

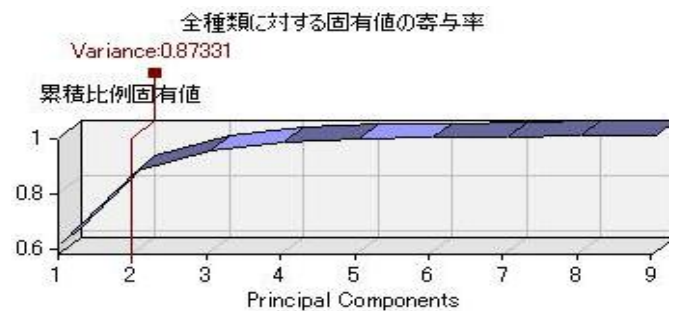
従来からある統計手法は、大きな母集団であっても、そこから無作為に抽出した小さな標本から、いかに効率的に母集団に関する結論を導き出すかということを中心としてきた背景がある。しかし、現在の情報機器の高度化やネットワークの普及に伴い、一見すると全数調査またはそれとほぼ同等の規模の標本であっても、同じような結論が求められることがよくある。例えば同時購買分析（バスケット分析）を考えると、変数を各商品としたら、販売されている商品すべてが変数の数になる。また標本であるコンビニエンスストアでも数千になる。データマイニングは、このような大量データ（POSデータ）から有用な情報を導き出すための方法であるが、それには探索的データ解析と統計的多変量解析法を柔軟に使いこなすことが必要となる。

実際にデータマイニングが用いられている分野として

- 流通サービス業における販売の予測
- 医療・バイオ産業における要因分析
- 製造業における品質管理
- 金融・保険業におけるリスクや顧客特性の分析

があげられる。具体的には、天候やイベント等の過去のデータから商品ごとの特性や購買数を分析したうえで、仕入れ量を決定し販売予測をすることに役立っている。

データマイニング手法としては、主成分分析・因子分析・回帰分析・クラスター分析など、通常が多変量データ解析で用いられている手法に加え、回帰分析のような明示的な関数を用いずに、例えばある商品の購買有無を「性別」「職業」「年齢」などの説明変数を、一連の手順に沿って、データを分けていくことにより予測や判別を行う決定木・回帰木。また上記で紹介したスーパーマーケットで買い物をする際に用いられる買い物かごにちなんでバスケット分析ともいわれる、



連関規則（アソシエーションルール）がある。これは、買い物かご単位で購入商品の傾向を分析することによって、消費者がある商品を買った場合に、別のある商品と一緒に買う傾向がどれだけあるのかを調べることができる。この分析手法が、データマイニングの代表的な方法の一つであることを示す、象徴的な例として、「ビールと紙おむつ」がある。

カナダのあるスーパーマーケットでバスケット分析を行ったところ、紙おむつを購入する人は、一緒にビールを購入する傾向（ルール：「紙おむつ→ビール」）が高いという意外な結果が出た。さらに調査をしてみたところ、実は赤ちゃんをもつ若い父親が、母親に頼まれてスーパーに紙おむつを買いに行く際、一緒にビールを購入する傾向が高いということだった。この結果を活かせば、例えば紙おむつの販売コーナーにビールを併置してみるということが一つ考えられる。ここで注意しなければならないのは、ビールの販売コーナーに紙おむつを置いてもあまり効果が見込めないということです。なぜならこのルールは「紙おむつ→ビール」であって「ビール→紙おむつ」ではないからである。

本講義では実際のデータに対し、計算機を用いて実際に解析を行える能力を養う。上記の手法に加え、ニューラルネットなどを中心に、理論的な理解およびソフトウェアを利用した実際の解析の両面を学ぶ。