

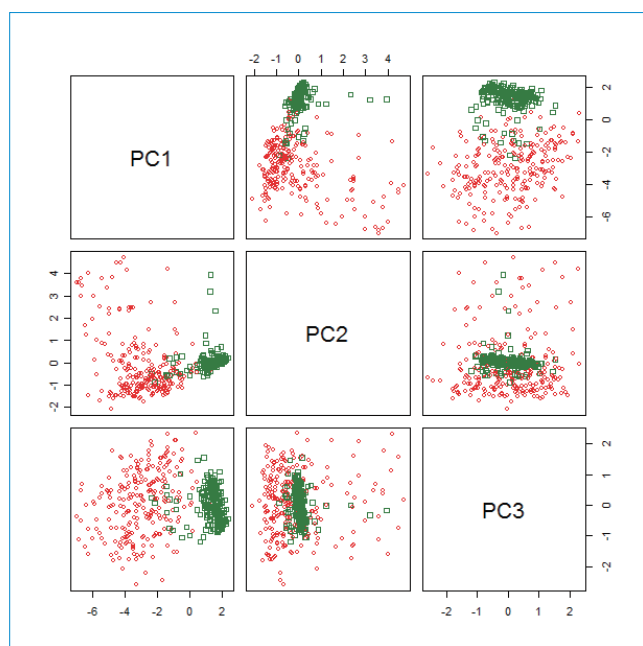
BIOVIA PIPELINE PILOT R STATISTICS

データシート

BIOVIA Pipeline Pilot R Statisticsを使用すると、統計的な解析や結果のグラフィック表示を行い、その結果に基づいた意思決定を行えるようになります。このコレクションには、Rによるクラスタリング等のデータ解析や、統計モデル構築などの統計手法を実装したコンポーネントが含まれています。基盤となる統計エンジンは、広く使用されているオープンソースパッケージであるRです。このコレクションを使用すると、Rの統計分析とデータ操作のメソッドをBIOVIA Pipeline Pilot のデータストリームに適用できます。ユーザーは、Rの出力結果をパイプラインに直接取り込み、BIOVIA Pipeline Pilotフレームワークのコンポーネントを使用して詳細に分析することができます。既存のRスクリプトをBIOVIA Pipeline Pilotのカスタムコンポーネントで使用できるため、別のプロトコルで再利用したり、組織内で共有することができます。

BIOVIA PIPELINE PILOT R STATISTICS を使用すると、次のことが実現できます。

- ヒートマップで値の相関を表示し、最も関連性の高いものを検出する
- Box Plotを使用してデータの分布を表示する
- 分散分析(ANOVA)を実行して複数のデータセット間の平均値に見られる差異を特定する
- ロジスティック回帰分析、サポートベクターマシン(SVM)、ニューラルネットワーク、その他10種類の学習メソッドのいずれかを使用してデータをモデル化する
- モデルが適切に適用されるよう支援する Model Applicability Domain(MAD)をサポートし、構築したモデルを適用して新しいデータセットを予測する
- 訓練データを任意のモデルに保存し、実験データの追加に伴って拡張できるようにする
- 多様なクラスタリング手法を適用する
- 独自のRスクリプトをBIOVIA Pipeline Pilotの個別のデータレコードやデータストリーム全体に適用する

