

基于拉曼光谱的已知混合物组分定量分析方法

颜凡¹, 朱启兵^{1*}, 黄敏¹, 刘财政¹, 雷泽民², 张恒², 张丽文², 李敏²

1. 江南大学轻工过程先进控制教育部重点实验室, 江苏 无锡 214122

2. 北京卓立汉光仪器有限公司, 北京 101102

摘要 利用拉曼光谱进行混合物组分定量分析一直是分析化学领域的一大难题。针对现有的基于机器学习(如支持向量回归机、偏最小二乘)的混合物定量分析方法存在的训练样本难以获得、模型推广性能差的问题,提出了一种基于拉曼光谱谱峰强度最小二乘拟合的已知混合物组分直接定量分析方法。该方法首先采集已知混合物及其各组分的拉曼光谱,利用连续小波变换和惩罚最小二乘法相结合的方法对采集的拉曼光谱进行去噪、基线校正等预处理。通过斜率比较法将预处理后的拉曼光谱分为多个光谱子区间,将各子区间的拉曼光谱看作是多个 Voigt 函数的线性叠加,并利用 levenberg-marquardt-fletcher(LMF)算法求解获得各谱峰的位置、强度、半高宽等表达系数。根据各组分参考光谱的谱峰位置,确定各组分对混合物光谱中每个谱峰的贡献度。依据朗伯-比尔定律中拉曼光谱的谱峰峰强与其所对应的浓度的正比关系,建立超定方程;最后利用最小二乘法拟合该超定方程得到各组分对应的系数,从而获得各组分的体积浓度。利用乙醇、乙腈、丙酮、环己烷、二丙酮醇、丙二酸二乙酯六种组分配置了 10 种三元混合物(每种三元混合物 9 个体积浓度比),采集了 90 组混合物及 6 种组分的拉曼光谱数据。在混合物及其组分参考光谱测量条件(功率和积分时间)相同情况下,所有组分的相关系数(r)均在 0.96 以上,均方根误差(RMSE)小于 6%,剩余预测偏差(RPD)均大于 2.5;在混合物及其组分参考光谱在不同测量条件下,各组分的 r 均大于 0.93,最大 RMSE 为 7.94%,RPD 均大于 2.0,证明了算法具有良好的准确性和鲁棒性。所提出的方法能够实现三组分混合物的快速、准确的直接定量分析,为混合物的定量分析提供了一种有效的途径。

关键词 拉曼光谱; 混合物组分; 定量分析; 拉曼谱峰强度; Voigt 函数

中图分类号: O433.4 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2020)11-3599-07

引言

近年来,基于拉曼光谱的检测分析方法由于具有无损、非接触、无化学污染、检测时间短、检测结果准确、重复性好、适用于大多数有机和无机化合物等优点,在微生物、法医分析,鉴定可卡因和其他非法药物的滥用、食品等领域得到了广泛的应用^[1-3]。在利用拉曼光谱进行混合物组分的定量分析时,不同成分的光谱信号相互重叠,并且由于物质分子之间的相互作用导致它们的振动方式与纯物质相比有所不同,这给定量分析带来了很大困难。

檀兵等^[4]利用 Voigt 函数对原始光谱进行拟合以解决多元素共存条件下的谱峰重叠问题;在此基础上,利用拟合谱峰的中心波长、光强、半峰全宽和谱峰质心构建特征参数向

量,解决了传统谱峰识别方法难以提取谱峰波形形态特征导致的识别率难以保证的问题。

刘燕德等^[5]利用 600~3 000 cm^{-1} 波段的拉曼光谱,并结合支持向量机(SVM),偏最小二乘(PLS)等定量分析模型分别对三组分食用调和油中各组分进行了快速定量检测,其中最优 PLS 模型的决定系数和 RMSE 分别为 0.990 4 和 0.018 8。高颖等^[6]利用非线性最小二乘法将天然气拉曼光谱分解为纯物质组分的拉曼光谱分量和若干个洛伦兹谱峰之和的形式,建立该组分特征峰面积和对应的浓度之间的模型。粟晖等^[7]根据混合物光谱向量与其组分参考光谱向量二者间的空间向量夹角值与组分的相对含量存在线性关系,通过少量已知样本拟合出夹角值与其含量的标准曲线,实现混合物组分的定量分析。近年来,深度学习方法在光谱定量分析中的应用也得到了初步的应用。Fan 等^[8]提出了一种基于

