

# ユビキタスなデータマイニング 流通業を中心に

専修大学ネットワーク情報学部  
江原 淳

## センサネットワーク 処理のクラウド化 ?

GDS(商品コード)やRFidやGPSなどセンシングデータがネットワークで提供されその処理が容易になりつつある。

そのようなデータからどう意思決定指標を出し環境適応につなげていくか

マイニングしながら新たな提案やサービスを生み出していくのが企業

他社開発BIツールだけでなく、自社ドメインでは行動ターゲティングの独自開発指標で競争優位を



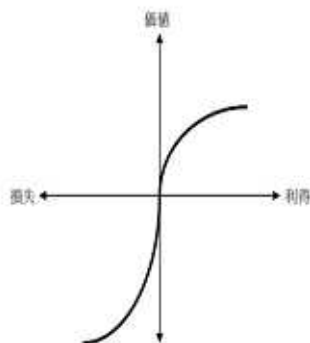
## 流通業でのデータマイニング

- オペレーションでは、マイクロマーケティングでのビジネスルールの更新
  - 顧客別需要調整 (FRM, CRM)、売場別需要調整、
  - カテゴリ別需要調整 (カテゴリーマネジメント)、SKU別需要調整、SP別需要調整、アカウント別需要調整、地域別需要調整
- 調達では、SCM最適化ルール (店舗レベルでも発注・在庫・品切れ等についてのルールの発見と更新)、マーチャンダイジング (MD) ルール、消費者行動の変化の発見など



## ポイントから個別クーポンへ

- ポイントは+の増大
- ポイントが二倍になっても、心理的価値は1.5倍?
- 100円の損は50円の損より3倍くらい? 回避したい
- 期間限定クーポンは使用率も高く額面以上の効果がある



## 顧客トランザクションデータを 行動ターゲティングデータに

- Webなら閲覧履歴で提示広告を最適化
- 検索語に連動した広告とリンク
- 協調フィルタリングによるリコメンデーション
- ID-POSの購買履歴に応じたクーポンプロモーション
- 個人別推薦アイテム・個人別価格
- 商品探索行動に応じた情報提供・サイネージ

## クーポンプロモーションの変化

時期	名称	内容	目的	データ
1966 -	グリーンスタンプ	クーポン	反復購買の促進	なし
1981 -	FFP	有休資産活用	顧客固定化	行動履歴
1985 -	FSP	ポイント	LTV増大	購買履歴
1997 -	コアリション	共通ポイント	効率・効果	複数企業での履歴
2010 -	?	?	?	ライフログ

