

20001年7月27日

SUJI-J2001

# ハイブリッド・コンポーネントの構築 ～多数モデルによる新しいデータマイニング～

三和銀行  
リテール業務部 金融エンジニアリング室  
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

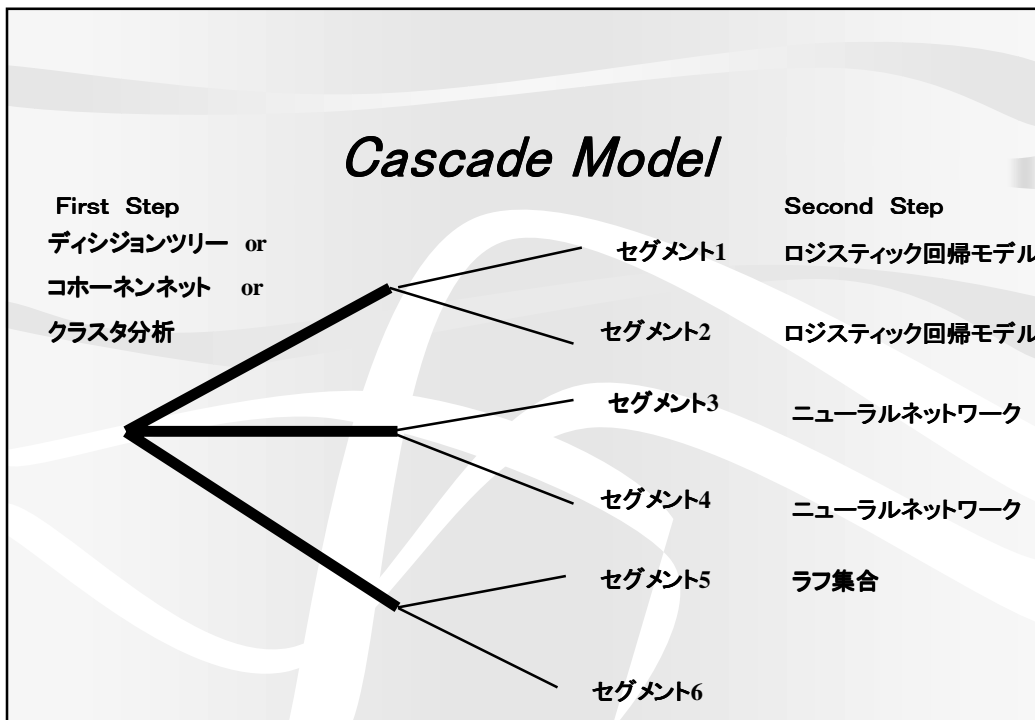
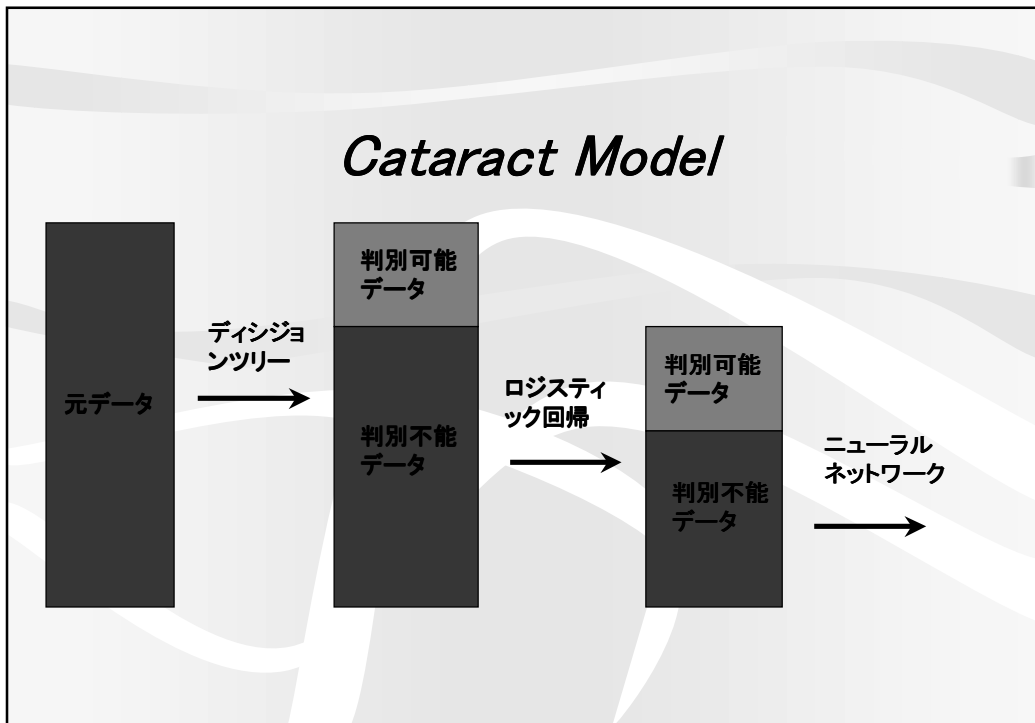
- 先行している会社の現状
  - マイニングのコンサルタントのノウハウを吸収
  - データクリーニング可能
  - マイニング・ツールによる簡単な分析可能
- 問題点
  - 分類器のパラメータ・チューニング
  - ハイブリッド&アンサンブル・モデルの自主開発の困難さ
- マイニング・ツールには意思決定機能が不十分である
  - レーシング・カーの性能を引き出すのは、ドライバーの腕次第
  - でも初心者用にチューアップは可能

