

Chemometricツール -Gateway to Chemometrics with NMR data-

関連製品:核磁気共鳴装置(NMR)

Chemometricsは、実験的に得られた大量の分析データから、次元の圧縮、判別、視覚化、回帰などのデータマイニング手法を適用して、情報を抽 出する学問分野です。NMRは、定量性および再現性にすぐれた分光法であり、また、試料調製が簡便で試料を非破壊で分析できます。この特徴を活か して、生体試料などの一連の試料のNMRスペクトルを入力情報として、多変量解析などを用いてデータマイニングする手法が活発に利用されています。 Delta NMRソフトウェア Ver. 6.0からChemometricsツールが実装されました。Chemometricsツールは、多検体の一連の一次元NMRスペクトルから、行 列へ変換し、探索的多変量解析をサポートします。多変量解析のエンジンとして、統計解析言語Rにおける"Chemospec"パッケージ[1]を用いていま す。"Chemospec"パッケージを別途インストールいただくことで、Deltaソフトウェアのユーザーインターフェイスから、シームレスに探索的多変量解析 を実行することができます。Chemometricsツールでは、まず、バケット積分機能から、複数のNMRスペクトルをデータマイニングに適した次元数の行 列へと変換します。また、Chemometricsツールでは、まず、バケット積分機能から、教師データを用いない探索的多変量解析である主成分分析およ び階層クラスター分析(オプションでヒートマップ表示可)をサポートします。Chemometricsツールでは、判別分析、回帰分析などのその他のデータマ イニング手法は、サポートしませんが、統計解析言語Rでは、世界中のユーザーや開発者が提供するパッケージにおいて、数多くのデータマイニング 手法をサポートします。

NMRデータの読み込みおよびバケット積分による入力行列の準備

NMRデータによる探索的データ分析のワークフローでは、まず、複数の一次元NMRデータを適切な次元(変数の数)の行列に変換し、その行列を 入力に多変量解析などを用いてデータマイニングします (Figure 1)。バケット積分が、最もシンプルな方法です。バケット積分は、一定の等間隔 (ex. 0.04 ppm) でスペクトルの興味のある領域を積分し、そのそれぞれの積分値を変量とします (Figure 2(a))。それぞれの積分範囲は、しばし ば "bucket" や "bin" とよばれ、バケット積分は、"bucketing" や "binning" とも呼ばれます。バケット積分することで、通常、数万点のデータ点から構 成される一次元NMRスペクトルを数百程度の次元に削減することができます。このことは相関のある変数をある程度統合し、また、測定環境の変化 による静磁場の均一度の影響(線幅の微妙な違いとしてスペクトルに現れます)、濃度、pH、イオン強度などによる試料間での化学シフトの微妙なず れによる強度変動の影響を受けにくなります。Deltaソフトウェアでは、スペクトルのローカルアライメントおよび非等幅のバケット積分などには対応し ません。



Figure 1 Deltaソフトウェア version 6.0以降におけるChemometricsツールの概要.

日本電子株式会社



Figure 2 Deltaソフトウェア (version 6.0以降) のChemometricsツールにおける複数のNMRデータの一括処理およびバケット積分. (a) バケット積分の概要. (b) Chemometricsツールにおける複数のNMRデータの一括処理およびバケット積分のユーザーインターフェース. (c) Phase Manuallyを選択時に、表示される位相補正ウインドウ.

ChemometricsツールにおけるBucket積分の設定画面をFigure 2(b)に示します。ChemometricsツールにおけるBucket積分では、未処理のNMR データ (FID) および処理済みのNMRデータ (周波数スペクトル) の両方に対応します。未処理のNMRデータの場合は、Preprocessingパネルの "Process Data"にチェックを入れることで、フーリエ変換および自動位相補正が実施され、その後、バケット積分が適用されます。"Phase Manually" にチェックを入れると、スペクトル ー枚ごとに手動の位相補正が適用可能です。"Baseline Correction"にチェックを入れると、dc_correct関数により ベースライン補正をおこないます。"Referencing"にチェックを入れ、補正したい化学シフトおよびその許容幅を、それぞれ"Chemical Shift"およ び"Tolerance"のテキストボックスに入力することで化学シフト補正を実施します。例えば、"Chemical Shift"および"Tolerance"が、それぞれ0[ppm]お よび0.4[ppm]の場合、-0.2-0.2 ppmの間で最も強度が高いデータ点を0 ppmに補正します。この処理は、DSS-d₆やTSP-d₄を化学シフトの基準物質とし て加えたときに役に立ちます。処理済みのNMRデータ(周波数スペクトル)から始める場合は、"Process Data"のチェックを外します。

