

2006年7月28日

SAS Forum

ユーザー会 学術総会

リテール金融工学の発展

三菱東京UFJ銀行
小 野 潔

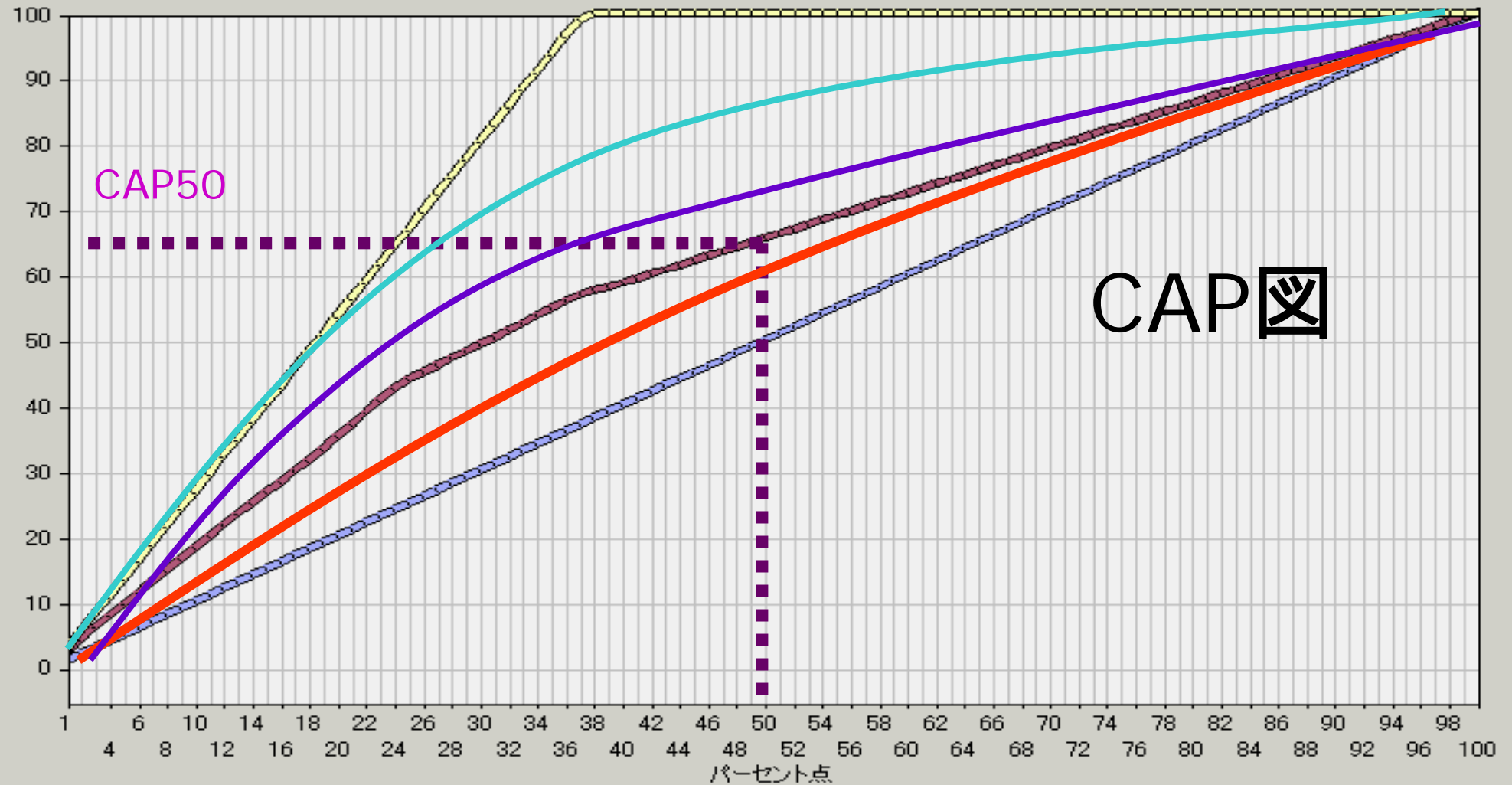
(本報告は個人的見解です)

金融業のデータマイニングの歩み

- 1992 米国IBMのレポートにデータマイニングが用いられる
- 1994 米銀Wells Fargoが中小企業向け与信モデルを運用
 - 複数のマイニングツールが販売される
- 1996 日本語マイニングツールの販売
- 1996～1999 先行する金融機関が開発
 - 住宅ローンモデル
 - 生命保険の解約分析(1998)
 - リスク細分型自動車保険(1999)
 - クレジットカード会社の不正利用検知モデル
- 2000～2004 一般の金融機関に拡大
 - CRMモデル
 - クレジット会社 & 消費者金融会社の与信モデル、都銀の中小企業の与信モデル
- 2004～現在 複雑な分野への適用
 - 新しい中小企業および小規模事業者のスモールビジネス用与信モデル
 - 不良債権回収モデル
 - 中古車の査定モデル

