

化学计量学在复杂样品分析中的数据处理策略探讨

戴华明

安徽新天地生物肥业有限公司 安徽蚌埠

【摘要】化学计量学在复杂样品分析中发挥着重要作用，但其数据处理仍面临诸多挑战，如高维度数据处理、信号重叠与干扰、模型泛化能力不足等。探讨了基于化学计量学的优化数据处理策略，包括降维技术、化学分解方法、机器学习与化学计量学的结合等，通过实际案例验证了这些策略在提高分析精度和模型适应性方面的显著效果。未来，化学计量学有望通过多学科交叉融合，进一步提升复杂样品分析的智能化水平，推动化学分析技术的持续发展。

【关键词】化学计量学；复杂样品；数据分析；优化策略；机器学习

【收稿日期】2025 年 1 月 23 日 **【出刊日期】**2025 年 2 月 26 日 **【DOI】**10.12208/j.jer.20250061

Exploration of data processing strategies for complex sample analysis using chemometrics

Huaming Dai

Anhui Xintiandi Bio-fertilizer Co., Ltd., Bengbu, Anhui

【Abstract】 Chemometrics plays a vital role in the analysis of complex samples, yet it still faces many challenges in data processing, such as high-dimensional data handling, signal overlap and interference, and insufficient model generalization ability. This paper explores optimized data processing strategies based on chemometrics, including dimensionality reduction techniques, chemical decomposition methods, and the integration of machine learning with chemometrics. These strategies have been validated through practical cases, showing significant effects in improving analytical accuracy and model adaptability. In the future, chemometrics is expected to further enhance the intelligent level of complex sample analysis through multidisciplinary integration, promoting the continuous development of chemical analysis technologies.

【Keywords】 Chemometrics; Complex samples; Data analysis; Optimization strategies; Machine learning

引言

随着科学技术的飞速发展，复杂样品分析在化学、生物、环境等领域的重要性日益凸显。复杂样品通常包含多种成分，且各成分之间存在复杂的相互作用，这使得传统的分析方法难以满足高精度、高效率的要求。化学计量学作为一门结合数学、统计学和化学的交叉学科，为复杂样品分析提供了强大的数据处理工具。然而，面对复杂样品的高维度、多组分特性，化学计量学在数据处理中仍面临诸多挑战。因此，探索优化策略以提升化学计量学在复杂样品分析中的应用效果，具有重要的科学意义和实际价值。

1 复杂样品分析中化学计量学的应用现状

在现代化学分析领域，复杂样品的分析已成为一项极具挑战性的任务。复杂样品通常包含多种成分，且各成分之间存在复杂的相互作用，这使得传统的分析方法往往难以准确、高效地获取全面信息。化学计量学作为一种结合数学、统计学和化学的交叉学科，为复杂样品的分析提供了强有力的工具。它通过数学模型和算法对大量实验数据进行处理和分析，从而提取有价值的信息，帮助研究人员更好地理解样品的组成和性质。化学计量学在复杂样品分析中的应用广泛且深入。在光谱分析中，化学计量学方法如主成分分析(PCA)和偏最小二乘法(PLS)

被广泛用于处理光谱数据,以实现样品的定性与定量分析。这些方法能够有效处理光谱数据中的噪声和重叠信号,提高分析的准确性和可靠性。

在色谱分析中,化学计量学同样发挥着重要作用。通过建立色谱峰的数学模型,研究人员可以实现对复杂色谱图的解析,分离和识别其中的各个组分。在多组分体系的分析中,化学计量学方法能够处理多变量数据,揭示样品中各组分之间的相关性,为复杂体系的研究提供了新的视角^[1-2]。随着计算机技术的飞速发展,化学计量学的应用范围不断扩大。机器学习和人工智能技术的引入为化学计量学的发展注入了新的活力。这些技术能够处理更复杂的数据结构,自动识别数据中的模式和规律,进一步提升了复杂样品分析的效率和精度。例如,神经网络和决策树等算法被用于处理高维度数据,能够有效解决传统方法难以处理的问题。化学计量学在复杂样品分析中的应用也促进了多学科交叉融合,推动了化学分析技术的不断创新和发展。

