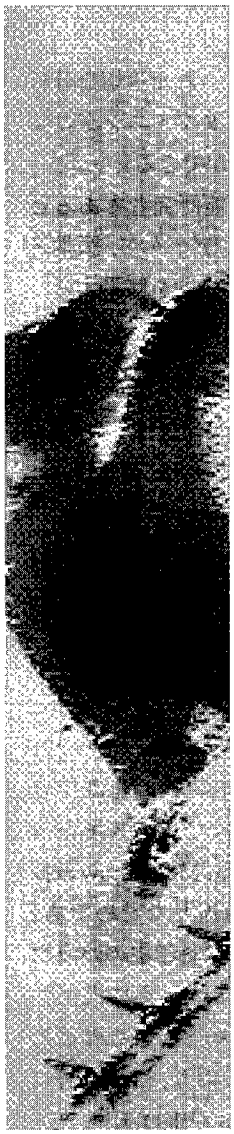

特集 データマイニングコンテスト

Data Mining Contests

編集にあたって

鈴木英之進（横浜国立大学工学部電子情報工学科）
suzuki@dnj.ynu.ac.jp
津本　周作（島根医科大学医療情報学講座）
tsumoto@computer.org

近年，データの取得，保存，および処理においてハードウェアの高性能化と低価格化が進み，大規模データの解析が盛んに行われるようになった．もっとも，これらのデータ解析は日常的で単純な業務であることが多く，有用な知識を発見するためにはより高度で知的な解析が求められている．このような背景を受けて，大量データからの有用知識の発見を目的とするデータマイニングが誕生した．データマイニングは，現実の大規模データと格闘するために，機械学習，統計学，およびデータベースなどの技術を総合して用いる研究分野であると位置づけられる．

データマイニングにおける課題は解決が困難である場合が多く，上記のように機械学習，統計学，およびデータベースなどの技術を総合して用いる必要がある．このためデータマイニングは研究分野としての歴史が比較的短いにもかかわらず，その基盤として種々の知識発見手法が提案されている．これらの手法は，大量データの高速かつ知的な解析を可能とするものであり，今後のデータ解析に大きな影響を及ぼすと考えられる．ただし知識発見手法は，その目的が適用領域における有用性であるために評価が困難であり，データマイニングの歴史が比較的短いこともあって優劣をつけることが難しい．対象問題に対して数多くの候補の中から適切な知識発見手法を選択するためには，評価基準の確立と実適用の経験が必要である．

データマイニングコンテストは，共通課題を与えて複数の知識発見手法を評価する試みであり，上記の背景もあって米国，欧州，およびアジアで盛んに開催されている．ただしデータマイニングコンテストは，データマイニングの目標である有用性が明示的に定義しづらいため，評価の場として重要であると同時に適切な設定が必要とされる．

本特集は2部の構成になっている．第1部（第1～4編）では，米国，欧州，日本・アジアの主要なデータマイニ

ングコンテストの主催者を招き，コンテストの実施に加え，基本方針や苦労話，得られた知見などを紹介していただいた．第2部（第5～6編）では，知識発見手法の社会応用に詳しい専門家を招き，応用面から見たデータマイニングコンテストについてその知見，問題点および展望について議論していただいた．

第1編では，昨年米国で開催されたKDD-Cup 2000の主催者たちが，eコマースの実データを用いた興味深いコンテストを解説する．問題の設定，データの準備，および投稿された論文と得られた結果の分析により，規模や影響力の面で重要であるデータマイニングコンテストの具体例を知ることができる．

第2編では，昨年と一昨年に欧州で行われたDiscovery Challengeの主催者が，金融データと医療データを用いたコンテストを解説する．機械学習の研究分野においては1990年代初頭より正答率を評価基準とした分類学習アルゴリズムの比較が盛んに行われているが，筆者はデータマイニングを総合工学としてとらえ知識発見プロセスの重要性を解いている．

第3編では，本特集のエディタである我々が日本・アジアにおけるデータマイニングコンテストを解説する．我々はこれらのコンテストに当初から開催者と参加者の両方の立場で参加しており，種々の知見を有している．本解説より，米国とは異なり仮説生成を重視するコンテストについて知ることができる．

第4編は，知識発見手法の標準問題についてのアーカイブを構築した経験に基づく興味深い解説である．執筆者たちはデータマイニングコンテストを開催したわけではないが，コンテストに使用できるいわば問題集を作成した．アーカイブのデータ集合は種々のデータ形式，発見・学習タスク，および応用領域にわたっており，今後機械学習において重宝されているUCI機械学習レポジトリと同様の発展が期待される．

第5編では，我が国の産業界におけるデータマイニング応用に詳しい研究者が，データマイニングコンテストの可能性と課題を解説する．執筆者は研究者としての実力と数多くのインタビューに裏付けられた適切な分析と提言を行っており，本解説よりデータマイニングコンテストが進むべき1つの道を知ることができる．

第6編では，医用統計学・人工知能研究者であるとともに医師でもある本特集のゲストエディタが，医学・医療からみたデータマイニングコンテストを解説する．執筆者は日本・アジア・欧州における仮説生成型コンテストの主催者でもあり，基礎研究と実領域の両方に詳しい立場からデータマイニングコンテストの現状と将来を論じている．

本特集はデータマイニングコンテストにおける最先端の動向を伝える目的から，半数の記事が海外の研究者によって執筆された．このため我々は翻訳やコピーライト交渉なども負担することになったが，これらの解説は興味深く労力に十分値した．ただし我々の心は晴れない．本特集の総説を執筆する予定だったノースキャロライナ大学のJan M. Zytkow教授が今年1月に急逝したためである．Zytkow教授は科学史のバックグラウンドを持つうえに人工知能において機械発見をその揺籃期から研究しており，この分野において国際会議（PKDD）やワークショップ（Discovery Challenge）を創設したことからも分かるようにデータマイニングには特別の思い入れがあったと思われる．彼と国境と世代を超えた良き友人だった我々は，彼の死にあたって涙するとともに，総説を入手できない不幸にも遭ったことになる．Zytkow教授のご冥福を心からお祈りするとともに，非力ながらこの分野の発展に引き続き努力することを誓いたい．

我々はデータマイニングコンテストにより知識発見手法についての性能面での知見が蓄積され，大量データの有効な知的解析が日常的になることを切に望む．本特集がそのきっかけになれば望外の喜びである．

（平成13年3月31日）