次にバケット積分の設定です。"Width"のテキストボックスからbinの幅を設定します。0.01から0.04 ppmが一般的によく用いられます [2]。"Start/Stop"のテキストボックスからバケット積分の始点と終点を指定します。"Normalized"にチェックをいれテキストボックスに数値を入力する ことで、スペクトルのbinの総和を指定した値になるようにノーマライズします。"Skip Region"にチェックし、テキストボックスから始点および終点を指 定することで、バケット積分に含まない領域を指定できます。例えば、水消しの消え残りの領域を取り除くために用いられます。"Process"のボタンを 押すことで、NMRのデータ処理およびバケット積分が実施され、データパネルにバケット積分の結果が表示されます。

データパネルに表示された行列は、csvファイルとして保存することができ、他のデータ分析ソフトウェアに読み込ませ、利用することができます。また、統計解析向けのプログラミング言語であるRの環境およびChemospecパッケージとその関連パッケージをインストールすることで、主成分分析および階層クラスター分析(オプションとしてヒートマップ表示)をDeltaソフトウェアのChemometricsツールからシームレスに実施することができます。 ここでは、チュートリアルとして、弊社アプリケーションノートNM170021″多変量解析および単変量解析を用いたNMRメタボロームデータの試料群間比較" < https://www.jeol.co.jp/applications/detail/1634.html> に用いた暗闇および光のある環境で育てたブロッコリースプラウトの極性抽出代謝 産物の¹H NMRデータ(未処理、暗闇および光のある環境でそれぞれ8試料、計16データ)を用いました。ここでは例として、"Process Data", "Phase Manually", "Baseline Correction", および "Referencing (0[ppm], 0.4[ppm])"にチェックを入れ、バケット積分の設定では、"Skip Region"を有効にして (開始値および終了値は、既定値の5[ppm]および4.5[ppm])その他のパラメータは、既定値に設定しました("Width": 0.04[ppm]、"Start/Stop": 8.5[ppm]および0.5[ppm]、"Normalize": 1000)。"Process"のボタンを押すと一つ目のデータの位相補正画面が表示されます。

Chemospecパッケージによる多変量解析による探索的データ分析

統計解析向けのプログラミング言語であるRの環境およびChemospecパッケージとその関連パッケージをインストールすることで、主成分分析および階層クラスター分析(オプションとしてヒートマップ表示)をDeltaソフトウェアのChemometricsツールからシームレスに実施することができます。まずは先ほど作成したデータパネルの行列の二列目(Cgroup)にサンプルの群情報を入力します(Figure 2(b))。主成分分析および階層クラスター分析は、教師なしの多変量解析であり、サンプルの群情報は分析に用いられませんが、プロットを作成するときの色分けに用いられます。

主成分分析は、多変量データにおいて情報を少ない次元に要約する統計解析の手法です。主成分分析では、相関のある多数の変数から、元の変数の線形結合であらわすことのできるかつ全体のばらつきを最もよく表す新しい変数(主成分)へと投影します。この投影では、第一主成分の分散を 最大化し、続く主成分はそれまでに決定した主成分と直交しかつ分散を最大化するようにして投影されます。結果として、少ない次元数(主成分の 数)において、データの大部分を要約することができます。ここで、主成分における座標をスコア、投影係数に相当する単位ベクトルをローディングと よびます。

DeltaソフトウェアにおけるChemometricsツールでの主成分分析の手順を説明します(Figure 3(a))。"PCA Processing"のタブにおいて"Option"のプ ルダウンメニューから"Classical"またが"Robust"をいずれかを選択します。"Classical"は、一般的な主成分分析であり、データが正規性を仮定し、 データのばらつきの指標として標準偏差を、代表値として算術平均を用いますが、一方、"Robust"は、ロバスト主成分分析であり、データのばらつき の指標としてMedian Absolute Errorを、代表値として1中央値を用い外れ値の影響を受けにくい性質をもちます [3]。"Scaling"は、変数のスケーリン グを指定するパラメータですが、"Classical"主成分分析では、オートスケーリング("Auto")、パレートスケーリング("Pareto")、およびセンタリングのみ ("No Scale")が選択可能です。ここでは、"Pareto"を選択しパレートスケーリングを採用しました。"Score"にチェックをいれると指定した主成分のスコ アプロット(Figure 3 (b))を作成します。"Loading"にチェックをいれると指定した主成分のローディングプロット (Figure 3(c))を作成します。この他にも、 スクリープロット(累積寄与率のプロット)や外れ値の診断用にスコア距離および直交距離をプロットする機能があります。詳細は、JNM-ECZ/ECZLシ リーズユーザーズマニュアル【溶液応用編】7章 Chemometricsツールを参照ください。"PCA" ボタンをクリックすると新しいウインドウが立ち上がり設 定したプロットが表示されます。

チュートリアルの第一主成分と第二主成分のスコアプロット (Figure 3(b)) からは第一主成分がサンプル群の差をよく反映していることが読み取れます。主成分においてスコアは、各変数の線形としてあらわされ、ローディングはその線形結合係数に相当します。したがって、例えば暗闇環境で生育したブロッコリースプラウト試料は、第一主成分のスコアが負ですが、同じように第一主成分のローディングが負の代謝産物が光環境と比較して暗闇 環境で生育した試料に多く含まれていたことが予想できます(Figure 3(c))。