然而,在实际应用中,化学计量学也面临着诸多挑战。复杂样品的多组分特性和数据的高维度特性不仅使得数据处理的难度大幅增加,还导致了数据冗余和噪声的累积,进一步干扰了有效信息的提取。化学计量学模型的建立和优化需要大量的实验数据和计算资源,这不仅增加了研究成本,也在一定程度上限制了其在实际分析中的广泛应用。模型的泛化能力不足以及对数据分布变化的敏感性,使得化学计量学在处理不同来源或批次的复杂样品时,往往需要重新调整参数甚至重新建模,这进一步增加了应用的复杂性和不确定性。尽管如此,化学计量学在复杂样品分析中的应用仍然是当前化学分析领域的重要发展方向。其不断发展的理论和技术,结合多学科交叉的优势,为解决复杂样品分析中的难题提供了有力支持,并有望在未来的分析实践中发挥更大的作用。

2 复杂样品分析中数据处理的瓶颈与挑战

在复杂样品分析中,数据处理环节是实现精准分析的关键,但同时也面临着诸多瓶颈与挑战。复杂样品通常具有多组分、高维度、强相互作用等特点,这使得数据呈现出高度复杂性和非线性特征。一方面,样品中各组分的浓度范围可能跨越多个数量级,导致数据的动态范围极宽,传统的数据处理

方法难以有效处理这种差异,容易造成低浓度组分信息的丢失或掩盖。

另一方面,复杂样品的光谱、色谱等信号往往存在重叠和干扰,增加了数据解析的难度,尤其是在多组分同时存在的情况下,如何准确分离和识别各组分的特征信号成为亟待解决的问题。数据的高维度特性也是复杂样品分析中的一个重要挑战。随着分析技术的不断发展,如高分辨率光谱、多维色谱等技术的应用,产生的数据量呈爆炸式增长^[3-4]。这些高维度数据虽然包含了丰富的信息,但也带来了数据冗余、计算复杂度高以及模型过拟合等问题。如何从海量数据中提取有价值的信息,同时避免计算资源的过度消耗,是当前复杂样品分析中亟需突破的瓶颈。

复杂样品分析中数据处理的另一个难点在于模型的泛化能力和适应性。由于样品的来源、性质以及实验条件的多样性,数据的分布和特征可能因批次、环境等因素而发生变化。现有的化学计量学模型在面对这些变化时,往往需要重新调整参数甚至重新建模,这不仅增加了分析成本,也限制了模型在实际应用中的广泛性和稳定性。复杂样品分析中数据的噪声和不确定性也不容忽视。实验过程中不可避免地会引入随机误差和系统误差,这些噪声会干扰数据处理过程,影响分析结果的准确性和可靠性。

3 基于化学计量学的优化数据处理策略

针对复杂样品分析中数据处理的瓶颈与挑战,化学计量学提供了多种优化策略,旨在提升数据处理的效率、准确性和适应性。在处理高维度数据时,降维技术成为关键手段之一。主成分分析(PCA)通过将原始数据投影到低维空间,提取主要信息,同时去除冗余和噪声,有效降低了数据复杂性。偏最小二乘法(PLS)结合了回归分析与降维的优势,能够在处理多变量数据时建立预测模型,提高定量分析的精度。为应对复杂样品中信号重叠和干扰问题,化学计量学引入了多种解析方法。

化学分解方法如自模型校正(SIMPLISMA)能够有效分离重叠信号,提取纯净的组分光谱。多维标度分析(MDS)通过可视化高维数据的相似性,帮助识别样品中的潜在模式和结构,进一步优化数据解析过程。在提升模型泛化能力和适应性方面,

机器学习与化学计量学的结合提供了新的思路。神经网络能够自动学习数据中的复杂非线性关系,通过调整网络结构和参数,实现对复杂样品数据的高效处理^[5-6]。支持向量机(SVM)通过寻找最优分类超平面,有效解决了高维数据分类和回归问题,提高了模型的泛化能力。随机森林等集成学习方法通过构建多个模型并进行投票或平均,进一步增强了模型的稳定性和适应性。

化学计量学还注重数据预处理和特征选择的优化。通过标准化、归一化等预处理方法,可以消除数据的量纲差异和动态范围问题,为后续分析提供更可靠的数据基础。特征选择算法则通过筛选出对分析目标最有价值的变量,减少数据维度,提高模型的计算效率和分析精度。这些优化策略的综合应用,为复杂样品分析中的数据处理提供了全面而有效的解决方案,推动了化学计量学在复杂样品分析领域的进一步发展。

