

より手軽に、より正確に、より大規模に

Visual Mining Studio

～ここまで進化した次期版 (V8.1) のご紹介～

(株) NTTデータ 数理システム 徐 良為

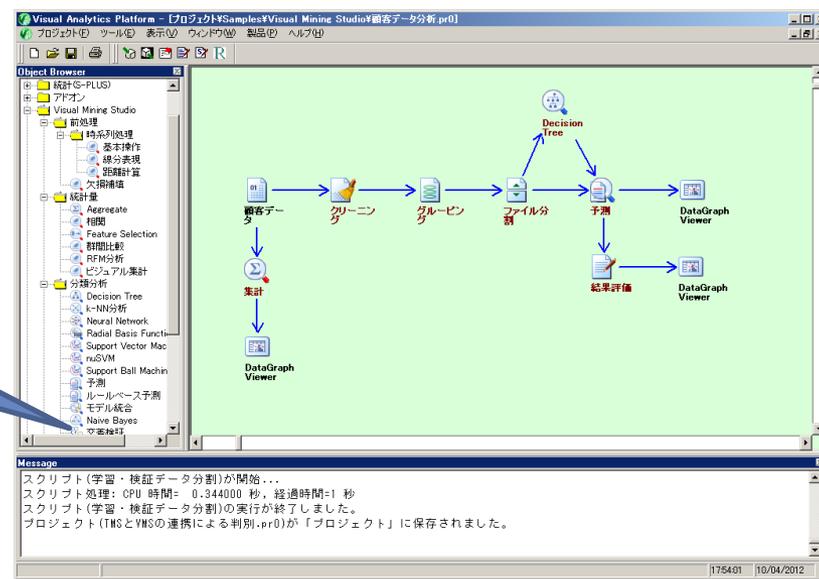
NTT DATA

株式会社 NTTデータ 数理システム

汎用データマイニングソフト
日本発・完全自社開発

● 特徴

- ◆ 手軽な試行錯誤環境
 - ビジュアルプログラミング
 - データ可視化
- ◆ 豊富なマイニング機能
 - 前処理・予測・データ探索・統計解析
- ◆ 多様な他システム連携機能
 - DBシステム・Excel・R・SPLUS・SAS・Python・MatLab



- 決定木モデル構築GUIの大幅刷新
- ランダムフォレストモデル機能
 - 複数決定木による集団学習
 - 高精度・高速度(マルチスレッド)
- 複数観測HSMMモデルの機能拡張及びアイコン化
 - HSMM=隠れセミマルコフモデル
 - 単一観測のみから、複数観測対応モデルへ
 - ユーザーインターフェース(アイコン)
- クラスタリング (KMeans)
 - クラスタ中心の計算方法の追加
- 前処理
 - 連携機能一台マシン上のMap/Reduce(マルチプロセス)並列計算仕組
 - BDM(Big Data Module)、VRPシステム連携機能

