ビジネス・インテリジェンスとCRM

－データマイニング・ビジネスの実際－

山端　博

1．はじめに

今更いうまでもないが，データマイニングという言葉が，米国IBMのアルマデン研究所員，Dr.ラケッシュ・アグラワルによって考案され広められたということは，よく知られた事実である．1993年当時，IBMではInductive LearningやNeural Networkのデータ・マイニング技術に対し，並列化を含めたいくつかのリサーチ・プロジェクトが走っていた．その中のメンバーの多くはKDD学会（Knowledge Discovery and Data Mining）に所属しニューロ・コンピューティングやマシン・ラーニングの博士号を持つ経験豊富なアナリストたちであったが，実際にはいくつかのパイロット・アカウントを抱え，実務に適用しプロジェクトを実証的に進めていた．このことからもデータ・マイニングが学問的な裏付けを持ちながらも，むしろビジネスの世界の中にあって，そこでの成果をもとに評価され，発展してきたことが伺える．

このことは，いま世にいわれているビジネス環境の変化と無関係ではない．当初より，適用分野はおもに銀行やスーパーなどの流通業が中心であったが，これはこれらの業界かすでに大量データを蓄積しており，データウェアハウスの構築に対し，すでに少なからぬ投資がなされていたので，その投資に対する成果が問われていた．さらに一方で，気ままな消費者を直接相手にしており，個々の消費者の購買行動を分析し理解した上でのマーケティングの展開が焦眉の急であった．以下に，その基本コンセプトから具体的ソリューションまで，順を追って紹介する．

2．ビジネス・インテリジェンスの5段階

ビジネス・インテリジェンスは，1996年に米国のGartner Groupによって提唱された，意思決定支援ソリューション体系をより広い視野から捉える新しい考え方であるが，IBMにおいてはこれをさらに整理し，「ビジネス・インテリジェンスの5段階」として定義している（図2－1）．

第1番目にデータウェアハウスに対する「データ検索」の段階があり，次に「統計解析」，そして3番目が「多次元データ分析（OLAP）」であり，「データマイニング」は第4番目のステージとして位置づけられている．5番目には，「最適化解析」の手法を置いているが，これらは現在，米国IBMのT.J.ワトソン研究所において主に，研究段階のものからパイロット・プロジェクトまで含め，様々な試みがなされている．図の横軸は，時間軸での発展段階を指すが，縦軸の高さはそれぞれの段階での機能の複雑性，または技術の新しさを表している．

一方で，この図の見方はもう1つの側面を示している，すなわち，各ステージの高さは，概ね分析技術の効果，すなわち「リフト」の順である．

ここで「リフト」の考え方に簡単に触れておこう．例えば，ある商品Aを買った顧客の中で，他の商品Bを買う確率が80％であったとしよう．これを，相関関係分析では，確信度80％といっている．一方で，商品Bを買うことを前提としない時，すなわち全顧客が平均的に商品Bを買う確率が40％であったとすると，上記の商品Aを購入した場合というのは，無条件の場合に比べ2倍もの確率で商品Bを買っていることが分かる．この2倍が，リフトである．このリフト値で，ビジネス・インテリジェンスの各5段階を評価してみよう．

まず，「データ検索」による分析を，上記に示したA商品とB商品の場合で考えてみよう．例えば，商品Bを売るためのクロス・セリング業務において，特に基準を定めずに販売促進策を講じると確率40％でしかないが，商品Aを購入した顧客を対象にすると購入確率が倍の80％に高まるので，その顧客リストを「データ検索」の機能により抽出し，ターゲットとして施策を講じたとしよう．ここで抜き出された顧客が