4 优化策略的案例验证与效果分析

在复杂样品分析中,优化策略的应用效果是验证其可行性和优越性的关键环节。通过对实际复杂样品的分析,可以直观地评估优化策略在数据处理中的具体表现。在光谱分析领域,采用主成分分析(PCA)结合偏最小二乘法(PLS)的优化策略,能够有效处理光谱数据中的噪声和重叠信号问题。PCA用于降维和特征提取,将复杂的光谱数据简化为少数几个主成分,同时去除无关信息;PLS则在此基础上建立定量模型,实现对目标组分的准确预测。这种组合策略显著提高了光谱分析的准确性和可靠性,尤其是在处理高浓度和低浓度组分共存的复杂样品时,展现出良好的适应性。在色谱分析中,优化策略的应用同样取得了显著效果。

通过引入化学分解方法如自模型校正(SIMPLISMA),可以有效分离色谱峰中的重叠信号,提取纯净的组分色谱图。这种方法在处理多组分复杂色谱图时表现出色,能够准确识别各组分的特征峰,为定量分析提供可靠依据。结合多维标度分析(MDS)对色谱数据进行可视化处理,能够直观地展示样品间的相似性和差异性,进一步优化分析流程。在高维度数据处理方面,机器学习与化学计量学的结合策略展现出强大的优势。

神经网络和随机森林等算法能够自动学习数据

中的复杂模式,适应复杂样品数据的非线性特征。通过构建神经网络模型,可以实现对复杂样品的高效分类和定量分析,模型的泛化能力显著提升^[7-8]。随机森林则通过集成学习方法,进一步增强了模型的稳定性和准确性,即使在数据存在噪声和不确定性的情况下,也能保持良好的分析效果。这些优化策略在实际复杂样品分析中的应用,不仅提高了数据处理的效率和准确性,还增强了模型的适应性和泛化能力。通过对比传统方法和优化策略的分析结果,可以清晰地看到优化策略在复杂样品分析中的显著优势,为复杂样品分析提供了更加可靠的解决方案。

5 结语

复杂样品分析中,化学计量学的应用不断拓展,其优化策略在提升数据处理效率、准确性和模型泛化能力方面展现出显著优势。从降维技术到机器学习的融合,这些方法为复杂样品分析提供了有力支持。然而,面对日益复杂的样品体系和数据结构,仍需进一步探索创新。未来,化学计量学的发展将更加注重多学科交叉融合,借助人工智能和大数据技术,进一步提升复杂样品分析的智能化水平,为化学分析领域带来更多突破性进展,推动相关技术在更广泛领域的应用。

参考文献

- [1] 曹勇,陆以云,杨波.多指标成分定量联合化学计量学评价不同产地独脚金质量[J].现代药物与临床,2025,40(02):341-348.
- [2] 罗晓宏,王楠希,陈红娟,等.基于化学计量学和近红外光谱法的油莎豆产地溯源[J].食品安全质量检测学报,2025,16(04):178-184.
- [3] 董平,彤霖,魏敏,等.HS-SPME-GC-MS 结合化学计量学分析湖北产区雪茄烟叶[J].烟草科技,2025,58(02):20-28.
- [4] 钱岩,黄逸文,贾昌平,等.基于UHPLC-Q-TOF-MS和化学计量学的金银花硫磺熏蒸前后化学成分差异性分析[J].中草药,2025,56(04):1146-1158.
- [5] 乔明锋,易宇文,胡金祥,等.基于电子鼻、气质联用技术结合化学计量学评价不同品牌鱼香肉丝气味差异[J].中国调味品,2025,50(02):209-219.
- [6] 周玲利,刘育辰,彭乐强,等.基于多指标含量测定及化学

- 计量学的黔产黑龙茎叶质量评价[J]. 药物评价研究, 2025, 48(02):455-461.
- [7] 王笑笑,胡卫南,王健,等.基于 HPLC 指纹图谱与化学计量学研究常山柚橙幼果与枳实的相似性[J].医药导报, 2025, 44(02):273-280.
- [8] 刘小春.测土配方施肥对桉树人工林土壤生态化学计量

特征的影响[J].绿色科技,2025,27(02):28-32.

版权声明: ©2025 作者与开放获取期刊研究中心(OAJRC)所有。本文章按照知识共享署名许可条款发表。

<https://creativecommons.org/licenses/by/4.0/>