【スクリプト】 新関数及び機能改善

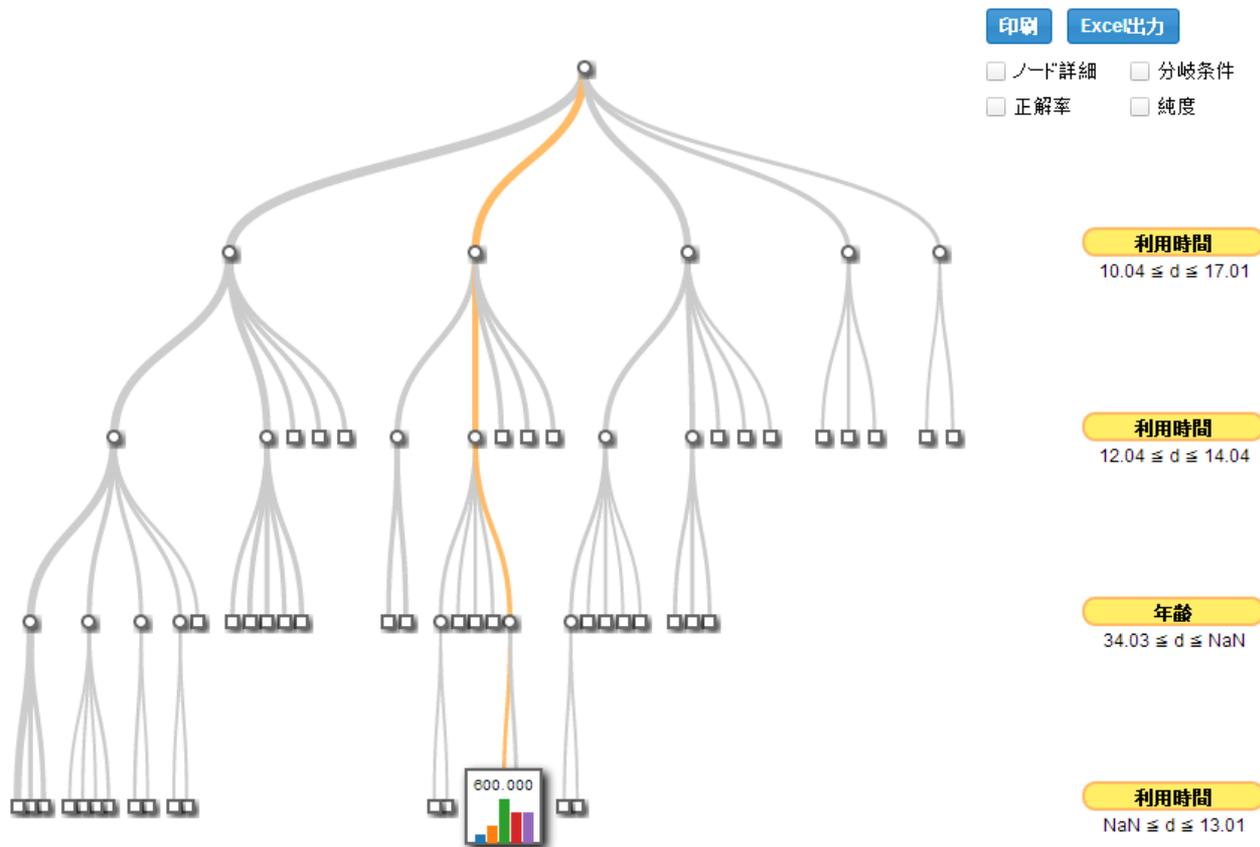
- `sys_proc_map_reduce` Map/Reduce機能呼び出す関数
- `merge_sorted_vdt` 二つキー順にソートされたテーブルをマージする
- `as.bd` テーブルをBDM(ビッグデータモジュール)形式へ
- `bind_bd_with_key` 二つキー列で分類されたBDMを一つに束ねる
- `optimize` 探索初期値設定、並列計算機能改善
- `run_sql_on_db_bd` ビッグデータモジュール対応
- `read_table_from_db_bd` ビッグデータモジュール対応
- `vrp_load_rdata` Rデータの読み込み
- `vrp_start_server` RServerチャンネル制御
- `vrp_close_server` RServerチャンネル制御
- 並列計算 大量同時計算時の不具合修正