ビジネス戦略とモデル精度の関係

正反応捕捉割合



ツールの名前

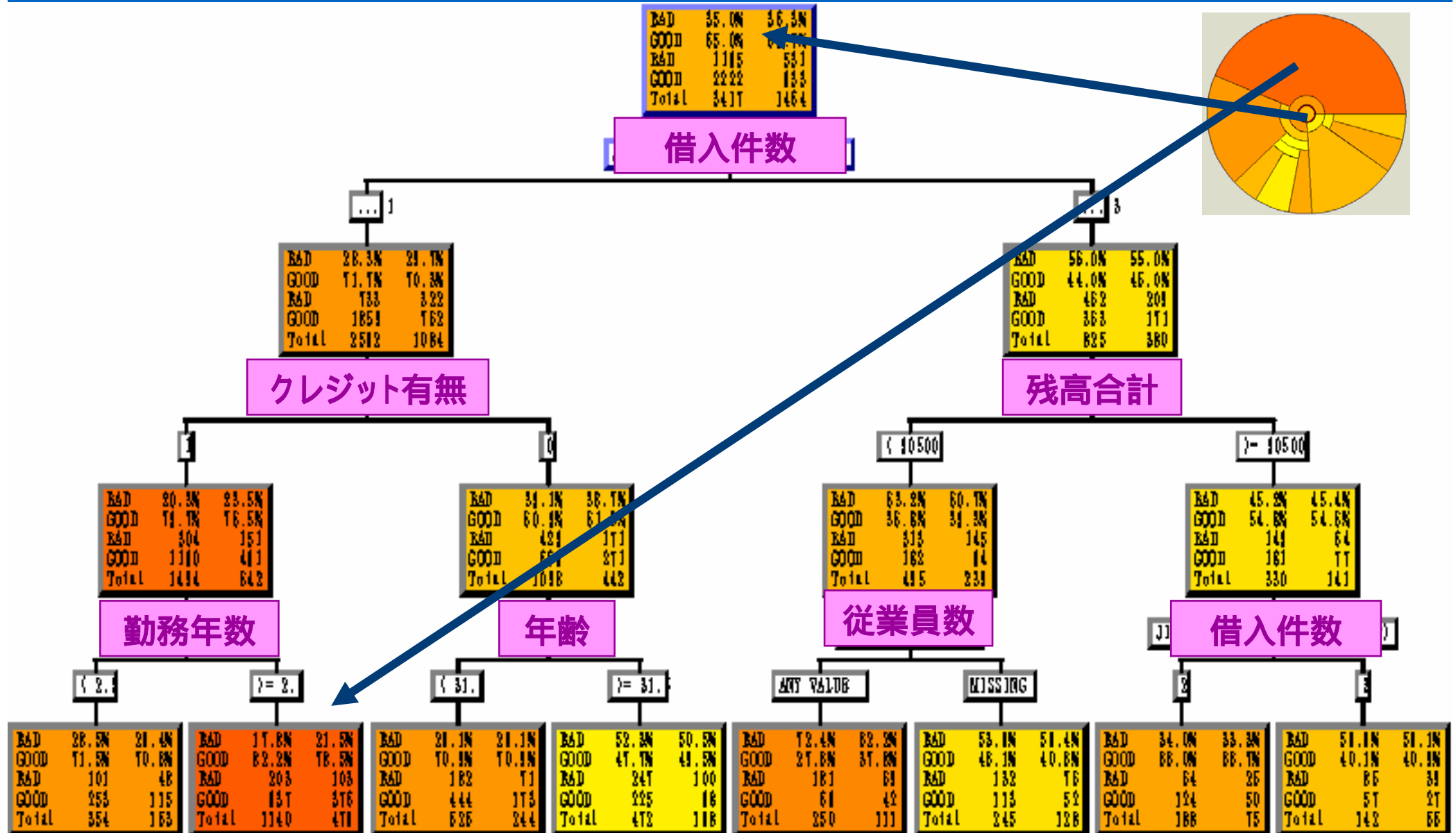
Baseline Tree Exact

収益の予測モデルの虚実

キャッシングモデルは収益でなく、デフォルトを推測する理由

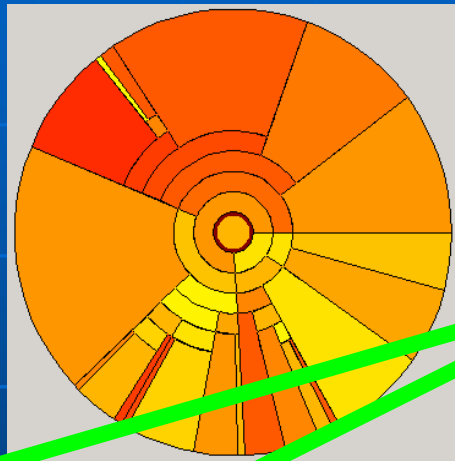
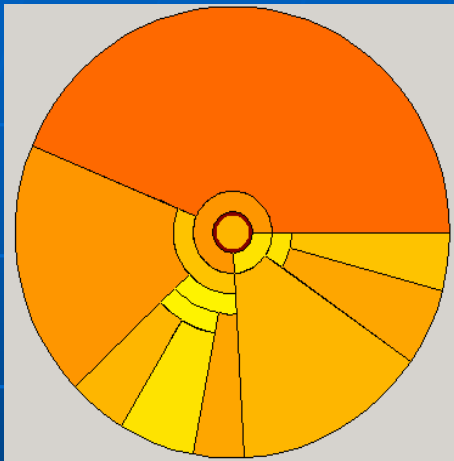
- Default Model vs Profit Model
- 目標変数: デフォルト / 正常 利益 / 損失
- 収益推測値: 事後計算 直接推測
- Profit Modelは日本では開発できない???
- 個人属性の変化を表わすリアルデータの入手不能

キャッシングのDefault Modelの構造

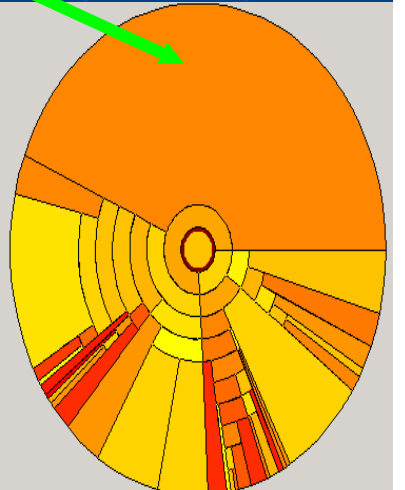
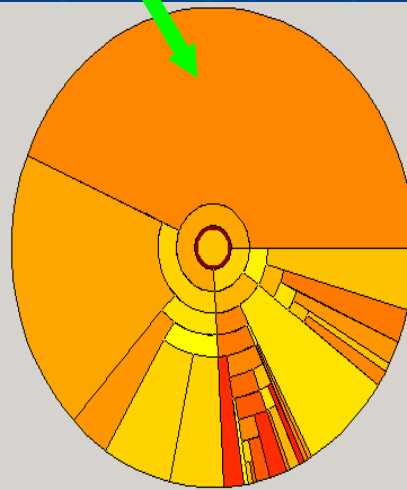
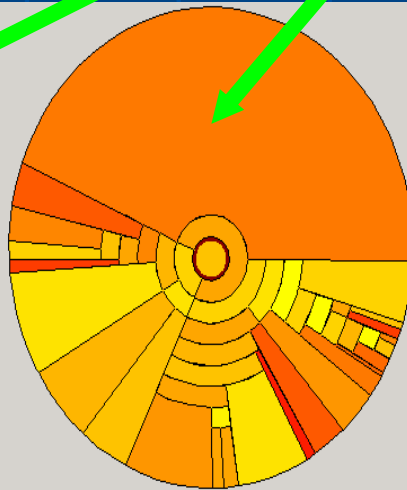
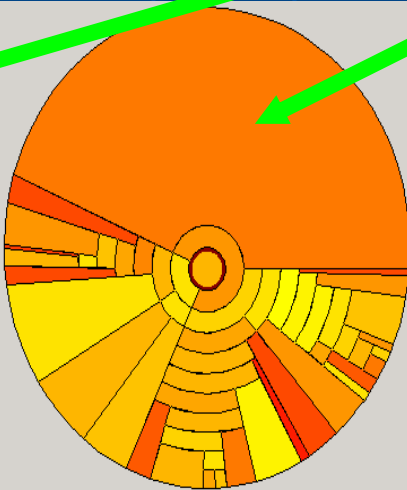
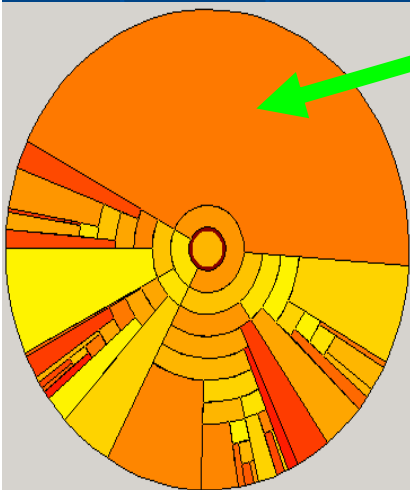


Default Model(上図 最終枚数 8枚、21枚)

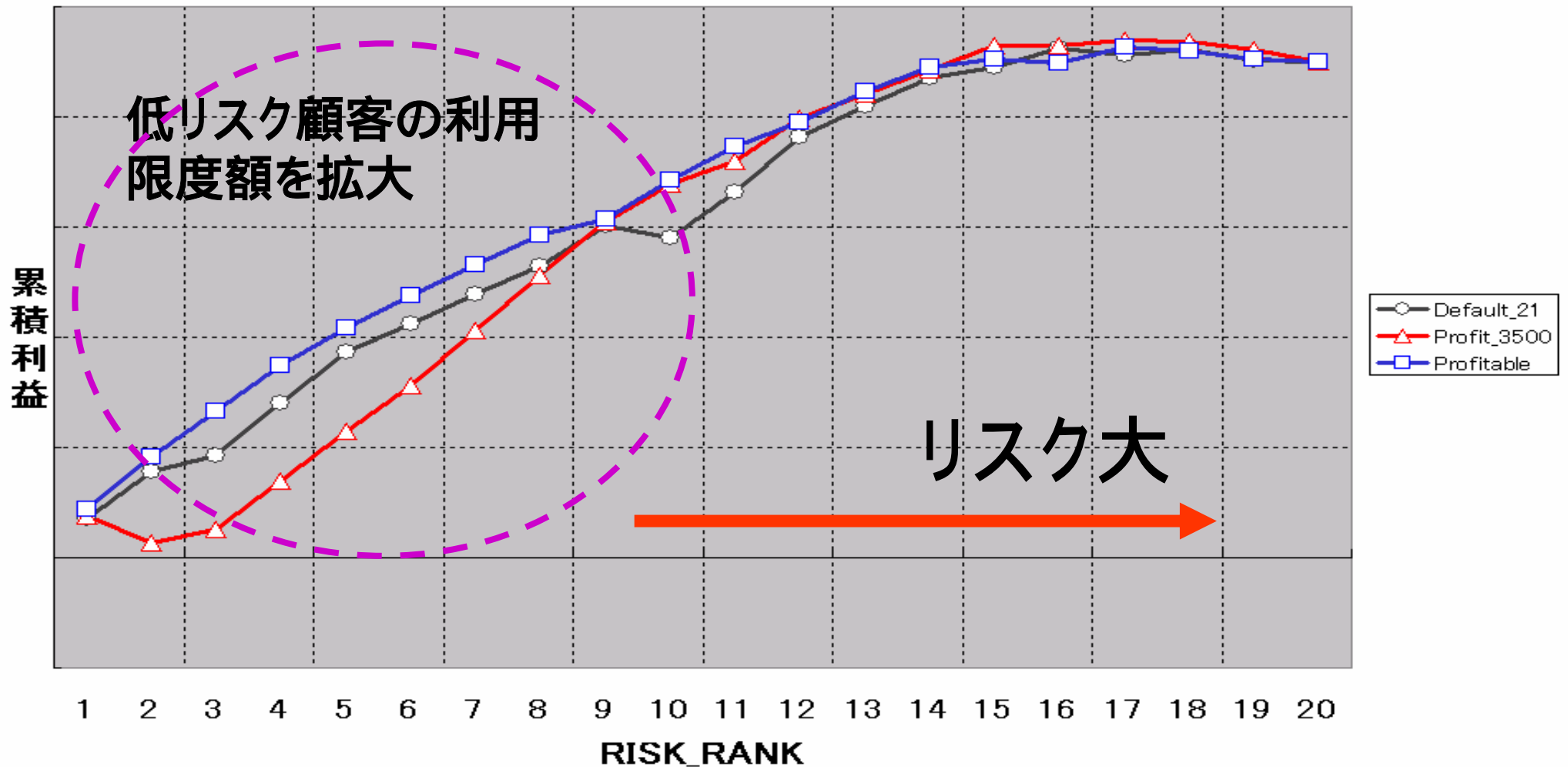
Profit Model(下図 閾値1,3500,6000,8000,1万円)