センシングデータから、どのようなアクチュエータを設計するのか？



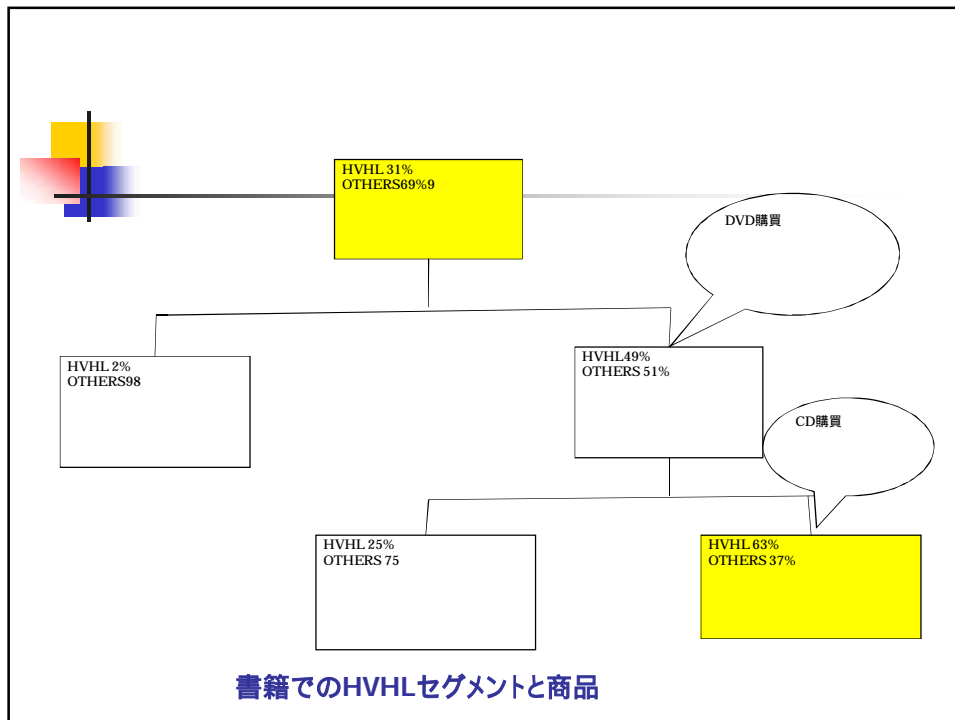
## モメンタムモデルでの顧客区分

HIGH LOYALTY	1	1	1	1		
(LOYAL)	0	1	1	1		
	1	0	1	1		
	1	1	0	1		
MEDIUM LOYALTY	0	0	1	1		
(ACTIVE)	0	1	0	1		
	1	0	0	1		
	1	1	1	0		
LOW LOYALTY	0	1	1	0		
(PASSIVE)	1	0	1	0	0	0
	1	1	0	0	0	1
	1	1	0	0	0	0
NEWCUSTOMER	1	0	0	0	0	1
NO SPENDER	0	0	0	0	0	0



## カスタマージャーニー

- HVHL LVHL HVLL LVLL HV&LVINFREQUENT NEW NOSPENDERS セグメント
  - ~3M 3~6M 6-9M 9-12M 12-18M 18-24M 24M~ リセンシー
  - MAINTAIN NEXTPURCHASE WINBACK 目的
  - 月次 隔月 四半期 半年 コンタクト
  - カタログ 手紙 葉書 eメール レジクーポン 手段
  - 何割引 フリーギフト カタログ 何もしない オファー
  - いくつ買ったら 期間限定 金額限定 基準
- これまで、金額管理レベルさえまともに実行されてこなかった
  - 単品管理レベルの顧客管理へ



## 流通業での事例1, 2

- レンタルビデオでのリコメンデーションメール
- 主成分分析による方法とアソシエーションによる方法
- 無意味なルールの除去がポイント
- ・既存ビジネスルールの見直し 関連陳列での神話等
- ・行事ではバーチャルアイテムの利用

スーパーマーケットでの顧客デシル上昇を意識したMD(マーチャンダイジング)・SP(販売促進)のための判別ツリー  
 成果変数でなく、原因となる変数で行う

FOCUS KPI	CRM VIEW	PRODUST VIEW
	CUSTOMER	PRODUCT
プライシング	一人当たりコスト ロイヤルティ 価格に敏感な顧客の比率 バスケットサイズ 購買部門数	一個あたりコスト 売上・利益 売上数量 カテゴリー売上 合計売上 製品売上 カテゴリーリフト
プロモーション	トライアル 既存客のストック ブランドスイッチ リピート	
新製品・品揃え	リピート購買 代替品 レパートリー伸び率 ブランドロイヤルティ	売上げ伸び率 スペース配分 売上げ回数
新業態・顧客サービス	キーカスタマーのニーズ キーカスタマーの使い分け (ストアレパートリー) 行動の変化を引出す	店舗売上げ スタッフ有効活用度

## 顧客のLTV(顧客生涯価値)を高めるには

新規顧客の獲得

既存顧客の維持率向上

既存顧客の購買回数の向上

既存顧客の購買単価の向上

~ を購買商品・購買パターンで説明してみる

## データマイニング手法

ニューラルネットワーク

決定木 ( $\chi^2$  によるCHAID, ジニ係数による  
CART, エントロピーによるC5.0等)

アソシエーションルール・ショッピングバスケット分析

記述的手法 記憶ベース推論, ベイジアンネットワーク

k-means法等のクラスタ分析, コホートネットワーク

リンク分析

重要サンプルの利用 SVM, GA

テキストマイニングの諸手法

多数決手法 ブースティング・バギング

## 小売業の情報化対応

- 調達システムの見直し(グローバル化と新PB)
- 物流・販売システムの最適化(SCM・CPFR)
- 販促のシステム武装 FSPとCRM
- SP 人(誰)、商品(何)、購買行動(how)
- メーカー参画の必然性と協同SPへ

データ/成果	金額	単品
レシートデータ	店舗SP	メーカーSP
購買履歴	FSP	メーカーFSP
他人の履歴	カテマネ	リコメンテーション

## リテンション手段

- データ統合
- FSP
- FRMと階層別管理
- LTVと目標の多元化
- OLAPによる迅速化
- マイニングによる高度化
- 顧客インタラクションのシステム化(ネットコミュニティ等)
- KMとデータ基盤(DWHに参照データとその履歴も)

