

太赫兹光谱技术在生物医学检测中的定性与定量分析算法

彭滢, 施辰君, 朱亦鸣**, 庄松林*

上海理工大学太赫兹技术创新研究院, 太赫兹光谱与影像技术协同创新中心, 上海市现代光学系统重点实验室, 上海 200093

摘要 提出了两种有效混合物成分的定性与定量分析算法。针对混合物中每种成分太赫兹光谱已知的情况, 提出了已知混合物成分光谱分析法, 该方法具有步骤简单、样本需求量少的优点。而针对成分过多或部分成分光谱无法有效获得的情况, 提出了未知混合物成分光谱分析法, 该方法适用于多成分混合物, 具有精度较高的优点, 样本需求量多于已知混合物成分光谱分析法, 但计算量相比传统算法大幅缩减。相关工作作为生物医学领域中关键物质的识别和测量提供了新的思路和方法。

关键词 太赫兹技术; 太赫兹光谱分析; 混合物算法; 定性与定量分析

中图分类号 O433.4

文献标识码 A

doi: 10.3788/CJL201946.0614002

Qualitative and Quantitative Analysis Algorithms Based on Terahertz Spectroscopy for Biomedical Detection

Peng Yan, Shi Chenjun, Zhu Yiming**, Zhuang Songlin*

Terahertz Technology Innovation Research Institute, Terahertz Spectrum and Imaging Technology Cooperative Innovation Center, Shanghai Key Lab of Modern Optical System, University of Shanghai for Science and Technology, Shanghai 200093, China

Abstract This study proposes two effective algorithms for the qualitative and quantitative analysis of components in a mixture. For cases wherein the spectrum of each component in the mixture can be obtained in advance, we develop a simple process, namely “known component analysis (KCA),” which requires a small number of samples. In contrast, for cases wherein a mixture contains many components or the spectra of some components cannot be obtained, we develop “unknown components analysis (UCA),” whose calculation time is considerably less than those of the traditional algorithms. Compared to KCA, UCA requires more samples to be analyzed to ensure the high accuracy rate. These proposed algorithms provide new ways to detect and measure key biomedical substances.

Key words terahertz technology; terahertz spectroscopic analysis; mixture algorithms; qualitative and quantitative analysis

OCIS codes 300.6495; 300.6170; 070.4790

1 引 言

太赫兹波指波长为 0.03~3 mm 的电磁波。相较于其他波段, 太赫兹波用于生物医学检测时具有众多优点: 1) 因其能量为毫电子伏特, 电离不会破坏物质本身的属性^[1]; 2) 太赫兹脉冲的典型脉宽为皮秒量级, 能够展示物质在皮秒量级的现象^[2]; 3) 大量

生物分子的转动和振动能级位于太赫兹频段^[3], 因此太赫兹光谱能反映分子种类和结构的细微变化, 使得它们具有类似指纹一样的唯一特点, 所以也被称为分子指纹谱; 4) 相对其他化学检测方法(如试剂盒法、质谱法等), 太赫兹光谱法是快速无损的样品检测方法。因此, 很多课题组已大力开展基于太赫兹波的生物医学物质识别和检测的研究工作^[4-5]。

收稿日期: 2019-01-24; 修回日期: 2019-03-07; 录用日期: 2019-03-15

基金项目: 国家重点研发计划“重大科学仪器设备开发”重点专项(2017YFF0106300)、国家自然科学基金(61771314, 61722111)、111 项目(D18014)、上海市启明星人才计划(17QA1402500)、上海市教育发展基金会“曙光计划”(17SG45)。