分割できない顧客は
どんな人達？



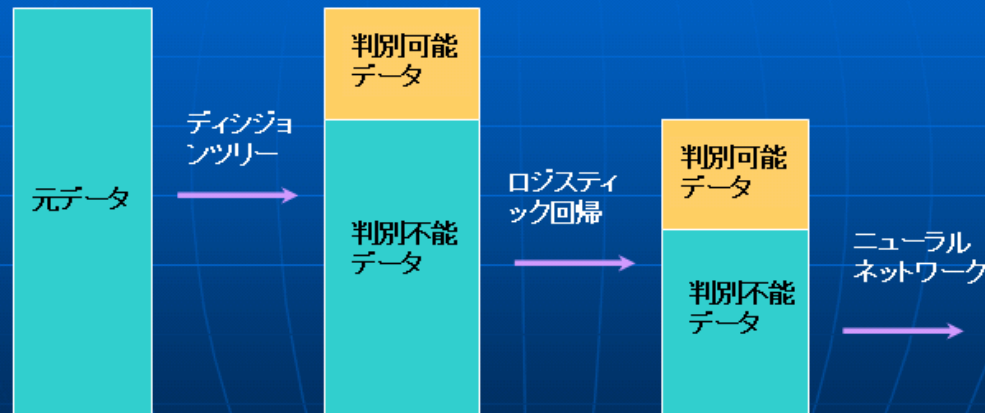
Profitable(青)/Profit(赤)/Default(黒) 累積収益チャート



