特集「パターン分類とクラスタリング」に寄せて

筑波大学 システム情報系 遠藤 靖典

人工知能（AI）によって、科学のみならず社会や私たちの生活に至るまで劇的な変貌を遂げようとしています。様々な出版物ではAIに関する解説や特集が数多く組まれており、インターネットの普及とも相俟って、それらの情報に容易にアクセスすることができるようになりました。

さて、AIの基幹技法といえば深層学習（ディープラーニング、DL）が真っ先に頭に浮かびます。2000年代に始まった第3次AIブームの原動力となったのはDLでした。AIの技法は、人が専門知識を基にしてルールを設定するルールベースと、大量のデータからルールを自動生成する機械学習とに大別されます。1980年代の第2次AIブームで盛んに研究されたルールベースでAIを構築する方法論は、データが膨大になるにつれてルール数も増大し、ルール相互の矛盾等が生じるようになり、人が専門知識をルールとして記述することの限界が露呈したため、下火になりました。人がルールを設定出来なければコンピュータに自動生成させればよい、との考え方から注目されるようになったのが機械学習です。教師あり学習の1つであるパターン分類や教師なし学習の1つであるクラスタリング、そしてDLは機械学習に含まれます。この機械学習の一手法であるDLが、ポストルールベースとして第3次AIブームの原動力となり、現在のAI研究の興隆の牽引役となったのです。

人が特徴量を設計するという点で、パターン分類とクラスタリングは特徴量を自動的に抽出するDLとは異なります。DLが脚光を浴びた理由はこの特徴量の自動的抽出にありました。データが大規模・複雑化していくと、人による特徴量の設計には限界があります。また、DLアルゴリズムの深化のみならず、コンピュータの性能向上もこの第3次AIブームを支えました。しかし、DLを効果的に用いるためには多くの学習データが必要となります。さらに、DLはある種のブラックボックスであり、何が計算されているかは分かっても、それが何を意味するのかの説明が困難、と言う問題もあります。（令和2年度の科学技術振興機構の戦略事業の戦略目標の1つに「信頼されるAI」が挙げられているように、AIの説明困難性が特に重要な問題ということは研究者間で共通認識となっています。）

そこで、DLに比べて学習データが少ない、計算過程や結果の説明が容易、数理的な深化や発展が期待できる等の点から、改めてパターン分類やクラスタリングの有効性が再認識され、DLに勝るとも劣らないAIを支える重要な技法として活発に研究されるようになっていきます。パターン分類、クラスタリング、DLにはそれぞれ長所・短所があり、どれが良い悪いと言う訳ではありません。今後もこれらの手法はそれぞれ活発に研究されていくでしょう。

そこで、今回の本学会誌「知能と情報」では、「パターン分類とクラスタリング」として特集号を組みました。本特集号では「膨大なデータを自動的に分類・要約して活用するためのデータ分類技術」に焦点を当てており、3件の論文と3件の解説記事の形で、特にクラスタリングに関する最新の研究成果が収められています。論文には、プライバシー保護、非負値行列分解、サイズコントロールといった、近年注目されている概念を取り入れた研究成果がまとめられています。また、解説にも近年得られた新たな知見がまとめられており、読み応えのある内容です。今回の本特集号が読者諸兄にとって有用なものとなることを願っております。

最後になりますが、ゲストエディタの大阪府立大学の生方誠希先生、近畿大学の濱砂幸裕先生をはじめ、本特集号に尽力いただきました多くの方々に厚くお礼申し上げます。本特集号がパターン分類やクラスタリングをはじめとした機械学習に関する研究のさらなる深化と発展の一助となることを祈念しています。