### 生涯価値LTVの例

	初年度	2年度目	3年度目	4年度目
既存顧客	1000	850	893	929
A新規顧客		200	200	200
B維持率	0.85	0.85	0.85	
C購買頻度	1.01	1.03	1.04	1.05
D客当り単価	50	52	54	56
売上げ	50500	56238	61383	66385
利益(25%と仮定)	12625	14060	15346	16596
割引率(5%と仮定)	1	1.05	1.1025	1.157625
利益の現在価値	12625	13390	13919	14336

## W.H.インモン「コーポレートインフォメーションファクトリ」、海文堂、1999

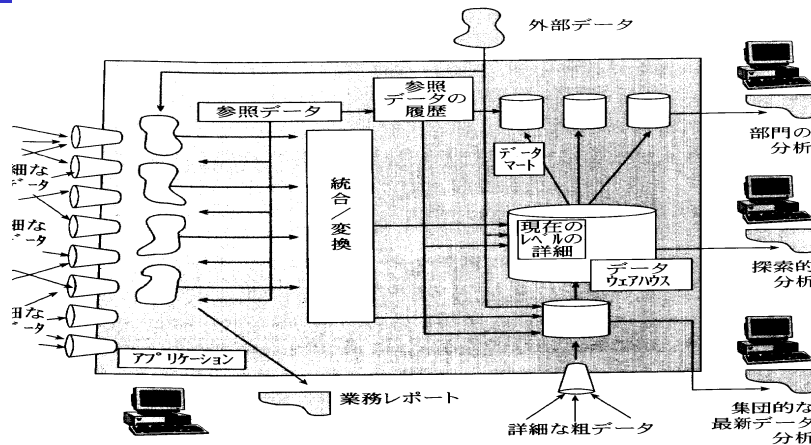


図 2.1 CIF の基本構造。



## コーポレートインフォメーションファクトリ

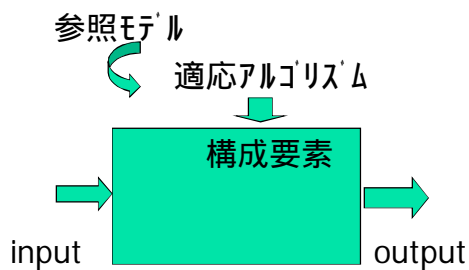
- 情報循環系としての企業
- DWHが基盤
- 部門の情報系とデータマート
- WebとDWHのリンクではODSが不可欠
- 事業ドメインの領域モデル・企業のロジカルデータモデル・DWH・データマート・ODS等々を通じたeビジネスと企業データの統合能力

## トランザクションデータを次のトランザクション(取引)の生成に活用する

- DBメンテはコスト、トランは副産物。トランの外部性を内部化する！データ販売よりマイニングを。
  - サマらないこと。水道料金 = 家族人数。この情報をすべての金融機関が捨てている。
  - ビジネスの方法が少しでも前進すればOK
- 真の原因や最適解である必要はない ビタミンC発見以前から英国海軍はライムを支給していた
- ・データから学べる適応的組織へ 要素調整！
- KMのキーとしての参照データ履歴
- DWHとユーザ部門の間にデータマート
- トランとDWHの間にODSオペレーショナルデータストアで履歴よりも直近のデータ統合

## 入力へのFBからルール生成・要素変更へ 情報循環系企業へ

入力調整・要素調整・モデル調整



- 営業システムや組織システムの出力から何をするか？
- 入力変更 (FB制御)
- 要素変更 (適応制御)
- モデル変更 (自己組織化)
- ビジネスルールの大半は適応アルゴリズム

## 数量・価格のFB制御から適応制御へ

- チェーンストアは、「調達」での規模の経済性、「オペレーション」での範囲の経済性、情報での集積の経済性の組み合わせ
- 調達システム、オペレーションシステム、組織システムの3つが必要
- 日本では卸売り業が多いのでそこに依存して調達システムが弱い小売りチェーンが多い。普通はSMはPB業態。DSがNB業態。
- マイニングすれば組織システムに依存したIYと同等に
- 米では、トイザらスをウォルマートがマイニングで圧倒

