

## 激光与光电子学进展

## 基于太赫兹光谱和机器学习算法的二元及三元混合物定量分析

马卿效, 李春, 李天莹, 蒋玲\*

南京林业大学信息科学技术学院, 江苏 南京 210037

**摘要** 利用太赫兹时域光谱系统测量了苯甲酸、山梨酸、木糖醇 3 种常用食品添加剂及其混合物的太赫兹吸收光谱, 并选取了偏最小二乘回归(PLS)、最小二乘支持向量机(LS-SVM)及反向传播神经网络(BPNN)3 种机器学习算法, 对食品添加剂二元及三元混合物进行了定量分析。研究发现, 在多元混合物的定量分析中, 非线性模型 LS-SVM 及 BPNN 较线性模型 PLS 更具优势, 且随着混合物成分的增加, 非线性模型的优势愈趋明显; 两种非线性模型中, LS-SVM 较 BPNN 建模步骤固定, 无需进行复杂的参数讨论与优化, 可高效地实现多元混合物的定量分析; 此外, 观察分析物光谱特征发现, 除算法适用性讨论外, 分析物的光谱特征也会在一定程度上影响定量检测的精度。

**关键词** 光谱学; 太赫兹光谱; 多元混合物; 机器学习; 定量分析; 食品添加剂

中图分类号 O433.4

文献标志码 A

DOI: 10.3788/LOP202259.1930003

## Quantitative Analysis of Binary and Ternary Mixtures Based on Terahertz Spectroscopy and Machine Learning Algorithm

Ma Qingxiao, Li Chun, Li Tianying, Jiang Ling\*

College of Information Science and Technology, Nanjing Forestry University, Nanjing 210037, Jiangsu, China

**Abstract** In this paper, we use the terahertz time-domain spectroscopy system to measure the terahertz absorption spectra of three common food additives, namely, benzoic acid, sorbic acid, and xylitol, along with their mixtures. In addition, we select three machine learning algorithms to analyze the binary and ternary mixtures of food additives, namely, the partial least squares regression (PLS), the least squares support vector machine (LS-SVM), and the backpropagation neural network (BPNN). We find that in the quantitative analysis of multivariate mixtures, the nonlinear models LS-SVM and BPNN are more advantageous than the linear model PLS. As the mixture composition increased, the advantages of using a nonlinear model for analysis become more obvious. Among the two nonlinear models, we find that LS-SVM has a fixed modeling step compared with that of BPNN and does not require complicated parameter discussion and optimization, which can efficiently realize the quantitative analysis of multivariate mixtures. Moreover, by observing the spectral characteristics of the analyte, it is found that in addition to the discussion of the applicability of the algorithm, the spectral characteristics of the analyte also affect the accuracy of quantitative detection to a certain extent.

**Key words** spectroscopy; terahertz spectroscopy; multivariate mixtures; machine learning; quantitative analysis; food additives

## 1 引言

太赫兹 (THz) 波是指频率为 0.1~10 THz, 介于微波与红外之间的电磁波, 与其他波段的电磁辐射相比, 其具有无损性、强穿透性、指纹谱性等诸多特性<sup>[1-3]</sup>, 基于这些特性, 利用 THz 光谱进行生物相关研

究成为了 THz 领域的前沿热点<sup>[4]</sup>。

食品添加剂是指可添加到食品中以改善食品质量、延长食品保质期或增加营养的合成化学品或天然物质, 但是, 一些非法贸易商为了利益而滥用食品添加剂, 因此, 进行快速准确的食品添加剂定量检测至关重要。近年来, THz 光谱技术结合机器学习算法实现定

收稿日期: 2021-08-02; 修回日期: 2021-09-14; 录用日期: 2021-10-13

基金项目: 国家自然科学基金(62001235)、江苏省自然科学基金(BK20161526)

