

Visual Mining Studio バージョン 8.3 の変更点

2017 年 1 月

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「Model Optimizer」の追加

「Model Optimizer」アイコンは、分析自動化の一環として、最適なモデル構築に必要なアルゴリズム、パラメータ、変数選択を機械的に行う最新技術です。分析対象データに応じて予め指定した評価基準に基づいて、交差検証を行いながら、予測精度の高いモデルを構築すると同時に、ベイズ最適化の基本原則を用いて、探索の効率化も実現しています。

VMStudio8.3 では、下記のモデリングのアルゴリズムを候補としています。

- Decision Tree
- Random Forest
- Neural Network
- Support Vector Machine



2. K-Means++機能の追加

クラスタ分析の K-Means アイコンで K-Means++が利用できるようになりました。

K-Means++は初期のクラスタ中心を決める際、もっとも近い他のクラスタ中心との距離が最大となるように初期クラスタ中心を決めるため、安定した収束速度ならびに最適解が得やすくなることが知られています。

KMeans のパラメータ指定 GUI も刷新しました。

注意：プロジェクトに配置されていたこれまでの **K-Means** アイコンは自動的に新しい **K-Means** アイコンに置き換えられないため、オブジェクトブラウザから再度配置する必要があります。



3. 新アイコン「離散化」「離散化(適用)」の追加

数値データを離散化し、カテゴリデータに変換する離散化アイコンを搭載しました。これまで、等数分割しかできませんでしたが、「離散化」アイコンを使うことで、よりバラエティに富む離散化が可能になったため、カテゴリ変数と相性の良い機械学習をより有効に活用しやすくなりました。離散化方法としては、以下の分割手法で離散化を行うことができます。

- 等間隔
- 等数
- X二乗
- Gini 係数
- 誤判別数
- 群内二乗和
- 手動

このうち、X二乗・Gini 係数・誤判別数では、設定した目的変数に関してその値が最もよくなるように分割点を定めることができます。

「離散化(適用)」アイコンは「離散化」アイコンで決定した分割点に基づいて、別データに対して同じ条件で離散化を行うことができます。

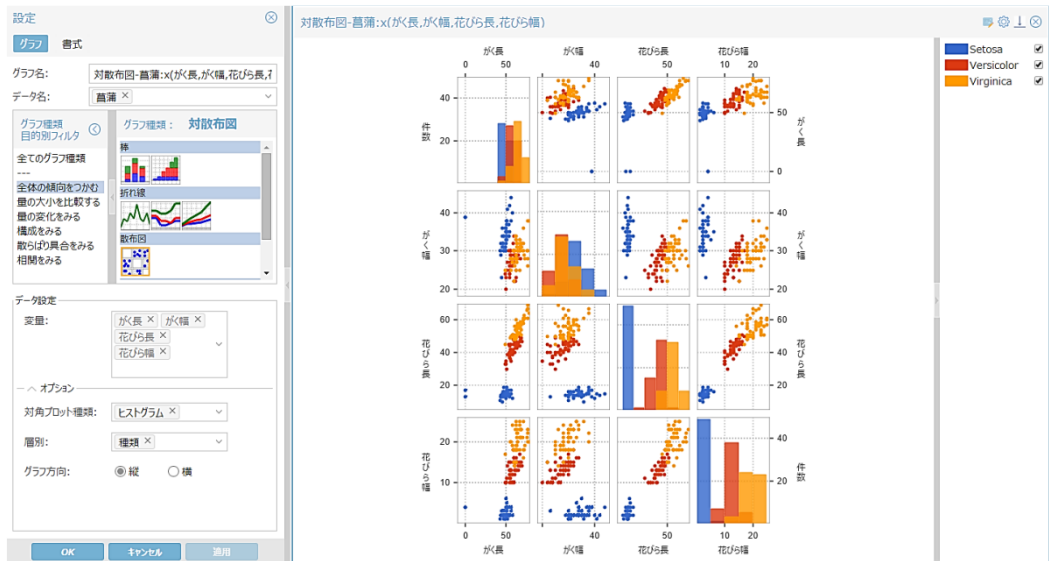


4. 新データビューアの搭載および「データ可視化」アイコンの追加

データビューアを刷新し、より簡単にグラフを作成できるようになりました。また、異なるグラフ間で対応するデータを確認することができるため、より深い分析も可能です。

各アイコンのデータビューアとして起動する他、データビューアを起動するアイコン「データ可視化」として利用することができます。

注意：初期設定では以前のデータビューが表示されるようになっています。各アイコンの右クリックメニュー「データビュー」の「旧型式で開く」のチェックを外すと新データビューアが利用できます。新データビューアを常に利用するには VAP のメニューバー「ツール」の「プロパティ」から表示される設定画面で可視化エンジンを切り替えてください。詳細はマニュアルをご参照ください。



新機能一覧

追加されたアイコン

Model Optimizer

ベイズ最適化を用いて、複数のモデル・パラメータ・変数から、より精度の高い組み合わせを探索します

離散化

数値データを離散化し、カテゴリデータに変換します

離散化(適用)

「離散化」アイコンで決定した分割点に基づいて、新しいデータに対して離散化を行います

データ可視化

グラフの作成が簡単になった新しいデータビューアを表示します

新しいスクリプト関数

summarize

高速に集計を行う summarize 関数を追加しました。従来の集計関数 aggregate と比較して、集計速度の向上に加えて、欠損効果の考慮および重み付き集計が可能です。

その他

Windows Server 2016 への対応

追加機能

改善された機能

K-Means

クラスタ中心の初期値選択に K-Means++が使えるようになりました。

Tree & RandomForest

各ノードの重みとノード番号が表示されるようになりました。

改善されたスクリプト関数

as.integer

16 進数などの文字列表記の n 進数を 10 進数へ変換できるようになりました。