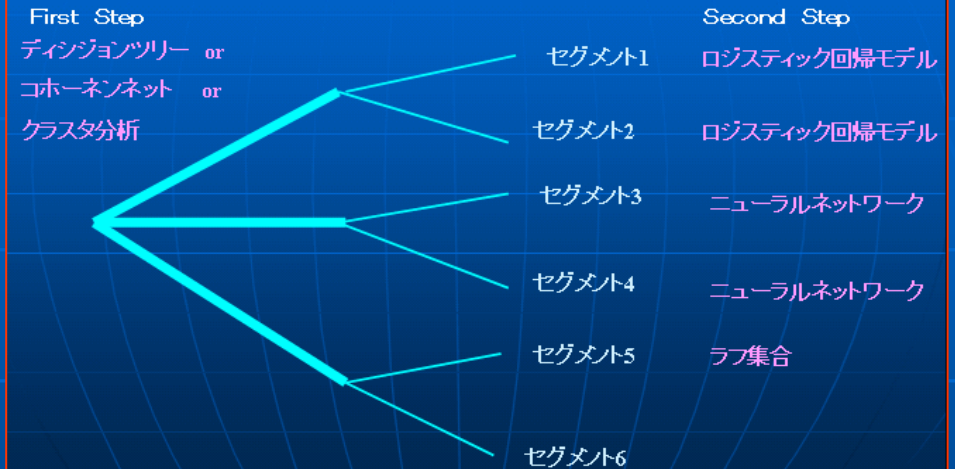
ハイブリッドモデルの真実

ハイブリッドモデルの例

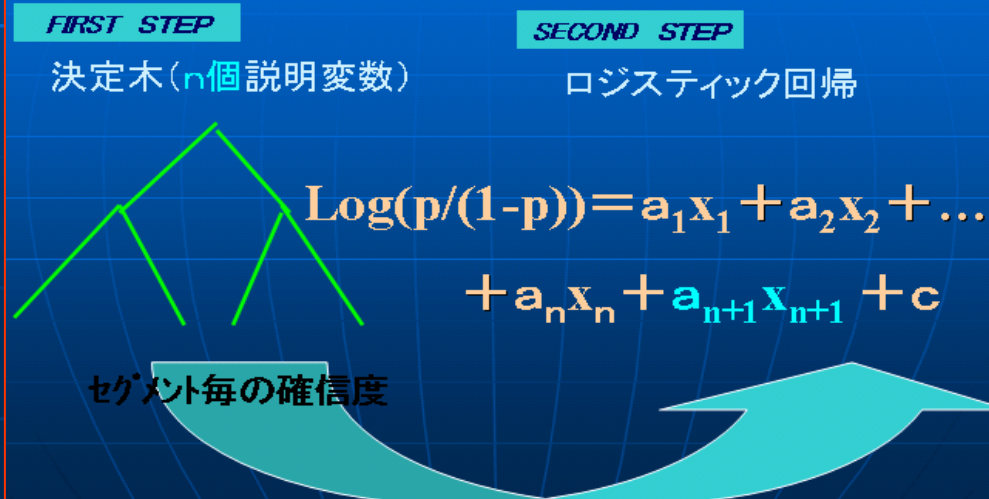
Cataract Model



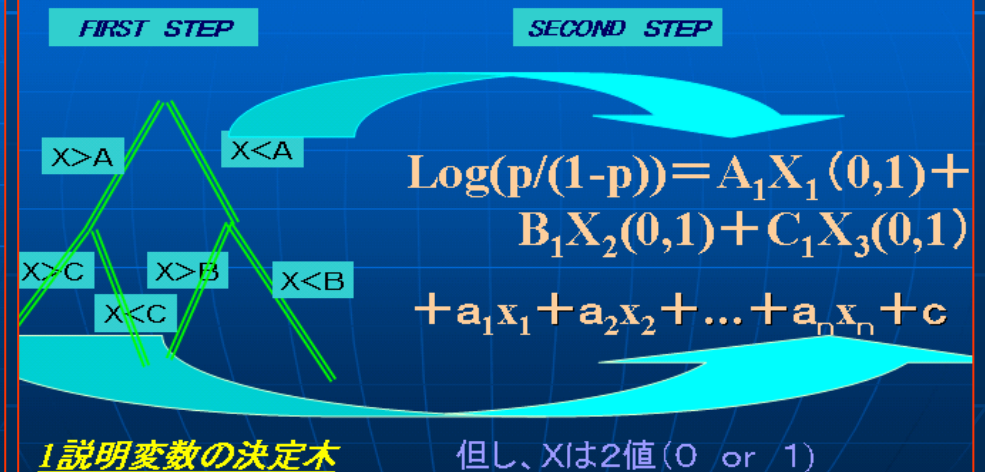
Cascade Model



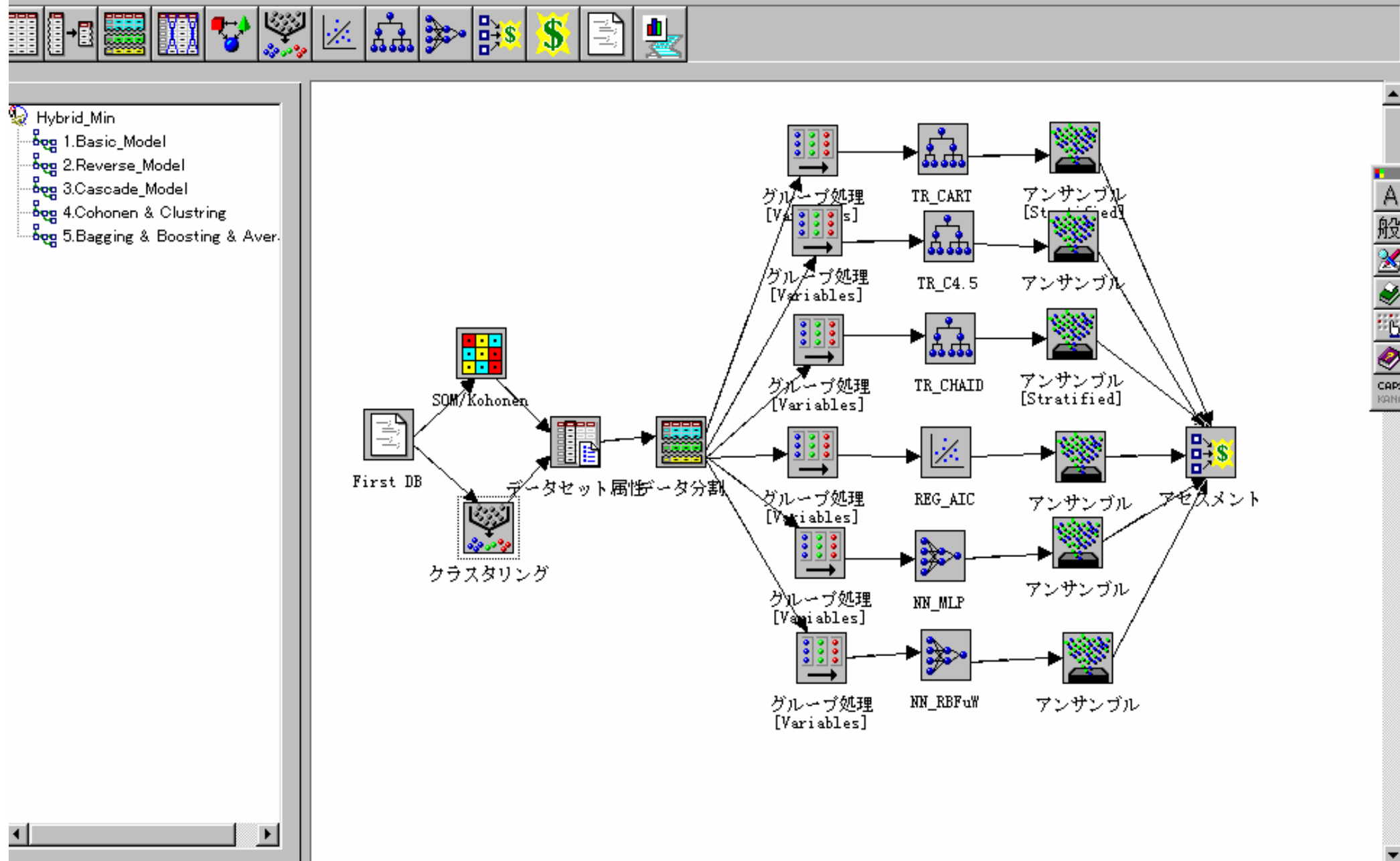
Reverse Model



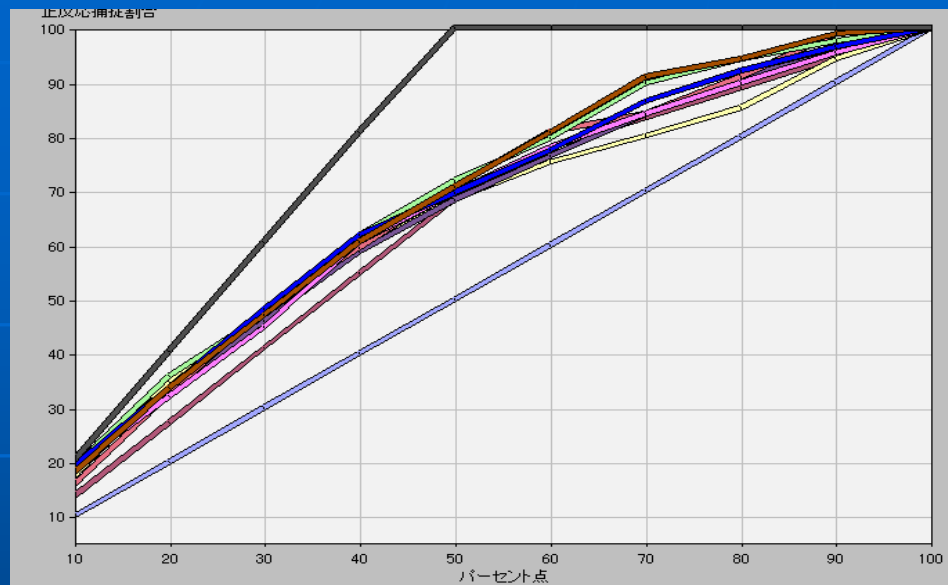
Category Flag Model



ハイブリッド・コンポーネント