通信作者: \*jiangling@njfu.edu.cn

量分析已广泛应用于该领域,2019年,Sun等<sup>[5]</sup>利用多元线性回归(MLR)算法实现了牛奶中三聚氰胺的定量分析,得到预测集相关系数为0.985,均方根误差为1.38%;同年,该研究者<sup>[6]</sup>比较了MLR、偏最小二乘回归(PLS)以及最小二乘支持向量机(LS-SVM)3种机器学习算法在小麦粉中苯甲酸的定量检测精度,结果显示LS-SVM获得了最理想的检测效果,预测集相关系数高达0.994,均方根误差仅为0.12%;2020年,胡军等<sup>[7-8]</sup>采用平滑、多元散射校正、基线校正和归一化等方法对原始光谱进行校正处理,并比较了常用的定量检测算法,如PLS、LS-SVM、反向传播神经网络(BPNN)和广义回归神经网络(GRNN)等模型,在奶粉中三聚氰胺及面粉中苯甲酸的定量检测,结果均表明BPNN可获得理想的检测精度,相关系数均达到了0.99以上;同年,Lian等<sup>[9]</sup>使用THz时域光谱系统研究了奶粉中山梨酸钾的含量,发现采用PLS算法,可实现单种物质成分的定量检测,且经过波长选取后,定量检测精度得到了进一步提升;2021年,Wei等<sup>[10]</sup>比较了PLS、BPNN、遗传算法-支持向量机回归(GA-SVR),改进灰狼算法-支持向量机回归(GWO-SVR)4种定量回归模型在大豆蛋白质定量检测中的表现,结果表明BPNN结合相关预处理可得到最理想的检测精度,检测相关系数高达0.9677。大量研究结果表明,基于THz光谱技术,选用合适的机器学习算法可高效实现食品中食品添加剂的定量检测<sup>[11-12]</sup>。然而,目前已报导的基于THz光谱的食品添加剂的定量研究多集中于比较不同算法在特定对象中的检测效果,遴选出具有最高检测精度的算法,但此方式下确定的最终模型仅适合该样本的定量分析,研究并未从光谱特性以及算法原理展开,深入分析不同算法下检测精度存在差异的原因;此外,目前基于THz光谱技术的食品添加剂定量分析多集中于单种成分的研究,鲜有同时含多种成分食品添加剂的研究,因此,实现基于THz光谱的多元混合物定量分析更具实际应用价值。

本文基于THz时域光谱(THz-TDS)系统,对苯甲酸、山梨酸、木糖醇3种食品添加剂的混合物样本进行光谱测量,讨论了PLS、LS-SVM以及BPNN这3种机器学习算法在食品添加剂混合物定量分析中的适用性。研究结果表明,非线性模型LS-SVM以及BPNN在多元混合物的定量分析中效果较优,且随着混合物成分的增加,由于光谱特征中非线性因素增加,非线性模型的优势愈趋明显;进一步,通过算法原理分析发现:两个非线性模型中,LS-SVM模型固定,参数单一,无需进行复杂的调参便可实现定量检测中浓度与吸收系数的完整映射,LS-SVM较BPNN定量检测效果更为理想;同时观察光谱特征发现,除算法外,不同的光谱特征也会在一定程度上影响检测结果。本文以食品添加剂为对象,研究混合物定量分析算法,旨在为实际生活中各类物质多元混合物的定量分析提供参考。

## 2 实验部分

### 2.1 THz-TDS系统

实验采用的THz-TDS系统来自Advantest公司,型号为TAS7500SP。该系统包含透射、反射和衰减全反射3个模块,本文采用透射模块。仪器分辨率为7.6 GHz,测量时需保持空气干燥,减少水分影响。TAS7500SP型THz-TDS系统采用双激光同步调制脉冲。激光器1将脉冲入射到发射器,在电场作用下产生THz信号,THz信号经抛物面到达样品表面,不同被测物质在THz脉冲下的吸收特性也是不同的,载有样品特征的THz波与激光器2的脉冲在探测器共线,通过模数转换与傅里叶变换等处理,提取样品吸收系数等特性。

### 2.2 样品制备

3种食品添加剂山梨酸,苯甲酸,木糖醇(>99%纯度)和聚乙烯(PE)均购自西格玛奥德里奇公司(Sigma-Aldrich),样品采用压片法制备,将样本经恒温干燥箱干燥后按图1制样的质量分数称取后,放入玛瑙钵中混合并充分研磨,在12 Mpa的压强下进行压片,保持样本直径为12 mm,厚度为1.2 mm。其中,每个样本质量为200 mg,其中75%为聚乙烯,其余25%为食品添加剂的混合物,制样完成后须放入恒温干燥箱,待测试取出,避免吸收水分,影响实验精度。

