

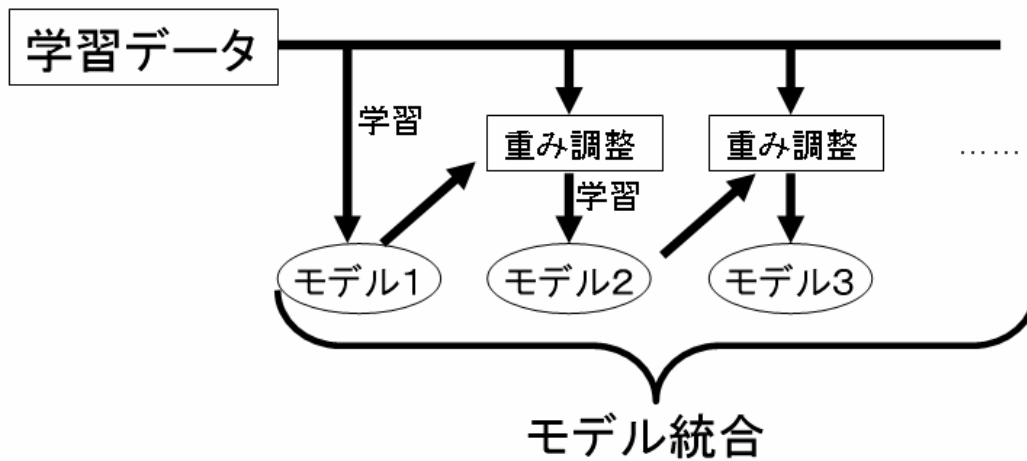
Visual Mining Studio バージョン 6.0 の変更点

2008年1月30日

(株)数理システム データマイニング部

1. 集団学習 Boosting/Bagging 機能

複数の既存の学習モデルを組み合わせることにより、高精度、優れた汎化能力を持つ判別モデルを実現



図：Boosting の学習メタアルゴリズム

特徴：

- VMStudio に既存の分類機能を弱学習モデルとして利用可能
- 重み調整、モデル統合機能のスクリプトによるユーザ定義が可能
- 他のスクリプトプログラムからの呼び出し（パラメータチューニング可能）

2. 分類モデル サポートベクトルマシンの改良

➤ nuSVM

サポートベクターの数を制御可能

➤ Support Ball Machine

クラスタリングによるデータ圧縮と Support Vector Machine を組み合わせた、大規模データに対する分類分析手法

3. その他データ探索機能

➤ Cluster Validation

クラスタリングの評価を行う

➤ 時系列分析機能 (S-PLUS、R)

S-PLUS、Rに含まれている時系列分析手法 (ARIMA など) に対するユーザインターフェースを提供

➤ データ可視化機能強化

データビューアで、大規模データの散布図を高速に表示

➤ パラメータ自動チューニング

各種のパラメータの組合せの生成から、モデル作成、モデル評価の一連の動作を自動的に行い、最適なパラメータを取得するためのスクリプト関数

sys_proc_group と sys_proc_map
が利用可能

4. 主なユーザビリティ強化機能

- | | |
|----------|--|
| 【データ表示】 | VMStudio 形式データファイルの表示
startDataSheetViewer.exe
startDataSheetViewerFront.exe |
| 【群間比較】 | ルール抽出時の条件入力等の機能改善 |
| 【システム全体】 | プロジェクト画面上のメモ画面のデフォルトデザインを設定可能 |
| 【DB連携】 | ODBC 経由で、MySQL との接続が可能 |
| 【決定木】 | 予測結果に新情報 (Node.重み、Node.全体占める割合) 追加 |
| 【システム関連】 | プロジェクト環境変数、ユーザ環境変数リストの取得
get_prj_env, get_usr_env |
| 【スクリプト】 | スクリプト定義文字列に対してのパーズ(parse)及び実行(eval) |
| 【スクリプト】 | スクリプト定義本文の取得、保存
get_script_def
read_from_text
write_to_text_file
print_text_file |

- 【スクリプト】 行選択関数 `sel` の拡張
`sel_case(<条件>,<結果>,...,<条件>,<結果>)`
- 【スクリプト】 G U I 画面によるテーブル表示
`view_data`
- 【スクリプト】 乱数発生シードを指定してのシャッフルが可能
`shuffle(table,seed)`
- 【スクリプト】 文字列、日付操作
`replace.all.str, char, format_str, format_date`
- 【スクリプト】 複数の外部ファイルを読み込み時に結合(列方向,行方向)
`read_binary, read_file`
- 【スクリプト】 一時ファイル、一時フォルダの取得
`get_a_temp_file, get_a_temp_folder`
- 【スクリプト】 外部ファイルの操作
`dirp`
`copy_file, copy_folder,`
`move_file, move_folder, make_folder,`
`folder_exist, file_exist`