## データマイニングはビジネスルールを データで更新すること

Time flies like an arrow.時蠅は矢を好むか？ 矢のような蠅を測れるか？

- リサーチ目的でない**大量データ**を扱う

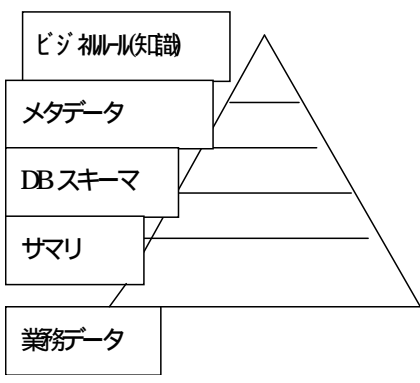
多変量解析でなくマイニング手法の組み合わせ

- **複雑なパターンを発見**する・一つの手法では限界  
多数決手法？の進展

統計処理でなくRDBとマイニングツールと人間の問題

- ・多数の試行錯誤・変数変換・視覚化
- ・目的は**ビジネスルール** 統計でなく**業務の専門家**が専任で

## ビジネスルールがなければ活用できない ルールやそのためのメタデータもDB化



•たとえば、量販営業なら、提案ノウハウ・陳列ノウハウなどがビジネスルール

•そのために、価格弾力性・陳列弾力性・商品間交互作用・販促の長期的効果などがメタデータ

•商品データベース・商品属性データベース・画像データベース・店舗データベース(フラットに)・商品分類

•サマリは各種データマートへも

•トランは捨てないでDWH化

•クライアントのビジネスルール・メタデータ(参照データ)を知る必要がある。クレーム内容のパターンとか、はその後の結果とか、何のためにどういう指標を求めてデータに問い掛けているのが、問題。

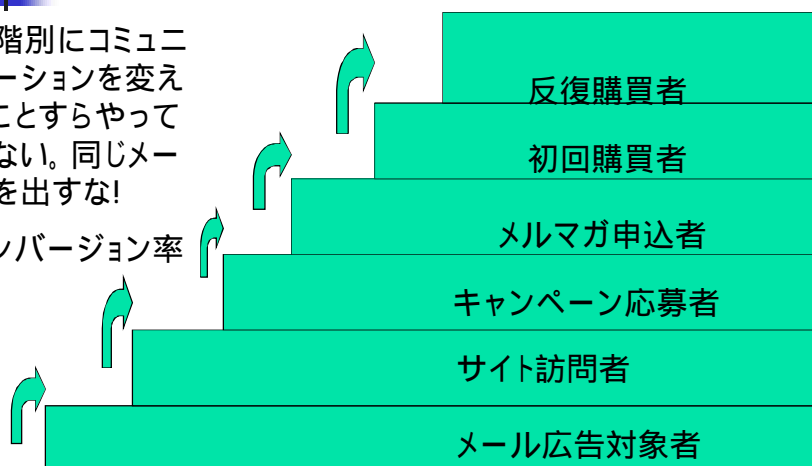
## CRMの深化 流通業も行動ターゲティングへ

- 既存メディアとID - POSの融合(TEL No.)
- Webログと問い合わせデータ(メールアドレス)
- オプトインメール・ワード広告・リコメンデーション
- トランザクションによる市場細分化が可能に！
- モバイルコミュニケーションへの始動(携帯へのオプトインやリアルタイムのオファー)
- アクセスログ・購買履歴による細分化

## ECもまともなマーケティングに...

段階別にコミュニケーションを変え  
ることすらやって  
いない。同じメールを出すな!

コンバージョン率



## ネット利用のマーケティング

- B to Bのマーケットプレイス……最適調達と外部I/fの標準化
- B to C の手段の多様化  
デジタル化によるコミュニケーション統合(CTIの次は...)  
リコメンメール・オプトインメール・ターゲットメール・そのBB化  
携帯オプトインの開始・リアルタイム調査等  
カスタマイズとエージェント・自己の知識と履歴活用リコメン  
ション・他人の知識と履歴活用
- Cインタラクション 知識どうしの統合プロセス  
コミュニティのテキストマイニング等

