第 43 卷 第 19 期/2023 年 10 月/光学学报 研究论文

光学学报

基于太赫兹光谱的三聚氰胺定量分析

郭以恒**，燕芳*，赵渺钰，卓炫

内蒙古科技大学信息工程学院，内蒙古 包头 014010

摘要 为实现对奶粉中非法食品添加剂三聚氰胺的精确定量检测，使用太赫兹时域光谱系统测定了三聚氰胺（质量分数梯度为0%～20%）与奶粉混合物的太赫兹吸收谱曲线。首先，利用主成分回归、支持向量机回归、偏最小二乘回归和最小二乘支持向量机回归（LSSVR）对混合物中三聚氰胺质量分数进行预测。结果显示，LSSVR的预测精度最高，预测集相关系数R_P为0.99838，预测集均方根误差f_{RMSEP}为0.41%。其次，为进一步提升LSSVR预测的精度，使用粒子群算法、遗传算法、布谷鸟算法、灰狼算法对LSSVR中正则化参数C和已确定核函数为高斯核函数后核参数γ进行参数优化。结果表明，经过4种算法优化后LSSVR预测精度均明显提高，其中经灰狼算法优化后的LSSVR对混合物中三聚氰胺的预测精度最高（R_P为0.99925，f_{RMSEP}为0.28%）。该算法可为食品添加剂的定量检测提供新的方法和思路。

关键词 光谱学；太赫兹时域光谱；定量分析；参数优化；食品添加剂

中图分类号 O441.4；O629.1 **文献标志码** A **DOI**：10.3788/AOS230607

1 引 言

近年来，食品添加剂在使用过程中的超范围及超量等违规使用及滥用等问题带来的食品安全隐患再次引起了社会的高度关注，研发无损、快速及准确高效的食品添加剂定性及定量检测技术成为目前学者的研究热点。传统的食品添加剂检测方法包括离子色谱法、液相色谱法、液质联用法、气相色谱法以及分子光谱法等，但上述检测方法的设备造价高昂、检测周期长、成本高、检测精度参差不齐，对技术人员的操作及有机溶剂的纯度要求较高[1]。传统方法还存在需要预处理、检测周期长、检测操作繁杂等局限性。与传统检测手段相比，新兴的免疫检测法、生物传感器法及光谱分析法等新技术的运用，推动了食品添加剂检测技术的发展。基于太赫兹（THz）光谱技术的检测手段具有无损、快速、高效的优点，已被广泛应用于食品、医药以及环境检测等领域。

由于许多生物大分子及其弱相互作用（氢键作用以及范德华力等）的振动能级多处于THz波段[2-3]，其THz光谱（包括发射、反射和透射谱）包含丰富的物理和化学信息特征指纹谱，利用THz时域光谱（THz-TDS）技术可以得到这些物质的特征谱，进而可分析物质的内部结构信息，确定物质的归属。此外，THz波可穿透许多绝缘材料，如衣物、纸张、塑料、皮革和陶瓷，且THz辐射具有很低的光子能量，不会在生物组织中产生有害的光致电离[4-6]。目前，国内外已有利用THz光谱技术对食品中添加剂进行检测的报道。曹恩达[7]利用THz技术检测食品药品，采用谱减法降低样品光谱水汽干扰，所提方法可以准确分辨样品种类，为THz-TDS探测技术在食品和药品的无损检测应用提供重要参考。Liu[8]采用THz技术结合化学计量法测定槐花蜂蜜的高果糖糖浆含量，PLS模型的校正集均方根误差f_{RMSEC}和预测集均方根误差f_{RMSEP}分别为0.0967和0.108，该结果证实了该技术的可靠性。Sun等[9]将THz-TDS技术与机器学习方法（广义回归神经网络（GRNN）、反向传播神经网络（BPNN））相结合，用于快速测定面粉中苯甲酸含量，实验结果表明：与BPNN模型相比，GRNN模型在苯甲酸含量预测的准确性和分析速度方面都更具有优势。胡军等[10]探索了不同预处理方法对奶粉中三聚氰胺THz光谱定量检测的影响，对数据进行平滑、多元散射校正、基线校正和归一化以及多元散射校正和归一化相结合等预处理方法后，利用BPNN和GRNN检测模型分析数据，实验结果表明：经多元散射校正结合归一化校正处理后的GRNN模型效果最佳，得到的预测集相关系数R_P为0.9967，f_{RMSEP}为0.0050。验证了THz光谱检测技术对奶粉中违禁添加剂三聚氰胺检测的可行性。

