

# Visual Mining Studio バージョン 8.1 の変更点

2015 年 1 月 29 日

株式会社 NTT データ数理システム

## 主な変更点

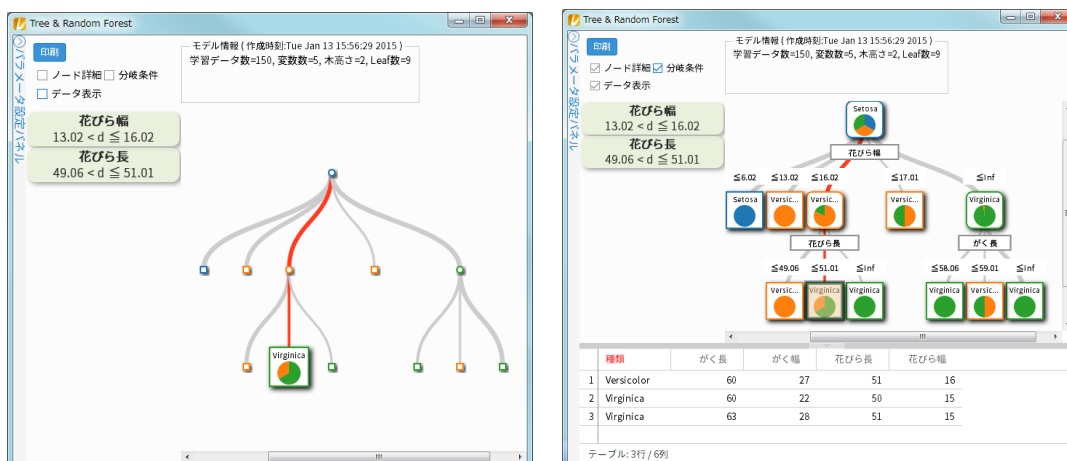
### 1. 新アイコン「Tree & Random Forest」の追加

決定木モデルを構築するユーザーインターフェースを大幅に刷新し、ランダムフォレスト機能を追加して、新たな分析アイコン Tree & Random Forest を作成しました。

