

特集にあたって

池辺 淑子 (東京理科大学)

近年、機械学習はもっとも進展が目覚ましく、注目度の高い研究分野の1つである。活用範囲は生命情報科学やロボット制御、金融工学やデータマイニングなど多種多様であり、多くの研究者が活躍している。

昨年10月に開催されたRAMPシンポジウムでも「機械学習における最適化」という最適化アルゴリズムに焦点を当てたセッションが組まれた。そのご講演を拝聴して機械学習分野のエネルギーや可能性に感銘を受け、編集委員の一人として、是非機関誌で取り上げたいと思い、この特集を企画した次第である。

2006年11月号はSVM周辺と領域判別を特集しているが、機械学習全般に関する紹介は機関誌ではまだなされていない。そこで本特集は、機械学習が活躍するさまざまな分野の紹介を目的として、5名の方々に解説記事をご執筆いただいた。

まず、杉山氏による「機械学習入門」では、データの統計的な性質を活用する統計的機械学習の基礎的な事項が、応用例とともに丁寧に解説されている。脳波による意志の伝達技術で用いられる非正常環境適応学習、部品のチェックで用いられる異常値検出、動画からのイベント抽出に用いられる二標本検定、ブラインド信号源分離などの手法である特徴抽出やクラスターリング、データの可視化で使われる条件付き確率推定に関して解説がされた後、それらは確率密度比推定と呼ばれる統一的な技術で扱えることが紹介されている。入門解説ではあるが、今後の発展の方向までが見え、本格的に学びたい人のために文献も多く挙げられている。

つづいて、瀬々氏の「生命情報解析で活躍する機械学習」では、ヒトのDNA配列の解析に使われる機械学習技術について解説されている。データの採取方法の紹介から入り、生命情報科学分野の解析の歴史やシーケンサと呼ばれる読み取り装置の近年の劇的な進歩、それに伴うデータの量や質の向上によって使われる機械学習の技術発展について、とても興味深く述べられている。生体から採取されるゆえに発生するデータのゆらぎの多さがもたらす問題とそれに対する取り組み、疾患や薬剤効果予測に期待が高いDNA配列の違いの発見を見つけるリシーケンシングなど、発展著しい分野の迫力が伝わる話題が満載である。

森本氏の「ロボット学習」では、人間型ロボットの制御における機械学習について解説されている。記事では、ロボット制御においては、HJB方程式と呼ばれる方程式が中心的役割を果たすこと、近年この方程式に対する新たな知見が、多くのシステム変数をもつ人間型のロボットの学習問題に対するアルゴリズムの提案につながったことが紹介され、さらに実際の歩行ロボットCB-iをモデル化した応用事例について述べられている。シミュレーション結果も示されており、実装化へ向けた開発が続いていることにも触れられている。人間型ロボットの实用化への期待が膨らむ記事である。

武田氏、後藤氏の「機械学習とポートフォリオ選択の素敵な関係」では、金融工学と機械学習の関わりについて解説されている。そこではまず、回帰モデルを機械学習で扱うアプローチについて、誤差の最小化をするときに、罰則項を考慮する正則化という機械学習の手法が過適合の防止など有用であることが説明されている。そして、金融工学のCVaRモデルを用いてSVMのあるモデルの妥当性が示されたこと、機械学習の正則化をポートフォリオの平均・分散モデルやCVaRモデルに採り入れることで予測精度のよい解が得られることが述べられている。異なる分野の方法が互いの発展に貢献する、まさに素敵な研究成果である。

宇野氏による「最適化からみたデータマイニング」ではデータマイニングにおける手法について、アルゴリズムとその高速化の技術が紹介されている。中心になるのはデータマイニングの頻出集合発見問題であるが、アプリオリ法とバックトラック法の2つの列挙法の枠組み、データベースの工夫による高速化について初歩的な事項から丁寧に解説されている。データマイニングは小売店の購買履歴など、非常に大量のデータを対象とするため、高速化が重要課題であるが、宇野氏ご自身の取り組みに関しても言及され、大変興味深い記事である。

著者の方々にはご多忙ななか、すばらしい解説記事をご執筆いただいたことに感謝したい。本特集をとおして、機械学習に興味を持って頂ければ幸いです。