* E-mail: slzhuang@yahoo.com; ** E-mail: ymzhu@usst.edu.cn

例如:Fischer等^[6]在2002年测得了构成脱氧核糖核酸(DNA)的基本单元(鸟嘌呤、腺嘌呤、胞嘧啶和胸腺嘧啶)以及其对应核苷的太赫兹光谱,并使用密度泛函理论进行了验证,证实了太赫兹波对分子氢键的响应性;Kawase等^[7]在2003年实现了基于太赫兹光谱对装在信封中的两种毒品脱氧麻黄碱和摇头丸的无损检测,为应对隐藏毒品提供了一种新的安检手段;岳伟伟等^[8]在2005年测得了构成人体蛋白质的基本结构芳香族氨基酸的太赫兹光谱,这项研究是进一步对人体蛋白质研究的基础;Chen等^[9]在2017年测得了脑胶质瘤中特异性物质(D-2HG和L-2HG)的纯品太赫兹光谱,并使用密度泛函理论对其分子振动模式进行了分析,为脑胶质瘤的诊断提供了一种新方法。这些研究都成功实现了对单一纯品的有效识别和检测。然而,在生物医学领域,多数物质为混合物,所含物质成分众多,而具有疾病标识性物质的含量通常很低。以脑组织为例,人体脑组织中含有上百种物质,但标志脑胶质瘤发生和病变程度的只有2~3种物质,且含量均在5%~10%内^[10]。用太赫兹波对疾病标识性物质进行识别时,大量物质的太赫兹指纹谱相互叠加,导致基于谱线对这些标识性物质定性及定量分析的难度极大。因此,基于太赫兹光谱法对混合物进行定性和定量分析成为一个重要的研究课题^[11-16]。

现阶段,多数混合物分析算法都是基于物质成分的纯品谱已知情况,例如:吉特等^[17]在2006年测试了混合物药品青霉素中的有效成分D-青霉素,采用的方法为最小二乘法,但只能在成分含量(质量分数)大于20%的情况下进行最低3%误差的D-青霉素浓度计算;Ueno等^[18]在2006年针对不同氨基酸的混合物,同样使用最小二乘法进行了三组分氨基酸混合物的定量分析,随着浓度从7%降低至3%,其计算结果误差从11%上升至20%;Zhang等^[19]在2007年针对广泛用于农药、有机化合物和染料的苯甲酸及其单甲基取代的衍生物,采用纯品光谱叠加和线性回归方法进行了二元混合物和三元混合物的定量分析。对于二元混合物,两种成分浓度分别为80%和20%,其测试误差约为5.5%;对于三元混合物,三种成分浓度分别为50%,30%和20%,其测试误差约为6.0%,高于二元混合物的分析误差。

此外,无法预先获得所有物质成分纯品谱的情况,是实际中更多出现和更需解决的难题,目前主要依赖偏最小二乘法、主成分分析和有效介质理论等

算法。例如:Hua等^[20]在2010年测试了农药中的主要成分吡虫啉,采用偏最小二乘法分析了吡虫啉和聚乙烯,以及吡虫啉和糯米粉的二元混合物,混合物中吡虫啉最高浓度分别约为50%和70%,其计算结果的误差分别是4.92%和4.58%;El Haddad等^[21]在2014年测试了果糖、 α 乳糖一水合物和柠檬酸的三元混合物,混合物中组分浓度的变化范围为0至100%,采用主成分分析和偏最小二乘法相结合的方法,可得其计算误差小于1%;Truong等^[22]在2018年,针对乳腺组织中的脂质、明胶和去离子水,制备二元混合物和三元混合物,三种成分的浓度分别是36%~82%,6%~13%和0%~54%,基于有效介质理论进行成分分析,其二元混合物的计算误差约为2%,而组分增加到三种后,误差上升到约5%;Palermo等^[23]在2007年对茶碱、乳糖、微晶纤维素与淀粉的混合物进行了成分分析,每个组分的浓度为20%~60%,使用偏最小二乘法先对前三种成分进行分析,其误差小于5.4%,但是对于4种物质的混合物,其误差上升到约7%。

综上所述,上述能保证准确率的研究最多只分析了4组分混合物,而且随着待测组分浓度的降低或者组分数量的增加,其识别准确率急剧降低。

针对上述问题,本文结合去噪算法和分析算法,提出了已知混合物成分光谱分析法和未知混合物成分光谱分析法两种算法,对混合物成分进行高精度定性和定量分析,这两种算法均适用于低浓度条件的样本。通过实例分析两种方法在不同数量样本下的误差变化,以证明利用太赫兹光谱定性、定量测试脑胶质瘤中特异性物质的可行性,为生物医学领域中关键物质的识别和测量提供了新的思路和方法。