Figure 3 Deltaソフトウェア (version 6.0以降) のChemometricsツールにおける主成分分析. (a) ユーザーインターフェース. (b) "Classical" 主成分分析のスコアプロット (PC1 vs PC2). (c) Classical"主成分分析のローディングプロット. 階層クラスター分析は、教師なしの探索的多変量解析の一種です。階層クラスター分析では、多変量データから計算した試料の間の距離に基づき、 デンドログラム(樹形図)を作成します。デンドログラムから、クラスター(変量のパターンが似た試料のあつまり)の数やクラスターを構成する試料を 分析します。また、試料間のデンドログラムおよび変数間のデンドログラムを用いて作成したヒートマップは、各試料のクラスターにおけるそれぞれの 代謝産物の分布を可視化します。

DeltaソフトウェアにおけるChemometricsツールでの階層クラスター分析の手順を説明します (Figure 4(a))。"Distance" は階層クラスター分析にお ける距離を定義します。11の指標を距離としてプルダウンメニューから選択可能です。ここではピアソンの積率相関係数を距離と同様に扱えるように 変換した指標である相関距離 ("Correlation")を選択しました。"Cluster"はクラスタリング法(クラスターとの距離の定義)を指定します。ここでは完全 連結法 ("Complete")を採用しました。距離の定義およびクラスタリング法の詳細は、JNM-ECZ/ECZLシリーズユーザーズマニュアル【溶液応用編】7 章 Chemometricsツールを参照ください。"Auto Scale" のチェックボックスにチェックすることで、変数はオートスケーリングされます。

"Heatmap"にチェックをいれることで、試料間の階層クラスター分析および変数間の階層クラスター分析によって並び替えられたヒートマップを作成 します(Figure 4(b), "Heatmap"にチェックを外すと試料間の階層クラスター分析のみ実施されます)。"Color Palette Settings"の項目からヒートマップ の色を設定できます。"HCA"ボダンをクリックすると階層クラスター分析が実行され、"Heatmap"のオプションを適用した場合、ヒートマップが新しいウ インドウに表示されます。今回の分析で作成されたheatmapは、データセットにおいて相対的に大きい変量が赤、少ない変量が青で表示されます。し たがって、heatmapにおいて赤色の集まりは、その試料群に豊富に含まれる変数、つまり代謝産物を示します。



Figure 4 Deltaソフトウェア (version 6.0以降)のChemometricsツールにおける階層クラスター分析. (a) ユーザーインターフェース. (b) 階層クラスター分析の結果をヒートマップとして表示.

Chemometricsツールのセットアップ

Chemometricsツールのユーザーインターフェースから、多変量解析を実施するには、統計解析向けのプログラミング言語であるRの環境、 Chemospecパッケージとその関連パッケージのインストール、およびDeltaソフトウェアの環境設定のExternal項の"Rscript Executable"に実行ファイ ル (Rscript.exe)のパスを設定する必要があります。詳細は、JNM-ECZ/ECZLシリーズユーザーズマニュアル【溶液応用編】7章 Chemometricsツール の7.2 構成およびChemoSpecパッケージの準備を参照ください。

参考文献

[1] B.A. Hanson (2021). ChemoSpec: Exploratory Chemometrics for Spectroscopy. R package version 5.3.21. https://CRAN.R-project.org/package=ChemoSpec, [2] A.H. Emwas, E. Saccenti, X Gao, R.T. McKay, V.A.P. Martins dos Santos, R. Roy, D.S. Wishart, *Metabolomics* (2018) **14**:31, [3] C. Croux, P. Filzmoser, and M.R. Oliveira, *Chemometr Intell Lab Syst.* **87** (2007) 218-225.

Copyright © 2022 JEOL Ltd. このカタログに掲載した商品は、外国為替及び外国貿易法の安全輸出管理の規制品に該当する場合がありますので、輸出するとき、または日本国外に持ち出すときは当社までお問い合わせください。



本社・昭島製作所 〒196-8558 東京都昭島市武蔵野3-1-2 TEL:(042)543-1111(大代表) FAX:(042)546-3353 www.jeol.co.jp ISO 9001・ISO 14001 認証取得



東京事務 〒100-0004 東京都千代田区大手刺2丁目1番1号 大手卸野村ビル **業務統括センター** TEL:03-6262-3564 FAX:03-6262-3589 **デマンド推進本部** TEL:03-6262-3560 FAX:03-6262-3577 SI営業本部 SI販促室 TEL:03-6262-3567 FAX:03-6262-3577 セミコンダクタ・ソリューションセールス部 TEL:03-6262-3567 産業機器営業部 TEL:03-6262-3570 MEソリューション販促室 TEL:03-6262-3571 SE専業戦略本部 SE営業グループ TEL:04-2542-2383 (本社・昭島製作所)