

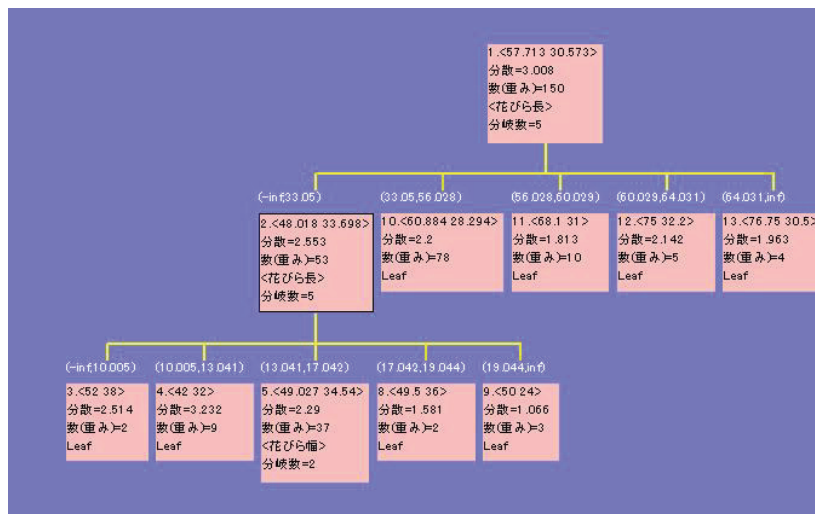
Visual Mining Studio バージョン 6.6 の変更点

2010年2月26日

株式会社 数理システム

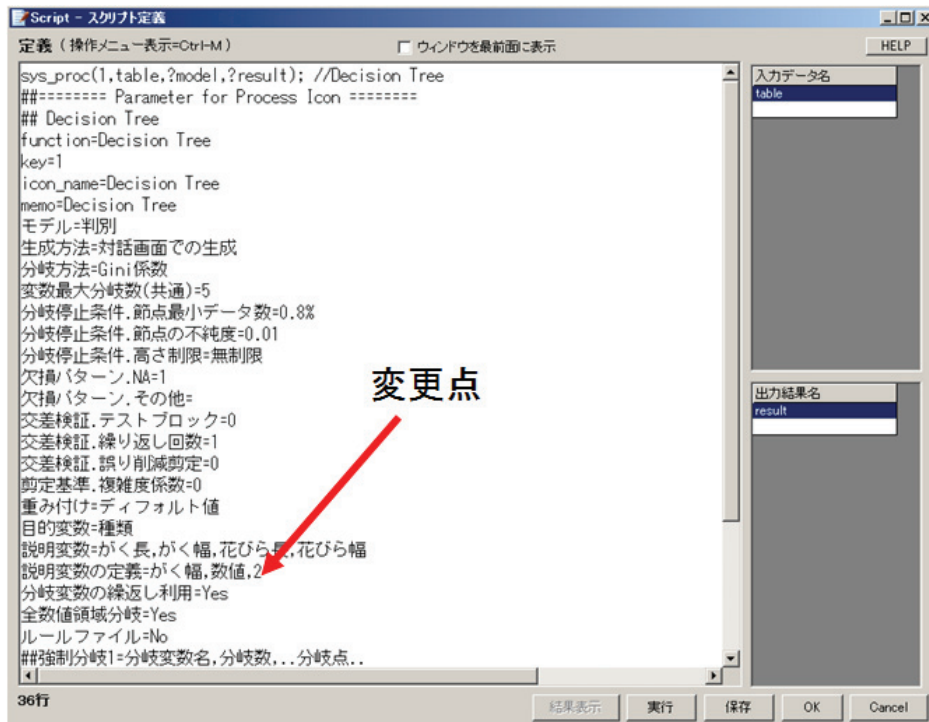
1. 決定木

決定木の目的変数に複数の変数を指定することができるようになりました。これにより、複数の項目を同時に予測することが可能になります。また、予測値に分位値で指定ができるようになり、データの距離の定義式が指定できるようになりました。



