

データサイエンス全体の歴史から気がついたこと

鈴木治郎*

Email: szkjiro@shinshu-u.ac.jp

*1: 信州大学全学教育センター

◎Key Words データサイエンス, 科学技術史,

1. はじめに

データサイエンス教育の扱う対象は、近年、急速な発展を迎えた人工知能（AI）技術をおおむね包含する領域である。とくに文部科学省による AI 戦略 2019 が提唱されて以降、高等教育機関ではそのカリキュラムを構築し、実施を進めている。一方でデータサイエンスの定義は人それぞれに違うという現状もある。このような状況を踏まえて筆者は、大学でのデータサイエンスリテラシー教育の基盤と考えるべきデータサイエンスの歴史を記述しようと構想した。そのために現代的視点から取り上げるべき対象を定めて、古代から現代までの、私たち人類のデータサイエンス的活動を一冊の書籍⁽¹⁾にまとめた。本発表ではその中でも、科学技術史として注目すべきなのに、科学史の文脈では取り上げられることの少ない事例を 4 点紹介したい。

2. 天動説とはデータサイエンス的に見て何であったか

プトレマイオスの天動説は、その依拠する地球中心モデルの誤りが、コペルニクスによる地動説の出版、ガリレイによる地動説を裏付ける事実の発見を通じて正されて、地動説へと交代すべき状況が醸成された。また、火星など惑星の公転軌道モデルは、ガリレイと同時代のケプラーにより楕円軌道であることが示され、それはニュートンにより単純な力学法則にしたがって導出されたことで幕を閉じたと通常は説明されている。ところでデータサイエンス的にはガリレイによる発見よりも、プトレマイオス、コペルニクス、そしてケプラーに至る幾何学モデルは、実際の観測値への説明能力がどうであったかは見逃せない。以下ではとくに火星の扱いに注目しながら整理したい。

2.1. プトレマイオスの幾何学モデル

プトレマイオス⁽²⁾は地球中心の円軌道に各天体を配置し、単純な円軌道では不十分であれば、周転円などの工夫により、当時として十分に高い水準まで天体運動を再現できた。とくに火星の場合、逆行（「惑星」という名のもとにもなった、天球上に想定した円上を逆に進む現象）を説明するために周転円が必須であり、後に楕円軌道が原因として解明された円軌道とのずれは、エカントと呼ば

れる円の中心をずらす工夫により修正できた。

2.2. コペルニクスの地動説モデルは十分か

コペルニクス⁽³⁾は地球も火星も太陽の周りを公転するとした地動説モデルにより、火星の逆行自体はプトレマイオスの周転円に頼らずに単純に説明できた。しかしながら、円軌道に固執した彼は、円軌道と楕円軌道との差異を埋める工夫であったエカントは採用せず、プトレマイオスモデルよりは少ないものの周転円を用いた地動説モデルを構築せざるを得なかった。しかも天体運動の説明能力はプトレマイオスモデルよりも低かった。コペルニクスの説は、パラダイム転換への抵抗感ばかりか、データサイエンス的に見ても説明能力の低いモデルであったことを覚えておきたい。

2.3. ケプラーによる楕円軌道

ケプラー⁽⁴⁾は、ティコ・ブラーエによる精度の高い大量のデータを引き継いだ。肉眼によるティコの観測精度は 1 分に届いていた。それを、円軌道モデルにより火星の軌道を計算したケプラーの与えた誤差は 8 分あり、観測精度を大きく超えていた。何回もの再計算の末、ケプラーは円軌道を捨て、楕円軌道を得たのである。それは周転円やエカントのような技巧を含まない美しいモデルであるとともに、データサイエンス的な説明能力の観点からも、プトレマイオスモデルを凌駕したのである。

3. マカロックとピッツの論文

私たちがニューラルネットワークの基本モデルとして目にするのは、ローゼンブラットによるパーセプトロンを表す次図であろう。この図が多くの入門書で「マカロックとピッツのモデル」と説明されている。