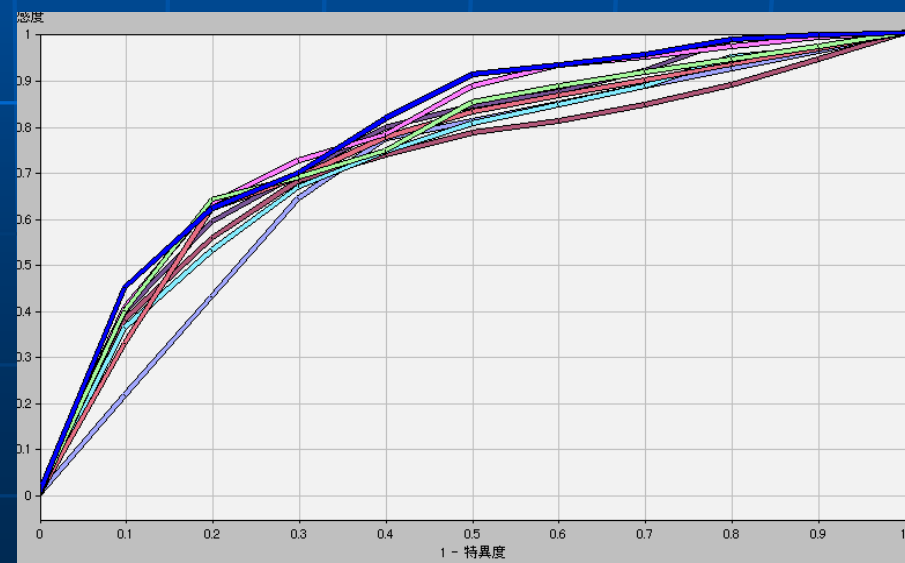
CAP図



ハイブリッドモデルは金融データマイニングにおいて、5%程度精度の向上に貢献する

それよりも第1分類器の失敗を第2分類器がフォローする作用がある

ROC図



CRMモデル最前線

クレジットカードの初使い分析

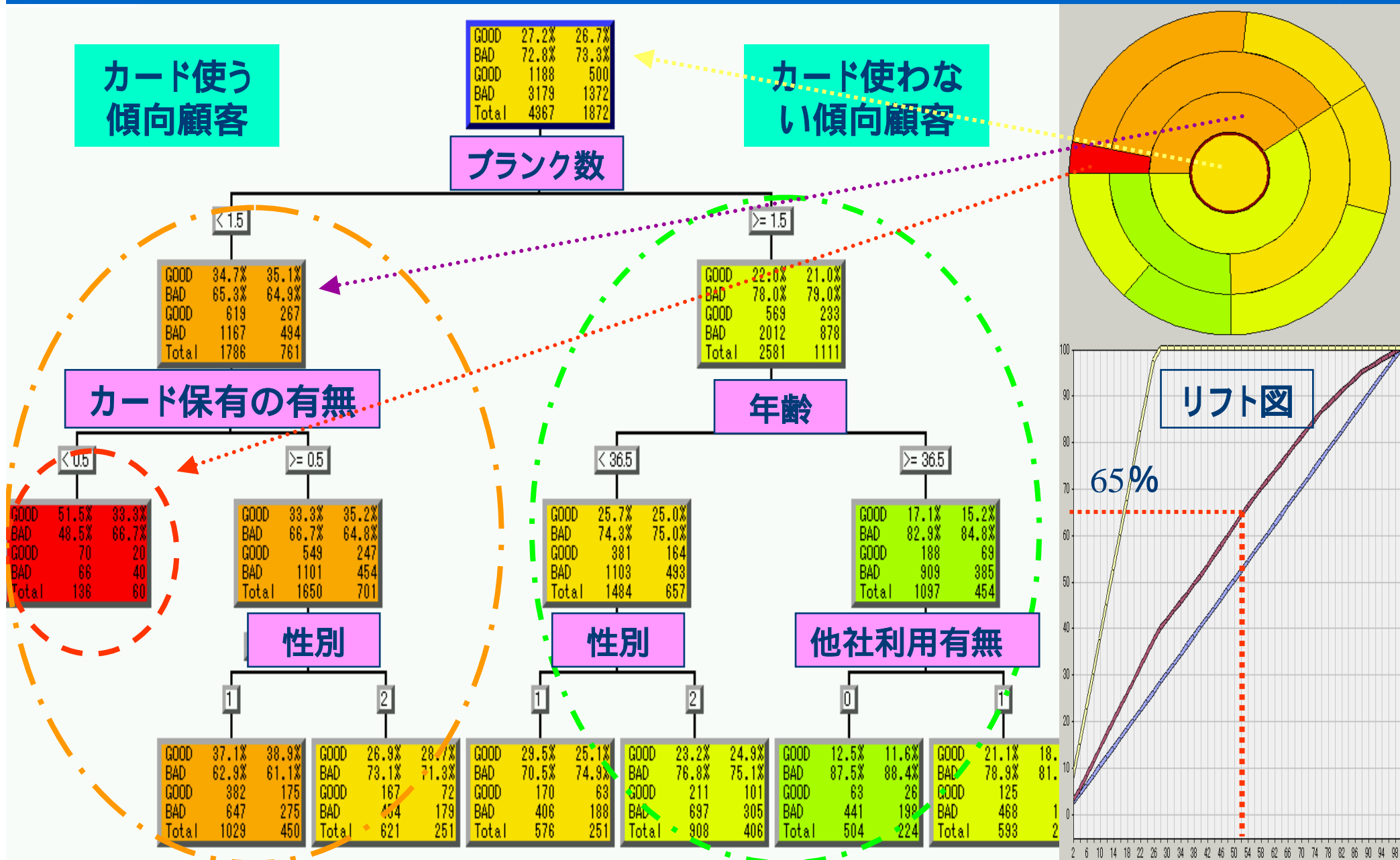
サイコグラフィックデータの利用したCRM

- デモグラフィックデータ(人口統計的データ)の限界
- サイコグラフィックデータ(心理的データ)の利用
 - 顧客の品物購入の動議づけ
 - 大手カード会社や旅行代理店が注目
 - ライフスタイルとアンケート