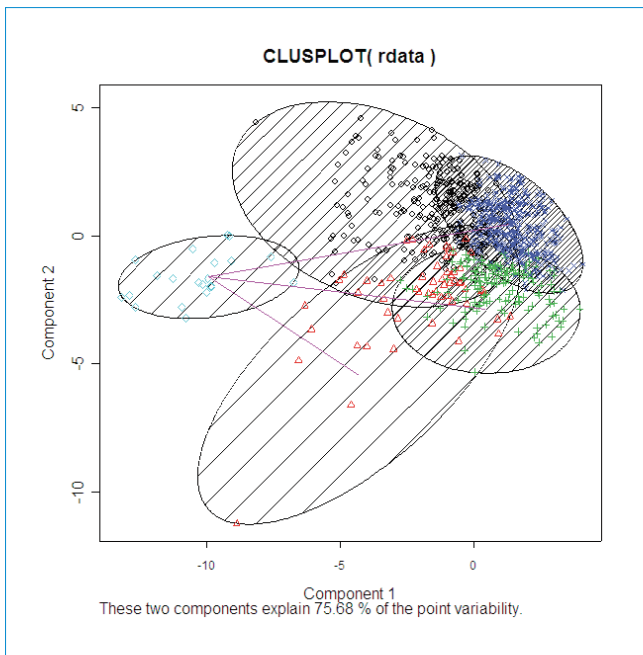


乳がんデータの主成分分析結果を組み合わせで表示

分析

複数の測定結果を比較するためにt検定やANOVAを使用して有意性分析を行い、異なる測定結果の平均が同じであるかどうかを確認できます。R Correlation Matrixコンポーネントでは、一連の記述子に関するマトリックスを生成して記述子の相関度を確認したり、ヒートマップを作成して記述子空間におけるパターンを視覚的に確認することができます。このコンポーネントには次のものが含まれています。

- R ANOVA
- R K-Nearest Neighbors
- R Correlation Matrix
- R Principal Components Analysis
- R Probability Distributions
- R One-variable Tests
- R Factor Analysis
- R Two-variables Tests



Cluster Plot

クラスタリング

Rを使用することで、あらゆる種類のBIOVIA Pipeline Pilotデータと併用できる様々なクラスタリング手法を提供します。たとえば、フィンガープリントをRの階層的クラスタリング手法やk平均クラスタリング手法の記述子として分子データセットに使用することもできます。このコンポーネントには次のものが含まれています。

- R Cluster Agnes
- R Cluster Clara
- R Cluster Diana
- R Cluster Fanny
- R Cluster PAM
- R Cluster K-Means

データ操作

データセットが不完全だったり、不要な情報が含まれていたり、その他の理由により不規則である場合、Rデータ操作コンポーネントを使用して欠落値を置き換えたりデータをスムージングすることができます。このコンポーネントには次のものが含まれています。

- R Missing Values
- R Loess Smoother
- R Remove Zero-Variance Properties
- R Spline Smoother
- R Smoother
- R Friedman SuperSmoother

チャート

チャートは統計結果の分析や報告において非常に重要な役割を果たします。このコンポーネントでは、HTMLビューアに表示したりレポートに表示したりレポートに埋め込むことができるPNG画像を作成します。このコンポーネントには次のものが含まれています。

- R 2D Plot
- R 3D Plot
- R Pairs Plot
- R Conditional Plot
- R Histogram
- R Parallel Coordinates Plot
- R XY Plot

学習モデル

BIOVIA Pipeline Pilot Data Modeling&Statisticsでは、ニューラルネットワーク、サポートベクターマシン(SVM)のほか、数種類の統計的学習理論のモデルタイプを追加しました。分類の問題と回帰の問題に対する様々な手法がサポートされています。このコンポーネントには次のものが含まれています。

- Learn R Linear Model
- Learn R Linear Discriminant Analysis Model
- Learn R Generalized Linear Model
- Learn R Neural Net Model
- Learn R Non-Linear Model
- Learn R Support Vector Machine Model
- Learn R Logistic Regression Model
- Learn R Partial Least Squares Model

多次元スケーリング

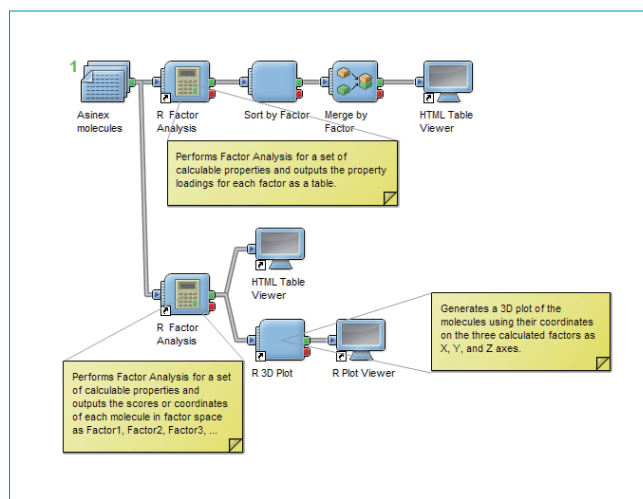
データセットの特性は、データアイテム間のペアワイズ距離を含むマトリックスで示すことができます。BIOVIA Pipeline Pilot Data Modeling&Statisticsでは、元の距離を可能な限り保持しながらデータを低次元空間に配置する多次元スケーリングを提供しています。このコンポーネントには次のものが含まれています。