2 算法模型

2.1 已知混合物成分光谱分析法

已知混合物成分光谱分析法适用于混合物成分较少的情况,要求每种物质的指纹谱事先已知,具有步骤简单、样本需求量少的优点。其具体步骤如下:

- 1) 测得混合物中每种纯品和混合物本身的光谱并截取有效区间的数;
- 2) 对测得的光谱做 Savitzky-Golay 平滑去噪^[24];
- 3) 基于去噪后的纯品光谱和混合物光谱信息,使用最小二乘法获得混合物光谱中每种纯品光谱的叠加比重,该比重即为混合物中每种物质的浓度信息。使用获得的比重,将纯品谱叠加计算出混合物

的光谱,并与实际测量的混合物光谱进行对比验证。

需要注意的是,已知混合物成分光谱分析法不适用于混合物成分较多或者部分物质光谱无法有效获得的情况。

2.2 未知混合物成分光谱分析法

未知混合物成分光谱分析法适用于混合物中成分较多的情况,不需要前期预测每种纯品的光谱,且精度较高。其具体步骤如下:1)测量已知成分信息混合物的光谱和待测混合物的光谱;2)截取光谱的有效区间;3)对测得的光谱做小波变换^[25],以此达到去噪和去基线的目的;4)将已知成分信息混合物的光谱作为特征向量,使用浓度信息作为标签向量,通过支持向量回归(SVR)建立模型;5)将待测混合物的光谱导入模型,预测混合物成分信息。

在混合物中成分增多的情况下,由于多类物质混杂引入了更多噪声,常规算法中的 Savitzky-Golay 平滑去噪法无法满足复杂去噪的要求,同时因成分增多引起的基线影响不可忽略;另一种去噪声算法即小波变换虽然步骤复杂,但是具有良好的时频定位特性和去相关性,可以去除不同尺度的基线和噪声,同时保留太赫兹频谱不同频率区域的有效信息(这些都是 Savitzky-Golay 平滑无法实现的),因此在多样本处理时选择小波变换进行去噪和去基线。需要注意的是,这种算法需要准备一系列关键成分比例关系已知的混合物样本光谱数据库,并建立模型,模型的准确率与这些样本的数量成正相关。

通过对比上述两种算法模型可知,对于混合物中每种物质的指纹谱事先已知的情况,未知混合物成分光谱分析法也适用,且识别准确率更高,但因样本需求量偏大,计算时间较长,可根据实际准确率要求在两种算法间进行选择。

3 样本配制与实验装置

本实验中所有样本纯品均从 Sigma-Aldrich 处购买,都为粉末状样品,使用时未再次提纯(纯度大于 99%)。由于纯品对太赫兹波的吸收很强,需要选择对太赫兹波吸收较弱的生物样品载体,实验选择高密度聚乙烯(Polyethylene,简称 PE)作为生物样品载体。将纯品与 PE 以不同比例均匀混合,在玛瑙研钵中研磨成细小粉末,在 3000 kg 压力下压制厚约 1 mm、直径为 13 mm 的圆片,保证前后表面平行、光滑,质量损失控制在 1% 以下。

本实验中采用德国 BRUKER VERTEX 80V 真空傅里叶变换红外光谱仪(FTIR)对所有样本进

行扫描。仪器在使用前先预热至稳定状态,使用样品架装载样品片,以纯 PE 样片作为参考样片,其扫描范围为 0.89~20.32 THz,分辨率为 0.001 THz,扫描 32 次并保存图谱。测试中使用全真空模式,以最大程度地消除外界环境影响。

将 FITR 采集到的频域数据转换为吸收谱线^[9],吸收谱线的表达式为

$$\alpha(\omega) = \frac{1}{d} \lg \frac{T_r(\omega)}{T_s(\omega)}, \quad (1)$$

式中: d 为样本药片的厚度,由游标卡尺测出; $T_r(\omega)$ 为参考样片(纯 PE,质量为对应样本中 PE 的质量)的频域信号; $T_s(\omega)$ 为样本片的频域信号; ω 为频率。

