

小特集・論文特集「アクティブマイニング」にあたって

元田 浩

(大阪大学産業科学研究所知能システム科学研究部門)

沼尾 正行

(東京工業大学大学院情報理工学研究科計算工学専攻)

小野田 崇

((財)電力中央研究所情報研究所)

河野 浩之

(京都大学大学院情報学研究科システム科学専攻)

データマイニングは多量のデータから有用な知識を発掘する技術の総称であり、ここ数年、シーズとしての技術面からも、ニーズとしての応用面からも著しい進展を見せており、確実に情報化社会の基盤技術として定着しつつある。その背景には、大容量記憶媒体の低価格化、計算機処理能力の向上、情報通信技術の急速な進展などにより、我々の情報収集能力が、ビジネス分野をはじめ政府機関や個人レベルに至るまで、社会のあらゆる局面で、飛躍的に向上したことがあげられる。しかし、その一方で、収集可能な情報を解析し、理解し、有効に活用する技術のほうに追いついていないという現実がある。情報の価値は、いかに多くを所有するかではなく、いかに効率的に運用できるかにある。

実際、情報の洪水に溺れてしまい、①膨大な情報空間の中で迷路にはまり込む、②脱出できても目的にかなった価値ある知識を簡単には取り出せない、③状況変化に即応できず頻繁な知識の更新に対応できない、などの大きな問題がクローズアップされている。情報収集・データ解析・目的設定変更のサイクルが高速回転し、個人も組織も情報洪水の中で疲弊しているのが現状である。

このような状況の打破には、単に機械学習や統計の諸技術を中心とするマイニング技術の開発だけでは不十分であり、領域専門家と計算機科学者が一体になり、情報収集、前処理、マイニング、視覚化、結果の評価、不足データの追加など、データマイニングの全サイクルを支援する総合環境の構築が不可欠である。これらは、個別の要素技術として従来からも鋭意研究されているが、全体として有機的に機能させる新しい取組みを必要とする。

本特集は、この問題に真正面から取り組むことを目的として、平成13年度から発足した、特定領域研究「アクティブマイニング」の概要と現時点までの成果、ならびに関連する研究成果を紹介することを目的に企画したものである。その概要と現時点までの成果については、領域全体の目標、それを達成するための10個の計画研究の取組みと進捗状況を一つの解説記事に、領域全体としてアクティブマイニングの実証問題として取り上げている肝炎データ解析に関する現時点までの成果を一つの解説記事にまとめてある。関連する研究成果の紹介については、個別の研究成果なので解説記事としてではなく、論文として公募した。20件の応募があり、その中から厳正

な査読の結果、6件を選択した。3件が情報収集、3件がマイニングに関するものである。学会誌にはアブストラクトしか掲載されていないので、個別の論文の詳細は電子出版されている論文誌のほうを参照していただきたい。

1. 情報収集に関する3件の論文

最初の「タンパク質の三次元モチーフ辞書」の論文は、タンパク質の構造マイニングに必要な、タンパク質構造中に特定の配置で存在する局所構造特徴(モチーフ)の辞書の作成に関するものである。構造データマイニングは化合物を中心とする化学の分野で注目を浴びている新しい分野である。2番目の「部分更新モニタリング」の論文は、ユーザに負荷をかけず、指定された場所に部分更新があった場合にのみ該当データを自動更新する手法を提案したものである。Webから収集する情報の中には価格情報のように頻繁に値が更新されるものがあり、非定常大規模分散知識源からの知識獲得の一助となる。3番目の「既存データベースを活用した情報収集」の論文は、既存のデータベースを訓練例として、特定の事項に関するデータを自動的に収集するための分類器の学習に関するもので、目的情報を効率的に自動収集するための一助となる。

2. マイニングに関する3件の論文

4番目の「ルール発見の最悪解析」の論文は、悪いルールの発見を回避するには、あるいは良いルールの見落としを回避するにはどのくらい事例が必要かを理論的に解析したものである。PAC学習理論の拡張にもなっており、データマイニングの理論的基盤形成に貢献するものである。5番目の「プロセス時系列のマイニング」の論文は、観測される連続時系列データからプラントのプロセス特性を得る方法に関するものである。従来の自己回帰モデルなどとは異なり、構造モデルを推定できる特徴があり、マイニングの新しい局面を示唆している。6番目の「帰納アプリケーションの並列合成」の論文は、マイニング対象データの性質に応じ最適なマイニングツールをオントロジーに基づき整理された部品から自動合成する手法の高速化と洗練に関するものである。多様なデータに対してきめ細かなマイニングが可能となることが期待される。

最後に本特集号にご協力いただいた執筆者の方々ならびに特集号編集委員の方々に感謝の意を表します。