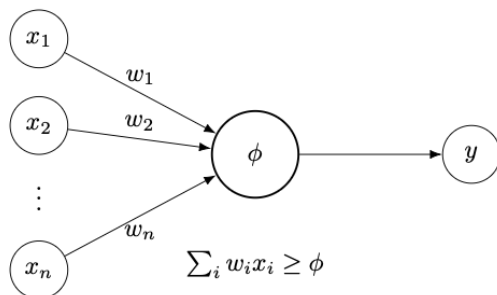


図 1 よく目にするマカロックとピッツのニューロンモデル

しかし、マカロックとピッツの論文⁽⁵⁾において、ニューロンモデルに学習のための重みは描かれてはいない。たとえば次図のようなものである。

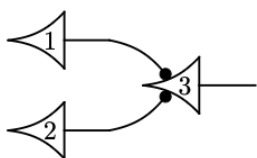


図 2 マカロックとピッツの論文にあるニューロンモデル（AND 素子の例）

この図にはローゼンブラットがパーセプトロンで設けた、学習のための重みはない。マカロックたちはニューロンによる AND, OR, NOT その他の論理素子を提案し、それを組み合わせれば脳を、ニューラルネットワークとして捉えられることを主張していた。図の数字は閾値の構成に対応し、この点はパーセプトロンにも受け継がれている。

3.1. ノイマンの EDVAC 報告書

ノイマンの EDVAC 報告書⁽⁶⁾に関しては、ストアドプログラム方式の発明者は誰であるかが議論の中心である⁽⁷⁾ように見える。以下は筆者が歴史概説とした文献⁽¹⁾よりも詳しく論じた論文⁽⁸⁾からの要約である。

ノイマンは真空管素子を有力候補としながら、逐次処理コンピュータを実現するための設計書を書き上げた。そこではマカロックたちのニューラルネットワークモデルに多く言及しながら、とくに以下を提案している。

- ・ニューロンの代わりの演算・記憶素子
- ・クロックに基づく強制計算の枠組み

設計書の記述においては、随所でマカロックたちの設計との対比があり、とくに記憶素子に関しては、当時の真空管は高価過ぎたためか、具体的な推奨はしていない。

クロックを持つ設計を特徴と捉えるべきなのは、マカロックたちの論文が、現代的視点から見たときデータフロー計算機に相当する設計を提案して

いたためである。

当時の電子計算機に至る道筋の中では、機械式計算機、継電器（リレー）式計算機の後継者として ENIAC などが登場したため、逐次計算機構そのものは、ノイマン以前には強調されていなかった。

4. チューリングとベイズ推定

アラン・チューリングは第二次世界大戦中のエニグマ暗号の解読作業において、暗号鍵候補の蓋然性の評価にベイズ推定の手法を用いていた。しかし、それはベイズの事前・事後確率の直接計算ではなく、オッズ比の形で進めていた⁽⁹⁾ことを解読グループの一人、グッドが説明している。

4.1. オッズ比の方法

一般に確率 P に対して

$$O = \frac{P}{1-P}$$

で定義される値がオッズである。ところで、仮説 H がデータ E で更新されるベイズの事後確率 $P(H|E)$ は、ベイズの定理により

$$\frac{P(H|E)}{1-P(H|E)} = \frac{P(H)}{1-P(H)} \times \frac{P(E|H)}{P(E|\neg H)}$$

という関係になる。これはオッズを用いれば

$$O(H|E) = O(H) \times B$$

と書き直せる。ここで B はベイズ因子または尤度比と呼ばれる。もともとのベイズ確率の式を扱う教科書の多くが、その計算は煩雑であるとしている。一方でオッズを用いれば、適当な事前確率から出発し、データ E が得られる度に事後確率に代わるオッズを計算することは容易である。

