

# ADVANCED DATA MODELING COMPONENT COLLECTION

データシート

BIOVIA Pipeline Pilot の Advanced Data Modeling では、再帰分割 (Recursive Partitioning; RP) 分類モデル、Genetic Function Approximation (GFA) 回帰モデル、および多目的パレート最適化(Pareto Optimization)用のコンポーネントを提供します。RPコンポーネントでは、単一あるいは複数の特性を記述するための、単一の決定樹あるいはフォレストモデルを構築する様々な手法を提供します。GFAコンポーネントでは、洗練された遺伝的アルゴリズムを用いて変数選択や複数のモデルの構築を行い、さらにそれらのモデルをコンセンサス、あるいはアンサンプルモデルに統合することにより、より精度の高い予測をすることができます。 Pareto Optimization Componentは、多目的最適化のための手法を提供し、複数の相反する目的間の最適な妥協点(トレードオフ)を与えます。

## RECURSIVE PARTITIONING COMPONENTを使用すると、次のことを実現できます。

- 分子フィンガープリントを付与された化合物ライブラリのような、多数の記述子によって表現された大きなデータセットに対して、高速にモデルを構築したりデータマイニングを実施したりできます。
- 構築した決定樹モデルを表示し、記述子の寄与を理解することができます。
- 変数の重要性を解析し、データ分割への寄与率が最も大きい記述子を同定することができます。
- 新たなデータセットに対する予測を高速に実施するとともに、Model Applicability Domain(MAD)を示すことによって、モデルが正しく適用されているかを確認することができます (これはGFAモデルについても同様です)。

## GFA COMPONENTで

### 以下のことが可能になります。

- 多くの試行の結果、単一の“最善”モデルではなく複数のモデルを返すことにより、さらなる検討のための多くの仮説を集めることができます。
- 複数のモデルを一つのアンサンプルモデルとすることができ、これはしばしばそれを構成するいかなる単一モデルよりも良い予測精度を示します。
- モデル集団の進化に従って、用いられている記述子の統計をプロットすることにより、モデルに対して最も寄与率の高い記述子に関する洞察を得ることができます。

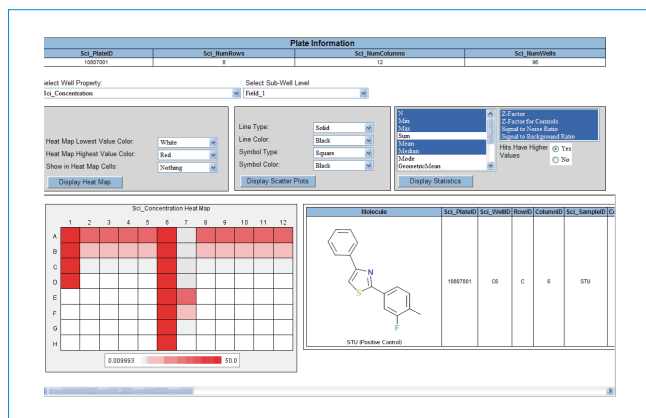
## PARETO OPTIMIZATION COMPONENTで

### 以下のことが可能になります。

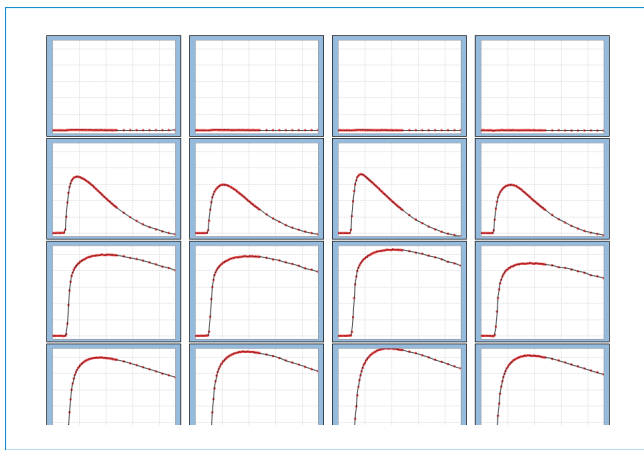
- コンビナトリアルケミストリーを用いたライブラリ設計における多様性最大化、製剤設計における成分調整最適化、および株式ポートフォリオにおけるリスク管理等の問題に、より適切に対処できるようになります。
- 与えられたデータセットの中で、目的関数間の最適なトレードオフを提供するサンプルを見つけることができます。
- より大きなデータセットの中で、トータルとして目的関数間の最適なトレードオフを提供するサンプルのサブセットを見つけることができます。

