MATHEMATICAL
SCIENCES
June 2019 Number 672

特集／データサイエンスの数理

# 巻頭言

山西 健司

## 1. データサイエンスを初めて耳にした衝撃

筆者が「データサイエンス」という言葉を初めて知ったのは，1999年の情報論的学習理論ワークショップ（通称：IBIS：Information Baseed Induction Sciences）でのことであった．筆者がプログラム委員長として「モデル選択と知識情報処理の将来」という特別セッションをオーガナイズし，当時慶應義塾大の柴田里程先生に招待講演して頂いた．その講演の中で，これからは「データサイエンス」という新しい学問が発展するのだということを仰った．理論統計学で大きな貢献をされた柴田先生が統計学という言葉を使わず，わざわざ「データサイエンス」というからには何か意図があるにちがいない．筆者は恥を忍んで，「統計学と何が違うのですか？」と失礼な質問をした．そして，正確な文言は失念したのだが，「データから価値を引き出す，モデリングの新しいパラダイム」が生まれるのだという趣旨の御回答を頂いたと記憶している．

以来，「データサイエンス」が妙に気になる言葉として耳に残っていた．それが新たなる響きをもって聞くことになるのは10年以上経ってからである．今やデータサイエンスは，ビッグデータ，データマイニング等の言葉を押しのけ，情報学，統計学，AI，機械学習等を包含する，新しい学問の世界的な潮流を成している．柴田先生が仰っていた世界はこのことだったのかと今更ながらその先見の明に敬服する．

## 2. データサイエンスの数理とは？

データサイエンスはモデリングの科学である．データからその振舞いを説明するモデルを数理を使って作り出す．一方で，既存のモデルで十分にデータを説明できなければ，モデルを記述する数理を逆に新しく生み出す．物理学や化学などの自然科学はできるだけ自然現象を精緻に記述できる法則を生み出そうという立場だ．これに対して，データサイエンスではむしろ人間の理解しやすい関係や構造を用いて認識しようという立場だ．まさに，人間主体のモデリングがそこにある．

この方法論は統計学のものに近い．しかし，昨今のデータサイエンスは機械学習やデータマイニングの発達のおかげで，従来の統計学よりも豊かな表現力とパフォーマンス手に入れた．自然科学や統計学が扱いきれない社会現象や生命現象等も人工モデルを使って学習してやれというのである．

では，データサイエンスは機械学習もしくはAIそのものなのか？というと，そうではない．機械学習はモデル化の後のアルゴリズムや解析の話がメインであるのに対して，データサイエンスはデータのモデル化そのものに包括的な責任を取る立場にある．そしてその差異が顕著になるのは，データサイエンスでは**Accountability**（**説明責任**）を要件としているという点である．深層学習を用いて予測や分類の精度が良かったからといって，その理由と情報処理の過程を明確に説明できなければ，データをモデリングしているとは言えない．