收稿日期: 2019-10-11, 修订日期: 2020-02-18

基金项目: 国家自然科学基金项目(61775086)资助

作者简介: 颜凡, 1995 年生, 江南大学轻工过程先进控制教育部重点实验室硕士研究生 e-mail: yanfanmail@163.com

* 通讯联系人 e-mail: zhuqib@163.com

卷积神经网络(CNN)模型的混合物组分预测方法,得到了比传统建模方法更优的结果。Zhang 等^[9]提出了一种无需数据预处理的光谱定量分析方法,该方法利用一个由三个卷积层,一个平坦层,一个全连接层和一个输出层组成的卷积神经网络来进行组分定量分析,并报道了比传统卷积神经网络更优的结果。以上方法的预测精度与模型的建立好坏有直接联系,要取得较高的预测精度则需要大量的已知样本进行训练。然而不同组分构成的混合物模型也存在着差异,一旦测量环境和测量对象的组成发生改变,则需要重新收集样本建立分析模型,因此这类方法难以满足实际应用需要。

针对传统机器学习方法存在的训练样本难以获得,模型推广性能差的问题,本研究基于混合物拉曼光谱及其各组分的关系特性,提出了一种非学习的已知混合物组分的拉曼光谱定量分析方法,解决了传统建模方法需要大量已知样本并

且模型不具有推广性等问题。

1 实验部分

1.1 仪器与样本

本研究的拉曼光谱数据由北京卓立汉光仪器有限公司的手持式拉曼光谱仪(Finder Edge)进行采集。其激发源为 785 nm 激光器,激光功率和积分时间可调,光谱范围为 150~2 700 cm^{-1} ,光谱分辨率为 8~10 cm^{-1} 。

本文采集了乙醇、乙腈、丙酮、环己烷、二丙酮醇、丙二酸二乙酯六种纯净物的拉曼光谱。利用这 6 种纯净物制备了十种不同类型的三元混合物,每种混合物样本依据其组分配置九种不同的体积比,共 90 个样本,表 1 为这 6 种纯净物构成的不同组分的三元混合物。

表 1 十种不同类型的三元混合物

Table 1 Ten different types of ternary mixtures

各组分光谱 测量条件	混合物光谱 测量条件	序号	乙醇	乙腈	丙酮	环己烷	二丙酮醇	丙二酸二乙酯	体积比
300 mW, 2 s	300 mW, 2 s	S1	✓	×	✓	×	✓	×	7 : 2 : 1
		S2	×	×	✓	×	✓	✓	
		S3	✓	×	✓	✓	×	×	
300 mW, 1.5 s	300 mW, 1.5 s	S4	✓	×	×	✓	×	✓	4 : 3 : 3
		S5	×	✓	✓	×	✓	×	2 : 1 : 7
		S6	✓	✓	✓	×	×	×	3 : 2 : 5
300 mW, 2 s	200 mW, 2 s	S6	✓	✓	✓	×	×	×	3 : 3 : 4
	200 mW, 1.5 s	S7	×	×	✓	✓	✓	×	1 : 7 : 2
	150 mW, 1.5 s	S8	×	×	×	✓	✓	✓	2 : 5 : 3
	250 mW, 1.5 s	S9	✓	×	×	×	✓	✓	3 : 4 : 3
250 mW, 1.5 s	S10	✓	✓	×	×	×	✓		

图 1 给出了这 6 种纯净物的原始拉曼光谱图。由于这 6 种物质在常温下都为无色的有机溶剂,具有相似的官能团,而拉曼光谱主要反应物质的分子结构,因此从图 1 中可以看出,任意两种物质的光谱都存在部分谱峰重叠的情况。

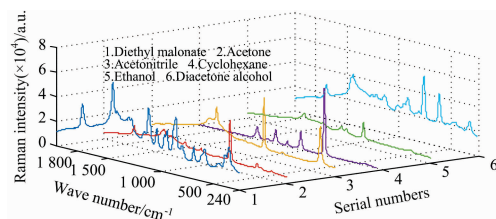


图 1 手持拉曼光谱仪采集的六种纯净物的原始光谱

Fig. 1 The raw spectra of six substances obtained by a handheld Raman spectrometer

1.2 算法流程与原理

本文提出的基于拉曼光谱谱峰强度最小二乘拟合的已知混合物组分直接定量分析方法主要包括光谱预处理、拉曼光谱谱峰分解、和分解后谱峰线性表示求解三个主要环节。其主要步骤说明如下。