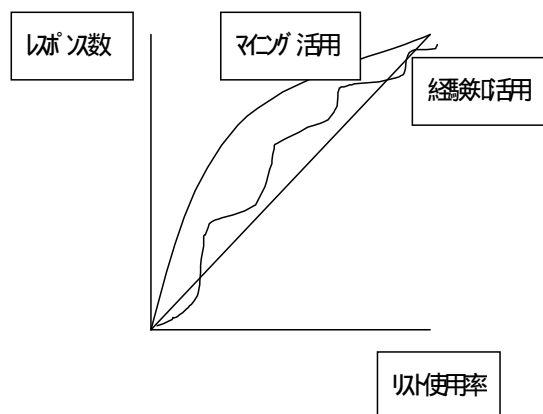
## パッケージ商品のデータ利用 適応ノウハウをマイニングで生成

マーケティング環境	特徴	内容	データ
販売組織形成期	市場対応	市場代表性	チャネル調査・消費者パネル調査
再編成	競争対応	エリアマーケティング	地域別・アカウント別実績
業態内効率競争	個店対応	売場管理・機能代替	個店別・フェス・日付調査
製販連携	コストと効果	カテゴリーマネジメント	取引先別マーケティングデータ

## インスタマーチャンドライジングに加えて インスタマーケティング

- 店頭でのサービス付加 位置・価格・POP
- 視線分析とレイアウト・陳列・パッケージ
- プロトコルなどスキーマ研究と関連陳列
- 店内動線調査の自動化とレイアウト
- 購買履歴に応じたサイネージ
- 購買履歴に応じたレシートクーポン
- 購買履歴に応じた個別価格
- カスタマイズされた販売促進(FSP)とリコメンデーション
- 大量陳列・価格・商品の組み合わせの最適化

データマイニングのコストは業務の高度化から産み  
出す必要 知識自体が価値あるわけではない





## マイニングできる条件

- DWHがありトラン収集できている
- 共有されたビジネス課題があり、新たなビジネスルールが求められている
- マイニングツールがあること
- 兼務でなく、専任である事
- 得られたルールの財務上の価値を評価・社内PRできること



## マイニングの実際

### データクリーニング

- データがあってもマイニングできない  
欠損値・異常値・矛盾・・・
- ルール上無意味(2,8月の残高変化・休眠顧客・高齢者  
意所発生時の記録はあるが正常時のデータがないことも(新規収集)

### 分類

- 従属変数の決定・作成・二値化と変数選択・・・
- 順序統計量利用によるカテゴライズとAIC
- ルール化が主目的なので、次元縮小的手法は最初だけ(全体的な configuration よりもマイ知な構造の方が問題、変数間非独立)

### 判別

- ビジネスルールは言語表現されるので、結局判別問題  
(MBR, 決定木, アソシエーションルール, NN・・・)

統計的有意性よりもマネジリアルな有意性



## 使えるデータマイニング手法

- 他のDBMSとの統合
- ユーザとの対話・相互作用
- ユーザによる知識のメンテナンス可能
- 結果の理解可能性・伝達可能性
- ノイズ・欠損値等の処理
- 統計的有意性の評価
- ...Understandability of Patterns



## OLAPからマイニングによるBPRへ

- OLAPによる成功事例→マイニングへ
- 英国自動車保険 ソフトハウスが1/3のシェア
- コンボスのRC化とSMDの消滅
- トイザラス(カテゴリキラー)よりウォルマート
- 消費者金融大手の判別ツリー与信管理
- アマゾンのデータマイニング利用と購買時点スキーマ(データ統合・顧客識別・CRMも)





## データフュージョン系

- フロー入出金・ATMデータからストックを予測・他行取引を予測
  - アンケート等とあわせてリスク選好・金融商品でのセグメンテーション
  - レアケース系
  
  - デフォルトの予測  
1-2%以下のケースの要因分析なので統計手法が異なる  
・ロンドンDMヒット率の予測  
モデルの複合化の勝負
- 通常の統計手法だけでは無力。問題は局所構造



## 判別ツリーの長所

- カテゴリカルデータと数量を同時に
- マイクロな構造なども、仮定をおかないで記述が可能
- ルールが求まり、既存知識と結合できる
- レア事象のためのルールも
- ツリーによるセグメントごとに他のモデルなど複合化して使える
- 説明しやすい
- ボイントは、専門的知見との接合

## ショッピングバスケット分析では

- $\Phi$ 係数、関連係数は簡単
- 定期的に出してバイヤーが眺めるだけ  
不足しているのはビジネスルール
- 目的は何なのか(売り場のインターフェイス設計・商品開発・情報提供ツール……)
- ポイントは、バーチャルアイテム設定  
(行事・時間帯・客層・ブランド・トレンド…)