## ハイブリッド・コンポーネントとは

- 9種類の分類器と、10タイプのハイブリッド&アンサンブル・モデルのテンプレートから構成
- 初心者でも高度な組合せモデルの設計可能
- 実務で使われる約100種類の分析テンプレート
- その前に、ハイブリッド&アンサンブル・モデルの方法論を紹介

## 組合せモデルの種類

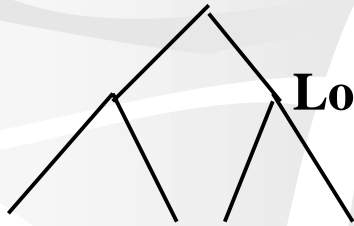
- 分類器の組合せ
  - ①ハイブリッド・モデル(Hybrid Model)
    - 複数の分類器の組合せ
      - Cataract Model
      - Cascade Model
      - Reverse Model
      - Category Flag Model
      - Un&Supervised Classification Model
  - ②アンサンブル・モデル(Ensemble Model)
    - 弱い分類器の仮説集合に投票権を与える
      - Type I (異種分類器)
      - Type II (同じ分類器、Committee Modelと呼ぶ)



## Reverse Model

### FIRST STEP

決定木(n個説明変数)



セグメント毎の確信度

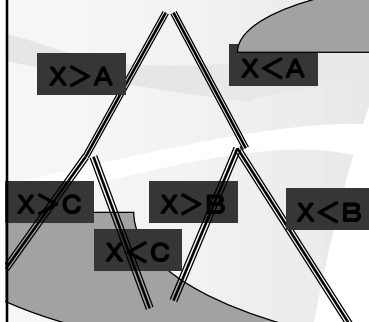
### SECOND STEP

ロジスティック回帰

$$\text{Log}(p/(1-p)) = a_1x_1 + a_2x_2 + \dots + a_nx_n + a_{n+1}x_{n+1} + c$$

## Category Flag Model

### FIRST STEP



1 説明変数の決定木

### SECOND STEP

$$\text{Log}(p/(1-p)) = A_1X_1(0,1) + B_1X_2(0,1) + C_1X_3(0,1) + a_1x_1 + a_2x_2 + \dots + a_nx_n + c$$

但し、Xは2値(0 or 1)

## アンサンブル・モデル

- 相異なる分析器あるいはパラメータやトレーニングデータが相異なる分類器をまとめる
  - Average Model                      Minimum Risk Model
  - Maximum Return Model          Majority Model
- 複数のランダムサンプリングしたトレーニングデータ適用した同じ分類器をまとめる
  - Bagging (データ・ウェイトが等しい)
  - Boosting (データ・ウェイトが相違する)

