

数据集 1: 汽车涂层样品的红外 (IR) 光谱分析

- 数据集:** 来自 Wiley “IR - Automobile Paint Chips” 数据库的前 200 个 IR 光谱
- PCA 参数:** 预处理 “均值中心化 (mean-center)”；Y 变换 “二阶导数 (11 个点)”
- 结果:** PCA 分析成功将涂层样品分为两个清晰的聚类，对应于搪瓷漆和清漆类型 (图 1)。3D 得分图清晰地展示了这种分离，不同涂层类别之间几乎没有重叠。该聚类结果支持通过投影到已建立的 PCA 空间，对未知涂层样品进行自动分类。

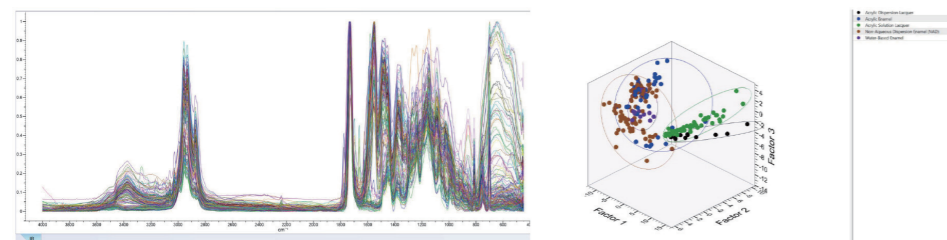


图 1: 汽车涂层碎片 200 个 IR 光谱的 PCA 分析
左: 200 个 IR 光谱
右: 显示搪瓷漆与清漆明显分离的 3D PCA 得分图

数据集 2: 血液样品的 ^1H NMR 分析

- 数据集:** 来自糖尿病患者和正常血液样品的 34 个 ^1H NMR 光谱
- PCA 参数:** 预处理 “均值中心化”；Y 变换 “按样本 2 范数归一化 (divide-by Sample-2Norm)”；采用 “智能分箱 (Intelligent binning)” 方法降低 ^1H NMR 光谱分辨率
- 结果:** PCA 分析在二维得分图中清晰地区分了糖尿病与正常血液样品 (图 2)。此外，通过载荷图可以识别与糖尿病相关的生物标志物。

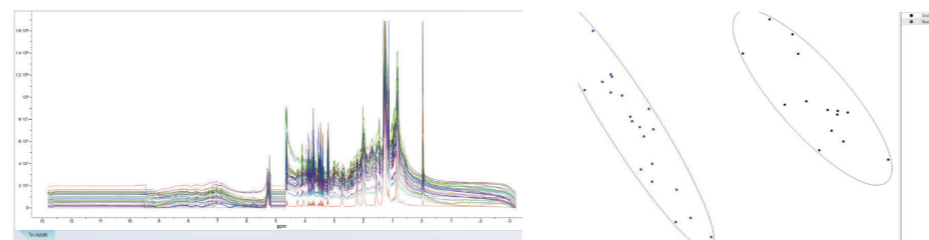


图 2: 糖尿病与正常血液样品的 PCA 分析
左: ^1H NMR 光谱数据
右: 显示样本分类的 PCA 二维得分图

数据集 3: LC-MS 串联质谱分析

- 数据集:** 本研究使用了来自 “Wiley Registry of Tandem Mass Spectral Data: MSforID” 数据库的 200 条相似结构检索结果
- PCA 参数:** 预处理 “均值中心化”；Y 变换 “按样本最大值归一化 (divide-by Sample Max)”
- 结果:** 3D PCA 得分图揭示了碰撞能量与光谱聚类之间的明确关系 (图 3)。较高碰撞能量的光谱呈现更大的分散性，而较低碰撞能量的光谱则紧密聚类 (黑点)。因此形成更加多样化且差异明显的质谱数据。该模式反映了在较高碰撞能量下分子碎片化程度的增加。

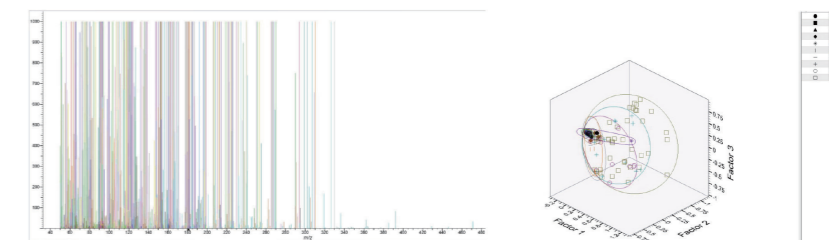


图 3: 串联质谱数据的 PCA 分析
左: 200 条 MS 光谱
右: 显示能量依赖聚类模式的 PCA 三维得分图

摘要 (Abstract)

本研究利用主成分分析 (PCA) 对多种分析技术下的光谱数据集进行了综合分析。结果揭示了在单独分析每个光谱时难以发现的潜在模式和相关性。研究表明，PCA 能够通过从高维数据中提取关键信息，从而促进科学发现。同时，结果显示 PCA 能够有效识别光谱聚类及特征之间的相关性，体现其在优化分析流程方面的重要价值。

引言 (Introduction)

主成分分析 (PCA) 是一种强大的统计方法，通过在保留大部分原始变异性的同时降低数据维度，从而简化复杂数据集。应用于光谱数据时，可揭示传统逐一分析方法难以发现的潜在趋势。本海报中，我们对多个数据库 (包括 IR、 ^1H NMR、MS) 进行了 PCA 分析。此外，该方法同样适用于 Raman、UV-Vis 以及色谱数据。

方法与结果 (Method & Results)

所有分析均使用 KnowItAll 2026 中的 TrendFinder 应用完成。通过分析三个不同数据集，展示了 PCA 在不同分析技术中的应用能力。

结论 (Conclusion)

PCA 是用于大规模光谱数据处理与分析的强大工具。它通过数学方法补充传统光谱分析，能够发现原本不可见的趋势。该方法将数学、计算机工程与化学相结合。本研究表明，PCA 作为一种辅助分析工具，在多种光谱技术中具有良好的适用性。

主要结论：

- 分类能力:** 无需先验知识即可区分不同样本类型
- 方法通用性:** 在多种光谱分析方法中均表现出良好效果
- 物理洞察:** 揭示实验参数与光谱特征之间的重要关系

研究结果表明，PCA 在通过模式识别与样本分类提升分析工作流程方面具有广泛的应用潜力。