単なる欠損値？

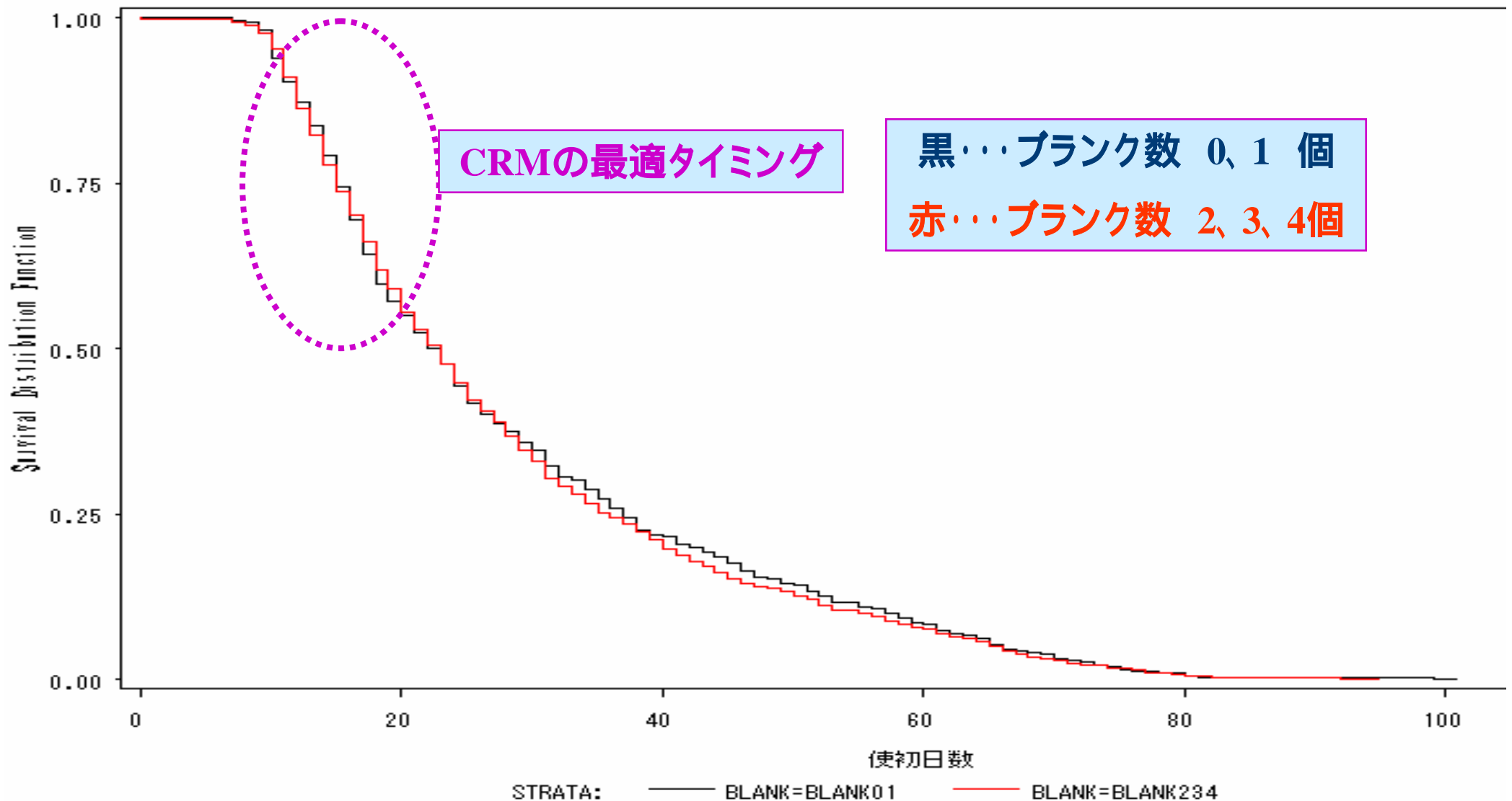
顧客の心理を反映！！

カード初使い1顧客のツリー



初使い1の期間分析

ANALYSIS FOR FIRST CARD USER



CRM最前線

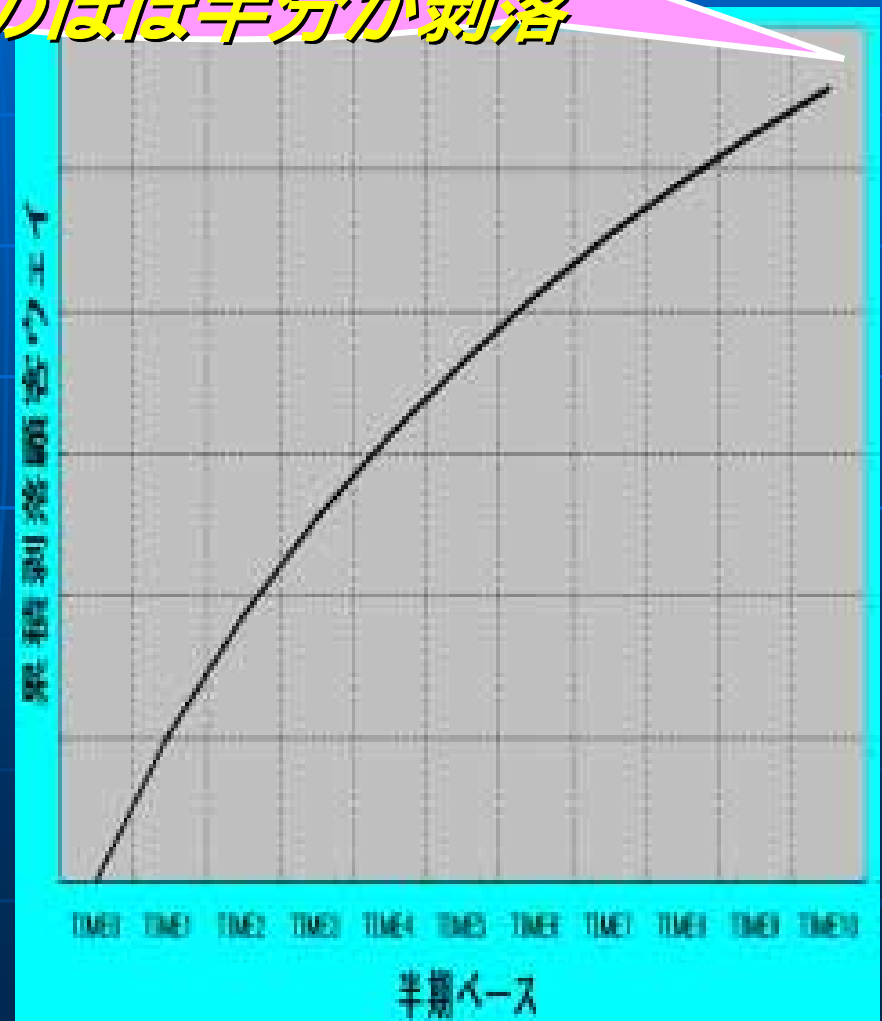
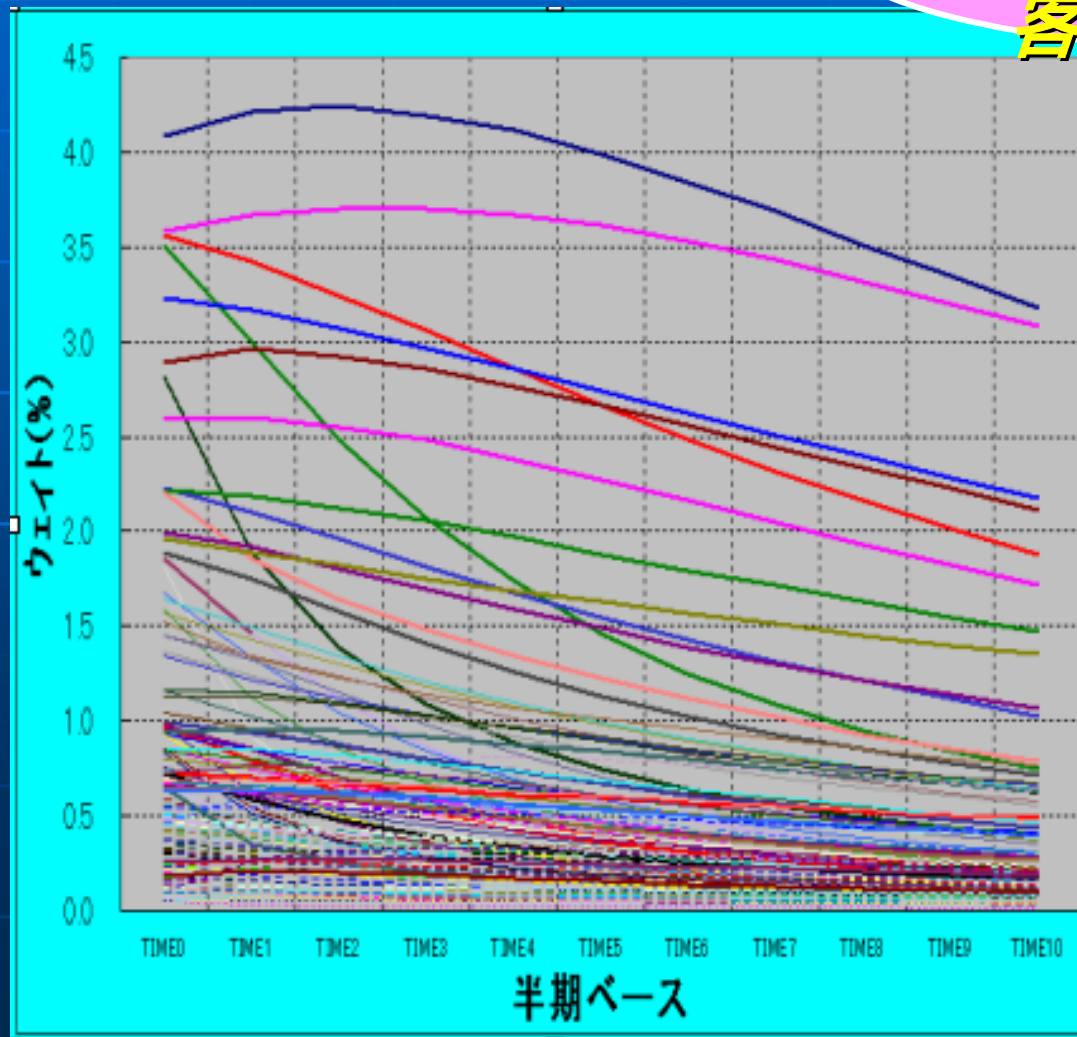
顧客の推移分析

金融顧客の推移分析

- 決定木のリテンション分析では手後れが多い
 - 剥落至る原因ではなく、兆候を見つける
- 顧客セグメントの時間変化
 - 吸収マルコフ連鎖モデルを適用
- 分析手順
 - 顧客のリテンション分析(決定木分析)
 - 顧客の生涯時間分析(吸収マルコフ連鎖モデル)

顧客セグメントの生存期間解析

5年間でアクティブ顧客のほぼ半分が剥落



中小企業向けの 新しい与信モデルの試み

企業情報を用いた 中小企業の倒産モデル

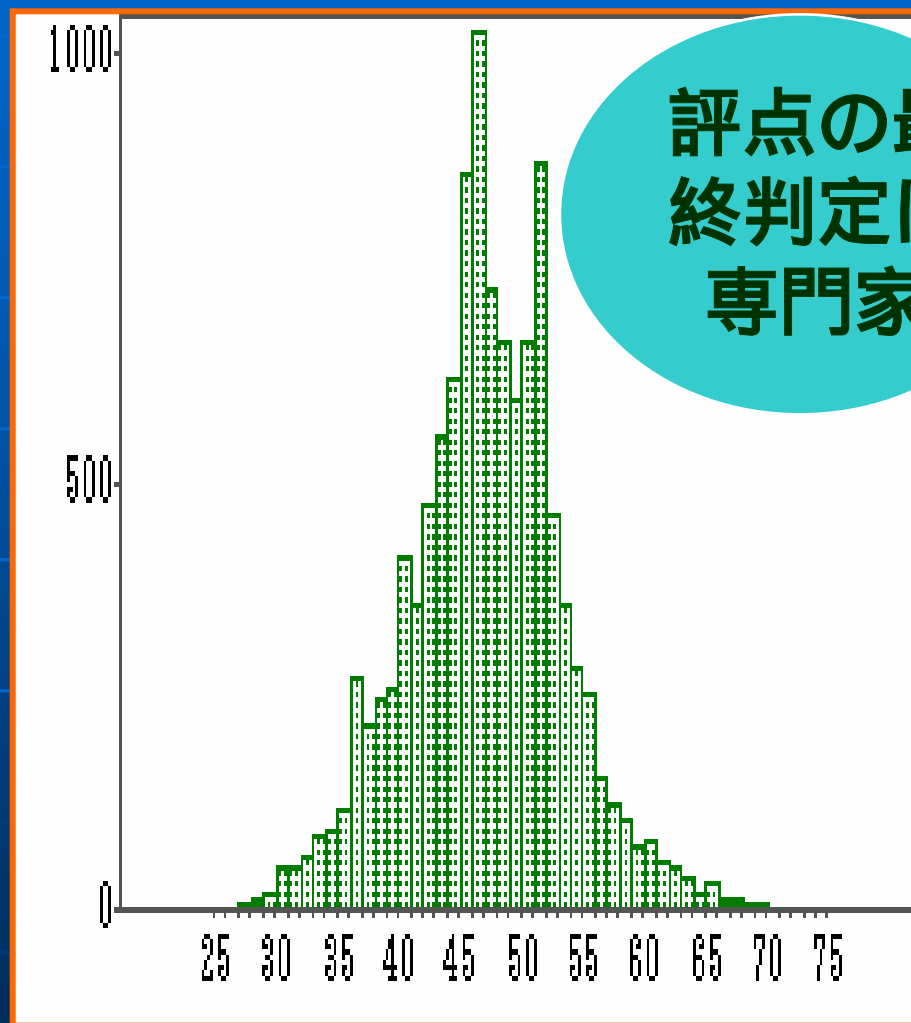
■ 金融機関の融資対象企業

- 日本の中小企業数 約500万社（中小企業庁定義）
- 金融機関が融資対象とする企業は数十万社
- 基準例：売上10億円以上、評点50数点以上
- 今、中小企業の与信対象を売上10億円未満に急速拡大へ