## 基礎分類器とテンプレート

基礎分類器 1

1	決定木	CART
2	決定木	C4.5
3	決定木	CHAID
4	ニューラルネットワーク	MLP
5	ニューラルネットワーク	RBF equal Widths
6	ニューラルネットワーク	RBF Unequal Widths
7	ロジスティック回帰	評価基準:SBC
8	ロジスティック回帰	評価基準:AIC
9	ロジスティック回帰	評価基準:Profit/Loss

基礎分類器2

10	クラスタ分析	Ward
11	クラスタ分析	k-means
12	ニューラルネットワーク	Cohonen Net

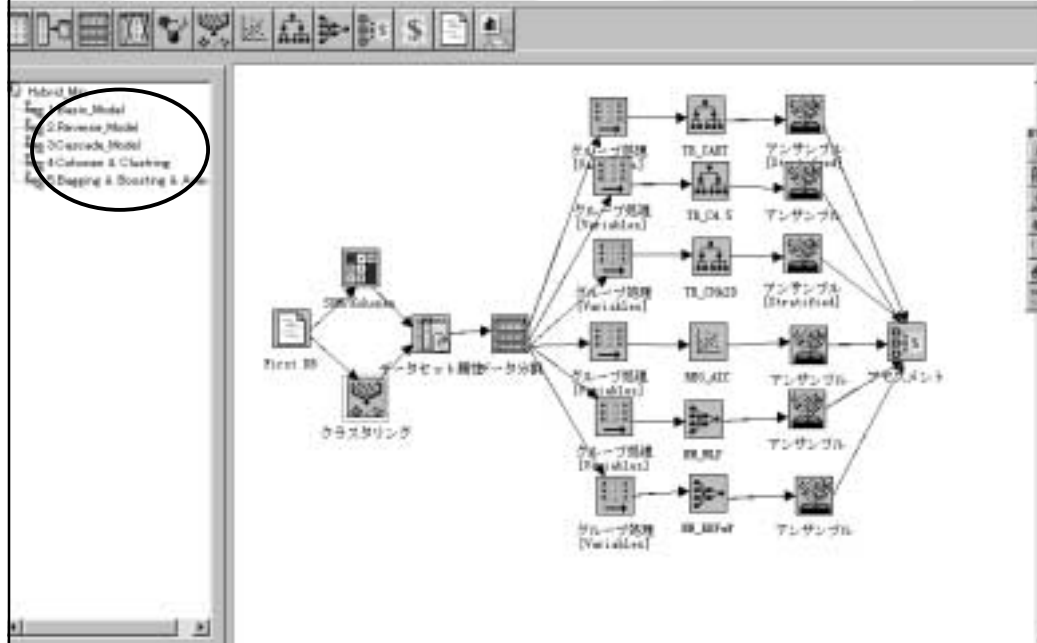
テンプレート 1

1	Hybrid Model	Cataract Model
2	Hybrid Model	Cascade Model
3	Hybrid Model	Reverse Model
4	Hybrid Model	Category Flag Model