1.2.1 光谱预处理

由六种纯净物的拉曼光谱图 1 可知,峰位主要集中于 240~2 000 cm^{-1} 区域内,因此选择该区域作为光谱区间。图 2(a)为 S1 样本在九种不同体积比下的拉曼原始光谱图,从中可以看出,拉曼光谱仪采集的原始光谱含有较强的荧光背景,为了减少荧光背景对定量分析的干扰,本文采用连续小波变换和惩罚最小二乘法相结合的方法进行背景校正。图 2(b)给出了背景校正后的拉曼光谱图,从图中可以看出,校正后的曲线消除了荧光背景产生的基线对物质拉曼光谱的影响,保留了光谱的有用信息。

1.2.2 拉曼光谱谱峰分解

根据朗伯-比尔定律,拉曼光谱的谱峰峰强和所对应的体积浓度成正比例关系,可以通过混合物的谱峰峰强与其组分参考光谱的谱峰峰强之比来确定该组分的体积浓度。由于拉曼光谱中不可避免会有谱峰重叠的现象,从而影响谱峰特征参数的提取^[10]。因此,需要将重叠光谱分解为多个单峰的叠加形式,以便提取峰位、峰强、半峰全宽等特征。若直接对整个光谱进行拟合,需要同时对几十个甚至上百个参数进行寻优,这可能令优化问题呈现病态无法寻到最优解。基于此,本文利用斜率比较法将整个光谱区间划分为仅包含单峰或多个谱峰的重叠峰区域。如图 3 所示,计算谱峰(ω_i^p , a_i^p)

左右两侧的光强极小值位置和强度 (ω_i^m, α_i^m) 、 $(\omega_{i+1}^m, \alpha_{i+1}^m)$ ，若其斜率 $K=(\alpha_{i+1}^m - \alpha_i^m)/(\omega_{i+1}^m - \omega_i^m)$ 小于给定阈值 T ，则划分为待拟合区域；若斜率 $K > T$ ，进一步计算斜率 $K=(\alpha_{i+2}^m - \alpha_i^m)/(\omega_{i+2}^m - \omega_i^m)$ ，直至满足斜率 $K \leq T$ ，得到拟合区域 $[\omega_i^m, \omega_{i+l}^m]$ ， l 为拟合区域谱峰个数。

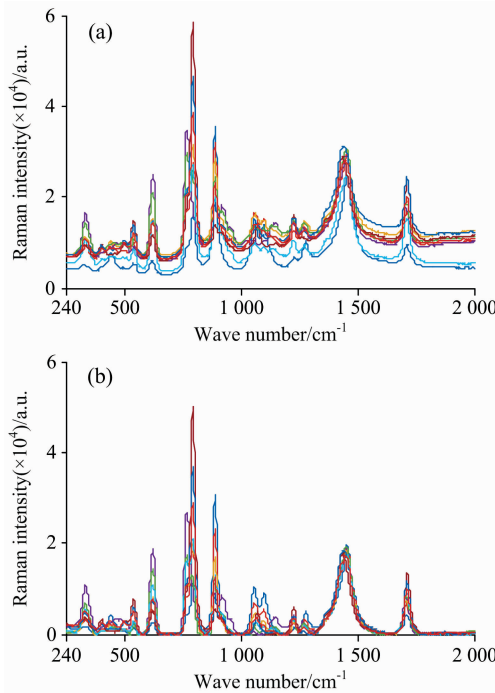


图 2 (a) S1 样本的拉曼原始光谱；
(b) S1 样本的预处理拉曼光谱
Fig. 2 (a) The raw Raman spectra of S1 sample;
(b) The preprocess Raman spectra of S1 sample

拉曼光谱的谱峰形态本质上是洛伦兹线型，由于受仪器和被测样本自身特性的影响，实际测量得到的拉曼光谱谱峰波形通常可由 Voigt 函数来近似表达^[11]

$$V(\nu, \alpha, \omega, \gamma, \theta) = \theta \alpha \exp\left[-\frac{4 \ln 2 (\nu - \omega)^2}{\gamma^2}\right] + (1 - \theta) \alpha \frac{\gamma^2}{(\nu - \omega)^2 + \gamma^2} \quad (1)$$