【その他】

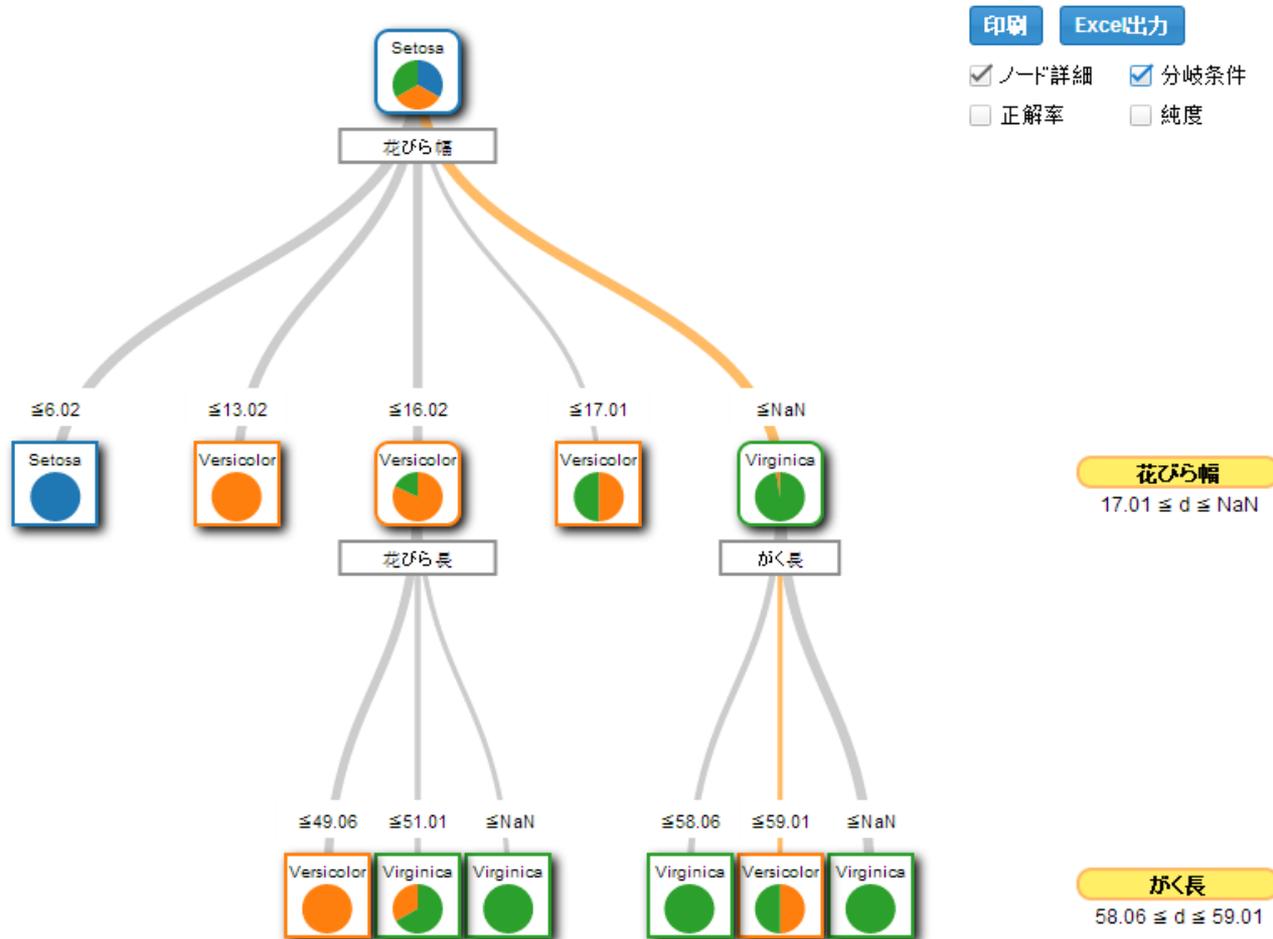
- VAP実行環境改善(多数)

- **決定木とは**
 - 木(Tree)構造を持つ予測モデル
 - データマイニングでよく用いられる手法の一つ
- **VMSの決定木は次の特徴を持つ**
 - 分類モデル(カテゴリ)、回帰モデル(連続値)両方に対応
 - 複数の分岐基準に対応 (Gini係数、InfoGain、分散など)
 - 柔軟な欠損対応
 - 多分岐対応
 - **ユーザ対話型(強制分岐、剪定など、木への直接的な操作)**
 - 多目的回帰モデルに対応
 - モデルからのルール抽出
 - C++文法に沿ったソースコード抽出