※ 従来の Decision Tree アイコンも継続して利用できます

### 決定木モデル構築ユーザーインターフェースの大幅刷新

より直感的な表示と操作性を実現しました。

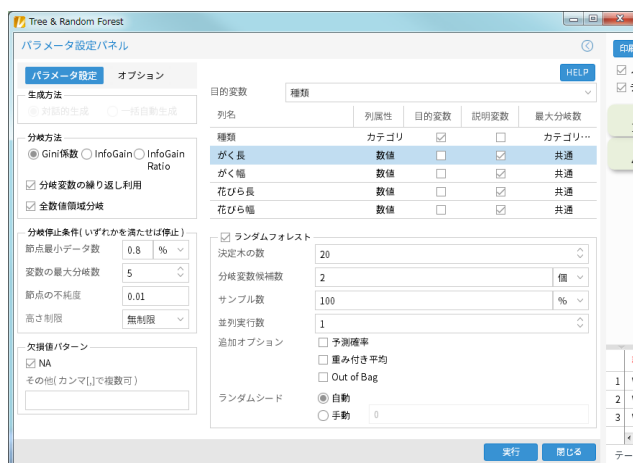


全体の俯瞰と詳細表示をシームレスに切り替え

選択されたノードに含まれるデータを同時に表示

### ランダムフォレストモデル機能の追加

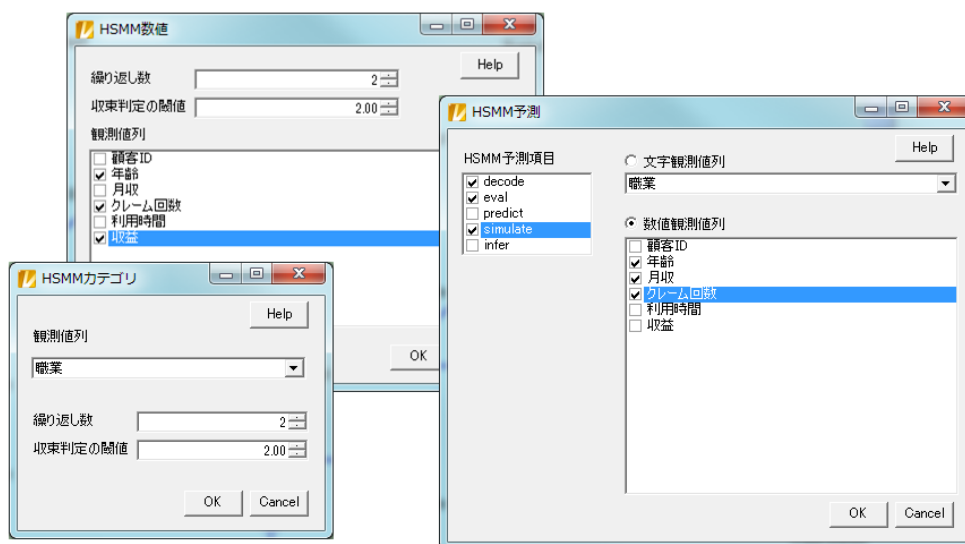
多数の決定木によって集団学習・予測を行う、ランダムフォレストモデル機能を追加しました。ブートストラップサンプリング、説明変数のランダム選択、並列実行により、高精度かつ高速なモデル構築を行います。チェックボックスにより Decision Tree と Random Forest の分析を切り替えることができるため、Decision Tree で検討したモデルをスムーズに Random Forest に拡張することができます。



Decision Tree と Random Forest の設定画面を一体化

## 2. 隠れセミマルコフモデルの機能拡張

隠れマルコフモデルの分析アイコンを追加し、マウス操作での実行が可能になりました。また、複数の時系列が観測されるモデルに対応しました。



## 3. Map / Reduce 機能の追加

Map/Reduce 機能により、1 台のマシン上で並列処理を行い、大規模なデータに対しても高速に処理を行います。分割されたデータの断片に処理を行い (Map)、それらの結果を束ねる (Reduce) 一連の作業を並列処理で行います。スクリプトによる Map 関数、Reduce 関数の記述が可能のため、複雑な処理も自在に記述することができます。

## 4. K-Means 法に新しい距離計算方法を追加

Cosine(spherical K-Means), Manhattan(K-Median) の2つを追加しました。距離計算方法に合わせて適切な Centroid の計算を行います。

## 5. アイコン画像の一部刷新

他のアイコンと同じ画像を使用していたアイコンの画像を刷新しました。

## 新機能一覧

### 追加されたアイコン

#### Tree & Random Forest

決定木モデル構築のユーザーインターフェースを大幅に刷新

Random Forest 機能を追加

※ Windows Server 2003 では Tree & Random Forest アイコンは利用できません

#### HSMM(隠れセミマルコフモデル)に関連したアイコン

マウス操作による HSMM 機能の利用

複数の時系列が観測されるモデルに対応

#### HSMM(カテゴリ)

HSMM のカテゴリモデルを作成する

#### HSMM(連続値)

HSMM の平均値モデルを作成する

#### HSMM 予測(カテゴリ)

HSMM のカテゴリモデルの予測を行う

#### HSMM 予測(連続値)

HSMM の平均値モデルの予測を行う

#### HSMM シミュレーション(カテゴリ)

HSMM のカテゴリモデルから観測データを生成する

#### HSMM シミュレーション(連続値)

HSMM の平均値モデルから観測データを生成する

### 追加されたスクリプト関数

#### sys\_proc\_map\_reduce

Map/Reduce 機能呼び出す

#### as.bd

テーブルを BDM (ビッグデータモジュール) 形式へ変換

#### merge\_sorted\_vdt

二つのキー順にソートされたテーブルをマージする

#### bind\_bd\_with\_key

二つのキー列で分類された BDM を一つに束ねる

#### lower\_case

対象列の文字列に含まれる大文字を小文字にする

upper\_case

対象列の文字列に含まれる小文字を大文字にする

run\_sql\_on\_db\_bd

データベース上で SQL を実行する

read\_table\_from\_db\_bd

データベースからテーブルを取得

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

## 改善されたアイコン

列属性変更

実数から文字列へ変換する場合の桁数の指定が可能に

クラスタ分析 (K-Means 法)

新しい距離計算方法を追加

## 改善されたスクリプト関数

HSMM に関連したスクリプト関数

複数の時系列データが観測されるモデルに対応

hsmm\_estimate

hsmm\_eval

hsmm\_simulate

hsmm\_decode

hsmm\_predict

hsmm\_infer

aggregate\_bd

重み付き集計オプションの追加

## その他

アイコン画像の一部刷新