式(1)中， ν 为波数； α 为谱峰强度； ω 为谱峰所对应的波数； γ 为谱峰的半峰全宽； θ 为高斯-洛伦兹系数， $0 \leq \theta \leq 1$ 。对含有 l 个谱峰区域的光谱 y 进行解析，实际上就是对 l 个 Voigt 峰参数 $\beta_i = (\alpha_i, \omega_i, \gamma_i, \theta_i)$ ， $(i=1, \dots, l)$ 进行拟合

$$r(\nu, x) = y(\nu) - f(\nu, x) \quad (2)$$

其中 $x = (\beta_1, \dots, \beta_l)$ ， $f(\nu, x) = \sum_{i=1}^l V_i(\nu, \beta_i)$ 。寻求最优解 $x = (\beta_1, \dots, \beta_l)$ ，使得残差平方和 $S = r(\nu, x)^T r(\nu, x)$ 最小。

利用 LMF 算法对方程(2)进行求解。图 4 为利用 LMF 算法将光谱分解成单个 Voigt 峰的示意图，从图中可以看出，分解后的 Voigt 峰能准确反映各峰的形态。

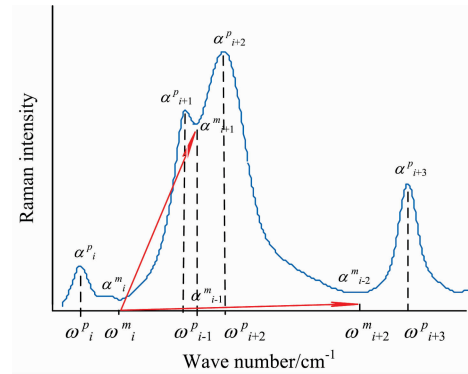


图 3 区域划分示意图
Fig. 3 Schematic diagram of area division

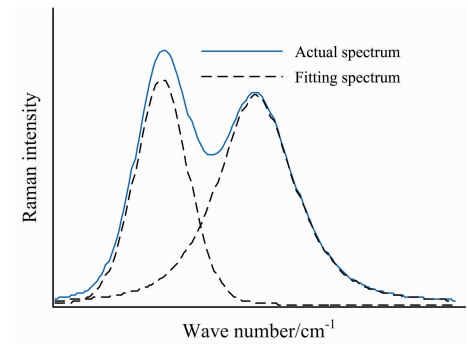


图 4 基于 Voigt 的光谱分解示意图
Fig. 4 Schematic diagram of spectral decomposition based on Voigt

1.2.3 谱峰的线性表示与求解

对于分解为 L 个 Voigt 峰的三组分待测混合物光谱 y ，可得 L 个方程

$$\begin{cases} c_1 \alpha_{M_1}^1 + c_2 \alpha_{M_2}^1 + c_3 \alpha_{M_3}^1 = \alpha_{y_1} \\ c_1 \alpha_{M_1}^2 + c_2 \alpha_{M_2}^2 + c_3 \alpha_{M_3}^2 = \alpha_{y_2} \\ \dots \\ c_1 \alpha_{M_1}^L + c_2 \alpha_{M_2}^L + c_3 \alpha_{M_3}^L = \alpha_{y_L} \end{cases} \quad (3)$$

其中 c_1, c_2, c_3 为各组分对应的系数，与其体积浓度成正比， $\alpha_{M_i}^j$ 为第 i 种组分在混合物 y 的第 j 个 Voigt 峰 V_{y_j} 所对应的贡献值。由于各组分所占体积浓度以及各波段拉曼峰的差异，若直接对方程(3)进行最小二乘拟合求解，所求得的系数更倾向于拟合拉曼峰强较强的方程，而忽略掉较弱的峰。因此，为保证每个方程在拟合时所占的权重一样，对方程(3)作峰强归一化处理，即

$$\begin{cases} c_1 \frac{\alpha_{M_1}^1}{\alpha_{y_1}} + c_2 \frac{\alpha_{M_2}^1}{\alpha_{y_1}} + c_3 \frac{\alpha_{M_3}^1}{\alpha_{y_1}} = 1 \\ c_1 \frac{\alpha_{M_1}^2}{\alpha_{y_2}} + c_2 \frac{\alpha_{M_2}^2}{\alpha_{y_2}} + c_3 \frac{\alpha_{M_3}^2}{\alpha_{y_2}} = 1 \\ \dots \\ c_1 \frac{\alpha_{M_1}^L}{\alpha_{y_L}} + c_2 \frac{\alpha_{M_2}^L}{\alpha_{y_L}} + c_3 \frac{\alpha_{M_3}^L}{\alpha_{y_L}} = 1 \end{cases} \quad (4)$$