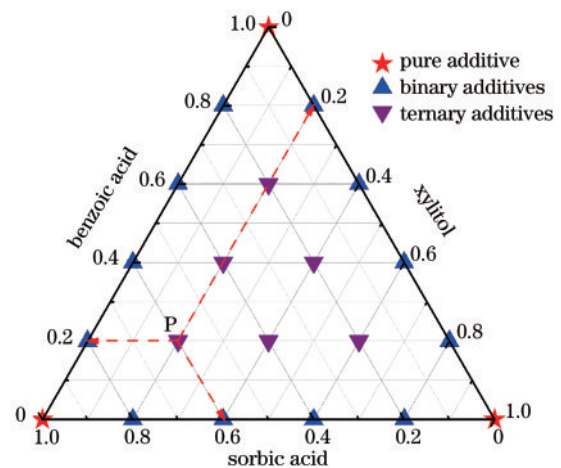


图1 食品添加剂混合物制样三角图

Fig. 1 Triangle diagram of food additive mixture sample preparation

21种食品添加剂样品可分为3类,即纯添加剂、二元混合物和三元混合物。图1三角形的3个角上展示了仅包含1种食品添加剂的3种纯样;3个边上标记了12种二元混合物样本,三角形内部展示了6种三元混合物样本。例如,样品P包含60%的山梨酸、20%的苯甲酸和20%的木糖醇。对于每种食品添加剂混合物,准备一式三份的样品,并对每个样品重复光谱实验3次,总共测试了189组数据集,用于后续定量分析,

数据集按 2:1 的比例随机分为训练集和预测集。

### 2.3 模型建立与评价

#### 2.3.1 PLS 算法

PLS<sup>[13]</sup> 是基于最小二乘回归的多元线性回归算法,适用于小样本与多重线性相关问题,PLS 模型的核心表达式为

$$y = \sum_{i=1}^n k_i x_i + b + \mu, \quad (1)$$

式中:  $y$  为模型中待检成分的预测浓度;  $x$  为光谱经主成分降维后的自变量;  $k$  为模型中回归系数;  $b$  为模型的截距;  $\mu$  为可接受范围内的误差。

#### 2.3.2 BPNN

BPNN<sup>[14]</sup> 是目前应用最为广泛的神经网络,其具有任意条件下的模式分类能力以及极强的多维函数映射能力,解决了线性模型不能实现的复杂映射问题,其隐藏层与输出层的计算如下:

$$H_j = f\left(\sum_{i=1}^n W_{ij} X_i - a_j\right), \quad (2)$$

$$O_k = \sum_{j=1}^L H_j W_{jk} - b_k, \quad (3)$$

式中:  $L$  是隐藏层节点的数量;  $f$  是隐层激励函数;  $W_{ij}$  是输入层和隐藏层中神经元之间的连接权重;  $X$  是输入序列;  $H$  是隐藏层的输出;  $a_j$  和  $b_k$  分别为节点  $j$  和节点  $k$  的阈值;  $W_{jk}$  是隐藏层和输出层中神经元之间的连接权重。影响 BPNN 性能的因素较多,包括隐藏层层数  $S$ 、隐藏层节点数  $L$ 、隐藏层之间激励函数  $f$ 、各层神经元之间的连接权重  $W$  等。BPNN 的参数选择以及结构确定至今尚无一种统一且完整的理论指导,一般只能由经验选定。网络结构选择过大,计算量对应增加,甚至出现过拟合现象;若选择过小,则又会造成网络不收敛,无法精准预测这一现象。