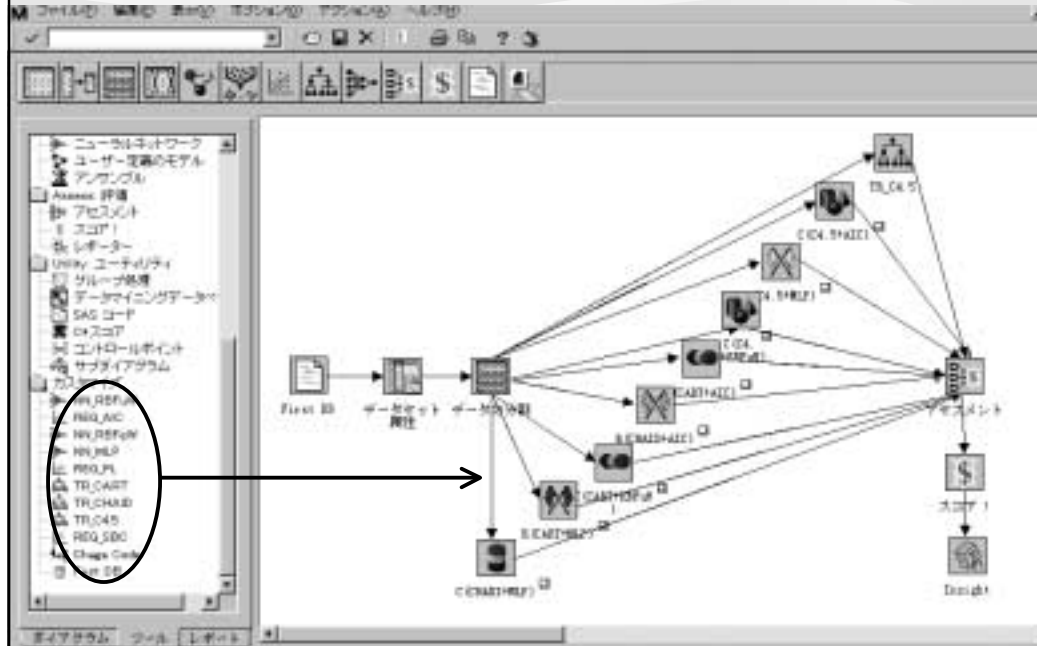
テンプレート 2

5	Ensemble Model	Average Model
6	Ensemble Model	Minimum Risk Model
7	Ensemble Model	Maximum Return Model
8	Ensemble Model	Majority Model
9	Ensemble Model	Bagging Model
10	Ensemble Model	Boosting Model