- R Classical MDS
- R Sammon
- R Nonmetric MDS
- R Self Organizing Map

カスタマイズ

BIOVIA Pipeline Pilot Data Modeling&Statisticsはすべて、サブプロトコルとして実装されます。Rスクリプトを記述できるのであれば、コンポーネントの変更やカスタマイズを行って、さらに多くのRの機能をBIOVIA PipelinePilotに取り入れることができます。また、次の2つのコンポーネントを使用すれば、BIOVIA PipelinePilotのデータストリーム全体に、またはコンポーネントに入力される各データレコードに対して、Rスクリプトを適用できます。

- R Custom Script
- R Custom Script for Each Data



Asinexデータの因子分析を示すプロトコルをBIOVIA Pipeline Pilot Data Modeling&Statisticsで構築

BIOVIA PIPELINE PILOTの概要

BIOVIA PipelinePilotは、さまざまな場所に保存されているデータから科学的価値を引き出し、科学的ワークフローを自動化して、より広範な科学コミュニティでのコラボレーションを促進することにより、研究開発組織の技術革新を支援する、拡張性に富んだ大規模サイエンティフィック・インフォマティクス・プラットフォームです。BIOVIA PipelinePilotのコンポーネントコレクションはプラットフォームの科学的な構成要素あり、科学的なカテゴリや機能でグループ化されています。コンポーネントをグラフィカルに組み合わせることで、データの取得、フィルタリング、分析レポート作成のワークフローを作成できます。

ダッソー・システムズの**3D**エクスペリエンス・プラットフォームでは、**12の業界**を対象に各ブランド製品を強力に統合し、各業界で必要とされるさまざまなインダストリー・ソリューション・エクスペリエンスを提供しています。

ダッソー・システムズは、**3D**エクスペリエンス企業として、企業や個人にバーチャル・ユニバースを提供することで、持続可能なイノベーションを提唱します。世界をリードするダッソー・システムズのソリューション群は製品設計、生産、保守に変革をもたらしています。ダッソー・システムズのコラボレーティブ・ソリューションはソーシャル・イノベーションを促進し、現実世界をより良いものとするためにバーチャル世界の可能性を押し広げています。ダッソー・システムズ・グループは140カ国以上、あらゆる規模、業種の約19万社のお客様に価値を提供しています。より詳細な情報は、www.3ds.com（英語）、www.3ds.com/ja（日本語）をご参照ください。



3DEXPERIENCE®

©2014 Dassault Systemes. All rights reserved. 3DEXPERIENCE, CATIA, SOLIDWORKS, ENOVIA, DELMIA, SIMULIA, SIMULIA, GEOVIA, EXALTED, 3D VIA, 3DSWIM, BIOVIA, および INETVIBES はアメリカ合衆国、またはその他の国における、ダッソー・システムズまたはその子会社の登録商標または商標です。その他のブランド名や製品名は、各所有者の商標です。画面による所定の承認が必要ですが、

 **DASSAULT SYSTEMES** | The **3DEXPERIENCE**® Company

CTC

▶ Challenging Tomorrow's Changes

伊藤忠テクノソリューションズ株式会社
流通・エンタープライズ事業グループ
ライフサイエンス事業部

141-8522
東京都品川区大崎1-2-2 アートビレッジ大崎セントラルタワー

TEL: 03-6417-6600
E-mail: ls-marcom@ctc-g.co.jp