

「特集 データマイニングと統計数理」について

樋口 知之¹・鷺尾 隆² (オーガナイザー)

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容易に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、現在では蓄積された膨大なデータをいかに分析し活用するかがあらゆる研究分野で最大の課題となっていると言っても過言でない。その有力な解決策の1つとして、大量のデータに散逸・埋没した有益な知識を発掘(マイニング)するデータマイニング技術が注目されているのはあらためて申し上げる必要もないであろう。

これまでデータマイニング研究の話題は、人工知能学会を中心にさまざまな情報科学系の学会や研究会で取り扱われてきたが、この分野における統計科学の研究実績と視点の重要性を鑑みて、データマイニングを主たる研究対象領域とする研究会「データマイニングと統計数理研究会」(以後、DMSM)を人工知能学会のもとに鷺尾・樋口を共同主査として新たに設立した。この研究会では機械学習を中心とする従来のアプローチに加え、統計数理的アプローチを中核的な研究対象に据え、人工知能分野と統計数理分野の研究者間交流を深めるとともに、両アプローチの融合による新たなデータマイニング技術の創出を目指している。現在活動をスタートして3年目になり、年3回の研究会を企画開催しているが、予想以上の参加者数をみると当初の目的はかなり達成されてきている。活動の詳細はDMSMのホームページ(“DMSM”で検索するとトップにくる)をご覧ください。

この研究分野のさらなる強化と国際化を目的に、2006年には科研費特別推進研究「半構造データマイニング」(研究者代表 有村教授・北大)とDMSMの共催で、DMSS06(Data Mining and Statistical Science)を9月末に札幌にて開催し好評であった。それではということで、2007年の秋に東北大で開催される国際会議ALT07/DS07(Algorithmic Learning Theory/Discovery Science)と同期して、直後に統計数理研究所で国際ワークショップDMSS07を開催することになった。この会議はDS07のサテライト会議として位置づけられた。なお、2008年にはDMSS08を東工大で開催し、引き続き国際的な活動を継続している。

DMSS07では、データマイニングと統計数理に関して、関連する研究論文を広く募集し、この学際分野の深い理解と発展を促すことを目標とした。具体的なテーマとしては、機械学習、確率計算、統計数理、データベース、半構造データマイニング、テキストマイニング、データ可視化インターフェイス、データマイニング固有技術等が挙げられる。DMSS07の開催の後すぐに、データマイニングと統計数理に資する研究を目的とする理論・手法・技術に関する基礎から応用までを対象とした研究論文を、DMSS07で講演したかどうかにかかわらず広く募集した。最終的には7編の論文が採択され、本特集号の掲載に至ったわけである。

データマイニング、統計数理、機械学習、さらにはデータベース等の研究領域は相互に密接な関係を持ち、学問全体としては統合に向かいつつある。これらの研究分野がお互いに刺激し合うことで、よりレベルの高い研究を行い、現代社会が直面するさまざまな難問の解決に資する総合科学に育つことを期待する。最後に、本企画の趣旨にご賛同くださった読者の方々をご

¹ 統計数理研究所：〒106-8569 東京都港区南麻布 4-6-7

² 大阪大学 産業科学研究所：〒567-0047 大阪府茨木市美穂が丘 8-1

の分野に参入し、近未来のデータ環境を見通して、そこで重要となる新しい研究領域を開拓してくださることを願ってやまない。