やまはた　ひろし　日本アイビーエム㈱
ビジネス・インテリジェンス事業推進部
〒103-8510　東京都中央区日本橋箱崎町19-21

© 日本オペレーションズ・リサーチ学会．無断複写・複製・転載を禁ず．

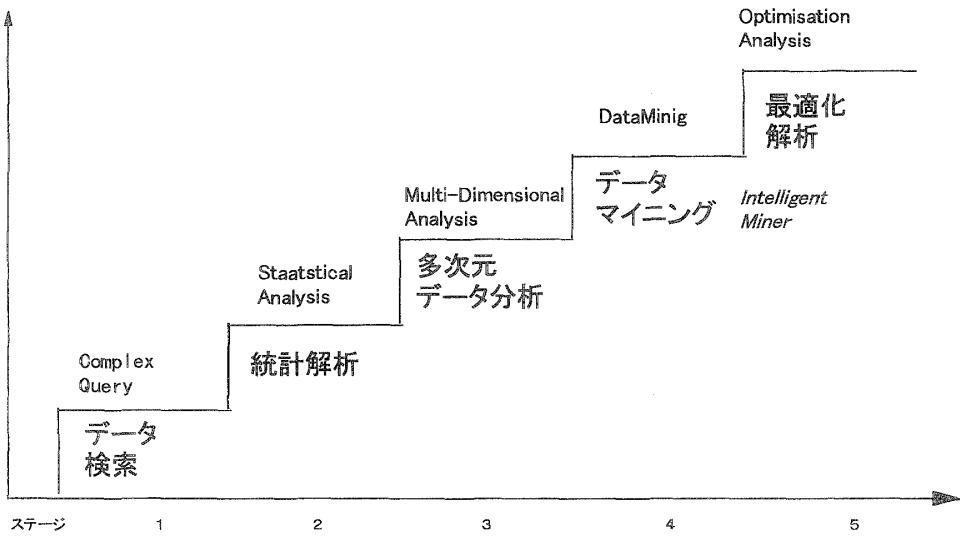

図2-1　ビジネス・インテリジェンス機能の5段階

1万件だとすると，その場合の効果は，図2-2のリフト・チャートに示すようになる．

また，「多次元データ分析」の場合は，上記の単純検索の場合に比べて，より多くの角度から順次視点を変えて分析しているため，特定の商品に対するクロス・セリングを考えた場合，さらに購入確率すなわち「リフト値」が向上する．この例を，一番目の条件を商品Aを買うことでリフトが2倍になり，年齢を20代に絞ることでさらに確率が3倍になったとすると，図のリフト・チャートでは，真ん中の折れ線で表されることになる．

これに対し，「データマイニング」の場合は，より多くの次元（＝項目）を同時点で総合的に分析し，最も高い確率で商品Aが買われるための条件と，その確率を提示することができる．例えば，商品Bを購入する確率の最も高いお客さまの条件は，商品Aの購入履歴があり，年齢20代の女性で，過去3年以内にユーザー会の会員になり，その間の月平均購入額が10万円以上のお客さまである……というように，極めて複雑な条件をも探しだすことができる．しかも，それは与えられたデータを使用して最も効率よくその条件を“自動的に発見”してくれている．したがって，もし，「データマイニング」によって顕著な条件や結果が得られなかったとしても，決して悲観することはない．それはあくまでも，その時点で得られるすべての情報を利用して得られる，最もリフト値の高い結果であり，また，データ空間の内部構造上，もともといかなる強い相関も存在しなかったという事実が，短時間にしかも効率よく分かったのだから．もし，この事実に到達するために，「多次元データ分析」表により試行錯誤を繰り返したとしたら，同じ“相関なし”という結論に到達するために膨大な時間を浪費したことになるだろう．この意味でも，データ・マイニングの与えてくれる恩恵は大きい．