4 算法实例分析

4.1 已知混合物成分光谱分析法实例分析

以人脑组织的 4 种主要成分为例进行分析,4 种主要成分分别是为 L-谷氨酸(L-Glutamic acid,简称 L-Glu)、D-肌醇(D-myoinositol,简称 D-MI)、肌酸一水合物(Creatine Monohydrate,简称 CMH)和 γ -氨基丁酸(γ -Aminobutyric acid,简称 GABA)。具体步骤如下。

1) 制备 4 枚单独纯品样片(纯品含量为 5 mg),以及 2 枚混合物样片(分别是 L-Glu、D-MI、CMH 各 5 mg 混合以及全部 4 种物质各 5 mg 混合的样片)。药片总质量为 100 mg,剩余质量使用 PE 填充。

2) 测量这 6 枚样片的光谱,截取包含特征峰的

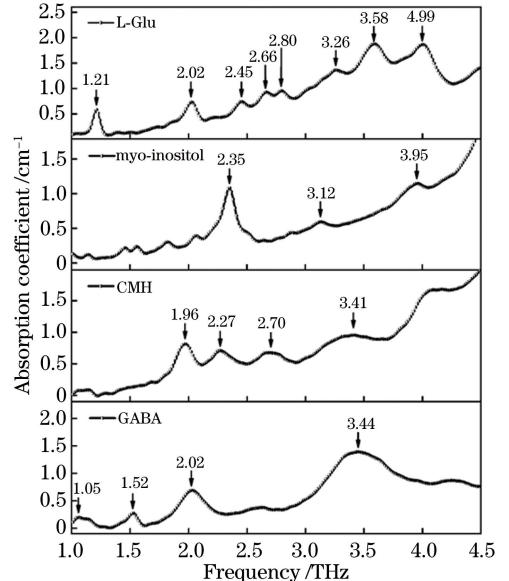


图 1 L-Glu、D-MI、CMH 和 GABA 4 种纯品的光谱

Fig. 1 Spectra of four pure samples of L-Glu, D-MI, CMH, and GABA

光谱范围(1.0~4.5 THz),通过 MATLAB 并使用 Savitzky-Golay 平滑去噪,去除光谱的噪声影响。处理后的光谱数据如图 1 所示。从图 1 可以看出,每种物质的特征峰位置不同,而且 Savitzky-Golay 平滑去噪消除了光谱的噪声抖动。

3) 基于去噪后的 4 枚纯品样片光谱和 2 枚混合物样片光谱信息,使用最小二乘法获得 2 枚混合物样片光谱中每种纯品光谱的叠加比重,该比重即为混合物中每种物质的浓度信息。混合物药片每个频率处的吸收系数可以用矩阵 A_m 表示:

$$A_m = k \times (A_L, A_D, A_C, A_G)^T, \quad (2)$$

式中: A_L 、 A_D 、 A_C 、 A_G 分别为 L-Glu、D-MI、CMH 和 GABA 纯品的吸收系数矩阵。 $k = (k_L, k_D, k_C, k_G)$ 为使用最小二乘法获得的叠加比重,即混合物中 4 种物质的浓度。根据纯品光谱得到的比重叠加计算出混合物光谱,并与实际测量的光谱进行对比,结果如图 2 所示,从图中可以看到计算的结果与实际测量的结果具有较高的吻合度。

4) 将求得的浓度信息与真实的浓度信息进行比较,两种混合物样片的结果如表 1 和表 2 所示。

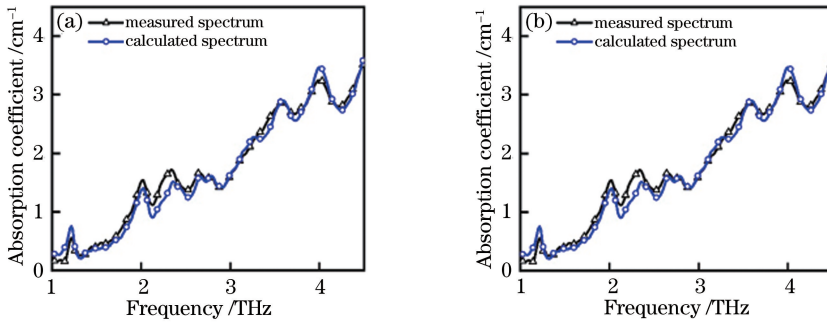


图 2 测量光谱和计算所得光谱。(a) L-Glu、D-MI、CMH 的混合物样片;(b) 4 种物质的混合物样片

Fig. 2 Measured spectra and calculated spectra. (a) Mixture sample of L-Glu, D-MI, and CMH; (b) mixture sample of all four components

