

New in Viscovery® SOMine 7.1 Data Mining Suite

Viscovery® SOMine 7.1 では以下の機能が追加されました。さらにバージョン 7.0.2 で起きていたエラーが修正されました。

Visual Cluster

- 統計ウィンドウの相関行列の表現が変更されました。属性の各対のリストだったのが、行列を表示するようになりました。行列の行と列の属性を別々に選択できます。
- ヒストグラムの調整ダイアログのレイアウトが変更されました：より多くの属性を同時に表示できるように、属性のリストが高さいっぱい表示されるようになりました。さらに下部パネルの高さも変更されました。
- ヒストグラムの調整ステップのレコードの消去ページのリスト・エントリーをドラッグして新しい位置に並べ替えられるようになりました。
- 表の列を表示/非表示できるようになりました。リスト・ヘッダを右クリックすると列をオン/オフできるポップアップ・メニューが出てきます。
- ワークフロー間の設定の違いをより詳しく表示するようになりました。さらに同じ先祖に直接つながっていない（が、間接的にはいくつかの先祖とつながっている）ステップ間の違いも表示されるようになりました。
- 名義値の定義ダイアログで、名義値と定義された別名の横に度数が表示されるようになりました。
- 前処理ワークフローの属性の定義ステップの新しい属性の定義ダイアログに、複製機能が追加されました。これは数式のコピーを作成するのに便利な方法を提供します。
- 数式で新しい関数が使用できるようになりました：
 - `find(what, where, start)` および `pos(what, where, count)` は、文字列が見つかった文字位置を返します。
 - `substr(str, pos1, pos2)` は、指定された位置の間の入力の一部を返します。
 - `mean(n1, ...)`, `sum(n1, ...)`, `count(n1, ...)` は、任意の数の引数を取って、平均、合計および非欠損値の数を計算します。すべての値が欠損値の場合、`mean` は欠損値を返し、`sum` と `count` はゼロを返します。
- 関数 `min(n1, ...)` および `max(n1, ...)` は、すべての非欠損値の最小と最大を計算するのに、任意の数の引数（最低 1 個）を与えることができました。すべての引数が欠損値の場合、結果は欠損値のみです。
- 関数 `min`、`max`、`mean`、`sum`、`count` は、区切り文字で文字列を分割でき、その部分で（それらを数字に変換した後）演算できるようになりました。
- マップ・ウィンドウやプロジェクト・ウィンドウでポップアップをどれだけ長く表示するかを設定するオプションがプリファレンス・ダイアログに加わりました。
- 前処理ワークフローの属性の定義ステップの名義属性の定義ダイアログで、データマート中で使用する属性の並び順が、1 個または複数のハイライトされた属性を順序内の新しい位置にドラッグすることでも定義できるようになりました。属性の順序を定義する機能は、すでに同じワークフロー・ステップの新しい属性の定義ダイアログで利用できていましたが、新しい機能がより便利です。
- 名義属性の定義ダイアログで列ヘッダをクリックして属性を並べ替える機能は、1 回のマウス・ク

リックで手動設定した属性の順序を壊してしまう事故を防ぐために取り除かれました。

- 前処理ワークフローのデータマートの書き出しステップの保持する属性の選択ダイアログで、属性の順序を配置する機能が取り除かれました。このダイアログで列ヘッダをクリックして属性が並べ替えられたとき、データマートでの順序は変更されません。属性の順序は、ワークフローの前のステップで出てくる名義属性の定義ダイアログで調整できます。
- 古いプロジェクト・ファイルがアップグレードされる時、(ユーザーが古いファイルを上書きしないように選択すると) 古いプロジェクト・ファイルがまだ以前のバージョンで使用できるように、参照される SOM モデル・ファイルが最新のファイル・バージョンにアップグレードされるようになりました。
- プリファレンス・ダイアログのタブが、活用パターンによりフィットするように再配置されました。
- スコア・チャートで属性の複合選択ができるようになりました。
- 散布図が、水平の線に沿った区間の中央ではなく平均値で傾向線のポイントを描くようになりました。
- チャート・リジェンドの整列を変更する機能が取り除かれました。テキストは、現在、常に左揃えになります。
- チャートの軸ラベルのサイズを自動で決定するオプションは、過度に大きな軸ラベルを避けるために取り除かれました。