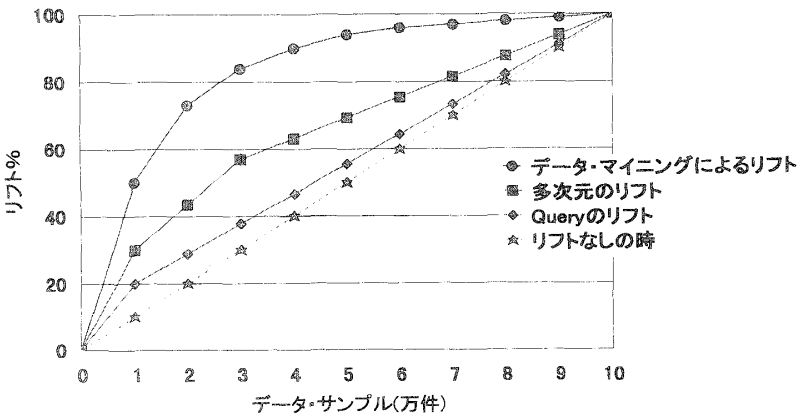

図2-2　分析効果のリフト・チャート

3．データマイニング・ビジネスの実際

データマイニングという言葉が世に出てから，もうかれこれ5年は経っているように思われる．そんな中で，現在その実状はどれほど変化し，また，進化してきただろうか？　最近，データマイニングの分析担当者の間で，顧客属性データとともに顧客の履歴データにより関心が移る傾向が感じられる．例えば「40代の女性が，20代の若い女性向けの衣料品を購入したならば，その方は当社にとって20代の女性というカテゴリーにセグメントされたターゲット顧客である．」という言葉の意味である．ここでいえることは，極論すれば，その女性が自分自身のためにその衣料品を購入したのか，誰か他人のために買ってあげたのかは，どうでもいいことである．重要なことは，そのようなお客さまは，今後もそのような商品を買い続ける可能性が極めて高い，ということなのである．

これは，まさに履歴データの本質を物語っており，顧客の購買行動とはまさしく

© 日本オペレーションズ・リサーチ学会．無断複写・複製・転載を禁ず．

Behavior であり，時系列データにしかその行動パターンは現れない．「市場と顧客の動向をみるなら履歴データを使え」ということである．このことはかなり以前から理解されてきたことであるが，近年この傾向は益々強まっている．

さて，データマイニングは，当初よりこのような見方をしていただろうか？　ここで，データマイニングの適用パターンがどのように推移してきたかを概観してみよう．

当初は，単純なマーケット・バスケット分布から始まり，また，OLAP（多次元分析）の使用に限界を感じた時，複雑なデータ構造から，モデルとしての判別ルールを明確に導出してくれる決定木などは使いやすく高度な手法として重宝がられてきた．しかし，データマイニングが，その性能上最も力を発揮するのは時系列データに遭遇した時であり，しかもデータマイニングを従来の手法と明確に区分するニューラル・ネットワークを使用した場合が多い．

1つの例を上げてみよう．ある米国のカード会社では顧客のカードの利用パターンからその顧客が“いつ”支払不能に陥るかを予測することに成功したが，その時は時系列データに Unsupervised Learning のニューラル・ネットワークを適用して支払不能に陥るパターンを発見した．結果として予測を行うことができたわけだが，データマイニングではその手法の特徴をよく理解して適用することにより，従来の限界を容易にブレーク・スルーすることができる．表3-1に代表的な「ニューラル・ネットワーク」の適用例を示しておく．

前記により，履歴データは何のバイアス（偏向）もかかっていない事実としての記録であり，貴重な情報であることが理解された．一方顧客が，最も直接的・明示的なメッセージとして企業にフィード・バックする情報として，テキスト（文章）がある．しかも，このテキスト・メッセージは，リアルタイムであり，時間の鮮度という観点からすると最も価値の高い情報であるにも拘わらず，今まで未開の宝庫として手付かずで放置されていた．ここに光を当て，新鮮な“顧客の意思”を明示的な形で浮き上がらせようというのがテキストマイニングである．テキストマイニングは，消費者が企業に向け直接発信したメッセージを分析し，その中に隠された要約情報や本質を探り出すための手法であり，すでに実用化に向けていくつかのパイロット・ケースが走り始めている（図3-2）．

この場合，メッセージはあらゆる顧客窓口を通じ，

表3-1　商用ニューラルネットワークのアプリケーション

アプリケーション	産業分野	機能
データベース・マーケティング	全産業	クラスタリング、クラス分類、モデル化
カスタマーリレーション管理	全産業	同情
不正検出	金融、保険、医療	クラス分類、モデル
光学文字認識	金融、小売	クラス分類
手書き文字認識	コンピュータ、金融	クラスタリング、クラス分類
売上予測、在庫管理	製造、卸小売、流通	クラスタリング、時系列予測
株式ポートフォリオ	金融	クラス分類、時系列予測
破産予測	金融	モデル化
工程スケジュール	製造/装置	最適化
プロセス制御	製造/金融	モデル化
債権格付	金融	クラス分類
担保引き受け	金融	モデル化、時系列予測
鉱物埋蔵調査	エネルギー	クラスタリング、クラス分類
医療診断	医療	クラス分類、モデル化
電力需要予測	ユーティリティ/製造	時系列予測
コンピューターウィルス探知	コンピュータ	クラス分類
音声認識	コンピュータ	クラスタリング、クラス分類
市場価格推定	不動産、金融	モデル化