表 1 L-Glu、D-MI 和 CMH 混合的混合物样片浓度结果

Table 1 Concentration results of mixture sample of L-Glu, D-MI and CMH

Component	Actual concentration	Calculated concentration	Root mean squared error E_{RMS1}	Average root mean squared error $\bar{E}_{RMS1} / \%$
	$n_{a1} / \%$	$n_{c1} / \%$		
L-Glu	4.65	4.65	0	4.65
D-MI	4.65	4.55	0.0215	
CMH	4.65	4.10	0.1180	

表 2 4 种物质混合的混合物样片浓度结果

Table 2 Concentration results of mixture sample of all four components

Component	Actual concentration	Calculated concentration	Root mean squared error E_{RMS2}	Average root mean squared error $\bar{E}_{RMS2} / \%$
	$n_{a2} / \%$	$n_{c2} / \%$		
L-Glu	4.90	4.50	0.0816	5.44
D-MI	4.90	4.90	0	
CMH	4.90	4.60	0.0612	
GABA	4.90	4.80	0.0204	

从上述结果可以得出,在纯品谱已知且混合物的成分数量较少的情况下,通过已知混合物成分光谱分析法可以获取混合物的成分信息。当混合物成分数量从 3 种上升到 4 种后,平均方均根误差(RMSE)略有上升(从 4.65% 上升到 5.44%),变化不足 0.8%,远优于传统算法。

4.2 未知混合物成分光谱分析法实例分析

以人体脑胶质瘤为例,脑部发生胶质瘤性质病变时,两种特异性物质(NAA 和 NE)会根据病情发生量

的改变。其中:NAA 是神经元的细胞结构,当患有脑胶质瘤时,NAA 浓度降低,反映神经元死亡;NE 是人脑中各种抑制性神经元和中间神经元的主要传递者,当患有脑胶质瘤时,其浓度增加,意味着胶质母细胞瘤细胞的侵袭^[26-27]。具体步骤如下:

1) 首先制备 10 枚混合物样片,将 NAA、NE 与人体脑组织中前 5 种主要成分以真实比例接近的不同比例混合,除了 NAA 和 NE,其他物质成分比例未知。10 例样本的成分信息如表 3 所示。

表 3 含有 NAA 和 NE 的混合物样本参数
Table 3 Parameters of mixture samples including NAA and NE

Sample number	NAA mass m_1/mg	NE mass m_2/mg	Mass of other five components m_3/mg
1	3.02	1.13	—
2	9.85	7.02	—
3	11.99	15.08	—
4	9.20	3.82	—
5	1.20	9.98	—
6	4.13	8.91	—
7	15.20	12.03	—
8	12.87	3.19	—
9	4.86	4.80	—
10	7.02	13.03	—

2) 测量这 10 枚样片的光谱, 截取包含特征峰的光谱范围(1.0~3.0 THz), 通过 MATLAB 使用小波变换, 去除光谱的噪声和基线影响。小波变换前后的光谱数据如图 3 所示。图 3 显示大量噪声的

存在, 使用小波变换进行处理, 消除了光谱的基线, 同时平滑了曲线并消除了噪声抖动。

3) 基于 LIBSVM^[28] 编写算法建立 SVR 模型。在 SVR 模型中需要选择核函数, 核函数包含 3 个参数: 正则化参数 c 、内核参数 g 和误差范围 e 。参数 c 确定训练误差和模型复杂度之间的最佳匹配, 参数 g 确定新特征空间中支持向量的分布, 参数 e 影响用于构建回归模型的支持向量的数量^[29]。本研究中, 由于无法确定样本中不同成分之间的比例关系为线性还是非线性, 且考虑到径向基核函数能够将非线性的数据升维, 使之线性可分^[30], 有利于后期光谱数据的处理与识别, 所以在模型中选择径向基核函数^[29]。对于上述 3 个参数的优化, 采用弃一法交叉验证^[31]来寻找最优参数: 使一个参数在允许范围内随固定步长变化, 而另外两个参数固定, 通过迭代循环, 得到方均根误差最小的参数, 并将其作为最优参数。迭代过程如图 4 所示, 最终得到的 c 、 g 、 e 参数值分别为 0.025, 0.01, 0.33。