在计算各组分贡献值 $\alpha_{M_i}^j$ 时，考虑到仪器本身与检测环境的

变化情况,混合物中对应其组分同一分子结构的拉曼峰位可能会存在少量的偏移,论文对这种峰位偏移做了校正处理。对于给定的混合物拉曼谱峰 V_{y_j} (峰强为 σ_{y_j}), 在各个组分中寻找与混合物谱峰 V_{y_j} 最接近的拉曼谱峰 V_{M_i} ($i=1, 2, 3$), 若拉曼位移差 $\Delta x = |V_{y_j} - V_{M_i}| \leq e$ (e 为许可的峰位偏移, 本文设置为 5), 则 $\alpha_{M_i} = \alpha_{V_{M_i}}$ ($\alpha_{V_{M_i}}$ 为 V_{M_i} 的峰强); 若 $e < \Delta x \leq \omega_{V_{M_i}}$ ($\omega_{V_{M_i}}$ 为 V_{M_i} 的半峰全宽), 则将谱峰 V_{M_i} 平移 $\Delta x - e$, 对应的谱峰强度即为贡献值 α_{M_i} ; 若 $\Delta x > \omega_{V_{M_i}}$, 则 $\alpha_{M_i} = 0$ 。

利用最小二乘法拟合方程(4), 获得组分系数, 并对组分系数进行归一化处理, 得到各组分的体积浓度。

1.3 算法评价指标

本文利用线性相关系数 r 、均方根误差 RMSE 和剩余预测偏差 RPD 作为算法的性能评价指标。其中 r 反应预测浓度与实际浓度之间的线性相关程度; RMSE 反应预测值与真实值之间的偏差; RPD 是真实值的标准差与均方根误差的比值, 证明模型的预测能力, 一般以 RPD 在 2 和 2.5 之间的值表示可以进行粗略的定量预测, 而在 2.5 和 3 之间的值对应较好的预测精度, 在 3 以上则表示极好的预测精度。

2 结果与讨论

2.1 数据处理

图 5 为纯净物乙醇、丙酮、环己烷以及由这三种组分构成的 S3 样本在 V3(4:3:3) 体积比混合下经过连续小波变换寻峰方法寻找到光谱的峰位并预估其谱峰特征参数, 基于此作为 LMF 算法的输入进行拟合分解, 峰位在图 5 中用空心圆标记, 从中可以看到主要的峰位被完全找到, 极少出现漏峰和伪峰的情况。图 6 为分解后的多个 Voigt 峰的叠加形式, 每个独立的 Voigt 谱峰都可以由某种特定成分的分子振动所产生。例如, 乙醇拉曼光谱分解后的主要特征峰的峰位 882 cm^{-1} (CC 骨架伸缩)、 1052 cm^{-1} (CC 伸缩)、 1096 cm^{-1} (CC 伸缩)、 1247 cm^{-1} (平面内变形)、 1453 cm^{-1} (CH_3CH_2 变形), 均能与原始峰位一一对应。从图 6 中可以看出, 混合物 S3-V3 的每个 Voigt 谱峰均能与其三种组分的至少一个谱峰相对应。

2.2 混合物和其组分参考光谱同积分时间和功率下的识别结果

对于分解为 L 个 Voigt 峰的待测混合物样本, 通过计算每种组分在混合物每个 Voigt 峰位光强的贡献值, 依据朗伯-比尔定律中光强与其所占浓度的正比例关系, 利用式(4)得

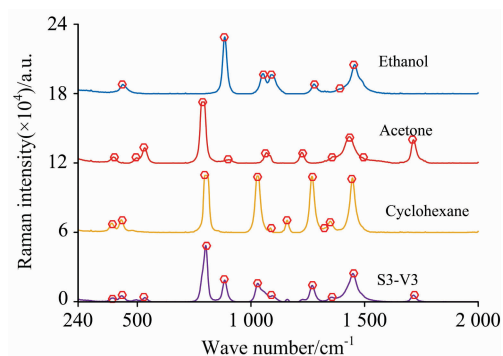


图 5 乙醇、丙酮、环己烷和 S3-V3 的预处理光谱和峰位

Fig. 5 The preprocessed spectra and peaks of Ethanol, Acetone, Cyclohexane and S3-V3