© 日本オペレーションズ・リサーチ学会．無断複写・複製・転載を禁ず．

インバウンドで入ってくることになる．この場合の仕組みを顧客とのコミュニケーションを双方向として体系化し，発展させたコンセプトがCRMである．

4．CRM（カスタマー・リレーションシップ・マネジメント）

全米の3大銀行の1つでは，カスタマー・センターからのテキスト情報などをはじめ，顧客を直接サポートするエンゲージメント・マネジャーからの契約情報を含め，関連情報を分析することにより20%のKey Issuesが全体の80%の問題を解決できることをすでに発見している．(数値は意図的に概略である．)これはカスタマー・アクセスから分析系までのすべてを1つのコンセプトのもとに体系化することによりはじめて可能になったといえる．この意味において今後の情報技術投資の主要な流れとなるのがCRMである．図4-1をご覧いただこう．今後は様々な価値観を持った消費者がインターネットやキオスク端末などの多様なアクセス手段を用いて企業とコミュニケーションを持つことになる．CRMはまさにこのような環境における激しい顧客獲得競争の状況下で競合上の優位性を保ち，差別化をはかるためのまったく新たなアプローチである．

そして，ここでの中心的な役割を演じることになるのが「カスタマー・センター」である．これはインフラとしてコール・センター機能を有し，顧客満足度の最前線ともいうべき，より広範囲な使命を持ったサービス・センター機能である．この「カスタマー・センター」は，個々のお客様の顔が見えるような仕組みを必要としているため，CRMの様々なサブ・ファンクションと繋がっているが，その1つに，「キャンペーン・マネジメント・システム」というコンポーネントがある．これは提案型マーケティングを強く意識し，お客さまに対して常に最適な形でのオファリングができるよう，また，お客さまからのリクエストに対しては，個々のお客さまの特性に合わせた対応ができるようにデザインされたシステムである．つまり，それぞれのキャンペーンに適した顧客リストや，お客さまに提案すべき最適な商品・サービスの情報，およびメディアやチャンネルなどが特定でき，HOW TO APPROACHにおけるワーク・フロー・ロジックをもコントロールすることが可能である．

これらはデータマイニングなど，「ビジネス・インテリジェンス」上のアプリケーションが，顧客データ・ベース上のデモグラフィック・データやテキスト情報を含む履歴データを分析し，結果をキャンペーン・ロジックに反映することにより可能となる．また，

