると同時に研究所を退職するのが平成が終わる直前の平成31年3月であるため、平成の30年間は私にとって研究者人生そのものである。私が入所した当時は、その10年ほど前から研究が活発化していたベイズモデリングの成熟期にあった。事前分布および尤度関数が線形・ガウス型の理論は確立し、応用の観点からも、いくつかの課題では本研究所の研究結果が標準的方法として認知されるなど、研究進展も著しかった。現象や既存知識のモデリングにおいては、計算上からくる線形・ガウス性の制約は、研究者の自然な思考を妨げる。そのため、その制約を“人手”でもって解析的に緩める論文が数多く出版されていた。その流れと平行して、計算機集約的な方法の開発も活発化していた。高い次元の潜在変数がつくる超高次元の分布から、マルコフ連鎖でもって直接的に実現値を得るマルコフ連鎖モンテカルロ法や、その分布を低次元の分布に分解し（いわゆる逐次ベイズフィルタ）、低次元の分布の逐次更新を数値的に得る「非線形フィルタ」などである。私は学位を宇宙プラズマ物理の領域で得たため、統計力学の自由エネルギーや転送積分とほぼ同一の発想にもとづく両手法を自然に理解でき、大いに魅惑された。

平成にはいつしか北川元所長は、「非線形フィルタ」の適用範囲を大幅に拡張するモンテカルロフィルタを提案された（1993,1996年）。この手法は今日において、粒子フィルタの手法群の中で、原始的ではあるがレガシーとして高く評価されている。北川先生のすぐ近くで研究していた自分は、そのアルゴリズムの大規模並列コンピュータへの実装容易性に衝撃を受け、さまざまな非線形・非ガウス型の時系列モデルへの適用を試みた。また、世界同時的に多くの研究者が、アルゴリズムの高度化と現実問題への応用をすすめ、21世紀早々（2001年）にシュプリンガーから書籍を出版できた。私も著者の一人として参加するチャンスを得、当時新進気鋭の海外の若い研究者と交流できたことは、その後の自分の研究者人生にとって大きな糧となった。この友人の多くは現在、後述する深層学習に研究の軸足を移しているが、そのことは偶然でなく、むしろモデリング技術の発展を考えれば極めて自然なことと私は考える。粒子フィルタは、時系列モデルが所与であれば、パラメータや状態変数の推定に関して原理的には万能である。もちろん、モンテカルロ誤差（有限サンプルによる表現限界）からくる、分布の表現能力の喪失（いわゆる退化問題）や、尤度値の不安定性など、数値的問題は避けられないが、問題や経験に即して“人”が対象を自由にモデリングできるようになった点は、モデリング技術の発展において大きな革新であった。

2000年代にはいつしか、非線形モデリングの観点からは、別の形で大きな進歩があった。それはカーネル法の登場である。1998年にGoogleが、また2004年にはFacebookが創業されるなど、ビッグデータを新しい情報サービスという価値に転換する企業が当時、続々と誕生していた。ビッグデータを用いた容易なタスクは、識別関数の構築を通じた判別や分別である。カーネル法は、データ空間で複雑な識別境界面を同定するのではなく、あえて超高次元の特徴ベクトル空間を構成し、そこで豊富に蓄積された線形のモデリング技術を適用するものであ

る。もちろん、データ間の類似度を規定するカーネル関数は“人”が設定しなければならないが、データ空間での非線形モデリングに求められる巧みの技の多くを、カーネルトリックに押しつけられた利点は大きい。また、パターン認識手法の多くはデータどうしの内積計算を含むため、カーネルトリックによって既存の線形モデリングの諸手法が非線形版に自然に拡張できた。2000年代はビッグデータの登場により、統計的モデリングの主たる興味である生成モデルの構築から、カーネル法と最適化の活用による、複雑な識別関数（識別モデル）の自動構築に、研究のトレンドがシフトした。それと同時に、機械学習の言葉で代表される研究者コミュニティが育っていったのもこの時代の特徴である。

2000年代半ばから2010年代にかけて、非ガウス型モデリングにおいても大きな発展がスパースモデリングの普及によってもたらされた。応用現場における統計的モデリングの成否は、膨大な説明変数群の中からタスクの解決に有効な特徴ベクトルの構築にあると断言できる。なお、画像、音声、自然言語の処理においては、現在、その課題は深層学習によって大幅に自動化されている。カーネル法は、データ空間から（暗に）高次元の特徴ベクトルを構成する方法なので、情報圧縮の文脈では逆センスの手法である。今、興味ある対象を膨大な説明変数の中から少数個で線形回帰表現する問題を考える。何次の回帰モデルとするのか、さらにどの変数を使うかで、膨大な数の回帰モデルが存在する。モデル数は組み合わせ爆発しており、AICなどの情報量規準による最適モデルの探索（変数選択）は機能しない。一方スパースモデリングでは、回帰係数にL1正則化を加えた上で推定を行う。最適化関数の形がL1であることから、L1正則化は非ガウスモデリングと言える。この最適化の結果として、重要な説明変数のセットが自動的に浮きあがる。このように、説明変数の選択といった、限定されたモデリング技術に関しては、スパースモデリングにより線形・非ガウス型モデリングが実現された。

2010年代にはいつの最大の衝撃は、特定のタスクでの深層学習の圧倒的パフォーマンスである。その性能の高さは、入出国時の自動顔画像判別、スマートスピーカーでの音声認識、多言語自動翻訳など、私たちの生活に身近な製品として既に具現化している。これらの技術は、人の働き方や社会の構造そのものにも直接的に影響を与えていることから、深層学習はこの四半世紀の情報科学技術における最大のブレイクスルーであると言わざるを得ない。深層学習で使われるニューラルネットワークは、層数が増えた以外、第二次ニューロブームの時のものと違いは無く、そのパラメータ学習アルゴリズムも、Back propagationを基本とする以前のものとは大差はない。層数が増えた結果、パラメータ数も爆発的に増え、学習アルゴリズムもさほど賢くなっているわけでもないのに、必然と計算リソースはこれまでとは桁違いに必要となる。ただし、上述したように、それまでの特徴ベクトル構成法は機械学習の『匠の技』と言え、“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が、その問題を特定領域では基本的に解決した点は画期的である。深層学習では、通常、最適化関数はL2であるため、誤差にガウス性を仮定している。よって、非線形・ガウス型モデリング技術は、深層学習により、少なくとも予測・判別性能の観点からはほぼ極みに達したと言える。今後は、深層学習の大きな成功事例である、入力データが画像、音声、テキストなどに、一般的な問題での入力データをあえて変換すれば、特徴ベクトルの選択問題を回避可能となる。

非ガウス・非線形モデリングを自動化する試みは、この四半世紀に大きな飛躍を遂げた。その一方で、ビッグデータ×AI（人工知能）を土台に大きな存在感を示す米中のプラットフォームに対する風当たりも強くなってきている。今後は、モデリング技術の向上を計る上で、モデルの“人”による解釈可能性、帰納法で構築された意思決定システムの説明責任、データのバイアスと社会的偏見の分離など、人に寄り添った視点が極めて重要になってくる。