

マーケティングリサーチ，データマイニング，統計 民間企業におけるデータの蓄積と利用

池田伸（立命館大学経営学部）

1. 本報告の課題

本報告では，近年の経営実務における統計的方法の適用に係わり，とくに新しい手法であるデータマイニングをとりあげる．データマイニングの発展および経営への導入の実態とその背景とについて検討する．それによって，データマイニングをめぐる統計実践を経営統計学・社会統計学上にどのように位置づけることが可能なのかについて考察する．

2. 経営におけるデータの獲得・蓄積

2.1 OLTP による業務記録の蓄積

近時 IT の飛躍的な発展にともない，企業等組織における情報処理の能力が飛躍的に向上している．それを利用して，経営情報処理が企業の競争優位の源泉となってきている．たとえば，大規模小売店では商品マスターのデータベース DB と照合された 1 件 1 品毎のアイテムが識別され，仕入から販売までの全取引がトラッキングされる．半自動で在庫の補充を発注することも含め，DBMS に支援された単品管理 OLTP が行われている（POS はこの一部である）．

これらの取引は業務記録として蓄積される．これらの記録は，日次や月次決算のための帳簿システムへの入力となるが，会計情報としては適宜集計を行うことが基本となる．また，証憑という点を除けば，OLTP の記録は「揮発」してもかまわない．

ところが，前世紀末から情報処理のハードウェア資源が安価になり（そのためのメーカーの販促でもあったが），TCP/IP によるネットワーク化が促進され，システムのにも並列処理が行

われるようになると，情報処理の資源が急速に安価かつ大規模に利用可能となった．このことは当然 OLTP を含めた基幹系の経営情報システムそのものを進化させることとなったが，他方で膨大になるこれら記録をデータ処理することによって経営上有益な情報が得られることが期待された．しかし，そのためには二つの課題が残されていた．

2.2 データウェアハウスによる DB 管理

最初の課題は，OLTP をデータとして利用可能とすることであった．すでに基幹系のシステムには OLTP の記録が TB 単位で蓄積されているが，業務用の DBMS のままではデータ解析に利用することはできない（図 1-1 の左側．ただし，本図は一つのモデルである）．

そこでシステム上基幹系とは別に情報系に解析用 DB のシステムの構築が必要となった．これをデータウェアハウス DWH という（宍戸（1996）参照）．DWH では，OLTP の記録を揮発させることなく，あまり処理は経ないが解析に耐えうるようには洗浄され正規化された「生データ」として DB に格納される．さらに，実際の解析にあたっては，研究目的に応じて DWH から抽出されたサブセットの DB であるデータマートを直接の対象とする（図 1-1 のなかほど）．

3. 経営における大規模 DB の利用

3.1 OLAP

DWH が構築されると，デーマートの DB を端末から操作することで，これまでは（品質管理でいう）「オフライン」で専門的に扱われていたさまざまな情報が，エンドユーザであるマネジャー自身によって文字通り「オンライン」で入手することができる．たとえば，マーケティングデータにつ

いて日付・商品・店舗等からなる 3 次元（以上の）データ（キューブ）の様子を知ることができる．関心に応じてその場で対話的に再集計したり掘り下げたり（キューブの

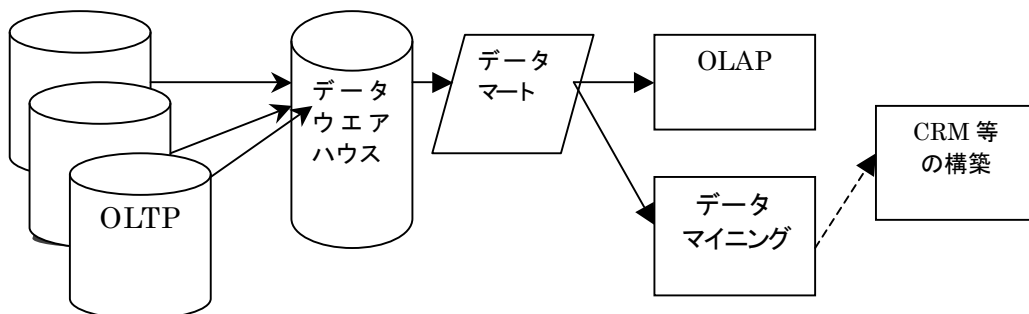


図 1-1 データマイニングにおけるワークフローの概要

スライス&ダイス)のクエリや結果の視覚化を指示することができる。これを OLAP といい、従来のオフラインの固定された製表によるレポートとは大きく異なる(ベリー, リノフ(2002)などを参照)。