2 实验部分

2.1 实验设备及样片制备

实验使用北京市工业波谱成像工程技术研究中心

收稿日期：2023-03-01；**修回日期**：2023-03-27；**录用日期**：2023-05-05；**网络首发日期**：2023-05-15

基金项目：内蒙古自治区关键技术攻关计划（2021GG0361）、内蒙古自治区直属高校基本科研业务费项目

通信作者：*0472yanfang@163.com；**guoyiheng2022@163.com

的透射式 THz-TDS 系统，THz 波的产生和探测方式为光电导天线结构法，MaiTaiHP 钛-蓝宝石飞秒激光器输出用以产生和探测 THz 波的超快红外激光，激光中心波长锁定为 800 nm，激光器红外光平均功率为 2.95 W，脉冲宽度低于 100 fs，重复频率为 79.3 MHz，GaAs 晶体为 THz 波辐射源，系统频带范围为 0.3～3.0 THz。在系统密闭光路中充入高纯度 N_2，实验时保持湿度＜4%。样品制备时使用的聚乙烯及三聚氰胺均购自阿拉丁化学试剂网，纯度均大于 99%，奶粉购买于学校附近某大型超市。实验样片采用压片法制备，样片中三聚氰胺质量比如表 1 所示。在玛瑙研钵中将称重后的样品粉末沿单一方向研磨 5 min，混合均匀后送入模具，使用压片机以 5 MPa 的压力压制 5 min 后取出，得到直径约为 13 mm，厚度约为 1.4 mm 的圆形薄片，用游标卡尺及电子天平测量并记录厚度与质量信息后装入样片袋，放入干燥箱内保存待测。按照上述方法每个质量分数梯度均制作 3 个样片，每个样片上架测量 3 次，将该样片的 3 次测量数据取平均后得到该样片最终的吸收谱数据；将实验所得的 63 个样片的吸收谱数据作为校正集，再次用同样方法制作三聚氰胺作为预测集，质量分数分别为 2%、6%、9%、11%、15%、17%、19% 的 7 个梯度共 21 个样片。

表 1　样品制备信息

Table 1　Sample preparation ratio information

No.	Mass fraction /%	Thickness /mm
1	0	1.36
2	1	1.36
3	2	1.37
4	3	1.36
5	4	1.37
6	5	1.38
7	6	1.38
8	7	1.38
9	8	1.37
10	9	1.39
11	10	1.38
12	11	1.39
13	12	1.39
14	13	1.40
15	14	1.40
16	15	1.38
17	16	1.40
18	17	1.39
19	18	1.40
20	19	1.39
21	20	1.40

2.2　数据采集

经过实验得到样片时域信号后，使用 Dorney 等[11]的参数提取模型计算获取样片的吸收系数谱，样片的折射率 $n(w)$、吸收系数 $\alpha(w)$ 可表示为

$$n(w)=\frac{\phi(w)c}{wd}+1, \tag{1}$$

$$\alpha(w)=\frac{2}{d}\ln\frac{4n(w)}{\rho(w)[n(w)+1]^2}, \tag{2}$$

式中：d 为样品厚度；c 为光速；$\phi(w)$ 为样品信号和参考信号的相位差；$\rho(w)$ 为样品信号和参考信号的振幅比；w 为频率。

考虑到系统漂移、仪器本身性能限制及样本对 THz 波的作用，在频率过高或过低时吸收谱会出现失真、信噪比低等现象，纯奶粉的吸收系数谱如图 1(a)所示，故截取频率在 0.5～2.8 THz 范围内的吸收谱数据如图 1(b)所示；因存在系统的标准具效应、样片实验时的散射和反射，对存在多重反射的样品时域信号进行傅里叶变换，得到振荡的频域样品信号[12]，采集数据后使用 Savitzky-Golay 卷积滤波器对数据进行平滑滤波，以此提高信噪比，减少吸收谱的噪声和散射干扰，降低后续定量分析的误差[13]。纯奶粉滤波前后的吸收谱、纯三聚氰胺的吸收谱及含 10% 三聚氰胺混合物的吸收谱如图 1(b)所示。

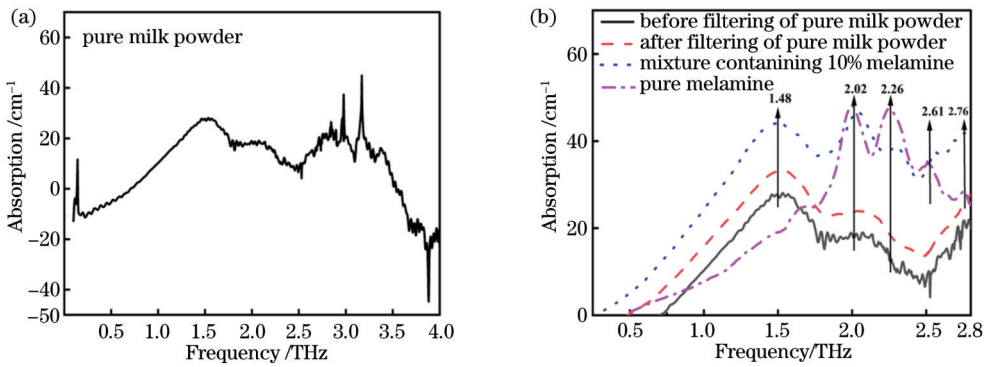

图 1　吸收谱数据。(a)纯奶粉吸收谱；(b)纯奶粉滤波、纯三聚氰胺、含 10% 三聚氰胺混合物的吸收谱

Fig. 1　Absorption spectra data. (a) Pure milk powder absorption spectra; (b) absorption spectra of pure milk powder, pure melamine and mixtures containing 10% melamine