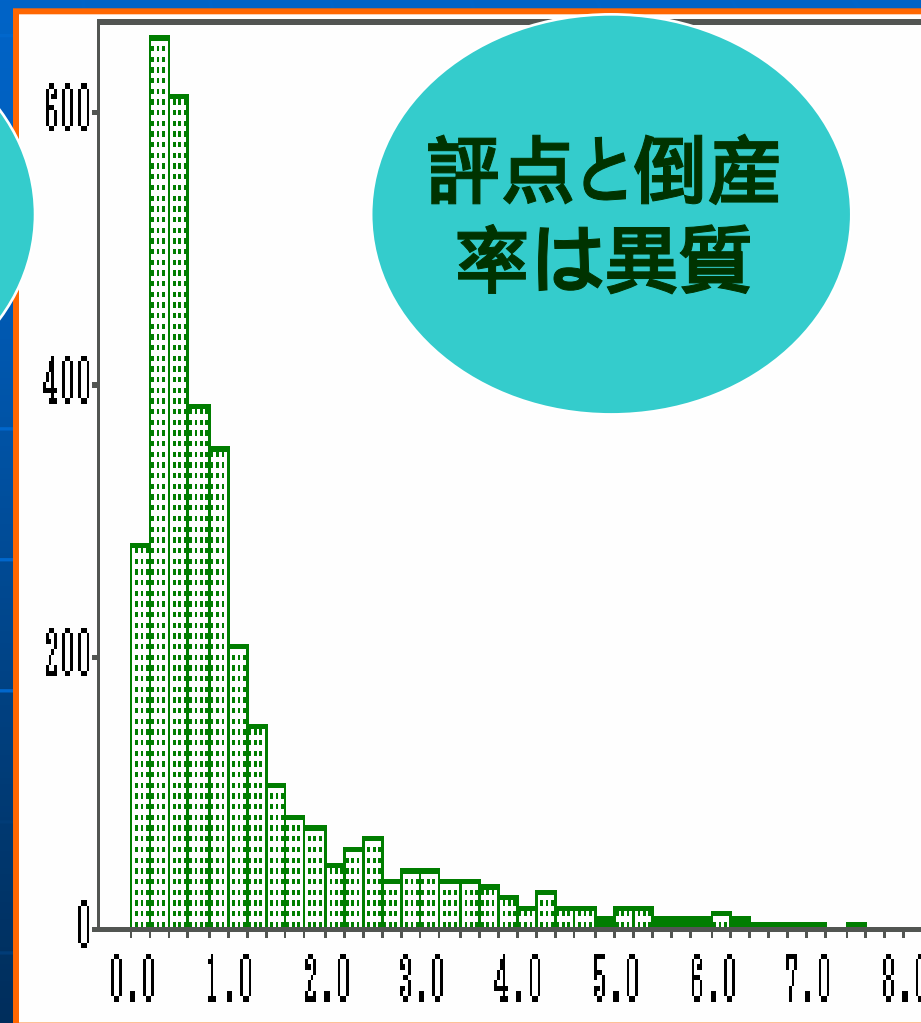
■ 再構築(Re-Construct)アプローチ

- 倒産率からモデルの推定する
 - 情報ベンダーは40万社の倒産率を公表している
 - ブラックボックスの倒産モデルの解明
- 説明変数に企業情報(約120万社)を採用した定性モデル
- 分析手法： 構造の解明 決定木

