

## はじめに

本書の主要なテーマは、人工知能とビッグデータの時代ともいわれる現代において新たな発想をつかむためにキーとなる、仮説を構築する方法についてである。

第5世代移動通信システム（5G）サービスの開始・普及に伴って、これまで以上に大容量かつ高速のデータが生成されるようになっていく。さらに一度に接続可能になるIoT（internet of things）デバイスなどの数も飛躍的に増え、実世界データ・ソーシャルデータ・オープンデータを含む多種多様なデータを組み合わせて取り扱う必要性がますます大きくなってきた。

このようなデータはビッグデータと総称される。ビッグデータを統合的に利用すれば、新たに有用な知識や価値が生まれ、それらを通して知的でダイナミックな社会を実現できるだろう。それによって期待される利用分野には観光、モビリティ、社会インフラ、医療、科学など社会基盤的な応用が含まれる。

しかしながらビッグデータが与えられて、それらに対して人工知能やデータマイニングなどの先進的なITを適用すれば、自動的に有用な仮説が得られるわけではない。特に先進的な応用では、複数種類のデータやすでに得られた仮説の組み合わせで統合的な仮説が得られる場合が多いが、異なるデータをどう組み合わせれば有用な仮説が得られるかは必ずしも自明ではない。

本書では多くのビッグデータのユースケース（具体的な応用例）の観察に基づいて、ビッグデータを活用する応用情報システムを構築するために必要な基盤技術となる統合的な仮説生成の方法を、データ分析（人工知能、機械学習、データマイニング）とデータ管理（データベース）の技術を調和的に利用したアプローチに基づいて説明する。

本書ではまず仮説の定義について述べたあと、仮説生成のヒントになるリサーチクエスチョンと、データを理解するために役立つデータの可視化の方法について説明する。基本的な仮説生成に役立つ推論形式として、演繹、帰納、もっともらしい推論、類推、問題解決の手法を具体的に解説する。また科学の諸

iv はじめに

分野（天文学や遺伝学）において仮説がどのように作られてきたかを、主に回帰を手掛かりにして、ケプラー、ガリレオ、ニュートン、ダーウィン、ゴルトン、ピアソン、ライト、メンデル、フィッシャーなどの場合を題材として説明する。

次に仮説生成の手法である機械学習について、クラスタリングとニューラルネットワークを中心に説明したあと、統合的仮説生成の方法論（差分、結合、重ね合わせ、和分）についてさまざまなユースケースを通して詳細に説明する。

本書は、ビッグデータと人工知能を応用した現代の情報システムの理解と構築のための参考書というだけでなく、広く仮説について身近において参照できる、いわば仮説大全のような書になることを願う。

最後に本書の執筆にあたりご協力いただいた共同研究者の皆様ならびに共立出版において企画から編集、出版までをお手伝いいただいた関係者の皆様に深く感謝する。

2021年8月

柿生にて 石川博