#### 2.3.3 LS-SVM 模型

LS-SVM<sup>[15]</sup> 是在支持向量机的基础上将不等式约束替换为等式约束,在回归中,它是将向量映射到一个更高维的空间里,在这个空间里建立一个最大超平面,使所有点都落在这个平面,通过对参数的调整,所有点到这个面的总距离最小,完整建立模型,这个平面就是所建立的预测映射方程。影响该模型较重要的参数是核函数,本文选用了径向基函数(RBF)作为核函数,表达式为

$$k(x, x_i) = \exp\left(-\frac{\|x, x_i\|^2}{2\sigma^2}\right), \quad (4)$$

式中:  $x_i$  是核函数中心;  $x$  是样本数据点;  $\sigma^2$  是内核函数。此算法中,核函数作为主要参数决定了最终检测精度,决定了该方法不但算法简单,且具有较好的鲁棒性。

#### 2.3.4 模型评价

3 种算法在回归中的表现均采用相关系数  $R$  和均方根误差 ( $E_{\text{RMSE}}$ ) 评价,  $R$  的范围为  $(0, 1)$  之间,  $R$  为 1

时,该模型建立完全准确,预测浓度与实际浓度完全一致,  $E_{\text{RMSE}}$  取值为  $(0, +\infty)$ ,  $E_{\text{RMSE}}$  为 0 时,模型精度最高,  $R$  及  $E_{\text{RMSE}}$  计算如下:

$$R = \frac{\sqrt{\sum_{i=1}^n (y'_i - \bar{y})^2}}{\sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}, \quad (5)$$

$$E_{\text{RMSE}} = \sqrt{\frac{\sum_{i=1}^n (y'_i - y)^2}{n}}, \quad (6)$$

式中:  $y_i$  为待检成分的实际浓度;  $y'_i$  为待检成分的预测浓度;  $\bar{y}$  为样本浓度的平均值。

## 3 分析与讨论

### 3.1 二元食品添加剂混合物定量分析

苯甲酸和木糖醇的二元混合物吸收光谱如图 2 所示,随着木糖醇浓度降低,苯甲酸浓度增加,吸收光谱由木糖醇特征逐渐转变为苯甲酸特征。从图 2 中可以看出,不同混合浓度的光谱特征并没较大区别,但吸收系数幅度存在差异,需选取合适算法建立定量模型表征吸收系数与食品添加剂浓度间的定量关系。

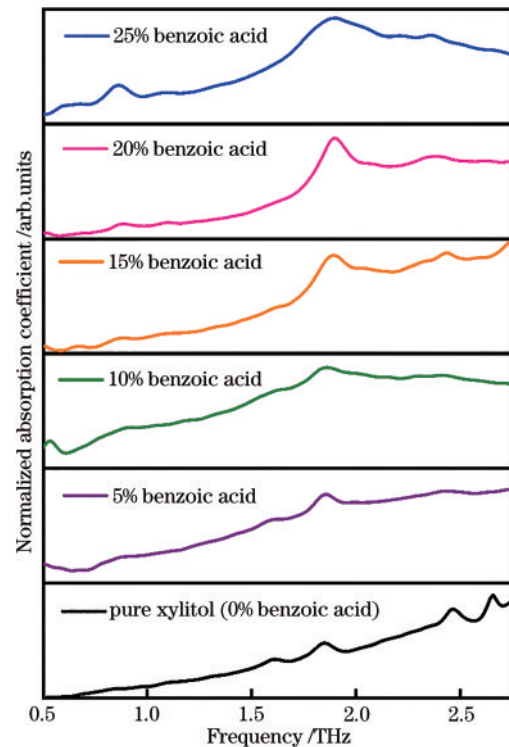


图 2 苯甲酸和木糖醇混合物吸收光谱

Fig. 2 Absorbance spectra of mixtures of benzoic acid and xylitol

图 3 展示了 PLS、LS-SVM 以及 BPNN 算法在二元混合物定量分析中的结果,为了让模型更加科学稳定,选用交叉验证(十折)用于模型构建与预测。在苯甲酸和木糖醇的混合物的定量检测中[图 3(a)],

