

# 特集にあたって

今倉 暁 (筑波大学)

近年の AI 技術の急速な発展に伴い、巨大企業や大規模コホート研究などが保有するビッグデータを中心に、さまざまな価値を生み出すことに成功している。一方で、巨大企業や大規模コホート研究以外においても、企業や自治体、医療機関などが独自にデータを収集・蓄積している。これら複数機関が分散保持するデータを集約し、新たな「分散ビッグデータ」として統合的に解析することで、従来になく新しい価値の創出に繋がることが期待される。しかしながら、企業や自治体もつデータは個人情報や企業秘密などのデータ秘匿性の観点から、集約することが困難である場合がある。

現在、わが国は、国際的に信頼性のある自由なデータ流通の促進を目指す「DFFT (Data Free Flow with Trust: 信頼性のある自由なデータ流通)」の実現を目指しており、そのための重要技術として、プライバシー保護とデータ共有・利活用のジレンマを解消し、プライバシーを保持したままデータの分析・演算や機械学習を可能にする技術「PETs (Privacy-Enhancing Technologies: プライバシー強化技術)」が挙げられている。

このような社会背景のもと、複数機関が分散保持するデータに対して個人情報や企業秘密などの問題点をクリアし安全な統合解析を実現する技術として、筑波大学では NEDO プロジェクト「人工知能技術適用によるスマート社会の実現/人工知能技術の社会実装に関する日米共同研究開発/データコラボレーション解析による生産性向上を目指した次世代人工知能技術の研究開発」を開発し、現在、実用化に向けた研究・開発を進めている。

本特集では、データコラボレーション (DC: Data Collaboration) 解析に着目し、基盤技術、応用技術および実応用例について紹介する。本特集を通じて、DC 解析を始めとした分散データ統合解析技術の発展および普及に貢献することを目的とする。

櫻井鉄也氏 (筑波大学) と今倉による「多機関分散データに対するデータコラボレーション解析」では、本特集で注目する DC 解析技術の概説として、技術概要、関連する研究の進展、および数値実験例が紹介されており、DC 解析の概観が理解できる内容となっている。

山城大海氏・面和成氏 (筑波大学) による「データコラボレーションへのプライバシー脅威とその対策に向けて」では、実際の事例を通してデータ解析におけるプライバシー問題とプライバシー保護手法について紹介されており、プライバシー保護手法の重要性が理解できる。また、DC 解析技術のプライバシーに関する解析および差分プライバシーの適用について紹介されている。

川上雄大氏・高野祐一氏 (筑波大学) による「データコラボレーション解析における統合関数の最適化」では、DC 解析技術の肝となる統合関数の最適化に着目し、2 種類の新しい統合関数の最適化法について紹介されている。最適化手法の改善により DC 解析の解析性能の向上が可能であることが示唆されており、今後の DC 解析技術の高性能化の指針となる内容となっている。

今倉による「多機関分散データに対する生存時間分析」では、DC 解析の応用技術として、医療分野で重要とされる生存時間分析に対する新しい DC 解析技術について解説されており、DC 解析技術が医療分野の重要タスクに対しても有用であることが示されている。

河又裕士氏・岡田幸彦氏 (筑波大学) による「分散機密データを想定した処置効果の推定手法の開発」では、DC 解析の応用技術として、医学・社会科学やマーケティングなどで利用される潜在アウトカムモデルに基づく処置効果推定に対する新しい DC 解析技術について解説されており、DC 解析技術の多方面への発展性が示唆されている。

鮭川矩義氏 (法政大学) による「健診データと検歴データの統合による糖尿病の発症予測」では、機械学習手法による糖尿病の発症予測に関する国内外の最近の事例とともに、つくば市と筑波大学附属病院が保有するデータに対する DC 解析の実施例について紹介されており、DC 解析技術の実データでの有用性が示されている。

本特集を一環として、日本発の技術であるデータコラボレーション解析を始めとした分散データ統合解析技術の研究と普及に貢献していくことができれば幸いです。最後に本特集に協力していただいた各執筆者と担当編集委員の高野氏に感謝いたします。