

- 研究テーマ
- データマイニング: 多量データからの知恵の発見
- キーワード
- データ解析、データマイニング、空間統計学
- 産業界の相談に対応できる技術分野
- 品質管理、マーケティング、医療情報解析



関 庸一 教授

連絡先 関庸一 TEL:0277-30-1806 FAX:0277-30-1806 e-mail:sekiyochi@gunma-u.ac.jp

研究概要

知識を採掘する

今、情報ネットワーク技術が進歩しています。これにより、低コストで、広範囲から即時に双方向で情報交換ができるようになり、デジタル化されたデータが、各所で大量に収集/蓄積されるようになってきています。

たとえば、スーパーやデパートなどのレジでは、バーコードで商品を調べて会計をします。そこでポイントカードやクレジットカード、電子マネーなどを使うと、カードの持ち主が特定され、お買い物の内容と突き合わされて、いつ、どの店で、どの客が、どんな商品を購入したかというデータが収集されます。この情報は、ネットワークを通じて集められ、在庫の管理や、売れ筋商品の把握だけでなく、個々の顧客の好みを知るためなどにも使われることになります。

また、国や地方公共団体が収集するデータもあります。医療保険や介護保険では、被保険者が受けたサービスに公的な補助が行われその記録が残されます。このデータも適正で効率的なサービス提供を考えるために有

益となります。

収集されるデータの一件一件はそのときの個人の気まぐれの結果だったりしますから、一件の記録だけを見ても意味のあることは何も引き出せませんが、多量に集まったものを上手に分析すれば、有益な情報が引出せる可能性があります。このように多量データから知識を引き出すことをデータマイニングと呼びます。マイニングは採掘するという意味で、データの山から金銀のように貴重な知恵を掘り出すという意味になります。

当研究室ではこのデータマイニングの方法の開発と活用に取り組んで、多量データの構造を理解する研究をしています。



図1 利用計算機

特徴と強み  
大規模データを腑分けする

先に挙げた大規模データは、会計手続きのために悉皆に収集されたもので、異なった趣旨での利用など、多様な事例がその中に含まれます。これから、集められた事例群の特徴を理解していくためには何が必要でしょうか。

よく、ものごとをわきまえていることを分別があるといいますが、多様性を含む多量のデータの分析で最も基本となるのは、それぞれができるだけ均質なグループになるよう、データを分別することです。そのような方法の一つにSOM(自己組織化マップ)があります。SOMは、人間の神経細胞に倣って開発されたニューラルネットの一つで、沢山のものから代表的な類型を構成して二次元平面に並べます。

図2は、介護保険制度の要介護認定結果から高齢者の心身状態を類型化したものです。また、図3は、CDを購入した顧客の音楽嗜好を類型化し、それぞれの類型にどのような年齢階層の顧客が含まれているかを図としたものです。このような分類があれば、類型ごとのサービス給付の標準を考えたり、音楽嗜好ごとにCDの売り上げを予想したりするための手掛かりとなります。

大規模データの中の多様な個人については、それを適切に分別・層別することができて初めて、従来の精緻な統計モデルを適用することが可能になると考えています。そこで、古典的な統計的方法論が適用できるような均質なグループにするための手法を発見して、実データに対して有効な方法論の研究を進めています。このような方法で分別のあるデータの分別を可能としていきたいと考えています。

今後の展開  
時間と空間からのデータマイニング

我々の研究室では、現在、データに位置と時刻が添付された履歴データについても取り組

んでいます。具体的な対象としては、大地の変形について国土地理院が公開している全国数千地点の電子基準点の数年に渡るデータや、医療保険のレセプトデータについて、地域的時系列的な変化を捉えることについて検討を進めています。

データとして観測された数値の背後にどのような仕組みがあったのか、実世界の生きたデータを可視化して理解できるようにすることで、無意味に見える数字の背後に潜む意味を発見することには感動があります。また、データが生まれて来た背後の仕組みに対し良いモデルを作ることで、現実を的確に説明し、予測できることは喜びです。今後も、多量データに挑んでゆきたいと考えています。多量データを抱えながら、分析の方法に困っている方がおられれば、相談に来てください。分析に協力できることがあると思います。

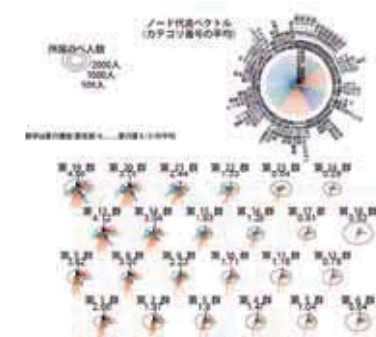


図2 高齢者の心身状態類型

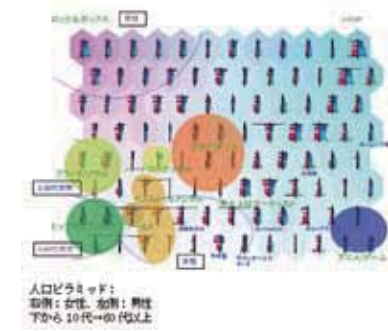


図3 CD購買層群ごとの人口ピラミッド