3.2 データマイニング

OLAP だけでも以前には考えられなかったシステムであるが、オフラインにおいてより集中的に DWH を解析するための方法論や組合せによる爆発を避けるようなアルゴリズムの開発が要求された。この課題と可能性とを一般化していうと、大規模 DB からの知識の発見 KDD であり、データマイニングとは KDD の手法の総称である*。

データマイニング手法には、次のような方法が含まれることが多い。連関規則とは具体的にはマーケットバスケット分析として POS のスキャンデータから同時購買のアイテムの組合せパターンを探ることでマーチャンダイジング等に反映しようとするのである。決定木は顧客のセグメンテーション等に用いられる。基準変数となる反応が平均よりもよい(リフト率が高い)グループを顧客属性によって識別・構成可能かを明らかにする方法である。

たとえば金融機関の与信管理においては、財務諸表からの予測は困難がともなう。ここで、貸出先がデフォルトになったケースとそうでないケースとにつき、これらデータを「教師」として学習させることで(人工的)ニューラルネットワークによるモデリングが行なわれる。適度に学習を終えたモデルは、与えられたケースのデフォルトのリスクを返してくる。

参考になる情報は OLTP によるだけではない。コールセンターに寄せられる顧客の苦情は当然マーケティングに有益な内容を含んでいるが、従来は組織的な知識の開発・利用は難しかった。最近では web のログも含めて、DB 化されたテキストに関して、テキストマイニングが行われる。コールセンターの例では、顧客の発するキーワードが CTI を通じて処理され、その結果に基づきオペレータは瞬時に経験を要さずに対応することができる。

3.3 CRM

これら KDD が競争優位に貢献するとすれば、商品の単品管理に止まらず個々の顧客の管理である 121 マーケティング、手法としては CRM に結びつきがちである。KDD では DB 全

* KDD は経営データのみにも適用されるものではないが、もっとも主要な適用分野といえる。本報告では、データマイニングの対象を経営、とくにマーケティングに限定している。

体でのパターンが発見をめざすが、その知見を有効利用できる可能性が高いのは CRM として実装することで競合への「乗り換え」(チャーン)防止やクロス/アップセリングに役立てられる場合である。

4. 社会統計学上の位置づけの試み

上記のデータ利用・解析の目的は、すべて収益や株主価値への貢献にあり、そのための意思決定の支援への有効性で効果は測られる。情報諸資源の活用によって単なる個人的経験主義的でない意思決定の費用対効果から DWH, データマイニングから CRM に至るシステム導入が考量される。経営上の統計実践としてのデータマイニングは学術研究ではない(製薬会社の研究開発などのケースは別である)。

そのことが、大規模で非線形性が強く調査目的収集されていないデータにおけるパターンを探索的に逐次的自動的非確率的に処理しておもに予測を行う点で、データマイニングを通常の統計学よりも機械学習や脳科学、EDA の伝統に近づけている。また経済統計学に関しては、観測数の少ない集計量の経済時系列解析や母集団・標本パラダイムにおいては事実上システムの安定性よりも解の一意性が重視され、クロスバリデーション等による過学習のチェックは等閑視されてきた(マイクロデータの解析に適用の可能性もある)。

社会的な統計実践としては、データマイニングの成果は CRM の有効性に依存するであろう。さらに条件付き確率による個人や集団の管理をおしすすめる傾向は強まるであろう。

5. 参考文献

- ベリー, リノフ『データマイニング手法』海文堂(1999).
宍戸周夫『データウェアハウス』B&T ブックス, 日刊工業新聞社(1996).
豊田秀樹『金鉱を掘り当てる統計学』講談社ブルーバックス(2001).
-----『非線形多変量解析』統計ライブラリー, 朝倉書店(1996).

付録：略語リスト

121	One-to-one (Marketing)
CTI	Computer Telephony Integration
CRM	Customer Relationship Management
DBMS	Database Management System
KDD	Knowledge Discovery in Database
OLAP	On-line Analytic Processing
OLTP	On-line Transaction Processing