

データマイニングについて

1. はじめに

データ解析手法の一つとして「データマイニング」という技術があります。本稿では、この「データマイニング」について簡単に紹介しようと思います。

2. データマイニングについて

データマイニングとは、蓄積された大量のデータから有用な知識や情報を抽出する手法のことです。KDD(Knowledge Discovery in Database)と同義で扱われることが多く、そのプロセスは大きく分けて、(1)データの収集、および解析を行うデータの選択、(2)異常値を除去する前処理、およびデータマイニング処理が行える形へのデータ変換、(3)データからパターンの発見、(4)データの解釈、および新しい知識の獲得、の複数の段階を踏みます。このうち(3)だけがデータマイニングと呼ばれることもあります。^{1) 2) 3)}

データマイニングで扱うデータは数値データだけでなく、テキストデータや画像データなども扱います。また、繰り返し処理が行われることから機械学習技術やパターン認識技術等も用いられます。このような点からデータマイニングは従来からある多変量解析等の統計解析だけではなく、様々な技術が融合された分野でもあります。

3. データマイニングの適用分野

データマイニングの考え方は以前から存在するため、既にマーケティングをはじめとした経済や金融、流通分野で多く利用されています。他にも、気象予測や通信、医療分野でも利用されています。製造業においても品質管理を中心に適用事例が増えています。最近、データマイニングへの注目度は高くなっており、今後適用事例が広がっていくことが期待されています。

4. データマイニングをとりまく環境

最近、データマイニングが注目されている理由として、安価で大容量の記憶装置やメモリ、高性能の演算処理装置が低コストで利用できるようになったことで、大量のデータでも高度な解析が行える環境が整ってきたことが挙げられます。また、無線をはじめとした通信技術の進展による通信コストの低下や、センサデバイスの高機能化や価格低下により、大量のセンサ情報を簡単に収集できる時代になりました。大量の情報を扱うのはデータマイニングの得意とするところであり、収集された膨大なセンサデータを使った解析への取り組みも始まっています。例えば、温度や湿度センサの情報を集めて省エネを目指したエネルギー管理を行う取り組み事例があります。

こうした技術の進歩に伴う環境の変化により、今まで取り組まれてこなかった分野への応用が考えられ始めています。

5. おわりに

データマイニングも万能な技術ではありません。解析精度や解析アルゴリズム等、課題を抱えている部分もあります。また、導き出された結果に対する判断は最終的には人間が行うものであることを忘れてはなりません。

今後、課題が克服されれば新たな事例も増えると予想されますし、利点や欠点を把握した上でデータマイニングを利用すれば、新たな可能性を秘めた強力なツールになるのではないかと思います。

参考文献

- 1) 元田浩 他：データマイニングの基礎
- 2) 石井一夫：図解 よくわかるデータマイニング
- 3) 豊田秀樹：データマイニング入門



工業技術部 自動車・機械技術室(旧室名 機械電子室) 浅井 徹
(現 産業労働部新産業課(052-954-6477))

研究テーマ：センサネットワークにおけるデータ処理システムの開発

担当分野：情報技術、EMC