2. 決定木アイコンのパラメータ取得時の結果

決定木アイコンを実行してパラメータ取得した結果の説明変数部分には、従来は説明変数の数と説明変数の名前、型、最大分岐数をすべて記述していましたが、今回のバージョンよりデフォルト値から変更した説明変数のみ、説明変数の定義として記述されるようになりました。



```
sys_proc(1,table,?model,?result); //Decision Tree
##===== Parameter for Process Icon =====
## Decision Tree
function=Decision Tree
key=1
icon_name=Decision Tree
memo=Decision Tree
モデル=判別
生成方法=対話画面での生成
分岐方法=Gini係数
変数最大分岐数(共通)=5
分岐停止条件. 節点最小データ数=0.8%
分岐停止条件. 節点の不純度=0.01
分岐停止条件. 高さ制限=無制限
欠損/バターン.NA=1
欠損/バターン. その他=
交差検証. テストブロック=0
交差検証. 繰り返し回数=1
交差検証. 誤り削減設定=0
剪定基準. 複雑度係数=0
重み付け=デフォルト値
目的変数=種類
説明変数=かく長,かく幅,花びら長,花びら幅
説明変数の定義=かく幅,数値,2
分岐変数の繰返し利用=Yes
全数値領域分岐=Yes
ルールファイル=No
##強制分岐1=分岐変数名,分岐数,...分岐点..
```

決定木のパラメータ取得結果

3. RFM 分析

Recency(最新の購買日)、Frequency(購入頻度)、Monetary(購入総額)を基に、良い顧客を見つけるための分析を行うアイコンです。RFM 分析によって、顧客をグループ化して優良顧客を発見します。購買意欲の高い顧客を見つけることで、売上の増大とコストの削減を行うことが可能です。

RFM分析

ヘルプ

対象列指定

顧客ID ID番号

購買額 金額

購買日 時間 日付形式
(日付形式=自動検出)

対象データ範囲

終了日

指定しない
 固定 2009/12/31

開始日

指定しない
 固定 2009/12/31
 終了日より 1 年前

スコア計算用重み

R 100

F 10

M 1

R(最新購買日) | F(累計回数) | M(累計額)

グループ分け

グループ数 5 取得件数 1

等間隔分割 等数分割 固定分割点

既存データからの再計算(等数分割)

ブロック 下限(≦) 上限(>)

結果列名の拡張子

OK Cancel

RFM 分析パラメータ設定画面

4. 新しいデータ操作アイコン

【日付時刻処理】

文字列型の日付時刻を読み込んで処理します。処理には、文字列から数値への変換、2つの日付の期間の計算、日付にオフセットを加えるという3種類の機能があります。

【行選択】

条件にマッチする行とマッチしない行に分けて抽出します

【日付選択】

条件にマッチする日付をもつ行のみを抽出します

【時系列操作】

基本操作	推移列とそれに対する統計量を計算します
線分表現	時系列データを線分化します
距離計算	線分化された時系列データのパターンとサンプルを比較して、パターンによく似たサンプルを抽出します

5. ユーザビリティ強化機能

【データインポート】

空行にもテキスト整形で対応

【集計】

列名に属性がつき、第 1,3 四分位点と歪度、尖度が計算可能に

【ファイル分割】

分割の際に用いる乱数の種が設定可能に

【改善されたスクリプト関数】

aggregate	第 1,3 四分位点と歪度、尖度が計算可能に
list	空のテーブル入力にも対応
sys_proc	スクリプトからアイコン名で呼び出せるようになった
guiInput	リストボックス、コンボボックスの画面が作成可能に
as.list	ソートを省略できるオプションを追加
format_file	2つのファイルを上下で結合できるようになった
save_plot	データビュー経由でグラフを外部ファイルへ出力

【その他の新しいスクリプト関数】

quantile	分位数を求める
chain_table	マスターキーがユニークである場合に、高速にマージを実行
dist_compute	2つのテーブルで各行の距離を計算する
is.list	データがリストかどうかを判定する
sys_proc_silent	sys_proc で出るメッセージを出力しない
ranking	テーブルのキー毎に順位付けを行う
date_proc_info	日付列の基本情報を出力
date_proc_period	2列の日付間の期間を計算する
date_proc_offset	日付列にオフセットを加えたものを出力する

****以前のバージョンへの影響****

Visual Mining Studio6.6 に新しく導入されたスクリプト関数と同じ名前は、すでに作成されたスクリプト内で、変数名として使用できなくなります。お使いの場合は置換機能などで変数名を変更してください。