为了容易分辨纯奶粉滤波前后曲线，将纯奶粉滤波后曲线向上平移5个单位。对比图1中纯奶粉滤波前后吸收谱线可知，滤波对吸收谱曲线趋势无影响，仅减少了干扰因素，这有利于提高后续定量分析精度。纯奶粉与纯三聚氰胺在整个频率范围内有良好的吸收特征，纯奶粉在1.48 THz处存在明显吸收特征，形成原因可能是奶粉中蛋白质和氨基酸等组分在微观层面相互作用后的影响[14]；在纯三聚氰胺吸收谱上，分子间H键相互作用、分子间大π键堆积以及分子与分子的共同作用分别形成了三聚氰胺在2.02、2.26、2.61 THz处的强吸收峰，排除实验操作与仪器设备两方面引起的合理偏差，上述实验结果与文献[10]基本一致。

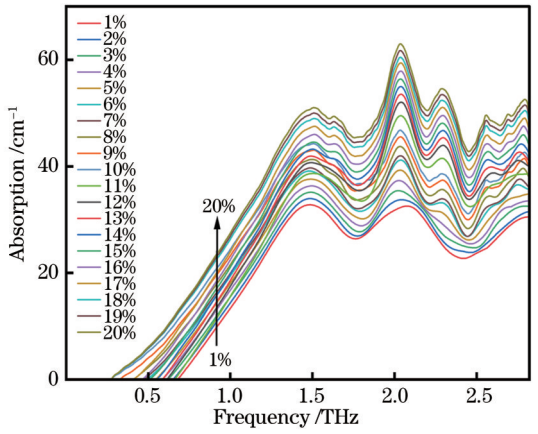

图2　奶粉混合三聚氰胺含量梯度吸收谱

Fig. 2　Gradient absorption spectra of melamine concentration in milk powder

图2为奶粉与三聚氰胺混合物（质量分数梯度为1%～20%）的吸收谱线。由图1、图2可知，混合物的特征吸收谱与奶粉单质及三聚氰胺单质的吸收谱特征明显不同，混合物共有4处吸收峰，分别位于1.48、2.02、2.25、2.63 THz，可见混合物吸收谱中同时具备了奶粉与三聚氰胺两种单质的吸收峰，THz波的指纹谱特性得到验证。由图2可知，混合物样片的吸收峰强度与样片中三聚氰胺质量分数成正比，此实验结果也符合朗伯比尔定律。吸收峰的强度以及吸收谱线的基线斜率都与样片中三聚氰胺质量分数成正比，这是回归预测定量分析的关键信息，也为利用THz-TDS信息精确定量检测三聚氰胺奠定基础。

3　回归预测及其优化

3.1　定量分析

探究混合物质量分数与THz光谱数据相关性，进而达到预测定量分析目的，需寻找数据多元变量数量关系变化的规律。对实验获得的高维度光谱数据，选择合理的预测方法非常重要，回归预测是数据分析中最常用的预测建模技术之一。基于光谱数据的维度特征，建立了主成分回归（PCR）、支持向量机回归（SVR）、偏最小二乘回归（PLSR）、最小二乘支持向量机回归（LSSVR）等4种回归预测模型进行数据分析[15-17]，比较获取的实验光谱数据在线性回归降维（PCR、PLSR）与非线性回归升维（SVR、LSSVR）两个相悖角度处理后预测的效果，并采用R_P、f_{RMSEP}作为模型性能评价指标。

$$R_P=\sqrt{1-\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2}}\times 100\%,\tag{3}$$

$$f_{RMSEP}=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2},\tag{4}$$

式中：y_i为预测集每个含量梯度3个预测值的平均值；$\hat{y}_i$为预测集每个梯度真实值；$\bar{y}$为预测集梯度真实值的平均值。

其中，PCR和PLSR的共同点都是对数据进行降维分析，用尽可能少的维度来表达最多的有效原始数据信息，达到剔除冗余信息，减少干扰因素，获得良好的预测结果。PCR属于非监督式学习，只对具有变化规律部分的自变量数据进行降维分析，而PLSR包含PCR的作用，还能将其余部分因变量数据降维作为响应回馈，利用特征约束F，达到监督式学习的效果，若充分利用已有数据建模，PLSR的预测结果理论上会比PCR更高，具体可表示为

$$X\xrightarrow{PCA}f(x)=Y,\tag{5}$$

$$X\xrightarrow{PCA}F(x)=y\xleftarrow{PCA}Y,\tag{6}$$

式中：X为自变量数据；Y为因变量数据；y为主成分分析后因变量数据；f、F为特征映射关系。

与PCR、PLSR相反，SVR、LSSVR的基本原理是对数据进行升维分析[18]。SVR通过核函数将非线性的光谱数据从非线性空间映射到高维线性可分空间，寻求使得分类间隔最大的最优超平面，同时最小错分样本数目，寻找并学习含量梯度的分类特征。