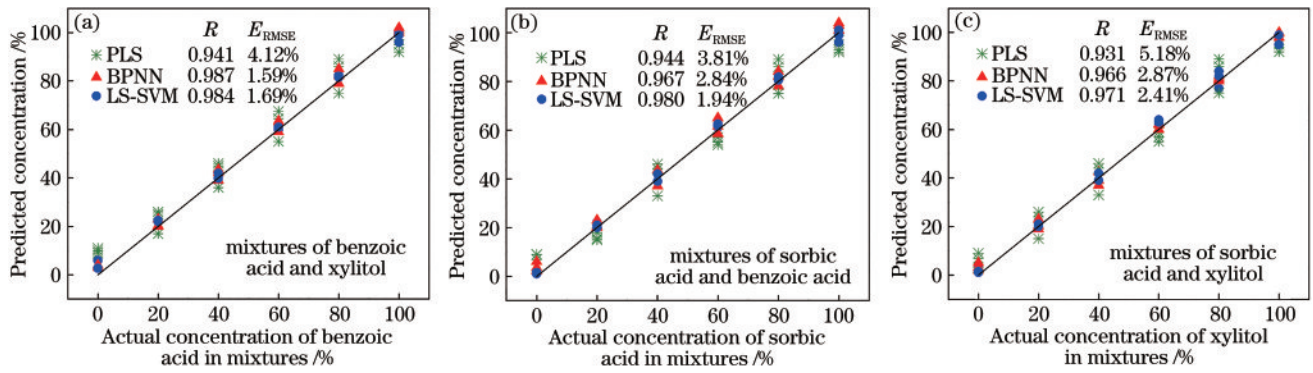


图3 二元食品添加剂混合物的定量分析结果。(a)苯甲酸与木糖醇的二元混合物定量分析结果;(b)苯甲酸与山梨酸的二元混合物定量分析结果;(c)山梨酸与木糖醇的二元混合物定量分析结果

Fig. 3 Quantitative analysis results of binary food additive mixtures. (a) Quantitative analysis results of binary mixtures of benzoic acid and xylitol; (b) quantitative analysis results of binary mixtures of benzoic acid and sorbic acid; (c) quantitative analysis results of binary mixtures of sorbic acid and xylitol

BPNN 展现了最好的分析性能,  $R$  为 0.987,  $E_{RMSE}$  为 1.59%, 同时 LS-SVM 也可提供较高的精度,  $R$  为 0.984,  $E_{RMSE}$  为 1.69%; 在山梨酸和苯甲酸混合物的定量分析中[图 3(b)], LS-SVM 预测浓度值更接近实际浓度,  $R$  为 0.980,  $E_{RMSE}$  为 1.94%; 山梨酸和木糖醇混合物的定量检测结果类似[图 3(c)], 与 BPNN 相比, LS-SVM 精度略高, 线性模型 PLS 误差最大。总体上看, 由于 BPNN 和 LS-SVM 具有较强的非线性映射能力, 在两种算法分别选取合适的网络参数和核函数后, 可以在一定程度上克服二元混合物中的非线性干扰, 所建立的非线性模型表现均优于 PLS 线性回归模型。此外, 定量分析的准确率除与模型特点有关外, 样本本身的光谱特性也会影响检测精度, 例如, 含山梨酸的混合物预测结果比苯甲酸和木糖醇的混合物的预测结果差, 归因于山梨酸具有相对较弱且较杂的吸收峰这一光谱特征。

### 3.2 三元食品添加剂混合物定量分析

在此基础上, 增加了样本复杂度, 进行了三元混合物的定量分析, 图 4 给出了 3 种食品添加剂及其三元混合物(苯甲酸-山梨酸-木糖醇)的光谱结果, 该混合物中 3 种成分含量比例为 1:1:1。光谱结果显示苯甲酸在 0.64、0.88、1.08、1.88 和 2.37 THz 处观察到 5 个峰, 山梨酸在 0.95、1.20、1.43、1.53、1.63、1.79、2.10、2.32、2.56 和 2.65 THz 处观察到 10 个峰, 木糖醇分别在 1.62、1.85、2.45 和 2.64 THz 处观察到 4 个峰, 而三元混合物的吸收光谱同时包含 3 种纯添加剂的光谱特征。值得注意的是, 本文中 3 种食品添加剂的吸收峰与已有报道中的检测结果基本一致<sup>[16-17]</sup>。

