# はじめに

データマイニングは一言でいえば，応用が対象とする大量のデータの中から，頻出するパタンや意味のある構造を発見することである．そのための基本タスクには相関ルール，クラスタリング，分類，外れ値検出がある．

データマイニングの伝統的な応用としては，バスケット分析や，顧客の分類，クラスタリングを基にしたマーケティング，クレジットの不正利用の発見などがある．

新しいところでは，インターネットや Web の普及に伴って，Web ページや XML ドキュメントの内容と構造の分析に基づき，それらの分類やクラスタリング，検索を行うという応用がある．また赤外線や温度，照度などのセンサの集まりからなるセンサネットワークは時系列データを生み出し，それらを空間情報も含めてマイニングすることにより人間の行動予測などが可能になる．さらにその結果を用いて機器の適応的制御を効率的に行うことで快適性を考慮した省エネルギー化に貢献することが期待できる．

地理情報に対しては，マーケティングや行政での地域にかかわるデータの利用から科学における衛星画像の利用まで幅広い応用が考えられる．さらに実空間を対象とするので，マイニングの結果を地図上に可視化することで．その有効性を増すことが期待できる．

科学の一分野である生物情報学においては，例えばアミノ酸配列に基づきタンパク質の構造や機能などを発見するということに，分類やクラスタリング，検索といった技術が適用される．

かようにデータマイニングの応用分野は拡大し続けている．

一方でしばしば現代はデータ洪水の時代と言われる．ではデータはどれだけ大量なのか．

パーソナルコンピュータにも内蔵され大量のデータを格納することのできる磁気ディスク装置は生まれてからおよそ 50 年がたつ．その間に磁気ディスクの記録密度は 1 千万倍以上にも増大した．一方で記憶されるデータの方も増加している．IDC（2008 年）の調査によれば，ごく最近の数年間だけを取ってみても，全世界のデータは，2006 年に 161 エクサバイトであったものが，2011 年には 1.8 ゼッタバイトにもなると見積もられている．ここでエクサは 10 の 18 乗，ゼッタは 10 の 21 乗である．しかも 2011 年には人間の生産するデータの総量が，人類が手にする記憶媒体の記憶容量の合計を一桁以上も上回るという予測がある．ちなみに 2007 年は，全世界のデータの総量（281 エクサバイト）が，人類が利用できる記憶媒体の記憶容量の合計にちょうど追いつき，そして抜き去っていったときである．

特に急速に増加している部分には，デジタルテレビ，監視カメラ，発展途上国でのインターネットアクセス，センサ，データセンター，ソーシャルメディア（ツイッターやフリッカーなど）由来のデータが含まれる．これからのデータマイニングは，こうした大規模データを対象

にしていかなければならない．

データマイニングで問題となるのはこうしたデータ量の大きさ (Volume) だけではない．データマイニングの応用分野が広がるにつれて，その扱うデータ構造の多様性 (Variety) も問題になりつつある．従来のデータマイニングは主として構造データを対象としてきたが，Web をはじめとするインターネットの発展につれてグラフや半構造データを扱う機会が増えつつある．またセンサネットワークから生まれるデータは本質的に時系列データであり，また GPS を利用すればデータに対して位置情報も付加される．静止画像，動画像や音声といった非構造のマルチメディアデータもデータマイニングの対象になってくる．

さらにセンサデータだけでなくソーシャルメディアの一つであるツイッターにも代表されるように，新しいデータの一部は，これまでデータマイニングが扱ってきたデータにくらべてより大きな速度 (Velocity) で生成されている．

本書はデータマイニングの基本概念や基本タスクとそのためのアルゴリズムを説明するだけでなく，現代のデータ（いわゆる“ビッグデータ”）の特徴である 3 つの V（大きさ，多様性，速度）を意識して，発展的な手法も合わせて説明する．さらに本書は最近注目されている集合知を，ソーシャルメディアに対するマイニングという観点から説明することを試みる．

そうした意味で本書では類書とは異なり，現代的なデータマイニングの全体像を伝えることを目指す．

本書が学生，若い技術者や研究者をはじめとして，現代的なデータマイニングに関心のある読者に幅広く利用されることを願う．また本書をまとめるにあたって大変ご協力を戴きました，情報系教科書シリーズ編集委員長の白鳥則郎先生，編集委員の水野忠則先生，高橋修先生ならびに共立出版編集部の島田誠氏に深くお礼を申し上げます．

2012 年 6 月

石川　博
新美礼彦
白石　陽
横山昌平