设训练集为$(x_i,y_i),i=1,\cdots,n,x\in\mathbf{R}^d,y\in\{+1,-1\}$，

$$\min\frac{1}{2}\|\boldsymbol{\theta}\|^2+C\sum_{i=1}^{n}\xi,C>0,\tag{7}$$

$$\text{st } y_i-[(\theta x_i)+b]\leqslant\varepsilon+\xi,i=1,\cdots,n$$

$$[(\theta x_i)+b]-y_i\leqslant\varepsilon+\xi,i=1,\cdots,n。$$

LSSVR基于统计理论，结合最小二乘法与支持向量回归，将不等式转换为等式问题[18]，具体可表示为

$$\min\frac{1}{2}\|\boldsymbol{\theta}\|^2+C\sum_{i=1}^{n}\eta^2,C>0,\tag{9}$$

$$\text{st } y_i-[(\theta x_i)+b]=\eta,i=1,\cdots,n。$$

式中：$\boldsymbol{\theta}$为最优超平面法向量；θ为最优超平面法向量的大小；b为最优超平面常数，用于确定最优超平面；C

为正则化参数；用于控制精度；ξ为松弛因子；ε为偏差；η为残差，用于反馈调整拟合效果。在确定使用的核函数为高斯核函数[19]后，使用交叉验证法确定超参数的选取。核参数对样本数据在映射空间中的分布复杂度有直接影响，而正则化参数则与模型对训练样本的拟合情况和模型的推广能力相关。

建立了上述4种回归预测模型，并将全频段的实验光谱数据导入模型计算，为了可视化模型的预测结果，使用了预测值(x轴)与参考值(y轴)的回归，并将斜率与和截距参数与1∶1线进行比较[20]。结果如图3所示。

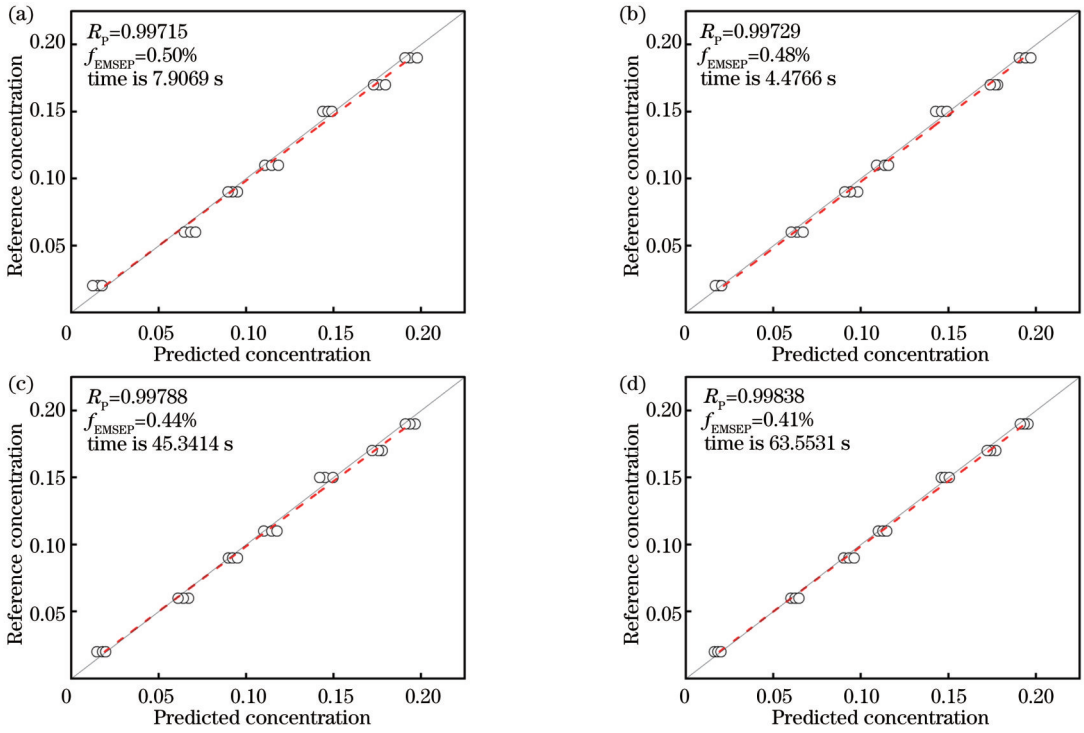

图3　4种回归模型预测结果。(a) PCR；(b) SVR；(c) PLSR；(d) LSSVR

Fig. 3　Prediction results of four regression models. (a) PCR; (b) SVR; (c) PLSR; (d) LSSVR

