

## 特集にあたって

喜連川 優 (東京大学生産技術研究所)

企業の業務を支える基幹系システムは、トランザクション処理データベースを中核に開発されてきた。永続データに対する操作の並行処理制御技術が核となり大きく発展してきたが、新しい日々の通常業務とは別に、企業の将来に向かっての戦略的な意思決定を支援する新しいデータベースの構築が求められる時代に入った。部門ごとに個別に運用されている各種のデータベースから、必要なデータをその形式を適切に変換し、整合のとれた統合的なデータの器であるデータウェアハウス、またさらに限定したデータマートの構築が進められている。ここでは、メタデータの活用とともに View 実体化技術が中心的役割を果たし、基幹系データベースから自動的にその変化が伝搬するように制御される。これにより、検索処理は大幅に性能が向上することになる。

データマイニング、OLAP はデータウェアハウスのフロントエンドツールと位置づけられる。関係データベースは、その非手的操作方法により従来の手続き型方法に比べて、大幅に記述のセマンティクスのレベルが向上されたものの、エンドユーザによる記述は容易とは言えず、OLAP ではデータを多次元キューブ状に配置し、簡単な操作で種々の視点から解析する環境を提供している。任意の軸を選択し、さらに階層的にドリルダウンすることにより詳細なデータを即座に得ることが可能である。このような高速なユーザインタフェースを実現するためには、事前に関係データベースから OLAP に最適化したデータ構造に変換することが必須となり、その構築には多大な計算が必要なことから、現状では長時間のバッチ処理により実現されており、効率的アルゴリズムの開発が活発になされている。

データマイニングは、データウェアハウスからの新しい知識の抽出技術の総称であり、近年大きく注目されている。決定木、分類、相関ルールなど、種々のアプローチが各種アプリケーションにおいて試みられている。KDD (Knowledge Discovery and Data Mining) なる国際会議は急成長し、1998年より ACM (Association for Computing Machinery) に SIGKDD が新設されるに到っている。また、アジア・オセアニア地区でも、PAKDD なる国際会議が昨年から開催され

ている。ニューヨークで8月に開催された KDD'98 は、データベースシステムの中で最も権威のある VLDB '98 なる国際会議と連続開催されていることから分かるように人工知能研究者とデータベース研究者が協力して活動している。莫大なデータの効率的な操作には、データベース技術の適用が不可欠であり、相関ルール抽出アルゴリズム (Apriori, DIC)、決定木生成アルゴリズム (SLIQ, SPRINT) など、データベース研究者からの効率的アルゴリズムの貢献が見られる。また、データマイニングが効を奏するためには、データのクレンジングが不可欠であり、データベース技術により構築されるデータウェアハウスの存在は大きな意義がある。マイニングにより得られた知識の価値は、最終的にはユーザによって決められるものであり、可視化ツール等の研究も盛んである。また、データウェアハウスに対する問合せ同様、データマイニングも負荷の重い処理であることから並列処理の研究も始まっている。データマイニングは、いまだ未成熟ではあるものの、通信、銀行、証券、小売り、医療をはじめとし、多くの業界ですでに広く利用され始めており、データウェアハウスを核として大きな発展が期待される。

本特集は、昨年、電子情報通信学会データ工学研究専門委員会により主催されたデータウェアハウス、データマイニングチュートリアルにおける講演者からの事例を中心に企画されている。ややデータベース寄りな紹介になっている可能性もあるが、今後より多くの分野の研究者によるさまざまな視野からの協調的研究が望まれる。

提案されつつあるデータマイニング技術は、OR の立場から見ても興味あるツールとなるのではなかろうか。たとえば、大規模なデータに対する分類木の作成が並列データベース処理によって提供されれば、従来想定していたのとは全く異なる規模の問題を現実的な時間内に処理可能となろう。データベース分野では2次記憶入力を含めた徹底した処理の効率化が達成されており、とりわけ巨大データの操作が自在に行える。データマイニングの現状の紹介を試みる本特集が OR 研究者に、いささかでもご興味を頂け、また、新しい試みのきっかけになれば幸いである。