

データセット 1: 自動車塗料サンプルの IR 分光分析

- **データセット:** Wiley 「IR - Automobile Paint Chips」 データベースから最初の 200 件の IR スペクトル
- **PCA パラメータ:** 前処理 「平均中心化」、Y 変換 「2 次微分 (ポイント数 11)」
- **結果:** PCA 解析により、塗料サンプルはエナメル塗料とラッカー塗料に対応するふたつの明確なクラスターに分離された (図 1)。3D スコアプロットではこの分離が明確に示されており、塗料カテゴリ間の重なりは最小限である。このクラスタリングにより、既存の PCA 空間への投影を通じて、未知の塗料サンプルの自動分類が可能となる。

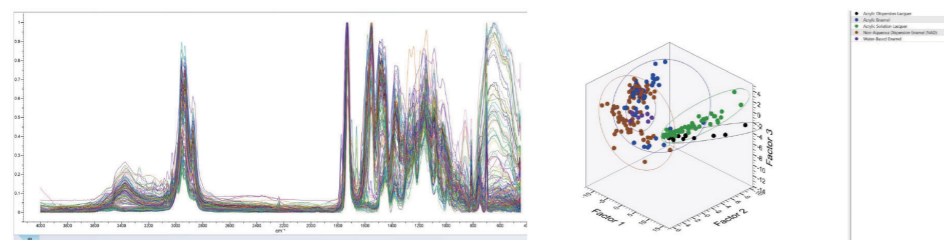


図 1: 自動車塗料チップ 200 件の IR スペクトルに対する PCA 解析
左: 200 件の IR スペクトル
右: エナメル塗料とラッカー塗料が明確に分離された 3D PCA スコアプロット

データセット 2: 血液サンプルの ^1H NMR 分析

- **データセット:** 糖尿病患者および健常者の血液サンプルから取得した 34 件の ^1H NMR スペクトル
- **PCA パラメータ:** 前処理 「平均中心化」、Y 変換 「サンプルの 2 ノルムによる正規化 (divide-by Sample-2Norm)」, および ^1H NMR スペクトルの分解能を低減するための 「インテリジェントビニング」 手法を使用
- **結果:** PCA 解析により、2 次元スコアプロット上で糖尿病患者と健常者の血液サンプルの明確な分離が確認された (図 2)。また、ローディングプロットを用いることで、糖尿病状態に関連するバイオマーカーの特定が可能となる。

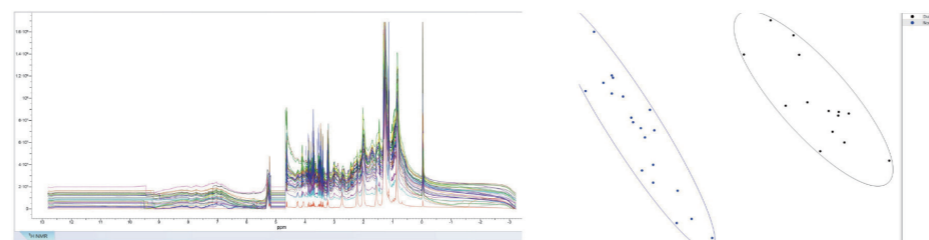


図 2: 糖尿病患者と健常者の血液サンプルに対する PCA 解析
左: ^1H NMR スペクトルデータ
右: サンプル分類を示す PCA 2 次元スコアプロット

データセット 3: LC-MS タンデム質量分析

- **データセット:** 「Wiley Registry of Tandem Mass Spectral Data: MSforID」 データベースから、類似構造検索ヒットリスト 200 件を使用
- **PCA パラメータ:** 前処理 「平均中心化」、Y 変換 「サンプル最大値による正規化 (divide-by Sample Max)」
- **結果:** 3D PCA スコアプロットにより、衝突エネルギーとスペクトルクラスタリングの明確な関係が確認された (図 3)。高い衝突エネルギーのスペクトルはより広く分散し、低い衝突エネルギーのスペクトルは密にクラスタリング (黒点) した。その結果、より多様で差異の大きい質量スペクトルが得られた。このパターンは、高い衝突エネルギーにおいて分子のフラグメンテーションが増加することを反映している。

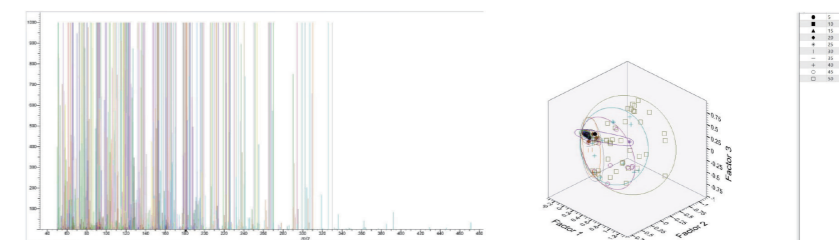


図 3: タンデム質量分析データの PCA 解析
左: 200 件の MS スペクトル
右: エネルギー依存的なクラスタリングを示す PCA 3 次元スコアプロット

概要 (Abstract)

本研究では、主成分分析 (PCA) を用いて、複数の分析手法にわたるスペクトルデータセットを包括的に解析した。その結果、個々のスペクトルを単独で解析した場合には見えにくい潜在的なパターンや相関関係を明らかにした。本研究は、高次元データから有用な知見を抽出することで、PCA が科学的発見を加速できることを示している。また、PCA がスペクトルのクラスターや特徴間の相関を効果的に特定できることを示し、分析ワークフローの改善に寄与する価値を明確にした。

序論 (Introduction)

主成分分析 (PCA) は、元データの変動性を維持しつつ次元を削減することで、複雑なデータセットを単純化する強力な統計手法である。スペクトルデータに適用することで、従来の 1 件ずつのスペクトル解析では見逃されがちな潜在的トレンドを明らかにすることができる。このポスターでは、IR、 ^1H NMR、MS を含む複数のデータベースに対して PCA を実施する。また、本手法は Raman、UV-Vis、クロマトグラフィーデータにも適用可能である。

手法と結果 (Method & Results)

すべての解析は、KnowItAll 2026 の TrendFinder アプリケーションを用いて実施した。PCA の異なる分析手法における有効性を示すため、3種類のデータセットを解析した。

結論 (Conclusion)

PCA は、大規模データセットのスペクトル処理および解析において非常に有効なツールである。従来のスペクトル解析を補完し、可視化されていなかったトレンドを抽出することができる。このアプローチにより、数学、コンピュータ工学、化学の融合が実現される。

本研究は、PCA が複数の分光分析手法において有効な補完的解析ツールであることを示した。

- **分類能力:** 事前知識なしで異なるサンプルタイプを正確に識別
- **汎用性:** 多様な分光手法において有効性を実証
- **物理的洞察:** 実験パラメータとスペクトル特性の有意な関係を解明

本結果は、パターン認識およびサンプル分類を通じて分析ワークフローの高度化に寄与し得る PCA の広範な適用可能性を示唆している。