SOM モデルの編集

- マップ・ピクチャの上にマウス・ポインタを置いたときに、複数のドキュメント (画像) のサムネイルを表示できるようになりました。サムネイルのサイズ、数、配置は、プリファレンス・ダイアログの表示タブで調整できます。サムネイルの順序は、データレコード・ウィンドウでのレコードの並べ替えに影響されます。
- 新しい SOM が最初に表示される時に示されるウィンドウの配置をカスタマイズできるようになりました。
- マップ・ピクチャ内のラベルのデフォルト・フォント・サイズが、より読みやすく大きくなりました。
- SOM の統計ウィンドウが、データマート内にあるすべての属性を分析するようになりました。その結果、データマートを記述統計についても精査できるようになりました (すべての記述統計がすでにモデル中に存在するので、以前はこれが必要ではありませんでした)。
- 列を表示ダイアログがデータマート属性に名前を変更され、表示 | データマート属性から呼び出せるようになりました。このダイアログで定義される属性の選択と順序が、データレコード・ウィンドウ内の列と、統計ウィンドウのさまざまなタブに一覧される属性に適用されるようになりました。
- 統計ウィンドウの記述統計または PCA タブで属性の一覧が並べ替えられた場合、その順序がデータレコード・ウィンドウの列にも反映されるようになりました。この順序が配置の中で保持されます。
- マップ・ピクチャをその視覚的な類似性で並べ替える機能 (マップ | 属性、類似度ボタン) が、表示マップ・ピクチャを優先させるようになりました。非表示マップ・ピクチャも類似度を分析しますが、最も類似した表示マップ・ピクチャの後に順序づけされます。
- 以前のバージョンでは、SOM モデルが多数の属性 (数千) と多数のクラスを持つ場合、クラス

タ・ウィンドウを開くのに時間がかかりました。このウィンドウを開く時間がかなり短縮されました。

- データレコード・ウィンドウを開いた状態で、SOM モデルで後処理の数式を定義するとき、数式が直ちに表示されるようになりました。以前は、データレコード・ウィンドウを閉じて、再び開いてから、これが表示されていました。
- マップのデータレコード・ウィンドウで表示されるデータをソートする属性は、レコードの一覧を変更したり、ウィンドウを閉じたりしてもリセットされなくなりました。さらに、ソートの状態が配置の中で記録され、異なる配置に切り替えたときにリストアされるようになりました。
- モニタリングの際（マップ | モニタ）、現在のレコード番号が表示されるようになりました。
- 以前はノード・データを（編集 | ラベル、コピー、パスから）コピーして、同じマップに貼り付けて戻したとき、ノードの位置が正確にはリストアされていませんでした。これが現在は忠実に位置を保持するように修正されました。また“置換を使用しますか?” という質問もなくなりました。
- 通常、グループ・プロファイル・ウィンドウのプロファイル列を p 値と組み合わせてみることに意味があるので、プロファイル列が右に 2 列分ずらされました。
- ユーザー・インタフェースは、現在、近傍ではなく最近接ノードという用語を使用するようになりました。なぜなら近傍という用語は、データ分析や統計では特別な意味を持つからです。

Explore and Classify

- 分類ワークフローで作成された SOM モデルが、ファイル | 名前を付けてコピーを保存を用いて新しい名前で作成できるようになりました。このコピーは、他の SOM モデルとして読み込んで修正できますが、その修正は元のワークフローには影響しません。これは、最初にデータ・サンプルから SOM モデルを作成して、そのモデルを他のデータレコードに適用して、たとえばビッグ・データのための SOM データ表現を作成することを可能にします。
- 分類ワークフローの分類をエクスポートとスコア・ワークフローのスコアリングのエクスポートの結果は、常にアプリケーション・データマートに書き戻されていました。これは現在、分類ワークフローとスコア・ワークフローのモデルの選択ステップで行うようになりました。したがって、分類やスコアリングの結果は、これらのワークフロー・ステップで、統計ウィンドウやデータレコード・ウィンドウを用いて調査できます。結果をエクスポートする必要はなくなりました。
- 分類ワークフローのモデルの選択ステップで、モデルに連想する属性を選択できるようになりました。数値属性に加えて、現在はデータマートで名義属性として定義された属性を選択することもできるようになりました。
- 分類ワークフローで、アプリケーション・データマートから収集したデータの統計量を含む新しいリード・オンリー・モデルを作成できるようになりました。以前は、モデル・データの統計にアプリケーション・データの統計を追加されていました。この新しいリード・オンリー SOM モデルでは、モデル・データの統計は提供されません。これは、いくつかのケースで生じる統計ウィンドウでの矛盾を避けます。以前提供されていた“(適用済み)”でマークされた属性の統計量は、現在はオリジナルの属性名で利用可能で、以前オリジナルの属性名で提供されていた統計量はもう利用可能ではありません。
- 以前、分類が評価される時、結果のデータで記録された実際のクラスタの名前が、モデルで使用された名前と同一である必要がありました。この制限が緩められました。異なるクラスタ名をデータ内で使用することができるようになりました。ただし、名義値の定義は、結果のデータを読み込

む前処理ワークフローの属性の定義ステップで確立されなければなりません。ここで、元の値がモデルのクラスタ名に定義されなければならず、別名はそれらがデータ内で発生するのと同じ名前です。

- 後処理の数式がタイプを割り当てられるようになりました。これらのタイプは、分類ワークフローの分類のエクスポート・ステップで、数式の結果がエクスポートされる時に使用されます。既存の SOM モデルの既存の後処理の数式でタイプを定義することは必須ではありません。そのようなケースでは、結果は文字列表現に変換され、文字列としてエクスポートされ続けます。
- クラスタ間での属性値チャートの属性のリストが配置替えされました。現在、チャートはより多くの色を使用しています。
- アプリケーション・データマートで属性に定義された変換が、分類ワークフローで属性がモデルに連想されるときに適用されるようになりました。

