

特集にあたって

宇野 裕之 (大阪府立大学)

ビッグデータを研究するとはいったいどういうことなのだろうか。

アルゴリズムの性能を評価する尺度の一つに、入力サイズが大きくなったときの漸近的な計算時間がある。入力はデータであり、その意味で、アルゴリズム理論はそもそも「ビッグ」なデータを想定し、その処理を効率化する研究を行ってきた分野とも言える。

一方で現在のビッグデータ研究は、個々の分野におけるデータホルダーが、機械学習などの既存手法にデータを与えて可能性の高い判断を得るという活用集中しているようにも見える。ビッグデータそのものの性質を知ることや、その解析・処理手法の開発はおろそかになっていないか。他にすべきことはないのか。このような状況で、ビッグデータに対するとくに理論研究は、どの方向を向いて何を目標せよいいのか。

その指針を与えていただくべく、関西支部では2018年11月に企業や学術界などから多様な立場の方を招きシンポジウムを開催した。本特集は、その5名の講演者に新たな1名を執筆者に加えた合計6編の記事から構成する。

最初の2編はビッグデータ利活用の事例である。平手勇宇氏 (楽天株式会社 楽天技術研究所) は、自社が提供する各種ウェブサービスから得られる膨大なデータを利用し、処理結果が顧客サービスにフィードバックされるサイクルを、おもに潜在顧客抽出やレコメンド・クーポン配信を対象に解説される。日本有数のデータホルダーによるサービス展開の実際を知る貴重な機会である。江頭瑠威氏と山川茂孝氏 (株式会社電通) は、Twitterが発する情報をリアルタイムデータと捉え、マーケティングに活用する事例を紹介する。これまでの蓄積された静的なデータの利活用とは異なり、マーケティングモデルを作り出す試みと合わせて次世代の活用方法の可能性を提示している。いずれの事例でも、用いられる主要な要素技術は機械学習であることがわかる。

データ解析の代表的な手法に確率や統計によるものがある。岩田具治氏 (NTT コミュニケーション科学基礎研究所) は、データに隠された構造を説明する確

率的潜在変数モデルを基礎から解説したうえで、筆者らの提案による人流推定など、データのクラスタリングやマイニングへの適用例を紹介する。現場で発生する問題がどのようにモデル化されるかの実際が興味深い。

人工知能やビッグデータ研究が諸外国に後れをとる中、科学技術振興機構が所管する戦略的創造研究推進事業 CREST の一領域「ビッグデータ統合利活用のための次世代基盤技術の創出・体系化」(研究総括:喜連川優氏 (国立情報学研究所)) が、国家戦略として実施されている。平成 25~27 年度に 11 の研究課題が採択され、それぞれ約 5.5 年の期間で進行中である。次の 2 編は、いずれも平成 26 年度採択課題の研究代表者によるものである。

宇野毅明氏 (国立情報学研究所) (課題名「データ粒子化による高速高精度な次世代マイニング技術の創出」) は、データマイニングにおけるクラスタリングを例にノイズ除去の手法であるデータ研磨の効果を解説されると同時に、自身の豊富な体験にもとづき、OR における現実問題の重要性と理論研究が乖離する現状を強い問題意識として提示される。私もメンバの一人である加藤直樹氏 (兵庫県立大学) (課題名「ビッグデータ時代に向けた革新的アルゴリズム基盤」) のプロジェクトは、ビッグデータ処理を可能にする劣線形時間アルゴリズムという計算パラダイムを提唱している。記事ではそのプロジェクトにおける研究成果の全貌を紹介する。

ビッグデータ CREST の領域アドバイザーでもあり、他にも数多くのコンピュータ・サイエンス関連のプロジェクトに同様の立場で関与してこられたのが徳山豪氏 (関西学院大学) である。その経験にもとづき周辺分野も含めた研究動向を熟知のうえで、ビッグデータ・人工知能技術に対する安易な盲信に警鐘を鳴らしつつ、研究(者)の進むべき方向性が期待を込めて語られる。

本特集が読者のみなさまに有益であることを願うとともに、今後ビッグデータにどのように向き合って研究していけばよいかを見つめ直す機会としたい。