

論文特集「データマイニングと統計数理」

鷺尾 隆

(大阪大学産業科学研究所)

近年、社会の情報化が進み、我々は多種多様な情報を計算機ネットワークを通じて入手し利用できるようになった。一方で、生成・収集・蓄積されるデータの量は急激に増加しており、膨大なデータをいかに分析し活用するかが課題となっている。その有力な手段の一つとして、大量データに埋没した有益な知識を計算機によって発掘するデータマイニング技術が注目されている。データマイニングでは人工知能分野で長年にわたり研究されてきた機械学習や知識ベースに関する手法が中心的役割を担う。これを背景として人工知能学会では、知識ベースシステム研究会や人工知能基本問題研究会、AIシンポジウム、全国大会をはじめとするさまざまな場でデータマイニングの話題を取り扱ってきた。一方、データを解析するというデータマイニングの本質から、統計数理に基づく手法開発も非常に重要であり、近年ますます、人工知能的手法と統計数理的手法の併用、融合が必要となってきた。

この状況を受け、データマイニングと統計数理手法の融合による新たな人工知能および統計数理の理論・技術の創出を目指し、人工知能学会第2種研究会である「データマイニングと統計数理研究会」を設立した [1]。そして、この研究会の中核的研究者を中心に、内外研究者による本分野研究および交流の推進を目的として2006年9月25、26日にThe International Workshop on Data-Mining and Statistical Science (DMSS2006)が開催された [2]。また、それと並行して人工知能学会論文誌に「データマイニングと統計数理」論文特集を企画し、特集論文とワークショップ論文を同時に受け付ける形で、機械学習、確率計算、統計数理、データベース、

テキストマイニング、データ可視化など、データマイニングが網羅する幅広い分野にわたって論文募集が行われた。全部で28件の投稿があり、23名からなる論文特集編集委員会を組織して厳正な審査を行った。その結果、ここに掲載する12件の論文が採択された。

最初の3件は、データマイニングの中心的テーマの一つであるバスケット分析原理を拡張し、変数従属性判定やアイテム集合導出・インデキシングを効率的に行う手法に関する論文である。次の3件は、構造データや数値データに関する新たな分類手法やその応用を扱っている。さらにその次の3件は、XML構造の効率的な可到達性判定や多変数非定常データ、Webリンク構造に関する時間的変化分析、予測に関する手法である。最後の3件はいずれも単語や自然文、会話などから、概念や意見、話題などを抽出する手法に関する論文である。本特集に採択された論文の内容は多岐にわたり、いずれもが従来のデータマイニングや統計数理的手法の枠組みを着実に拡大するものである。

これらの論文が提案する手法には、離散構造を扱うアルゴリズム的な内容のもの、連続数理モデルに関するものも含まれるが、過半がアルゴリズム的な手法と確率、統計、数理的手法を組み合わせ、融合させたものであることに注目したい。今後のデータマイニング研究が、計算機科学や統計数理の枠組みの相互影響によって発展していく可能性を強く示唆する論文特集といえる。

[1] URL http://www.med.shimane-u.ac.jp/med_info/sig-dmsm/

[2] URL <http://www-ikn.ist.hokudai.ac.jp/dmss2006/>