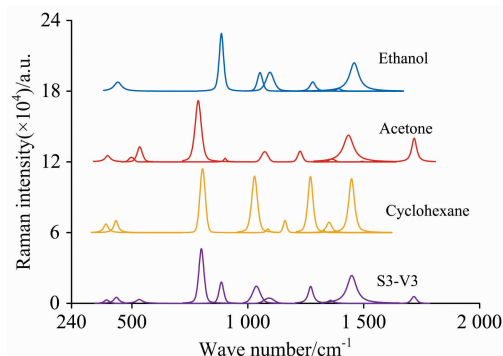


图 6 基于 Voigt 的乙醇, 丙酮, 环己烷和 S3-V3 的分解光谱

Fig. 6 The spectra decomposition of Ethanol, Acetone, Cyclohexane and S3-V3 base on Voigt

到混合物各组分的体积浓度。采用本文提出的算法对混合物及其组分在同一积分时间和功率下的光谱进行定量分析, 即选择表 1 的 S1—S5 共 5 种混合物(每种混合物各 9 个体积浓度比), 共 45 个样本。表 2 为根据组分划分的预测体积浓度与实际体积浓度的 R , RMSE, RPD 以及含有该组分的样本个数。从表 2 中可以发现, 六种组分的预测结果相关系数均在 0.95 以上, 均方根误差都小于 6%, 并且 RPD 指标仅乙醇相对较低, 为 2.921 5, 其余五种组分 RPD 指标均在 3.0 以上。图 7 为 6 种组分的散点图, 横轴为实际体积浓度, 依据本研究选用的体积比, 其取值为 0.1, 0.2, 0.3, 0.5 和 0.7, 纵轴为通过算法预测的体积浓度比。从图中可以看出, 六种组分的浓度与其拉曼谱峰光强都呈现较强的正相关。

表 2 混合物和其组分参考光谱同积分时间和功率下的结果

Table 2 The quantitative analysis results of the mixture and its component reference spectra under same integration time and power

	乙醇	乙腈	丙酮	环己烷	二丙酮醇	丙二酸二乙酯
r	0.950 7	0.990 9	0.972 6	0.971 0	0.966 2	0.991 0
RMSE	0.059 3	0.030 8	0.040 8	0.053 6	0.055 4	0.046 9
RPD	2.921 5	5.852 7	4.221 7	3.261 7	3.128 8	3.730 5
Number of samples	27	9	36	18	27	18

