

Visual Mining Studio バージョン 8.4 の変更点

2018年1月

株式会社 NTT データ数理システム

主な変更点

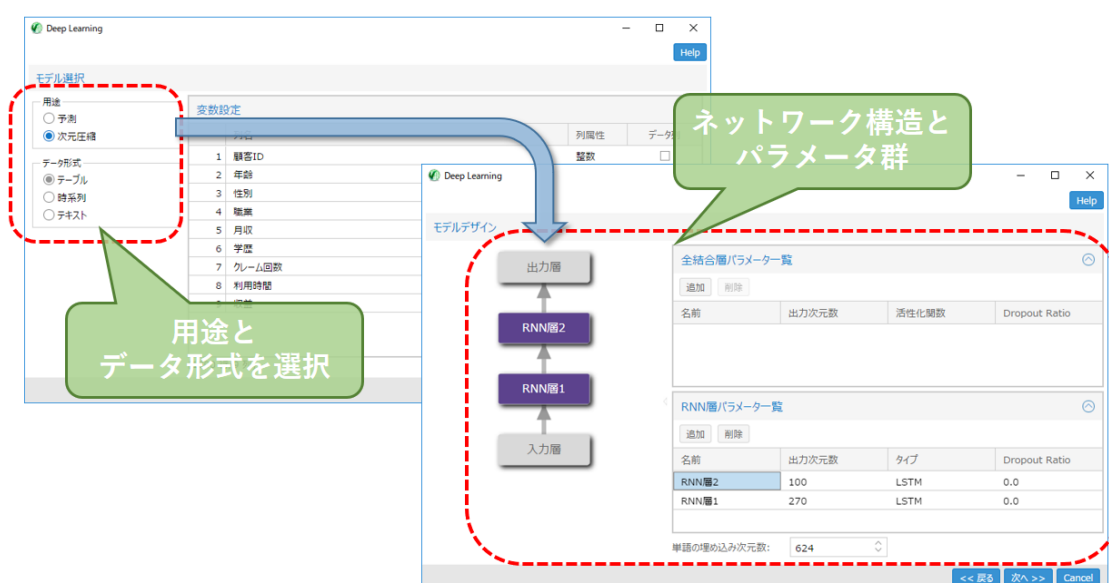
1. Deep Learning 機能

多層ニューラルネットワークによる教師あり学習・教師なし学習を行う Deep Learning 機能を追加いたしました。

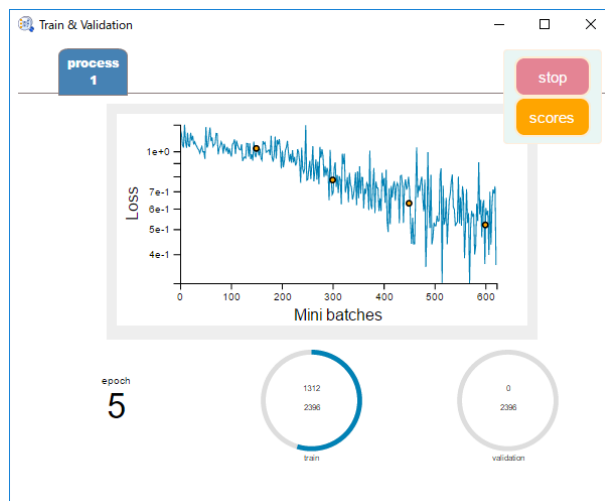
教師あり学習では、カテゴリ値の予測については判別モデル、数値の予測に対しては回帰モデルを構築します。教師なし学習では、データを次元圧縮し低次元化された表現を得ることができます。入力とするデータは、1行1件のデータである通常のテーブル形式に加え、可変長の時系列データや弊社テキストマイニングツール **Text Mining Studio** で分かち書きされたテキストデータも扱うことができます。

多層パーセプトロン、Autoencoder、Recurrent Neural Network (RNN)、Encoder-Decoder モデル、といった構造のネットワークを備えていますが、パラメータ設定画面上でモデルの用途とお手持ちのデータの形式を選択していただくと、その目的に適ったモデルの構造を自動的に選択いたします。

ネットワークの各層は多様なパラメータを持ち、様々なモデルを作成することが可能ですが、最良なパラメータを自動的に探索する Model Optimizer 機能をあわせて搭載しており、パラメータチューニングの効率化と省力化を実現しています。