- VMSの新しい描画技術用いたユーザ対話IF画面
 - より直感で、ユーザーフレンドリー
 - 本格的なTreeモデル分析が対話的に可能



決定木の詳細表示画面



- **多数の決定木による集団学習予測モデル**

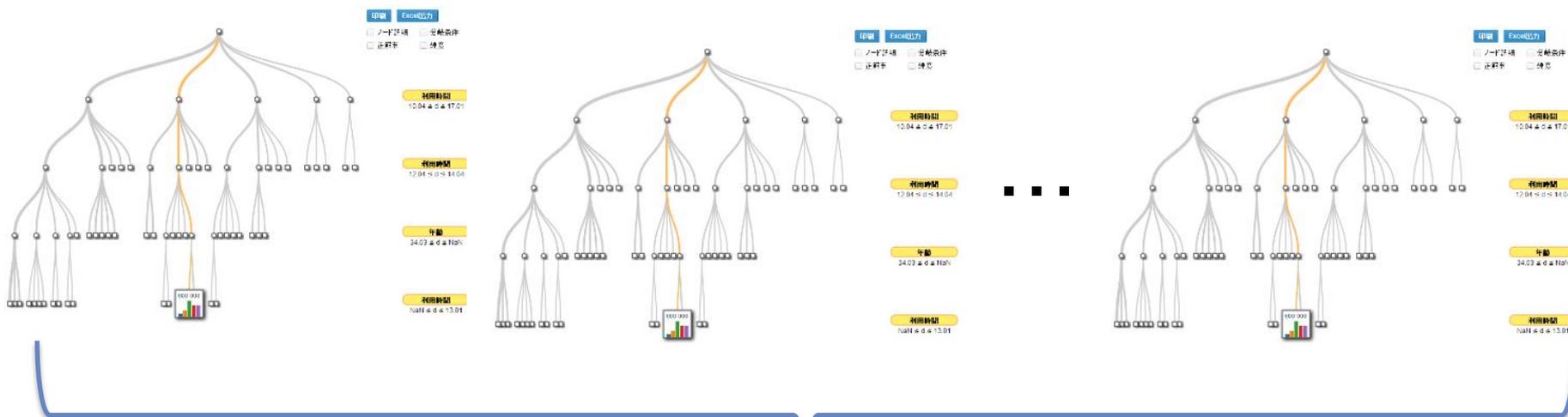
- **特徴**

- **ブートストラップサンプリング、説明変数のランダム選択**
- **モデル構築・予測高速**
- **説明変数寄与度の算出が可能**
- **分類モデル予測値中の各クラスの出現確率を算出可能**

