

プロローグ

科学において世界を知る唯一の頼り
「データ」を使って未来を予測する

IT技術者の視点から見た データ・サイエンスの世界

中島 隆夫

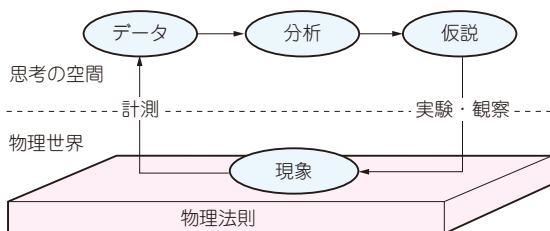


図1 データ・サイエンスとは…仮説検証のプロセスだとされている

データ・サイエンティストが「21世紀でもっともセクシーな職業」であると言われてはや10年が経ちました注1。その後深層学習を中心としたAIブームと相まって、データ・サイエンスは社会の中で確実に浸透してきたと感じています。

しかし、データ・サイエンスとは一体何者かと問われると、いささか心もとないのも事実です。百花繚乱の様相を呈している技術的な広がりを考えれば、その理解は百人百様あって然るべきですが、本稿ではIT技術者である筆者の視点からデータ・サイエンスの世界を概観してみます。

科学としてのデータ・サイエンス

データ「サイエンス」と言うからには、データ・サイエンスも科学の一味になります。では科学とは何でしょうか。

一般的には図1のような仮説検証のプロセスだとされています。現象に何か法則性がありそうだと気付くところから始めて、仮説とその検証を繰り返しながらその法則性を明らかにしていく試みです。そこで自然現象が起きる物理世界と人の思考をつなぐのがまさにデータです。

現象学の言葉を借りれば、私たちは事象そのものを認識することができませんから注2、データを通して世界を見る(見たつもりになる)ことになります。科学において世界を知るにはデータだけが頼りなのです。

●天文学者ケプラーは世界初の本格的データ・サイエンティストだった

▶データから現象を解き明かす「逆問題」

その事情は自然現象を原理原則からひもとく科学の王道、物理学においても同様です。

私たちが物理学を勉強するときにはニュートンの運動方程式が与えられた上で振り子の運動を計算するわけですが、ニュートンは現象(≒データ)からスタートして彼の名を冠した微分方程式を導きました注3。運動方程式のような法則が与えられた上で現象を説明することを順問題、現象からその法則を推定することを逆問題と呼びます。物理学を理論的に支えているのは(学校での勉強とは違って)逆問題の成果であると言えます。

▶神学的・哲学的だったかつての自然科学

しかし、物理学の黎明期には計算機がまだなかったので扱えるデータ量に限界がありました。神学的要請から提案された仮説を哲学的に議論することが物理学や天文学の主な関心事で、データは主張を補強する程度の従属物でした。

そのような状況の中で、ケプラー(図2)は彼の師であるティコ・ブラーエが残した膨大な惑星の観測データを使って惑星の軌道が楕円軌道であると結論づけたのでした。当時は天動説と「コペルニクスの転回」で有名なコペルニクスの地動説の間で決着がついていませんでしたが、いずれにしても円軌道を前提とした仮説でした。全知全能の神が創造した世界ですから、完ぺきな対称性のある円こそが神の完全性を象徴するのにふさわしいというわけです。

注1: Thomas H. Davenport: Data Scientist: The Sexiest Job of the 21st Century, Harvard Business Review, 2012年10月号。

注2: 現象学は哲学的方法論の一種で、「事象そのものへ(Zuden Sachen selbst!）」というスローガンを掲げて認識することの本質に迫ろうとしたドイツの哲学者フッサールは、事象そのものを認識するためには判断を一旦停止すべし(エポケー)という驚愕の方法を提唱した。

注3: 「自然哲学の数学的諸原理」(1687年)によって運動方程式や万有引力を含む一連の理論体系が発表された。