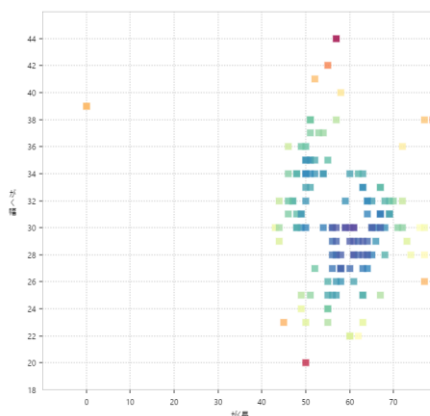
また、学習の進行中は、リアルタイムでその進捗状況が表示されるため、一般に長い処理時間を必要とすることの多い Deep Learning の学習時間を見積もることができます。



本機能は、2018年9月30日までご利用いただけます。2018年10月以降は、さらなる機能拡充を行ったうえで、単体の製品としての販売を予定しております。ご期待ください。また、本機能は64Bit OS をご使用の場合のみお使いいただけます。

2. Isolation Forest

外れ値・異常値検出を行うための新手法です。外れ値は他の点と離れており、空間をランダムに順次分割していったときに比較的早く孤立しやすい、逆に正常値は近隣に似たような点が多数存在しているため分割を何度も繰り返さないと孤立しない、という傾向を前提としています。ランダムな分割を繰り返し試行し、各々の点について「データが孤立するまでの分割数」の平均値からデータの異常度を算出します。個々の点の間の距離計算が不要であり、非常に高速に動作いたします。下図は、Isolation Forest により求めた異常度に基づいて赤(異常)～青(正常)で点の色付けを行った例です。

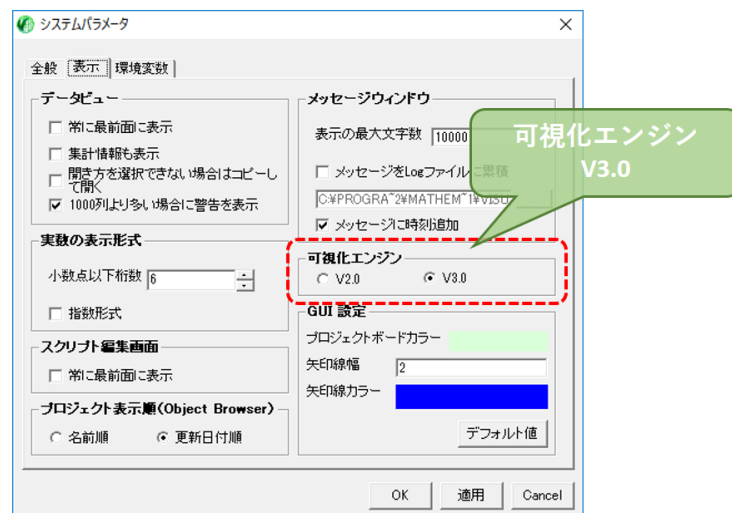


3. 「データビュー」の動作変更

各アイコンから右クリックメニューの「データビュー」で表示される機能が更新されました。従来「データ可視化」アイコンの機能になっていた新しい可視化機能が、各アイコンの「データビュー」から呼び出されるようになりました。

利用方法の詳細については、メニューの ヘルプ(H)→データ可視化マニュアル をご参照ください。

旧バージョンの「データビュー」機能とは メニュー の ツール(E)→プロパティ→表示「可視化エンジン」項目で切り替えることができます。”V3.0” という項目が新しい可視化エンジンに相当いたします。使い勝手の向上と洗練されたグラフ表示を是非ご体感ください。



新機能一覧

追加されたアイコン

Deep Learning

Deep Learning 予測

多層ニューラルネットワークを用いて予測・回帰・次元圧縮を行います。

「Deep Learning」アイコンでモデルを構築し、「Deep Learning 予測」アイコンで予測対象のデータにモデルを適用させることが可能です。

Isolation Forest

Isolation Forest 判定

外れ値検出を行います。「Isolation Forest」アイコンで外れ値検出のためのモデルを構築し、「Isolation Forest 判定」アイコンで判定対象データに異常度を算出して付与することができます。

その他

「データビュー」の動作変更

各アイコンの右クリックメニュー「データビュー」で表示される表・グラフの表示機能が新しくなりました。従来「データ可視化」アイコンとして独立に提供していた可視化機能が、各アイコンの「データビュー」から起動するようになりました。