TDBの評点と倒産率の度数分布



評点 120万社



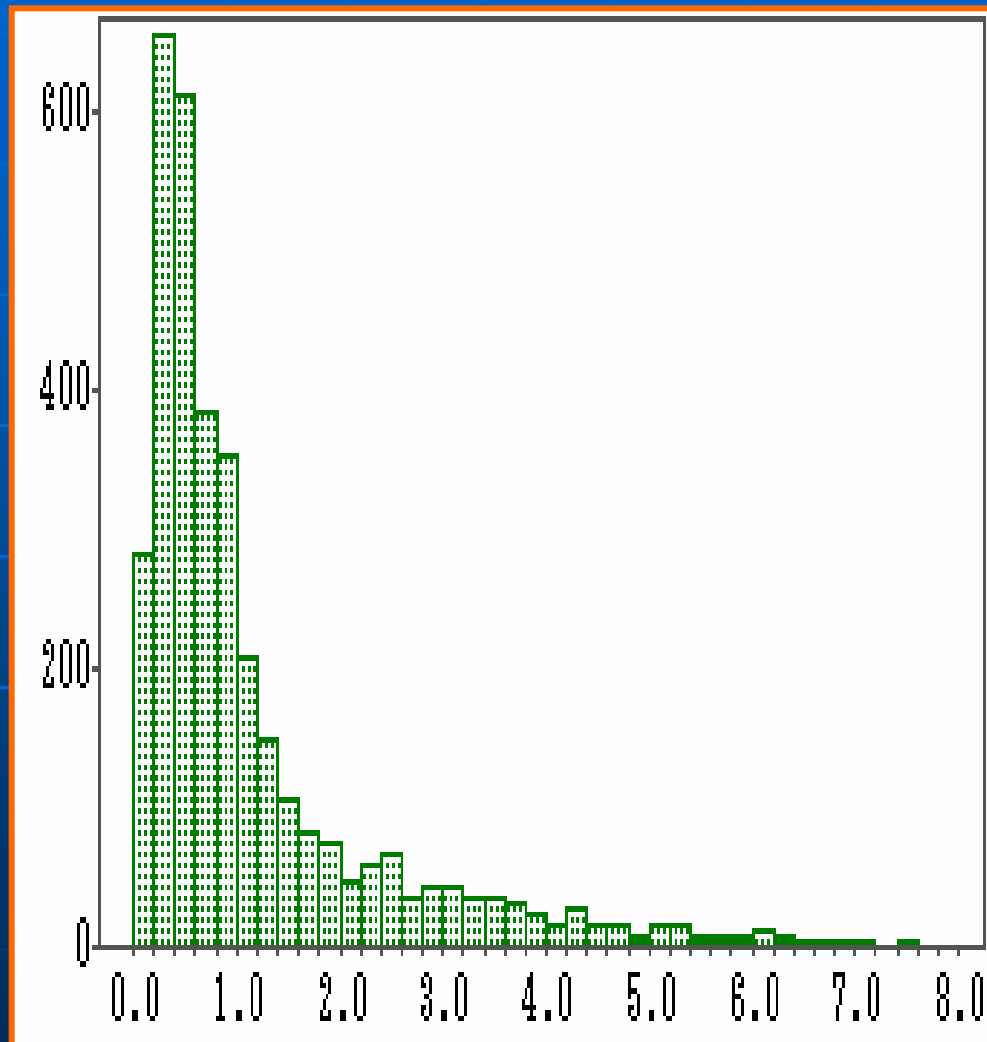
CCRモデル 40万件

決定木の構成要素

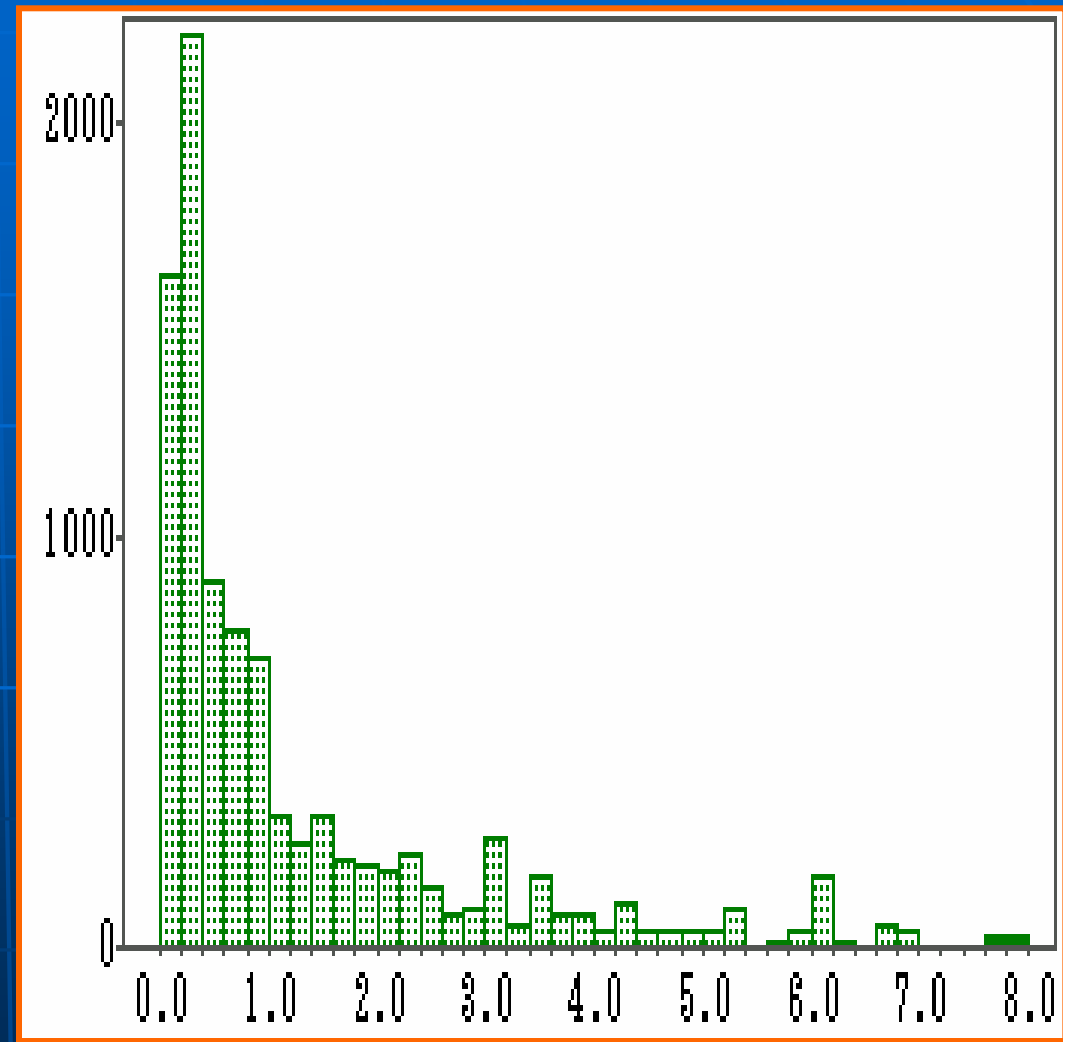
情報公開企業を
優遇する変数

	変 数	平均分割度		変 数	平均分割度
1	評点	5.00	11	株主数	1.39
2	2分類P	3.70	12	最新期業績売上	1.22
3	最新期業績自己資本比率	3.20	13	税引後利益の正数P	1.02
4	税引後利益減少数P	3.10	14	全国業種ランキング相対位置P	0.93
5	法人申告所得額増加数P	2.34	15	前々期からの平均売上高伸率P	0.89
6	自己資本比率増加数P	2.30	16	最新期業績税引後利益	0.64
7	設立年数P	2.16	17	法人格コード	0.17
8	業績利益存在数P	2.01	18	前期売上高伸率P	0.12
9	決算書存在数P	1.54	19	前期平均売上高伸率P	0.07
10	最新期業績決算書有無P	1.41		(変数名にPがつくものは加工変数)	

倒産率の度数分布



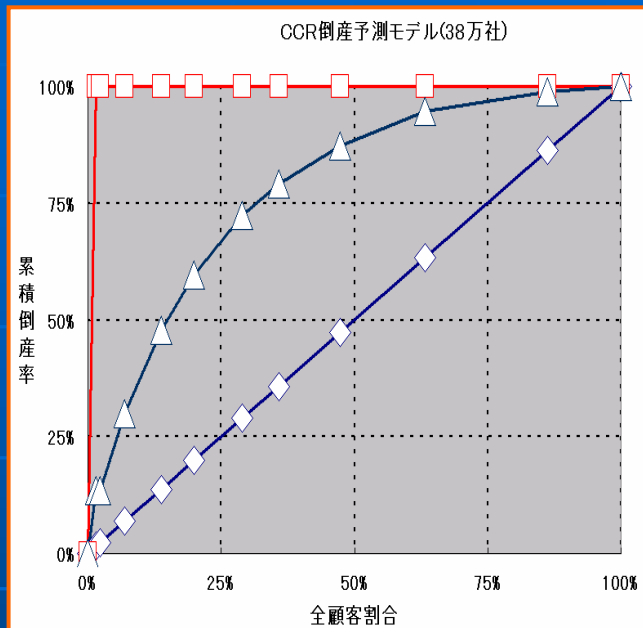
CCRモデル (40万社)



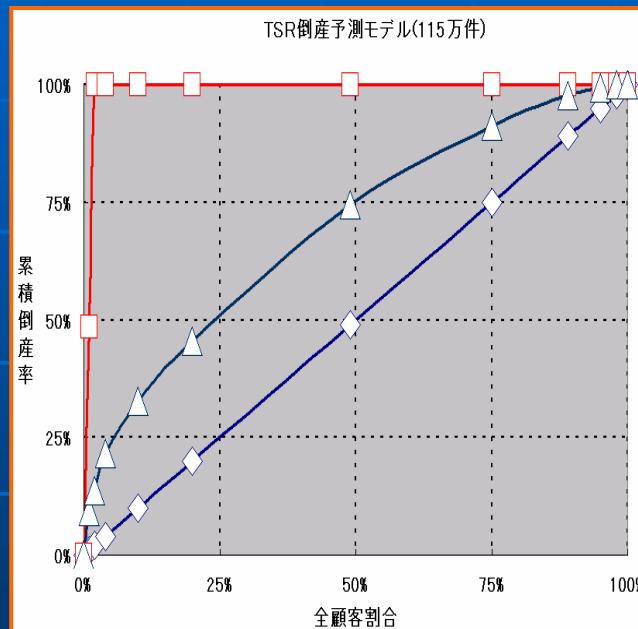
再構築モデル (120万社)

倒産率(%)

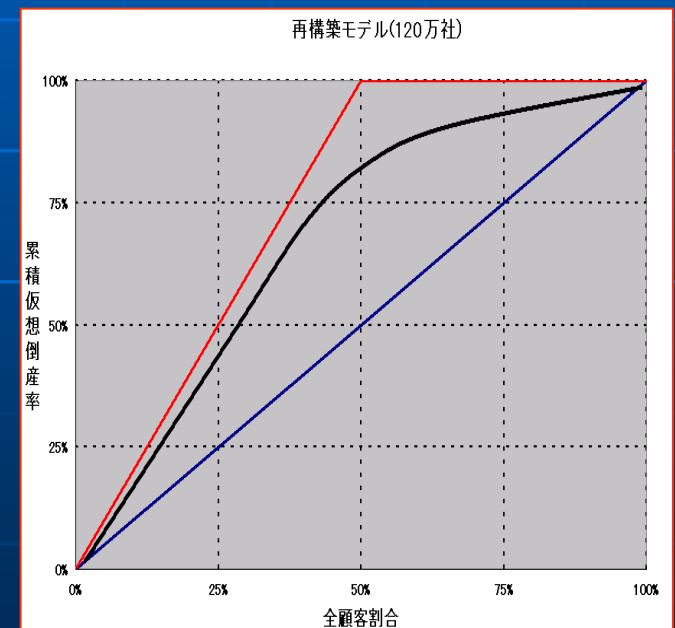
倒産モデル比較



TDB社の倒産モデル
(40万社)



TSR社の倒産モデル
(115万社)



再構築モデル
(120万社)

ま と め

- リテール金融工学の基礎理論の確立
 - データマイニングと金融工学の融合
 - 与信モデルの基礎理論の確立
 - CRMはサイコグラフィックデータの活用へ
- モデルの戦術から戦略利用への飛躍
 - リテールの断片的な情報を収益拡大へ
 - リスクとリターンさらに時系列変化