## 学習機

このコレクションは、RPコンポーネントを含んでおり、決定樹モデル、cross-validateされた決定樹モデル、およびフォレストモデルを構築できます。このコンポーネント中のパラメータを調整することにより、ランダムツリーによるフォレストのサイズだけでなく、各ツリーの大きさや深さ、分割や枝刈り、重みづけの方法をコントロールすることが可能です。目的関数は一つだけでなく、複数の目的関数に対するモデルを構築することもできます。GFAコンポーネントを用いると、一般的な特性予測モデル、分子の特徴に対するモデル、あるいは処方設計における混合モデルを構築することができます。全ての学習機はModel Applicability Domain (MAD) を提示することができ、予測をする際にモデルが適切に適用されているかを確認することが可能です。



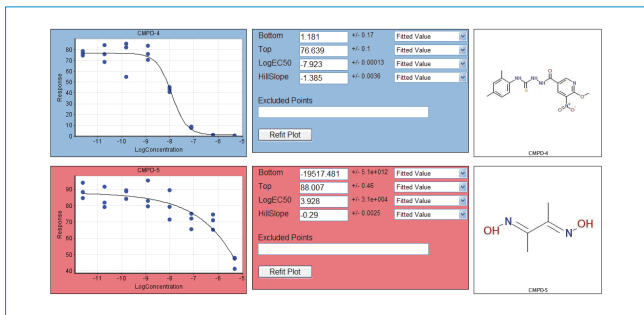
単純な部分集合最適化におけるパレートフロントの進展



各成分の量と特性に拘束条件をかけた上での製剤設計最適化プロトコル

## VIEWER

Tree Model Viewerを用いると、一つ、あるいは複数の決定樹をWeb画面上でインタラクティブに視覚化することができます。フォレスト中にある決定樹、あるいは決定樹中のある部分を支持してブラウズすることも可能です。Tree Displayは各分岐点において利用されている記述子(Fingerprint)化されている分子フラグメントのグラフィカルな表示を含む)や分割の割合、そのルールを表示します。決定樹のノードをクリックすることによって、そのノードにアサインされているサンプルや、そのノードに至るためのルールを表示させることができます。これらのルールは自動的にPilotScriptに翻訳され、これをCustom Filter Componentに貼り付けることによって、同じルールに合致する別のデータレコードを同定することが可能になります。



記述子とその応答の関係を理解するための決定樹の可視化

## OPTIMIZERS

このコレクションには、Pareto Sorting、Pareto Subset Optimization、Pareto Combinatorial Library Optimizationが含まれています。

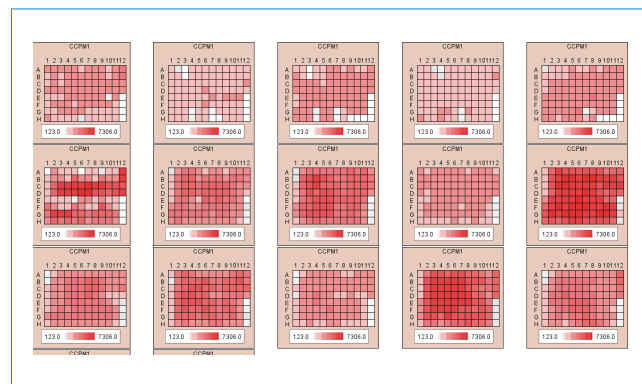
Pareto sortingは、パレートスコアに従って解をランク付けします。各プロパティやそれらがとるべき値の観点から、スコア関数に寄与する閾値を設定することになります。

Pareto Subset Optimizerは、設定したゴールの中で最適なトレードオフとなり得るパレートフロント上の解の部分集合を見つけ出します。

Combinatorial Optimizerでは組み合わせによる制約を設定することができ、これによって合成後の化合物の集合ではなく、合成前の試薬の集合として最適なものを選択することができます。コンビナトリアルケミストリーによるライブラリデザインの例として、最大限多様で、かつ可能な限りdrug-likeな8x12x20のコンビナトリアルライブラリを、より多数の100x100x100の可能な試薬の組み合わせから選択することができます。

## 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・インフォーマティクス・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。



各成分の量と特性に拘束条件をかけた上で最適化された製剤設計

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](http://www.3ds.com) (英語)、[www.3ds.com/ja](http://www.3ds.com/ja) (日本語) をご参照ください。



©2014 Dassault Systemes. All rights reserved. 3DEXPERIENCE, CATIA, SOLIDWORKS, ENOVIA, DELMIA, SIMULIA, GEVOIA, EXPLER, 3D V&R, 3DSWIM, BIOVIA, および INETWORKS 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商標です。ダッソー・システムズまたはその子会社の商標を使用する際には、書面による承認が必要です。