基于三元混合物的 189 份不同浓度光谱数据, 采用 3 种机器学习算法结合交叉验证预测了三元混合物中苯甲酸的浓度, 如图 5 所示, 从图中可以看出, 随着混合物成分的增加, PLS 线性模型的性能愈趋下降, 浓度预测点明显偏离 0 误差线, 而非线性模型 BPNN 和

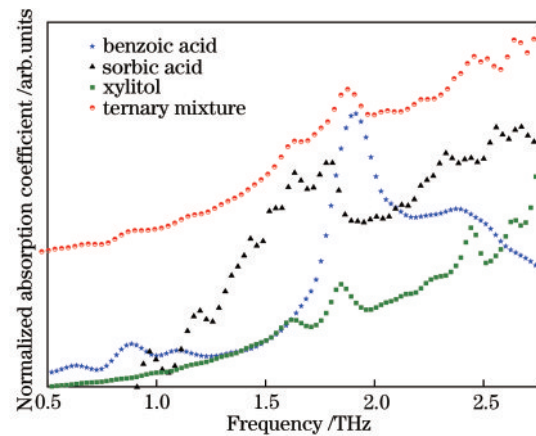


图4 苯甲酸、山梨酸、木糖醇以及三元混合物的吸收光谱  
Fig. 4 Absorbance spectra of benzoic acid, sorbic acid, xylitol, and ternary mixture

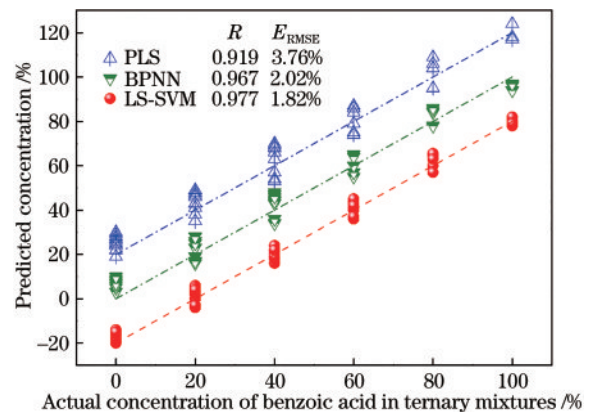


图5 三元食品添加剂混合物中苯甲酸的定量分析结果  
Fig. 5 Quantitative analysis results of benzoic acid in ternary food additive mixtures

LS-SVM 仍可获得较高的检测精度。其中, PLS 是线性回归中最经典的算法, 但是在对三元混合物的分析中, 由于光谱数据的非线性以及算法本身的局限性, 几乎没有优势; 而 BPNN 和 LS-SVM 两种模型仍然获得

了较高的检测精度,其中 LS-SVM 结果最佳,分析算法原理发现, BPNN 在其训练过程中,需进行多参数的同时调整,如隐藏层层数  $S$ ,隐藏层节点数  $L$ ,隐藏层之间激励函数  $f$ ,各层神经元之间的连接权重  $W$  等,仅选取默认参数难以获得令人满意的结果;而 LS-SVM 模型需要调整的核心参数便是核函数,选取合适的核函数便可准确建立定量模型,本文中核函数选取了处理非线性问题中常用的 RBF。

表 1 列出了不同算法对 3 种食品添加剂成分的预测结果,其中训练集的相关系数普遍高于预测集相关系数。比较不同算法可以发现,利用 LS-SVM 模型进行定量预测的结果最优,对苯甲酸、山梨酸以及木糖醇 3 种食品添加剂的检测相关系数  $R$  分别达到了 0.977、0.986 和 0.990。相比二元食品添加剂混合物定量分析,三元食品添加剂的定量分析结果更能体现出非线性模型的优势,可以推测,混合物组成越复杂,非线性干扰越大,非线性模型相比非线性模型更具实际应用价值。

表 1 三元混合物中各食品添加剂成分的定量分析结果  
Table 1 Quantitative analysis results of various food additives in ternary mixture