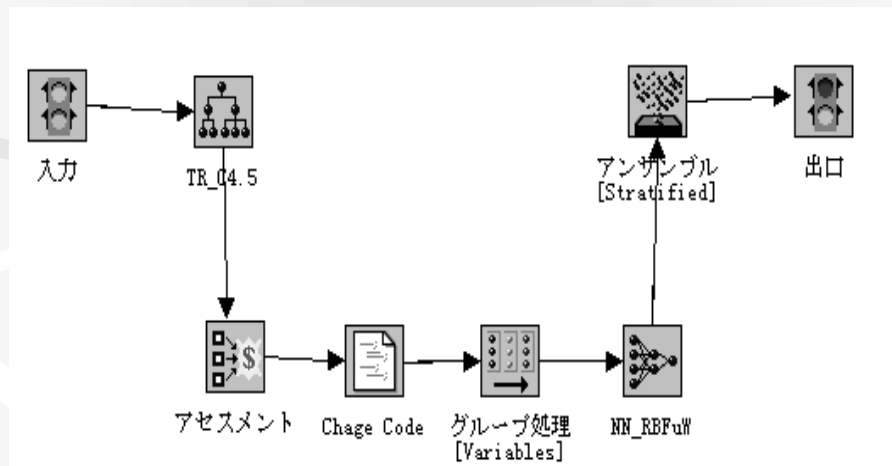
## ハイブリッド・コンポーネント I



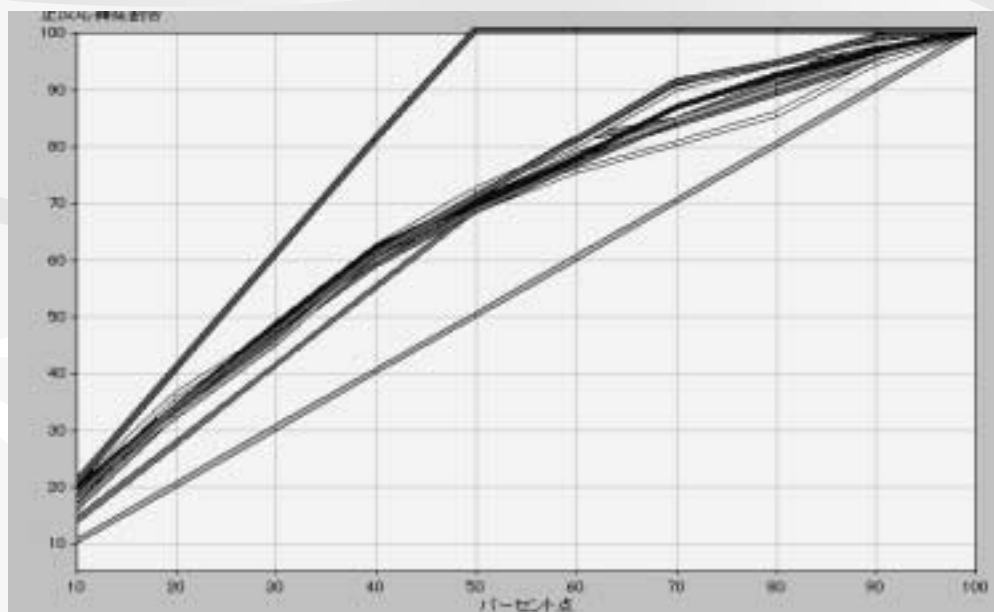
## ハイブリッド・コンポーネント II



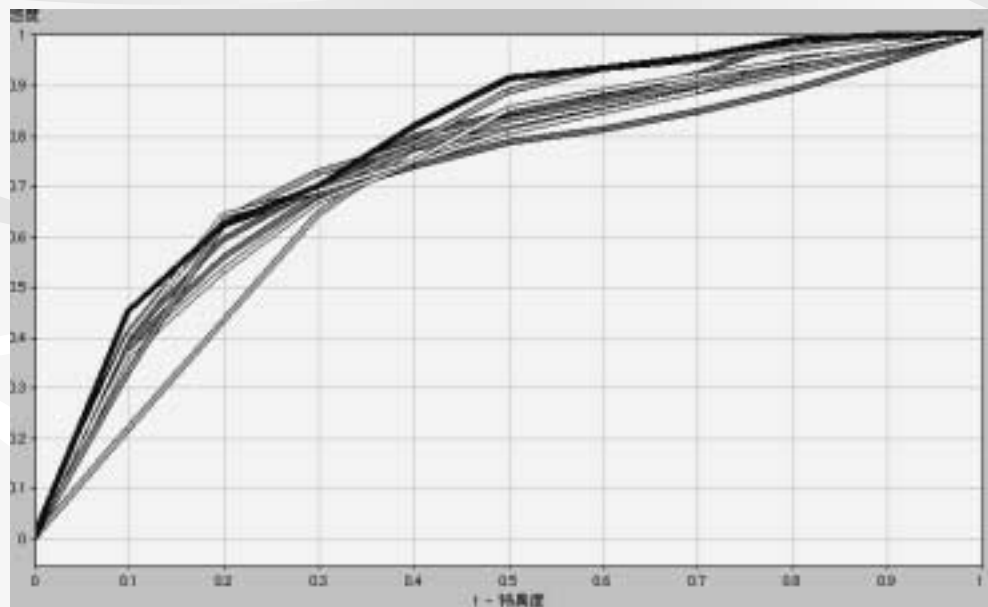
## 分類器ノード(Cascade Model)



## 反応捕捉割合チャート



## ROCチャート



## ハイブリッド・モデルの評価

同じ精度を有するモデルが多数発生する！！  
どのモデルが最適でしょうか？？？

### 選択基準

- テストデータによる評価
- 反応(捕捉)割合チャート、ROCチャート、クロスリフトチャートによる比較
- いき値変動による安定性の評価
- 実務運用に適しているか！！



## ま と め

- ハイブリッド&アンサンブル・モデルの体系化
- マイニング・ツールを利用した新しいマイニング技術
  - 約100種類の分析テンプレートの開発
  - 5~15%の精度向上
- ハイブリッド化は互いの分類器の弱みを補足しあうため  
同じ精度を有することが多い
  - 初心者でも精度の高いモデル構築可能
- モデル開発期間の短縮
- 専門家のノウハウの共有化

## Question & Answer