4.1. オッズ比による推論例

次の仮説 H を考えよう。

H : 「この暗号文には WE ATTACK AT DAWN が含まれている」

1. 事前確率 $P(H) = 0.10$ (ベイズ推定では適当な根拠のもとに事前確率を設定する) とすれば、

$$O(H) = \frac{P(H)}{1-P(H)} = \frac{0.10}{0.90} \approx 0.111$$

になる。

2. 暗号文の一部を復号してみたら、

E_1 : 短い語の並びが WE に見えた。

$P(E_1|H) = 0.60$ を仮定し、また間違っそう思う確率 $P(E_1|\neg H) = 0.20$ と仮定すれば、ベイズ因子は

$$B_1 = \frac{P(E_1|H)}{P(E_1|\neg H)} = \frac{0.6}{0.2} = 3$$

になる。新たなオッズ O_1 は

$$O_1 = O \times B_1 = 0.111 \times 3 = 0.333$$

になる。つまり仮説 H が正しいとするオッズは 3

倍になった。オッズから確率を計算すれば 0.25 と、最初の 0.1 と比べて大きくなっている。

以下「ATTACK に関する確率の仮定」, 「AT に関する確率の仮定」, 「DAWN に関する確率の仮定」と続けて、計算を進めることは容易である。

このようにオッズで考えるとき、新たなデータを用いてオッズを更新するのは電卓を用いた簡単な作業になるのである。

5. 暦の改定

グレゴリオ暦の改定は、制定から 1500 年ほど経過したユリウス暦（閏年は 4 年ごとに設けるもの。現在の西暦＝グレゴリオ暦に比べて簡単である。）が、たとえば春分の日において 10 日近くのズレを生じるようになった当時、取り組まなければならない課題であった。

この暦の改定事業は、当時のヨーロッパ諸国に関わる一大事業となり、データの公開、改暦モデルの公開と検討など長年の月日をかけてなされた。オープンサイエンスの先駆けの一つといってよい事例である。

5.1. 日本の改暦

日本でも江戸時代前期、春分の日に 2 日のズレを生じる暦の改定が課題となっていた。その改定を請け負ったのが渋川春海である。そのことを描いた小説（沖方丁『天地明察』）は映像化作品にもなったので、知っている人もいるだろう。渋川自身は、その暦が日食を正しく予測できるまでに失敗も繰り返しながら誠実に成し遂げたわけだが、ヨーロッパの改暦と比べると、一人の人間に重大な事業を託してしまった進め方は、文化の違いと考えるべきなのか興味深い。

6. おわりに

以上、科学技術史の事例のいくつかを、データサイエンスの視点から見ると、今まであまり注目されてこなかったいくつかの事実気づく。その中には「オッズによるベイズ推定」のように、教育手段としても興味深い例もある。文献⁽¹⁾に取り上げた事実の多くは、筆者が調べた限りでは、エッセイやデータサイエンスの演習問題としてのみ公表されているものも少なくない。

データサイエンスを進める営みは、数値化された事実に向かい合う私たち人類の歴史でもある。だから、より興味深い話題を発見できることがまだまだ潜んでいると思われる。文献に取り上げたリストを大きくできれば幸いである。

参考文献

- (1) 鈴木治郎：“データサイエンスの航海”，1-160，学術図書（2026）。
- (2) プトレマイオス（薮内清訳）：“アルマゲスト”，1-610，恒星社恒星閣（1993）。

- (3) ニコラウス・コペルニクス（高橋憲一訳）：“完訳 天球回転論”，1-728，みすず書房（2017）。
- (4) Ioannes Keplerus（ケプラー）：“新天文学（*Astronomia Nova Aιτιολογητός seu Physica Coelestis, tradita commentariis de motibus stellae Martis*）”，ラテン語／日本語未訳（1609）
- (5) McCulloch, W. S., & Pitts, W.：“A logical calculus of the ideas immanent in nervous activity”，115-133，*Bulletin of Mathematical Biophysics*, 5(4)（1943）
- (6) バークス著：“誰がコンピュータを作ったか”，1-545，工業調査会（1998）
- (7) John von Neumann：“First Draft of a Report on the EDVAC”，Moore School of Electrical Engineering, University of Pennsylvania. June 30 (1945)
- (8) 鈴木治郎：“マカロックとピッツの論文について”，オンライン，第 19 回国際 ICT 利用研究学会研究会（2026）。
- (9) Good, I. J.：“Probability and the Weighing of Evidence. London”，Charles Griffin (1950)