

DATA MODELING COMPONENT COLLECTION

データシート

BIOVIA Pipeline PilotのData Modelingは、一連の学習やモデリングの機能、統計フィルター、可視化ツール、大規模データに最適化されたクラスタリングコンポーネントなどを提供します。このコレクションによってBIOVIA Pipeline Pilotの通常の機能を拡張し、データマイニングのための統計や予測モデルの機能を取り込むことが可能になります。

DATA MODELINGを使用すると、次のことを実現できます。

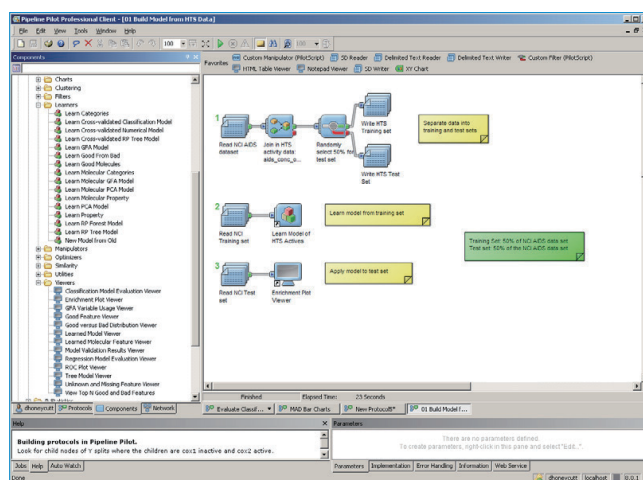
- 迅速なクラスタリング
- ベイジアン統計を用いた分類モデルの構築
- 主成分分析(PCA)
- 線形、部分最小二乗法(PLS)、k近傍法(kNN) による回帰分析

作成したモデルの品質を、ROCプロット、エンリッチメント・プロットなどでビジュアル化 Chemistry Collection を一緒に用いることにより、次のことが可能になります。

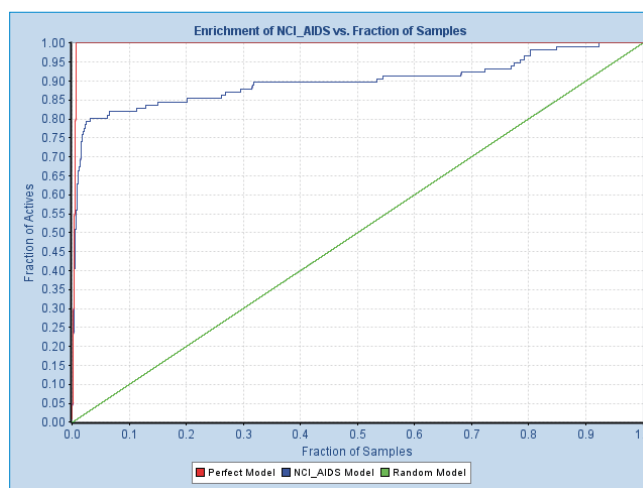
- 構造活性相関モデリング
- 化合物のクラスタリング
- 類似化合物の検索

ベイジアンモデリング

BIOVIA Pipeline Pilotのベイジアンモデルによって、非常に大きなデータセットに対する分類学習が可能になります。非常に早く予測モデルを構築することができ、そのモデルを化合物ライブラリに適用することによって、今後試験すべき化合物を同定することができます。Chemistry Collectionに含まれる化合物フィンガープリントの計算機能と組み合わせることにより、化合物取得やHTSのための理想的な環境を提供します。ベイジアンモデルでは、単純な二値分類（たとえば活性あり、なしなど）だけでなく、3つ以上のクラスや多数の評価項目への分類を一つのモデルで実施することができます。例えば、化合物データベース全体に対するあらゆる活性を予測するモデルを構築し、次にそれぞれの評価項目に対する活性値を予測することによって、一群の化合物をプロファイリングすることが可能になります。



Data Modelingを用いて予測モデルを構築し、データに含まれる重要な特徴を理解したり、リードとなる可能性のある化合物を発掘したりできます。



ここに示したEnrichment Plotのように、構築したモデルを解釈するために作られた多くのViewerを含んでいます。

回帰モデル

連続的に変化するデータに対する回帰モデルを構築し、公表することは容易な作業です。ベイジアンモデルと同様 BIOVIA Pipeline Pilot上で構築したモデルをドラッグアンドドロップにより、あるいはメールに添付することにより、社内に配布できます。線形回帰、部分最小二乗法(PLS)、k近傍法(kNN)など、様々な手法を選択することができます。回帰や分類モデル構築のためのさらなるコンポーネントが、Advanced Data ModelingやR Statistics Collectionsによって利用可能です。

モデル評価

ベイジアン学習機は、モデルの質を評価するための高速なcross-validation機能を備えています。特別に用意されたcrossvalidation、コンポーネント群を利用することにより、学習パラメータ(PLS、コンポーネントの数など)を調節したり、別の手法を用いたモデルを評価したりすることが可能です。モデルの種類によってRegression Model Evaluation ViewerやClassificationModel Evaluation Viewerを使い分けることにより、複数のモデルの予測精度を比較することができます。全ての学習機がModel Applicability Domain (MAD)機能を備えており、これによって予測対象モデルの適用範囲にあるのか否かを確認することができます。学習データはモデルと一緒に保存することができるため、新たな実験データが得られた際、モデルを構築し直すことが可能です。

主成分分析

主成分分析(PCA)は、データの次元数を削減する標準的な手法です。データを圧縮したり、変数間の相関を排除したり、あるいはデータを可視化したりするために用います。

類似構造検索

Chemistry CollectionをData Modeling Collectionと共に用いることによって、化合物ライブラリに対する類似構造検索を分子の特徴や分子フィンガープリントを用いて行うことができます。ユークリッド距離やマハラノビス距離など、いくつかの方法で分子間の距離を定義し、類似構造を検索することができます。

データの可視化

Data Modeling Collectionは、構築したモデルを解釈するために作られたいくつかのViewerを含んでいます。これらのコンポーネントには、ROCプロット、エンリッチメント・プロットなど、視覚的にモデルの質を評価するための複数のViewerが含まれています。Chemistry Collectionをお持ちであるなら、ベイジアンモデルにおいて最も活性に寄与する構造的特徴をハイライトするGood Feature Viewerなど、さらにいくつかのViewerを用いることができます。

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。

BIOVIA PIPELINE PILOTの概要

BIOVIA Pipeline Pilot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 Pipeline Pilot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であり、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。