• VMSのRF特徴

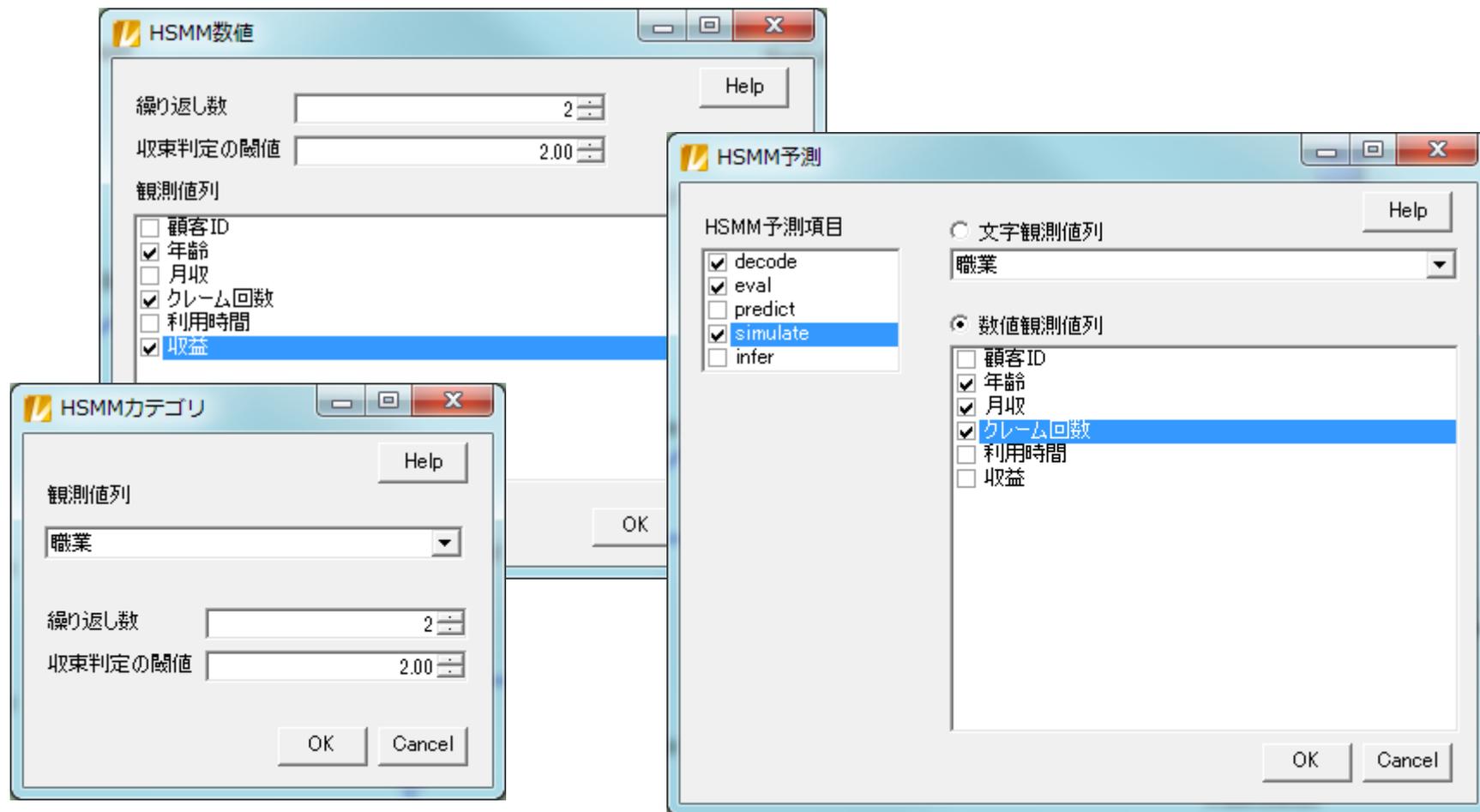
- RFの共通特徴
- 決定木の拡張(Treeパラメータ指定共有)
- 並列計算

同一マシン上のマルチスレッド計算



多数決 or 平均

1. 多観測値: $O = (\text{obs}1, \dots, \text{obj}N)$
2. ユーザーインターフェース



• Map/Reduce

- 大量データの並列分散処理用フレームワーク
- Map と Reduce 処理を組み合わせる処理を行う
 - Map = 分割されたデータの断片に何らかの加工を施す
 - Reduce = Mapの結果を束ねる
- 上記作業をできる限り並行処理を行う
- 例： 並列計算用 MergeSort

• VMSのマルチプロセスMap/Reduce機能

- 1台のマルチコアマシン上で行う
- VMSのスク립トでMap関数、Reduce関数の記述が可能に

・ 開発目標

より高速、より手軽に、より正確に

- **新しいデータ分析・機械学習機能の追加**
- **分析結果の展開ツール
手軽にドリルダウン・ドリルアップなどの
仕組み構築**

**引き続き、皆様のデータ分析のお役に立てるよう、
開発者一同頑張っています。
ご期待ください！**

お問い合わせは:

株式会社NTTデータ数理システム 営業部

TEL:03-3358-6681

vmstudio-info@msi.co.jp