对比R_P和f_{RMSEP}可知，4种算法的预测效果均较好，模型也较为稳定。其中，基于PCR回归模型的R_P最小(0.99715)，f_{RMSEP}最大(0.50%)，经分析后判定其原因可能为该模型在降维过程中丢失了一些光谱特征信息，只能提取有限的信息；同为基于光谱数据的降维模型，PLSR模型比PCR模型的评价系数更好，表明PLSR模型相对PCR模型更能充分利用有限的数据。而基于LSSVR模型的预测对数据进行升维分析，使光谱特征信息更明确，但计算数据量较大，故耗时也较长。由此可见，若仅从光谱数据降维或升维的角度并不能得出最优模型。综合R_P、f_{RMSEP}指标判断更为客观合理。经过比较可知，基于LSSVR算法的回归模型对光谱数据的定量预测精度最高，更适用于基于THz光谱的添加剂定量检测技术中，但LSSVR模型的性能还受到许多因素的影响，如：样本数据预处理、核函数、模型超参数等。未来将继续针对此模型进行参数优化，提高回归算法对混合物质量分数预测的精度，进一步提升基于THz光谱技术对食品添加剂的检测能力。

3.2　参数优化

采用Savitzky-Golay卷积滤波器对实验数据进行预处理，LSSVR模型的核函数则选用了适应能力强的高斯核函数。对LSSVR模型进行参数优化的关键是选取模型超参数(核参数γ、正则化参数C)。

针对LSSVR模型，传统的利用交叉验证的方法选取超参数，继而进行模型优化时会面临难以构建数学模型及求解的难题，而基于梯度的优化算法则需满足目标函数为可导函数且导数连续。近年来，群智能优化算法因其自身优势在参数优化领域得到众多研究者关注与应用，该算法是人类观察群居动物个体间相互协作、信息共享的群体自适应行为，而提出的一种随机智能算法。与传统优化算法相比，群智能算法具有理论简单、计算精度高、适应能力强、稳定性好，对数据要求低等特点[21-23]。目前较为经典的群智能算法包括粒子群算法(PSO)[24]、遗传算法(GA)、布谷鸟算法(CS)和灰狼算法(GWO)。利用群智能算法进行参数优化时首先应确定种群，即确定某一参数解向量的集合；其次要加入随机干扰，防止出现局部最优的情况；随后进行选择操作，保留最优解，加速系统收敛；最后利用迭代公式寻找出新的最优解向量，从而完成优化[25-26]。分别利用上述4种算法对LSSVR模型进行参数优化，结果如图4所示。

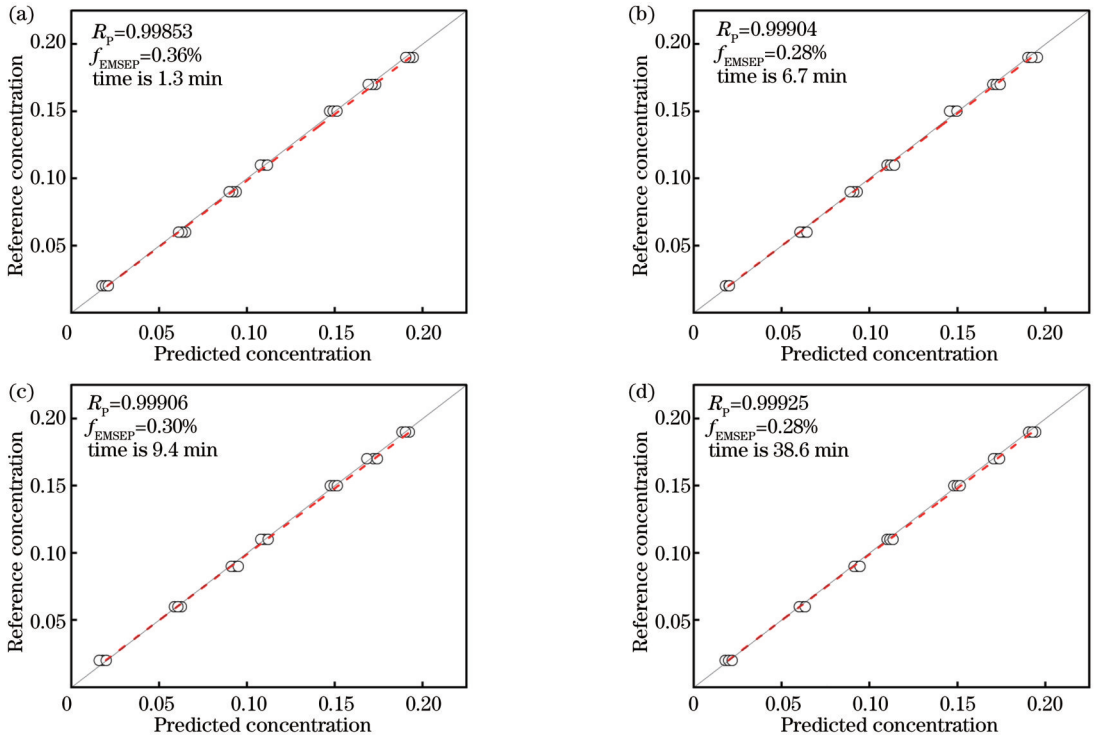

图 4　经 4 种算法优化后 LSSVR 模型预测结果。(a) GA-LSSVR; (b) PSO-LSSVR; (c) CS-LSSVR; (d) GWO-LSSVR

