

今月の

用語

隣に伝えたい

新たな言葉と概念

【データ・マイニング】

英 data mining

〈解説〉

「マイニング (mining)」という単語は「採掘」を意味する。すなわち、「データ・マイニング」とは、「データの山から採掘する。」という意味である。通常、鉱山から採掘するのは金や銀などの価値のある鉱物や金属だが、データ・マイニングにおいて発掘するものは、知識である。

データを解析する際に、通常その解析者は特定の仮説を持ち、その検証のために解析を行う。一方、データ・マイニングにおいては、解析の目的はもっと探索的である。例えば、「大腸がんの発生と関係がありそうな因子について、自分が持っているデータの中から発見したい。」というようなものになる。

医療機関でこの言葉がきかれるようになった一つの理由は、カルテの電子化等に伴う電子化されたデータの蓄積にある。現在の病院情報は実に大量かつ多様なデータが含まれている。これらをマイニングするには、専用のソフトとともに、格納されているデータを解析できるような状態に加工する情報技術が必要になる。

解析対象となる変数が膨大であり、現時点で特定の仮説が立っていないような場合にはデータ・マイニングという手法は有効である。また、データ・マイニングのひとつのかたちとして最近研究分野で「テキスト・マイニング」というものがある。これは、自由記載アンケートや対話の逐語録などをデータの山ととらえ、そこに格納されている単語や文節をとらえた上でマイニングを行う手法である。

(国立病院機構東京医療センター 臨床研究センター (感覚器センター) 臨床疫学研究室 尾藤 誠司)

本誌 4p に記載