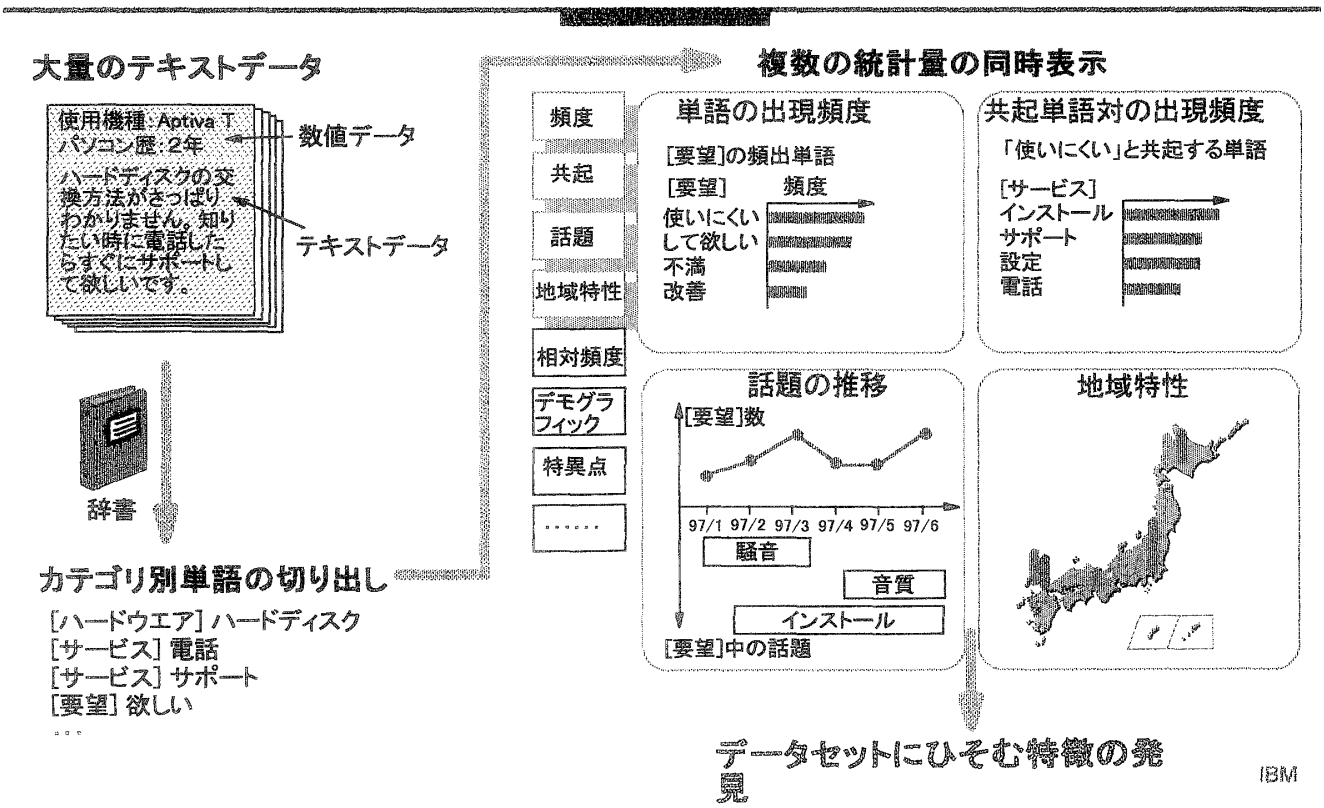

図3-2　テキストマイニングの概要

© 日本オペレーションズ・リサーチ学会．無断複写・複製・転載を禁ず．

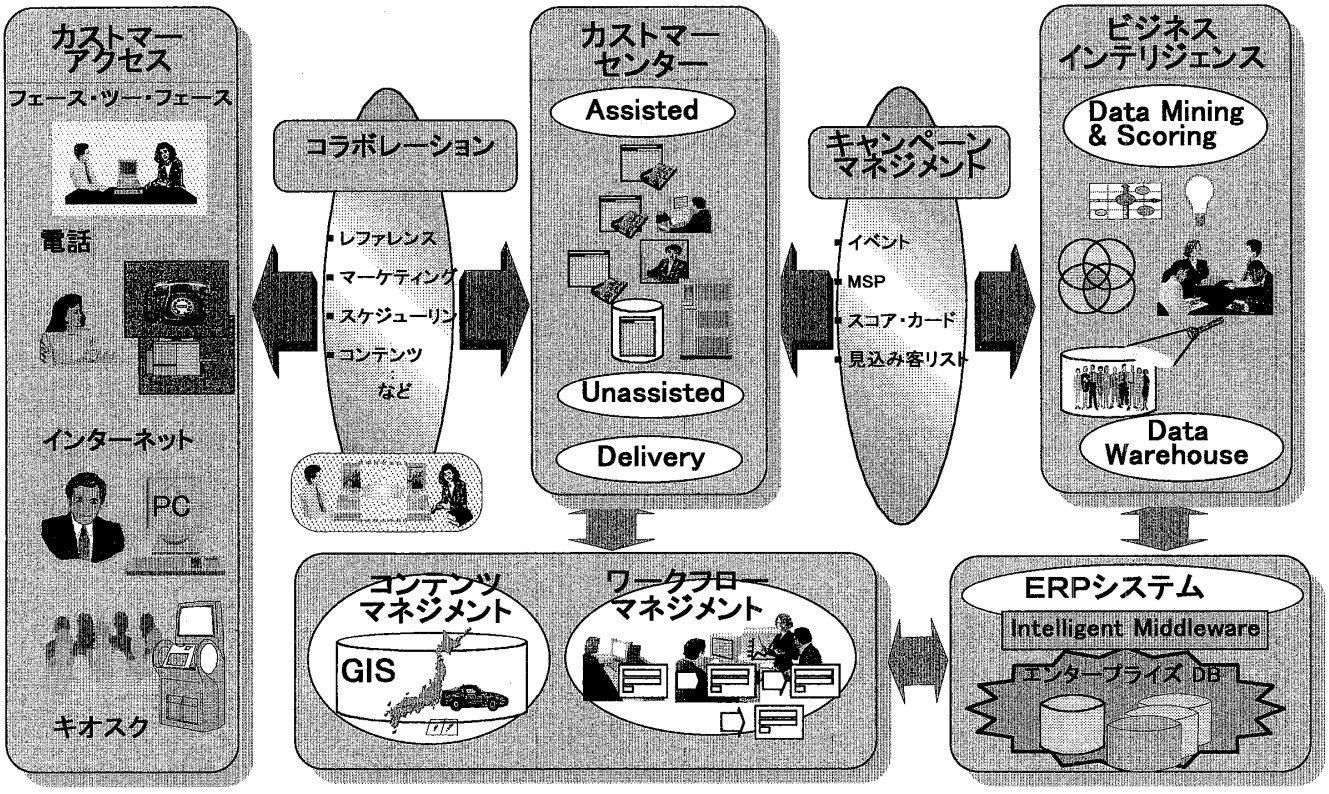

図4-1　カスタマー・リレーションシップ・マネージメント（CRM）ソリューション

CRMには「コラボレーション」というコンポーネントがある．これは，カスタマー・アクセスをサイバー空間で統合的に管理し，カスタマー・センターの自動化をはかり，負荷を軽減することができる機能である．ここから分かることは，これらの先端的システムのすべての機能が相互に有機的に連繋して働き，はじめて全体として体系的な1つのソリューションを形成することができる．これがCRMである．

キャンペーン・マネジメント・システムは，製品としては，Campaign Directorとしてすでに実稼働しているソリューションであり，またその他のサブ・コンポーネントも，すでに開発済み，もしくは開発中のものである．その意味で，CRMは，これからの顧客志向マーケティングを担う，最も新しく先端的なコンセプトであるが，実際上のビジネスにおいて，すでに現実のものとなっていることも忘れてはならない．

5．おわりに

以上，データマイニングをテーマの入り口とし，ビジネス・インテリジェンスと，さらにそれを活用したCRMというより大きなコンセプトの概要に触れることができた．そこでは，データマイニングは，顧客志向経営という大きな視点から，ビジネス・インテリジェンスやCRMという上位のコンセプトを構成するキー・テクノロジーの1つである．