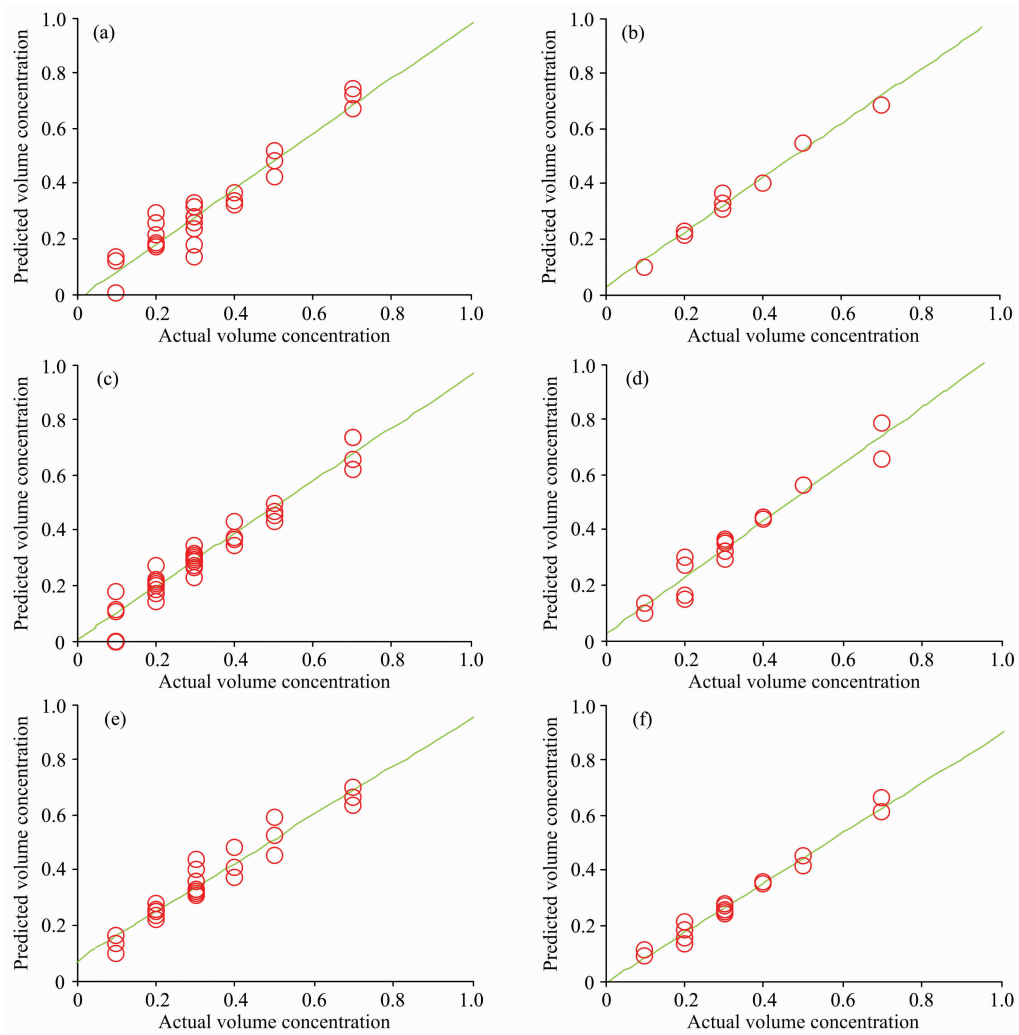


图 7 预测体积浓度与实际体积浓度的散点图

(a): 乙醇; (b): 乙腈; (c): 丙酮; (d): 环己烷; (e): 二丙酮醇; (f): 丙二酸二乙酯

Fig. 7 Scatter plot of predicted and actual volume concentrations

(a): Ethanol; (b): Acetonitrile; (c): Acetone; (d): Cyclohexane; (e): Diacetone alcohol; (f): Diethyl malonate

2.3 混合物和其组分参考光谱不同积分时间和功率下的识别结果

在实际应用中, 为了得到信噪比适当的混合物拉曼光谱数据, 往往需要对拉曼光谱仪的功率或积分时间进行调整, 从而导致混合物与各组分光谱测量条件的差异。为了验证算法的适应性, 论文也对不同积分时间和功率下的混合物识别结果做了测试, 即选择表 1 的 S6—S10 共 5 种混合物(每种

混合物各 9 个体积浓度比), 共 45 个样本, 组分的参考光谱统一选择光源功率 300 mW 和积分时间 2 s。表 3 给出了具体的定量分析结果。由表 3 可知, 六种组分的 r 均大于 0.93, 最大均方根误差为 7.94%, 除了组分乙醇、乙腈的 RPD 值小于 2.5(分别为 2.191 5 和 2.203 0), 其余 4 种组分的 RPD 值均接近或大于 3.0, 证明了该算法具有较强的鲁棒性。

表 3 混合物和其组分参考光谱不同积分时间和功率下的定量分析结果
Table 3 The quantitative analysis results of the mixture and its component reference spectra under different integration time and power

	乙醇	乙腈	丙酮	环己烷	二丙酮醇	丙二酸二乙酯
r	0.963 6	0.939 1	0.951 1	0.986 0	0.991 8	0.945 8
RMSE	0.079 0	0.079 4	0.058 3	0.028 9	0.033 0	0.059 4
RPD	2.191 5	2.203 0	3.000 0	6.045 7	5.240 9	2.913 6
Number of samples	27	18	18	18	27	27

3 结 论

提出了一种基于拉曼光谱峰强度最小二乘拟合的混合物组分直接定量分析方法。为了验证算法的可行性,本研究配置了 90 个三元混合物样本,其中样本与其组分参考光谱同功率和积分时间下的 45 个,不同功率和积分时间下的 45 个。结果表明,在同功率和积分时间下,所有组分的相关系

数都在 0.95 以上,均方根误差均小于 6%,RPD 最低为 2.921 5,其余组分 RPD 指标均在 3.0 以上;在不同功率和积分时间下,所有组分的相关系数均大于 0.93,最大均方根误差为 7.94%,RPD 指标相比同功率和积分时间下的略有下降,但仍大于 2.0,满足混合物组分快速定量分析的要求,证明了该算法良好的鲁棒性,在实际应用中具有重要的价值。