Fig. 4　Prediction results of LSSVR model optimized by four algorithms. (a) GA-LSSVR; (b) PSO-LSSVR; (c) CS-LSSVR; (d) GWO-LSSVR

由图 4 可知，相比于交叉验证选取超参数的 LSSVR，经过优化后的模型 R_P 均增大，f_{RMSEP} 均减小，优化效果较好，说明利用智能群算法优化选参对基于 THz 光谱定量检测三聚氰胺含量的精度有所提升，其中基于 GA 算法的优化 R_P 为 0.99853，f_{RMSEP} 为 0.36%，与其他 3 种优化方法进行对比，其耗时最短，但预测效果提升量却最小，这是因为 GA 算法的无记忆性导致解向量包含的信息丢失，故优化效果较弱。此外，其他 3 种优化方法在得到解向量时会保存相应信息。对比评价系数发现，基于 PSO 算法的优化 R_P (0.9904) 小于基于 CS 算法的优化 R_P (0.9906)，预测效果相对较差，但其 f_{RMSEP} (0.28%) 却小于后者 f_{RMSEP} (0.30%)，故基于 PSO 算法的优化模型稳定性较优，对数据有更强的适应能力，模型的推广应用能力更强。基于 GWO 算法的优化 R_P 最大为 0.99925，f_{RMSEP} 最小为 0.28%，在四者中拥有最好的预测效果，但耗时也最长。由图 3、图 4 可知，经过参数优化后的 LSSVR 模型有很好的预测效果与稳定性，但算法的耗时呈几何倍数增加。

上述研究结果表明，使用群智能算法优化模型参数的方法能提高回归模型对混合物质量分数预测的精度，提升了基于 THz 光谱数据对三聚氰胺定量检测的精度，推进了 THz 光谱技术在食品添加剂检测领域的应用；但模型的预测效果与稳定性也是相对的，且计算量以及耗时问题也应作为算法选取时综合考虑的重要因素。

4　结　　论

以聚乙烯为基质，混入奶粉和三聚氰胺粉末后制备样片，包括三聚氰胺单质样片、奶粉单质样片和在奶粉中混入质量分数呈梯度变化的三聚氰胺粉末混合物样片。使用透射式 THz-TDS 系统获取样片的吸收系数谱。由实验结果可知，奶粉、三聚氰胺及两者混合物均在 0.5～2.8 THz 的波段内具有特征吸收峰，验证了 THz 光谱对生物大分子的指纹谱特性和朗伯比尔定律，满足了利用 THz 光谱技术对奶粉中三聚氰胺进行定量检测的条件。采用 4 种不同的回归模型分别预测了奶粉中的三聚氰胺含量，使用 R_P 与 f_{RMSEP} 作为模型评价系数，比较 4 种模型的评价系数，得出 LSSVR 模型预测效果最好（R_P 为 0.99838，f_{RMSEP} 为 0.41%）。在此基础上，使用了性能明显优于传统方法的群智能算法分别对 LSSVR 模型的超参数选取进行优化，经 4 种算法优化后的模型预测精度均有所提高，其中 GWO-LSSVR 模型的评价系数最好（R_P 为 0.99925，f_{RMSEP} 为 0.28%）。利用优化后的 GWO-LSSVR 模型可以实现基于 THz-TDS 技术对奶粉中非法添加剂三聚氰胺的定量检测，为食品添加剂的定量分析提供了新的方法和思路。

参　考　文　献

[1]　林木生 . 食品中非法添加物检测与技术进展研究[J]. 中国食品

工业, 2022(13): 106-109.
Lin M S. Research on detection and technical progress of illegal additives in food[J]. China Food Industry, 2022(13): 106-109.

[2] Ferguson B, Zhang X C. Materials for terahertz science and technology[J]. Nature Materials, 2002, 1(1): 26-33.

[3] Walther M, Fischer B M, Ortner A, et al. Chemical sensing and imaging with pulsed terahertz radiation[J]. Analytical and Bioanalytical Chemistry, 2010, 397(3): 1009-1017.

[4] Wu Q, Zhang X C. Ultrafast electro-optic field sensors[J]. Applied Physics Letters, 1996, 68(12): 1604-1606.

[5] Rudd J V, Johnson J L, Mittleman D M. Quadrupole radiation from terahertz dipole antennas[J]. Optics Letters, 2000, 25(20): 1556-1558.

[6] Parrott E P J, Sun Y W, Pickwell-MacPherson E. Terahertz spectroscopy: its future role in medical diagnoses[J]. Journal of Molecular Structure, 2011, 1006(1/2/3): 66-76.