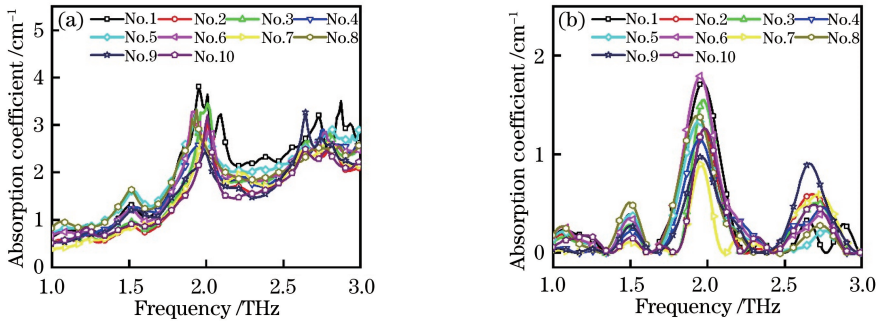


图 3 10 枚混合物样本的光谱。(a)小波变换前;(b)小波变换后

Fig. 3 Spectra of 10 mixture samples. (a) Before wavelet transform; (b) after wavelet transform

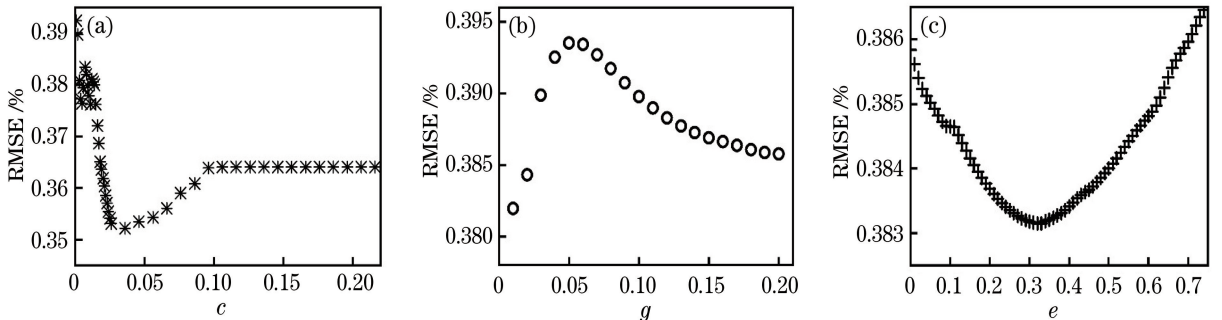


图 4 基于弃一法交叉验证的 SVR 模型三个参数的方均根误差。(a) $g=0.01, e=0.01$ 时的参数 c ;
(b) $c=0.25, e=0.01$ 时的参数 g ;(c) $c=0.25, g=0.01$ 时的参数 e

Fig. 4 RMSE of three SVR parameters based on leave-one-out cross validation. (a) Parameter c when $g=0.01$ and $e=0.01$; (b) parameter g when $c=0.25$ and $e=0.01$; (c) parameter e when $c=0.25$ and $g=0.01$

选择模型的核函数和设置参数之后,将样本光谱导入模型进行训练和预测。模型训练和预测的具体流程如下:首先将已知 NAA 和 NE 浓度的混合物光谱作为特征向量,浓度信息作为标签向量导入模型中进行训练;然后将待测的未知浓度混合物光谱导入模型,模型输出的标签向量就是预测的混合物成分信息。为了节省样本量,使用迭代的方法训练模型和预测,即每次从 10 例样本中选出 8 例用作训练模型的已知 NAA 和 NE 浓度的混合物光谱,