## 手法より使いこなし

- ショッピングバスケット分析では……
- 目的に応じて、年末年始・連休・クリスマスなどのバーチャルアイテムをレシートに追加すること
- ゴミのルールと有益なルールの識別
- アイテム選定は目的と出現率で分類を上下させて RDBでないとダメ。この分析はSQLで書け
- サポート(ルール支持件数)率・信頼性(条件と結果の同時出現率 / 条件の出現率)・改善率= $p(\text{条件かつ結果})/p(\text{条件})p(\text{結果})$  ランダムより改善



## SFA,B2Bなど、意味空間限定で

- HNC社のシステムの場合、280次元のベクトルで共起性を表現、検索語による関連文書検索・文書分類・自動応答などへ
- 含まれている単語のベクトルを合成するだけで文書全体の合成ベクトルが求まる
- アンケートの自由記入の分析等よりも、企業内文書など意味空間が限定されているものに向いている → 営業報告書の成功例と失敗例



## テキストマイニングは社内文書から SFAでの文書分類・事例交換

- ・エリア管理・テリトリー管理
- ・取引先管理
- ・販売計画
- ・文書自動記入
- ・チーム営業支援
- ・参照・引用資料データベース
- ・生産性分析ツール・発送文書の一括処理
- ・電話での販売・営業サイクルの管理
- ・営業方法・予測



## 製造プロセス改善

- 輪転グラビア印刷でのバンディング発生  
非発生時のデータがない 収集から 決定木
- ・新ガイドライン(ゴム溶解度は高く・インク温度は低く・  
インク粘性は高く)で発生減(英国海軍がライムを  
水兵に配布したのはビタミンC発見前であった。)
- ・オフセット印刷でのロール紙の節約  
RDBはあったが、マイニング用のデータ統合課題  
作業ルールを専門家を入れて再作成(決定木)



## 携帯のチャーン

- 明日なら全員ロイヤル・100年なら全員  
チャーター(従属変数を作ること自体がルール  
生成) DMツールなしにはできない!!
- 目標のレベル分け  
短期: 解約予想者リストとアクション  
中期: チャーン管理手法確率(CMA)  
長期: CRM



## 小売業での利用例

- ロイヤルティカードからの顧客の視覚化  
k-meansクラスタ それを属性で説明する決定木
- ・他業態使い分けのクラスターと決定木  
(成果と原因の間に、顧客タイプを入れる)
- ・ターゲットマーケティングとFSPとの融合へ
- ・リコメンデーションシステムの開発



## マイニングプロセス自体のツール化

- 作業の大半はデータクリーニング・データ変換・データ選択である
- ツールを使わないとできない複雑なモデリング・モデルの組み合わせ
- それら手順のテンプレート化による共有！



## 他の情報処理方法との関係

- データベーススキーマ既知
  - 更新と定型クエリ OLTP(オンライントランザクション処理)
  - 無更新非定型クエリ
  - 構造的クエリ OLAP(オンライン分析処理)
  - 半構造的クエリ 多次元データベース
  - 非構造的クエリ データマイニング
- データベーススキーマ未知
  - 概念情報の定義 アウトライニング等
  - 概念情報間の構造化 テキストマイニング



## マイニングの応用領域

- 関係性の発見 バスケット分析・オファー修正等
- 分類 チャネルセグメンテーション・市場細分化・顧客リストへのアクション設計・リスク評価・品質管理
- モデル予測 価格設定・SP計画・スコアリング
- 構造の推定 広告計画・製品開発...
- KDD手法の前にOLAPで仮説を絞ること
- ポイントは、「ビジネスルールから」、手法からではない



## 流通チャネルとデータマイニング

- チャネル内情報・営業報告書等のDB化と知識発見・共有化、チャネルとしてのKM
  - 情報共有からマイニング能力、知識発見能力による新たなチャネルキャプテン？
  - C P F R 的な企業枠を超えた情報共有とKMでは、新たな中間媒介者が？
- (コアリッションでのクーポンCRM、データウェアハウス・データマートとOLAP・マイニングのASP、FSPや決済・与信のASPなど)



## メーカーとデータマイニング

- 製品開発での消費者の意思決定スキーマ情報、チャネル情報等の双方向性がビジネスルール水準で可能となる
- 苦情処理部門・広報等ではテキストマイニングが先行
- 品質管理のかなりの部分がデータマイニングへ
- 営業管理やSFAの基幹システムではOLAPが先行する テキストマイニングはツール開発次第か
- 販売促進や広告ではマイニングによる効率改善へ一挙にシフトする可能性がある。