Predict and Score

- スコア・ワークフローのアプリケーションの確認ステップで、新しく最適利得チャートが利用可能になりました。このチャートは、予測レスポンスが最大化されるように、データレコードをキャンペーン用にグループ分けすることができます。
- ローカル・モデルの計算ステップの属性の重みづけダイアログで、既存の SOM モデルから重要度を初期化できます。以前は、選択できる SOM モデルが、同じグローバル・モデルの計算ステップに由来する兄弟ワークフローで作成されたそれらに制限されていました。この制限が解除されました。現在はその SOM モデルでも選択できます。重要度は先行するグローバル・モデルの計算でのグローバル・モデルで使用されたのと同じ名前を持つ SOM モデル内のそれらの属性から取得されます。
- スコア・グループの定義チャートで分割線を指すと、スコア・グループの限界の正確な位置を表示するようになりました。
- スコア・グループ・セパレータをドラッグすると、分割される場所のレコード数をポップアップで表示するようになりました。
- グローバル回帰の結果を一覧する表が、列ヘッダをクリックして並べ替えできるようになりました。
- 既存のマップ・オプションを用いて（ローカル・モデルの計算ステップの最適化戦略の定義ダイアログのマップに基づくローカル・モデルのセクション）、1つのマップ・トレーニングを保存できるようになりました。以前は、同じグローバル・モデルの計算ステップに由来するワークフロー・ステップで作成されたマップのみが可能でした。この制限が解除されました。他のマップが同じ変数名を持ち、グローバル回帰モデルで使用されたそれらの変数の前処理を持つかぎり、他のマップを使用することができます。
- 以前は、2とおりの前処理オプションでスコア・チャートが計算されました：モデル内に存在する属性について、モデルの名義値定義とデータマートの値の置換が使用されました。現在は、名義値定義も値の置換もモデルから取得されます。
- ローカル回帰モデルの係数ウィンドウ（予測ワークフロー）が品質パラメータの複数のピクチャを表示するようになりました。これらのピクチャでの白い背景色の使用が、パラメータに“意味がない”ときに一律に示すために、幾分変更されました。
- 以前のバージョンでは、いくつかのチャートで軸ラベルの位置がオプションではありませんでした。これが修正されました。

- 予測ワークフローでのさまざまなステージで表示される最小値と最大値が、以前はデータセット全体から計算された値を表示していました。しかしながら、平均、標準偏差および欠損値の数は、モデル・データ分割のみから計算されていました。これが、最小値と最大値がモデル・データ分割からのみ計算されるように修正されました。
- ローカル・モデルを伴う SOM が、予測ワークフローのローカル・モデルの計算ステップで以前に保存された名前を用いて開かれているとき、非線形診断とローカル回帰要約が表示されませんでした。これらが現在は利用可能になりました。ただし、Viscovery SOMine,の以前のバージョンで作成されたモデルでは、その情報は現在も表示されません。

Enterprise Data

- 極めて大きな Microsoft Excel files (*.xlsx) からのデータ・インポートが可能になりました。
- データマートのエクスポート手順が、現在選択されているワークフロー・ステップに応じてエクスポートするデータマートを提案し、目的のファイルのファイル名を提案するようになりました。

Workflow Automation

- Workflow Automation service が、パスワード認証を必要とするメール・サーバーから通知メールを送信できるようになりました。ただし、メール・サーバーを設定するために Workflow Automation によって送信されるメッセージは暗号化されません。

互換性の注意

- 予測ワークフローのデータマートの選択ステップでデータをモデル・データ、テスト・データ 1、テスト・データ 2 にランダムに分割するアルゴリズムが変更されました。以前は、100 レコードのバッチがシャッフルされて、与えられた正確なパーセンテージで分割に割り当てられていました。新しい方法は、パーセンテージで与えられた確率に応じて、各レコードを別々に分割に割り当てます。
- ファイル | エクスポート | マップのノード値機能が、再構築された名義値ではなく、名義属性の 2 値のノード値を書き出すようになりました。一般的に、これがより便利です。再構築された名義値によるノード値は、編集 | すべてを選別 の後、ファイル | エクスポート | 選別 でエクスポートできます。
- 関数 $\min(a, b)$ および $\max(a, b)$ が欠損値の引数を無視するようになりました。以前は、1 つでも引数が欠損値だと欠損値を返していました。この変更は、Viscovery SOMine の以前のバージョンで作成されたプロジェクトやモデルにも適用されます。
- 関連補償オプションが現在はデフォルトでオフになっていることに注意してください。プロジェクトが読み込まれると、そのプロジェクトで記録された設定がまだ留意されています。
- 未完了（黄色または白色）のワークフローでの選択のリセットは、続くすべてのステップ（すなわち、そのステップの右側のそれら）をリセットします。