[7] 曹恩达, 于勇, 宋长波, 等. 基于太赫兹时域谱分析的常见包裹物屏蔽下食品药品检测方法[J]. 激光与光电子学进展, 2021, 58(1): 0112002.
Cao E D, Yu Y, Song C B, et al. Method of food and drug detection under shielding of common wrappings based on terahertz time domain spectroscopy[J]. Laser & Optoelectronics Progress, 2021, 58(1): 0112002.

[8] Liu W, Zhang Y Y, Han D H. Feasibility study of determination of high-fructose syrup content of Acacia honey by terahertz technique[J]. Proceedings of SPIE, 2016, 10030: 100300J.

[9] Sun X D, Liu J B, Zhu K, et al. Generalized regression neural network association with terahertz spectroscopy for quantitative analysis of benzoic acid additive in wheat flour[J]. Royal Society Open Science, 2019, 6(7): 190485.

[10] 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.

[11] Dorney T D, Baraniuk R G, Mittleman D M. Material parameter estimation with terahertz time-domain spectroscopy [J]. Journal of the Optical Society of America A, 2001, 18(7): 1562-1571.

[12] 张天尧. 基于时域光谱测定的固体太赫兹吸收及介电性质表征[D]. 北京: 北京科技大学, 2019.
Zhang T Y. Terahertz absorption and dielectric properties of solid-state materials characterized by time-domain spectroscopy [D]. Beijing: University of Science and Technology Beijing, 2019.

[13] 宁鸿章, 谭鑫, 李宇航, 等. 空-谱维联合 Savitzky-Golay 高光谱滤波算法及其应用[J]. 光谱学与光谱分析, 2020, 40(12): 3699-3704.
Ning H Z, Tan X, Li Y H, et al. Joint space-spectrum SG filtering algorithms for hyperspectral images and its application [J]. Spectroscopy and Spectral Analysis, 2020, 40(12): 3699-3704.

[14] 燕芳, 张俊林, 毛莉程, 等. 基于太赫兹辐射的糖类异构体信息提取方法研究[J]. 光谱学与光谱分析, 2022, 42(1): 26-30.
Yan F, Zhang J L, Mao L C, et al. Research on information extraction method of carbohydrate isomers based on terahertz radiation[J]. Spectroscopy and Spectral Analysis, 2022, 42(1): 26-30.

[15] Yu H Y, Niu X Y, Lin H J, et al. A feasibility study on on-line determination of rice wine composition by Vis-NIR spectroscopy and least-squares support vector machines[J]. Food Chemistry, 2009, 113(1): 291-296.

[16] 张燚, 丁建丽, 张子鹏, 等. 光谱配置对最优波段组合算法预测土壤有机质和电导率的影响[J]. 激光与光电子学进展, 2021, 58(21): 2128001.
Zhang Y, Ding J L, Zhang Z P, et al. Effect of spectral configuration on soil organic matter and electrical conductivity predicted by optimal band combination algorithm[J]. Laser & Optoelectronics Progress, 2021, 58(21): 2128001.

[17] 李扬. 最小二乘法、ε-支持向量回归机与最小二乘支持向量回归机的对比研究[D]. 上海: 华东师范大学, 2018.
Li Y. The comparation and study on least square method, ε-support vector regression and least square support vector regression[D]. Shanghai: East China Normal University, 2018.

[18] 阎辉, 张学工, 李衍达. 支持向量机与最小二乘法的关系研究[J]. 清华大学学报(自然科学版), 2001, 41(9): 77-80.
Yan H, Zhang X G, Li Y D. Relation between a support vector machine and the least square method[J]. Journal of Tsinghua University (Science and Technology), 2001, 41(9): 77-80.

[19] 段会川. 高斯核函数支持向量分类机超级参数有效范围研究[D]. 济南: 山东师范大学, 2012.
Duan H C. Research on the effective range of super parameters of Gaussian kernel function support vector classifier[D]. Jinan: Shandong Normal University, 2012.

[20] Pan S B, Zhang H, Li Z, et al. Quantitative determination of sucrose adulterated in red ginseng by terahertz time-domain spectroscopy (THz-TDS) with monte carlo uninformative variable elimination (MCUVE) and support vector regression (SVR)[J]. Journal of Spectroscopy, 2022, 2022(4): 1-10.

[21] 卫佳敏. 群智能优化算法及其在分数阶系统参数辨识中的应用研究[D]. 北京: 北京交通大学, 2020.
Wei J M. Research on swarm intelligence optimization algorithms and their applications to parameter identification of fractional-order systems[D]. Beijing: Beijing Jiaotong University, 2020.

[22] 续婷. 基于群智能算法与机器学习的预测与分类研究[D]. 太原: 中北大学, 2021.
Xu T. Research on prediction and classification based on swarm intelligence algorithm and machine learning[D]. Taiyuan: North University of China, 2021.

[23] 陈国良, 王熙法, 庄镇泉, 等. 遗传算法及其应用[M]. 北京: 人民邮电出版社, 1996.
Chen G L, Wang X F, Zhuang Z Q, et al. Genetic algorithm and its application[M]. Beijing: Posts & Telecom Press, 1996.