References

- [1] Stöckel S, Kirchoff J, Neugebauer U, et al. *Journal of Raman Spectroscopy*, 2016, 47(1): 89.
- [2] Fernandes D, Pacheco M, Lednev I, et al. *Journal of Raman Spectroscopy*, 2016, 47(1): 28.
- [3] Liu Z F, Huang M, Zhu Q B, et al. *Journal of Food Engineering*, 2019, 258: 9.
- [4] TAN Bing, HUANG Min, ZHU Qi-bing, et al(檀兵,黄敏,朱启兵,等). *Chinese Journal of Lasers(中国激光)*, 2018, 45(8): 0811002.
- [5] LIU Yan-de, JIN Tan-tan, WANG Hai-yang(刘燕德,靳昙昙,王海阳). *Optics and Precision Engineering(光学精密工程)*, 2015, 23(9): 2490.
- [6] GAO Ying, DAI Lian-kui, ZHU Hua-dong, et al(高颖,戴连奎,朱华东,等). *Chinese Journal of Analytical Chemistry(分析化学)*, 2019, 47(1): 67.
- [7] SU Hui, PAN Hao-ran, YAO Zhi-xiang, et al(粟晖,潘浩然,姚志湘,等). *Spectroscopy and Spectral Analysis(光谱学与光谱分析)*, 2019, 39(6): 1742.
- [8] Fan X, Ming W, Zeng H, et al. *Analyst*, 2019, 144(5): 1789.
- [9] Zhang X, Lin T, Xu J, et al. *Analytica Chimica Acta*, 2019, 1058: 48.
- [10] LI Bao-qiang, LI Cui-ping, HUANG Qi-bin, et al(李宝强,李翠萍,黄启斌,等). *Journal of Chinese Mass Spectrometry Society(质谱学报)*, 2015, 36(3): 199.
- [11] LIU Ming-hui, DONG Zuo-ren, XIN Guo-feng, et al(刘铭晖,董作人,辛国锋,等). *Chinese Journal of Lasers(中国激光)*, 2017, 44(5): 0511003.

Quantitative Analysis Method for Mixture With Known Components Based on Raman Spectroscopy

YAN Fan¹, ZHU Qi-bing^{1*}, HUANG Min¹, LIU Cai-zheng¹, LEI Ze-min², ZHANG Heng², ZHANG Li-wen², LI Min²

1. Key Laboratory of Advanced Process Control for Light Industry, Ministry of Education, Jiangnan University, Wuxi 214122, China

2. Beijing Zolix Instruments Co., Ltd., Beijing 101102, China

Abstract Quantitative analysis of mixture components by Raman spectroscopy is a difficult problem in analytical chemistry. The existing quantitative analysis methods of the mixture based on machine learning (such as support vector regression, partial least squares) have the problems of difficult to obtain training samples and poor generalization performance of models. A direct quantitative analysis method for mixture with known components based on Raman peak intensity and the least square fitting algorithm was proposed in this study. Firstly, the Raman spectra of the mixture and its components were collected respectively, and the noise reduction and baseline correction for the Raman spectra were conducted by using the combination of continuous wavelet transform and penalized least square method. Secondly, the preprocessed spectra were divided into several spectral subintervals by slope comparison method, each subinterval was regarded as the linear superposition of several Voigt functions, and the positions, intensities and half-widths of the peaks were obtained by Levenberg-Marquardt-Fletcher (LMF) algorithm. Thirdly, the contribution value of each component to the spectral peak of the mixture was determined based on the peak position. Finally, the over determined equation was established based on Lambert-Beer law that the peak intensity of mixture is proportional to its concentration of components, the coefficients corresponding to each component were obtained by fitting the equation with the least square method, so that the volume concentration of each component was obtained. In this study, ten kinds of ternary

mixtures (9 volume concentration ratios of each ternary mixture) were prepared with 6 components, including ethanol, acetonitrile, acetone, cyclohexane, diacetone alcohol and diethyl malonate, and Raman spectral data of 90 mixtures and 6 components were collected. If the spectra of mixtures and their components were acquired at the same measurement conditions (power and integral time), the obtained correlation coefficient (r) for each component was above 0.96, the root means square error (RMSE) was less than 6%, and the residual prediction deviation (RPD) was greater than 2.5. If the spectra of mixtures and their components were acquired at the different measurement conditions, the correlation coefficient (r) was above 0.93, the RMSE was less than 7.94%, and the RPD was greater than 2.0, this proved that the algorithm has good accuracy and robustness for quantitative analysis of mixture based on Raman spectroscopy. The proposed method can achieve a rapid and accurate direct quantitative analysis of ternary mixtures, which provides an effective way for the quantitative analysis of mixtures.

Keywords Raman spectroscopy; Mixture component; Quantitative analysis; Intensity of Raman spectral peak; Voigt function

(Received Oct. 11, 2019; accepted Feb. 18, 2020)

* Corresponding author