Analyte	Model	Calibration sets		Prediction sets	
		$R$	$E_{\text{RMSE}}/\%$	$R$	$E_{\text{RMSE}}/\%$
Benzoic acid	PLS	0.940	2.61	0.919	3.76
	BPNN	0.981	1.63	0.967	2.02
	LS-SVM	0.988	1.30	0.977	1.82
Sorbic acid	PLS	0.932	2.81	0.905	4.26
	BPNN	0.961	2.06	0.947	2.49
	LS-SVM	0.994	1.06	0.986	1.34
Xylitol	PLS	0.927	3.06	0.914	3.91
	BPNN	0.986	1.27	0.985	1.52
	LS-SVM	0.995	1.04	0.990	1.21

此外,LS-SVM 是解决高维特征回归问题的有效手段,在特征维度大于样本数时依然有很好的效果;同时,该算法仅仅使用一部分支持向量来做超平面的决策,无需依赖全部数据,运算量相对需要多参数选取与优化的 BPNN 算法较小;当样本量不是海量数据的时候,如本文中 189 份光谱数据 ( $189 \times 302$ ),LS-SVM 算法可以准确地表征此类非线性回归问题,泛化能力较强;在本文多元混合物的定量检测中,相较于 BPNN,LS-SVM 建模步骤固定、调整参数较少、无需进行复杂的参数讨论与优化,便可实现较高精度的定量检测,是实现此类多元混合物定量检测的理想算法。

除算法适用性讨论外,基于相同算法,对不同分析物进行定量分析的精度也存在差异。本文中,木糖醇在 1.62、1.85、2.45 和 2.64 THz 处的 4 个吸收峰较为明显,光谱特征显著;苯甲酸在 1.88 THz 处有明显的吸收峰,而在 0.64、0.88、1.08 和 2.37 THz 处的 4 个吸

收峰并不突出;山梨酸的吸收峰较为杂乱,吸收系数强度较弱,吸收峰形均不完整。因此运用 LS-SVM 或 BPNN 算法,在多元混合物的定量检测中,木糖醇检测精度最高,苯甲酸次之,而山梨酸最低。可以发现,同一算法下,分析物光谱特征会对预测结果产生一定的影响。

实际应用中,线性回归算法即可实现小样本单种物质成分的定量检测,无需采用复杂的非线性回归模型;LS-SVM 建模步骤固定,调整参数较少,适用于小样本多元混合物的定量研究;而 BPNN 模型参数较多,且需要大量样本数据支撑,以构建完整的网络模型,更适用于大样本多元混合物的定量研究。此外,后期研究中,为探寻合适的定量检测模型,极端梯度增压 (XGBoost)、随机森林 (RF)、梯度提升决策 (GBDT) 等传统机器学习算法也可用于不同样本量多元混合物定量检测的对比研究分析。

## 4 结 论

以食品添加剂为研究对象,选取 PLS、LS-SVM 和 BPNN 这 3 种机器学习算法,基于 THz 光谱实现了多元混合物的定量分析。结果表明,光谱中非线性干扰会随着混合物成分的复杂性增加而增加,在多元混合物的定量检测中,与 PLS 线性模型相比,LS-SVM 和 BPNN 两种非线性模型均具有较高的预测精度。此外,由于建模步骤固定、调整参数较少、无需进行复杂的参数讨论与优化,在多元混合物的定量检测中,LS-SVM 较 BPNN 更具优势,BPNN 的预测集相关系数为 0.967、0.947 和 0.985,而 LS-SVM 对应预测集相关系数高达 0.977、0.986 和 0.990。验证了机器学习算法与 THz 光谱相结合在多元混合物定量分析中的应用潜力,实现了食品添加剂多元混合物成分含量的高效、快速、无损检测,有望推进其在生物、食品、药物等领域的实际应用。

## 参 考 文 献

- [1] Feng H Z, Mohan S. Application of process analytical technology for pharmaceutical coating: challenges, pitfalls, and trends[J]. AAPS PharmSciTech, 2020, 21(5): 179-183.
- [2] Huang Y X, Singh R, Xie L J, et al. Attenuated total reflection for terahertz modulation, sensing, spectroscopy and imaging applications: a review[J]. Applied Sciences, 2020, 10(14): 4688-4706.
- [3] Jain R K, Hoffman A J, Jepsen P U, et al. Mid-infrared, long-wave infrared, and terahertz photonics: introduction [J]. Optics Express, 2020, 28(9): 14169-14175.
- [4] Peng Y, Shi C J, Wu X, et al. Terahertz imaging and spectroscopy in cancer diagnostics: a technical review[J]. IEEE Transactions on Terahertz Science and Technology, 2018, 8(6): 696-707.
- [5] Sun X, Zhu K, Hu J, et al. Nondestructive detection of melamine in milk powder by terahertz spectroscopy and