一方では，データマイニング自身は，無限の可能性を持った情報技術の一分野である．表3-1にあるように，その応用すべき対象は製造や通信などの分野も含めて多岐にわたっている．例えば，IBMのネット・マイニングは，インターネットのWebログを解析し，その参照パターンにより，利用者のセグメンテーションを行ったり，利用者に合せてホームページの構成や表示順をダイナミックに変えられるようにする分析サービスである．このように，その時々の焦点の当て方により様々な使われ方をするが，今まで想像もできなかったような，まったく新しい問題領域への応用も考えられる．個々のテクノロジーの知識と併行して，常に柔軟に発想するBI（ビジネス・インテリジェンス）リテラシーが求められる所以である．

今後，益々当分野が発展することを念じて止まない．

参考文献

[1] ジョゼフ・P・ビーガス著，㈱社会調査研究所・日本アイビーエム㈱共訳（1997），
「ニューラル・ネットワークによるデータマイニング」
日経BP社

[2] Rock Angier, Kurt Kovacic, Dana Morris, Tom Piercce, Jim Quinn, Mike Reyman, Kathy Smith (1998, Sep),
「Customer Relationship Management」
White Paper IBM Global Services

© 日本オペレーションズ・リサーチ学会．無断複写・複製・転載を禁ず．