

特集「AIとデータ —データに基づく意思決定と 社会イノベーション創出—」にあたって

城戸 隆

(株式会社 Preferred Networks)

早矢仕 晃章

(東京大学)

近年、さまざまな分野で蓄積された膨大なデータを利用し、さまざまなAI技術を用いて発見した価値を意思決定に役立てようとする動きが活発になり、「ビッグデータ」と呼ばれる世界的なブームも起きてきた。スマートフォンなどのパーソナル情報端末の普及により、今まで取得困難であった個人の行動履歴が取得可能となり、マスマーケティングから、「個」のマーケティングへ進展してきた。また、それらの現象に呼応してIoTやAIの流行も発生し、これらが相乗効果をもたらして多様なトレンドを生み出してきている。そのような社会背景の中で、AIとデータの連携の需要は高く、さらに、既存の分析技術や統計学では扱えないデータの登場など、今までAIが主要技術ではない分野においても注目が集まっている。AI研究は間口の広いテーマを扱っているが、根本にある技術はデータに基づく意思決定であり、それに対して正面から取り組むことが重要であると考えられる。そこで、本特集号では、AIの学習のために必要なデータをいかにして、つくり・集め・守り・生かすのか、また学習済みデータをどう流通させるのかについての知識、AIとデータによる社会イノベーションの創出について、人工知能分野の最新の研究を特集としてまとめ、学会員および読者らに提供することを狙いとする。

AIとデータを取り巻く環境、社会からの要請、トピックスは非常に広範囲にわたるが、本特集では以下の6編の記事を寄稿していただいた。また、執筆者に依頼するにあたり、大学・企業の双方の意見を取り入れるように構成した。

No. 1 樋口の「データ関連の数理技術の変遷：深層学習は春秋戦国時代の秦となるのか？」は、統計学を専門とする著者の長年の経験から、データに関わる諸数理技術の変遷を振り返り、深層学習隆盛の現在の人工知能研究を俯瞰している。深層学習も統計的機械学習の一手法であり、内在するパラメータ推定（学習）にまつわるさまざまな根源的弱点を避けて通ることができないことを指摘したうえで、深層学習が人工知能技術の「度量衡」となりつつある理由として、パラメータ推定法、計算プラットフォーム、成果発表の場の三つをあげている。深層学習の登場と秦の「度量衡」のアナロジーを用いなが

ら、深層学習適用の今後の有望分野やさまざまな課題の位置付けを明確に語る俯瞰的論考である。

No. 2 丸山、城戸の「機械学習工学へのいざない」は、深層学習の技術の進化に伴い帰納的システム開発の特徴を捉えた、新しい開発方法論「機械学習工学」の必要性を論じている。システムの仕様を定義し、先験的な知識に基いてそれをモデル化し、段階的に詳細化していくという今までの演繹的なシステム開発に対して、機械学習に基づく帰納的なシステム開発は、仕様を訓練データの形で表現し、実装は学習によって行うという形を取る。帰納的なシステム開発においては、仕様定義の方法、実装の方法、テストの方法、運用の方法などが今までのシステム開発と大きく異なり、そのため演繹的システム開発を前提としたソフトウェア工学の方法論がそのままでは適用できない。演繹的システム開発と帰納的システム開発の共通点と相違点を明らかにし、ソフトウェア工学から得られる知見と、新たに機械学習のために必要な項目について議論した論考である。

No. 3 麻生、本村の「実社会ビッグデータを活用する次世代人工知能技術」は、今後社会で新たに出現するビッグデータ（実社会ビッグデータ）を活用する次世代人工知能技術について論じている。これまでの人工知能技術はインターネットサービスで収集される大規模なデータや知識を利活用することで発展してきたが、今後は、各種のIoTデバイスやロボットなどの技術と連携して、実社会でのサービスや生活行動を通じて収集されるビッグデータを活用することが重要になると洞察している。産業技術総合研究所人工知能研究センターで取り組んでいる五つのプロジェクトと人工知能技術コンソーシアムの取組みを紹介しながら、次世代人工知能技術の現状と今後の課題を解説した論考である。

No. 4 早矢仕、大澤の「データジャケットを用いた異分野データ連携」では、データ概要情報記述手法「データジャケット」を用いたデータの価値発見と異分野データ連携の実例を示し、データ流通および交換を促進する諸技術について紹介している。異なる領域のデータを組み合わせる新しい価値を創出し、意思決定に役立てることへの期待は高まっているものの、プライバシーやデータの使途が明らかとなっていないことによる機会損失な

ど、社会的障壁が多く存在する。組織横断的な協働を促進するためには、データのみならず、データから得られる情報や知識の交換が重要となる。データをどう使うか、データを使うことによりどのような問題が解決可能であるのか、さらに、データの価値を発見し、流通させ、新しいデータ駆動によるイノベーションを創出させるためのデータ利活用方法を検討する支援技術についても解説している。

No. 5 橋田の「データ循環としての知能」では、個人のレベルでも社会のレベルでもサイクル(意味・価値)が知能の基本的な構成要素であるべきことを述べ、そのようなアーキテクチャによってシステムの複雑性を低減させることがモジュール構造をもつ知識や技能を学習するための前提条件であることを指摘する。そしてデータ循環の観点から、人工知能およびスマートソサエティ(人工知能と融合した社会)の課題について論じている。人工知能の実運用においてリアルタイムのデータ循環を効率良く行うには、個別サービスの特定の相手や状況に関するリッチなデータが簡単に得られなければならない。GDPの大半を個人消費が占めるため、リッチなパーソナルデータが本人の意思で自由に活用できる社会の構築がAIの社会実装に不可欠である。特にパーソナルデータが個別サービスにおいて本人の意思に基づき円滑に一次利用できるようにすることの重要性を指摘している。AIとデータに関して人工知能の基礎から、実社会にお

ける問題点とそれらを解決するためのアプローチについて、理論から実践まで広く解説する。

No. 6 坂本の「AIに学習させやすい知覚データの取得方法」は、データとして取得が難しいとされてきた消費者の知覚にフォーカスを当てた論考である。データのプラットフォームが海外に握られてしまっている中で、リアルデータを価値ある形で利活用し、革新的製品やサービスの実現につなげるリアルデータ利活用サイクルの重要性を論じている。その中で、背景知識や嗜好の異なる個人のデータをどのように取得し、計算機可読性を高めて分析に役立てるかということについて、オノマトペ(擬音語・擬態語の総称)に着目し、モノづくりに携わる人や商品としてのモノに接する消費者の評価をビッグデータ化できる可能性について基礎技術を紹介しながら議論している。欧米の言語には少ないとされるオノマトペを用い、モノづくりの現場や市場の消費者の感覚を丁寧に拾い上げ、リアルデータから新たな価値を生み出すことが我が国の強みであり、チャンスであると述べている。

本特集を企画するにあたり、多くの先生方にご協力いただいたことをこの場をお借りし、深く感謝したい。

本特集が、日本におけるAIとデータ関連研究を盛り上げるための一助となることを願っている。