- correlation analysis algorithm[J]. *Journal of Applied Spectroscopy*, 2019, 86(4): 661-665.
- [6] Sun X D, Zhu K, Liu J B, et al. Terahertz spectroscopy determination of benzoic acid additive in wheat flour by machine learning[J]. *Journal of Infrared, Millimeter, and Terahertz Waves*, 2019, 40(4): 466-475.
- [7] 胡军, 刘燕德, 孙旭东, 等. 基于BP神经网络的太赫兹时域光谱对面粉中苯甲酸的定量检测研究[J]. *激光与光电子学进展*, 2020, 57(7): 073002.  
Hu J, Liu Y D, Sun X D, et al. Quantitative determination of benzoic acid in flour based on terahertz time-domain spectroscopy and BPNN model[J]. *Laser & Optoelectronics Progress*, 2020, 57(7): 073002.
- [8] 胡军, 徐振, 李茂鹏, 等. 基于神经网络算法与太赫兹光谱检测技术的奶粉三聚氰胺含量测定[J]. *激光与光电子学进展*, 2020, 57(22): 223001.  
Hu J, Xu Z, Li M P, et al. Determination of melamine content in milk powder based on neural network algorithm and terahertz spectrum detection[J]. *Laser & Optoelectronics Progress*, 2020, 57(22): 223001.
- [9] Lian F Y, Fu M X, Xu D G. Qualitative and quantitative analyses of potassium sorbate in milk powder using terahertz spectra[J]. *Journal of Applied Spectroscopy*, 2020, 87(4): 764-772.
- [10] Wei X, Li S, Zhu S P, et al. Terahertz spectroscopy combined with data dimensionality reduction algorithms for quantitative analysis of protein content in soybeans[J]. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 2021, 253: 119571.
- [11] Tang M J, Xia L P, Wei D S, et al. Rapid and label-free metamaterial-based biosensor for fatty acid detection with terahertz time-domain spectroscopy[J]. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 2020, 228: 117736.
- [12] Zhao R, Zou B, Zhang G L, et al. High-sensitivity identification of aflatoxin B1 and B2 using terahertz time-domain spectroscopy and metamaterial-based terahertz biosensor[J]. *Journal of Physics D: Applied Physics*, 2020, 53(19): 195401.
- [13] 管爱红, 李智, 葛宏义. 红薯淀粉中添加剂明矾的定性和定量太赫兹时域光谱技术检测[J]. *光谱学与光谱分析*, 2018, 38(1): 267-270.  
Guan A H, Li Z, Ge H Y. The qualitative and quantitative detection of potassium alum in sweet potato starch based on terahertz time-domain spectroscopy[J]. *Spectroscopy and Spectral Analysis*, 2018, 38(1): 267-270.
- [14] Aghelpour P, Mohammadi B, Biazar S M. Long-term monthly average temperature forecasting in some climate types of Iran, using the models SARIMA, SVR, and SVR-FA[J]. *Theoretical and Applied Climatology*, 2019, 138(3/4): 1471-1480.
- [15] Nithya A A, Lakshmi C. Enhancing iris recognition framework using feature selection and BPNN[J]. *Cluster Computing*, 2019, 22(5): 12363-12372.
- [16] Lepodise L M, Horvat J, Lewis R A. Terahertz (6-15 THz) spectroscopy and numerical modeling of intermolecular vibrations in benzoic acid and its derivatives[J]. *Applied Spectroscopy*, 2015, 69(5): 590-596.
- [17] Jiang Y Y, Li G M, Lv M, et al. Determination of potassium sorbate and sorbic acid in agricultural products using THz time-domain spectroscopy[J]. *Chinese Physics B*, 2020, 29(9): 098705.