[24] 白鹤轩, 杨峰, 李丹阳, 等. 基于表面增强拉曼光谱的多组分物质分类识别[J]. 光学学报, 2021, 41(20): 2024001.
Bai H X, Yang F, Li D Y, et al. Multi-component substance classification and recognition based on surface-enhanced Raman spectroscopy[J]. Acta Optica Sinica, 2021, 41(20): 2024001.

[25] 春花. 基于群智能算法的K-均值聚类研究[D]. 大连: 大连理工大学, 2019.
Chun H. Research on K- means clustering based on swarm intelligence algorithm[D]. Dalian: Dalian University of Technology, 2019.

[26] 智慧. 群智能优化算法的研究及应用[D]. 西安: 西安电子科技大学, 2020.
Zhi H. Research and application of swarm intelligence optimization algorithm[D]. Xi′an: Xidian University, 2020.

Quantitative Analysis of Melamine Based on Terahertz Spectroscopy

Guo Yiheng**, Yan Fang*, Zhao Miaoyu, Zhuo Xuan

School of Information Engineering, Inner Mongolia University of Science &Technology, Baotou 014010, *Inner Mongolia, China*

Abstract

Objective In recent years, the food safety hazards caused by the illegal use and abuse of food additives during their use have once again attracted much attention from society. The research and development of non-destructive, fast, accurate, and efficient qualitative and quantitative detection technologies and methods for food additives have become a research hotspot for scholars. At present, traditional detection methods for food additives include ion chromatography, liquid chromatography, liquid chromatography-mass spectrometry, gas chromatography, and molecular spectrometry. The traditional methods have obvious shortcomings, such as high equipment cost, long detection cycle, high cost, uneven detection accuracy, high requirements for the operation of technicians and the purity of organic solvents, and complex detection operations. Compared with traditional detection methods, the application of new technologies such as immune detection, biosensor, and spectral analysis has supplemented and improved old technologies, thereby promoting the development of food additive detection technology. The detection method based on terahertz spectroscopy technology is non-destructive, fast, and efficient, and has been widely applied in fields of food, medicine, and environmental detection in recent years.

Methods Firstly, we construct melamine samples with concentration gradients and obtain experimental training and testing sets by a transmission terahertz time-domain spectroscopy system. For the high-dimensional spectral data obtained from the experiment, a Savitzky-Golay convolutional filter is adopted for preprocessing to reduce quantitative prediction errors. Secondly, based on the dimensional characteristics of spectral data, we build four regression prediction models including PCR, SVR, PLSR, and LSSVR for data analysis. The obtained experimental spectral data are compared in terms of the predictive performance after linear regression dimensionality reduction (PCR, PLSR) and nonlinear regression dimensionality enhancement (SVR, LSSVR), which are processed at opposite angles. The correlation coefficient R_P of the prediction set and the root mean square error of the prediction set (RMSEP) are employed as indicators for model performance evaluation. Finally, according to the optimal evaluation index, we find that the prediction effect of the LSSVR model is optimal. We leverage particle swarm optimization (PSO), genetic algorithm (GA), Cuckoo search algorithm (CS), and grey wolf optimization (GWO) to calculate the regularization parameter C in LSSVR and the kernel parameter after the determined kernel function is Gaussian kernel function γ for parameter optimization.

Results and Discussions The filtering preprocessing operation for spectral data yields sound effect (Fig. 1). We employ four different regression models (PCR, PLSR, SVR, and LSSVR) to predict the melamine content in milk powder, and adopt the correlation coefficient of the prediction set and RMSEP as the model evaluation coefficients. After comparing the evaluation coefficients of the four models, it is determined that the minimum correlation coefficient of the linear model PCR's prediction set is 0.99715, the maximum RMSEP is 0.50%, and the nonlinear model LSSVR has the best prediction performance. Its prediction phase set relationship number R_P is 0.99838 and RMSEP is 0.41%, which indicates that the nonlinear model has better detection performance for terahertz spectral data (Fig. 3). On this basis, we utilize swarm intelligence algorithms (PSO, GA, CS, and GWO) whose performances are significantly better than those of traditional methods to optimize hyperparameter selection of LSSVR model respectively. The prediction accuracy of the model after optimization by the four algorithms has been improved. Among them, the evaluation coefficient of the GWO-LSSVR model is the best, with R_P of 0.99925 and RMSEP of 0.28% (Fig. 4).

Conclusions Results show that nonlinear models can be better applied to the detection of food additives by terahertz technology. The optimized GWO-LSSVR model can improve the accuracy of regression models in predicting mixture concentration and the quantitative detection accuracy of melamine based on terahertz spectral data. Additionally, it can promote the application of terahertz spectral technology in food additive detection and provide new methods and ideas for the quantitative analysis of food additives. However, the predictive performance and stability of the model are also relative, and the issues of computational complexity and time consumption should also be considered important factors in algorithm selection.

Key words spectroscopy; terahertz time-domain spectroscopy; quantitative analysis; parameter optimization; food additive