其余 2 例用作待测的未知浓度混合物光谱,迭代 5 次即可得到 10 例样本的预测数据。预测的浓度信息与真实的浓度信息如图 5 所示。为了验证模型的准确率,对两种物质浓度的预测结果和真实结果之间的均方根误差和相关系数进行了计算,结果如表 4 所示。可以看出,未知混合物成分光谱分析法能够非常准确地预测混合物中的成分信息,两种物质的平均方均根误差只有 0.0040,平均相关系数可达 0.99135。

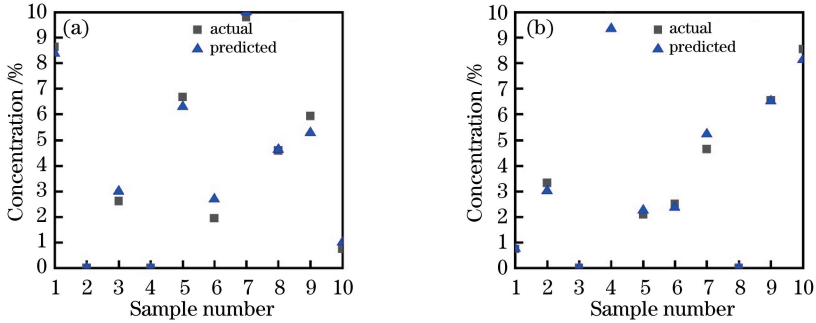


图 5 10 枚混合物样本的实际浓度和预测浓度。(a) NAA;(b) NE

Fig. 5 Actual and predicted concentrations of 10 mixture samples. (a) NAA; (b) NE

表 4 NAA 和 NE 浓度预测结果和真实结果之间的均方根误差和相关系数

Table 4 RMSE and correlation coefficient between predicted and actual concentrations of NAA and NE in mixture

Component	RMSE E_{RMS}	Average RMSE \bar{E}_{RMS}	Correlation coefficient R	Average correlation coefficient \bar{R}
NAA	0.0040	0.0040	0.9913	0.99135
NE	0.0040		0.9914	

为了比较样本数量对模型准确率的影响,从 10 种样本中随机选取 6 种和 8 种进行分析。如表 5 所示,随着样本数量从 6 种上升到 10 种,其模型的方均根误差下降,相关系数上升,证明该算法模型的准确率与样本的数量成正相关。

表 5 不同样本数量下算法模型的准确率

Table 5 Accuracy of algorithm models under different sample numbers

Number of samples	RMSE E_{RMSn}	Correlation coefficient R_n
6	0.0125	0.9026
8	0.0055	0.9853
10	0.0040	0.9914

本研究还尝试使用常规算法中的偏最小二乘法和 BP(back propagation)神经网络法替代未知混合物成分光谱分析法中 SVR 算法来建立模型,结果如表 6 所示。结果证明,相比这两种算法,SVR 算法具有更高的准确率,将其归因于:偏最小二乘法更擅长分析线性数据,而本研究中数据存在不确定的非

线性关系,直接限制了其有效性;对于 BP 神经网络法,实例中的混合物样片只有 10 例,不满足该方法所需要的样本数量。SVR 可将线性不可分的数据升为使之线性可分的核函数,同时对样本的需求量小于 BP 神经网络法,因此对于样本量不够庞大的混合物进行分析具有更高的准确率。

表 6 不同算法的预测准确率

Table 6 Prediction accuracy of different algorithms

Algorithm	RMSE E_{RMSa}	Correlation coefficient R_a
Partial least squares	0.0231	0.8052
BP nerve network	0.0175	0.8353
Support vector regression	0.0040	0.9914

5 结 论

针对混合物成分分析方法提出了两种算法,分别是已知混合物成分光谱分析法和未知混合物成分光谱分析法。已知混合物成分光谱分析法适用于混合物成分较少的情况,具有步骤简单,样本需求量少

的优点;未知混合物成分光谱分析法适用于混合物成分较多、部分物质光谱无法有效获得的情况,具有精度较高的优点。通过实例证明了这两种算法的准确性,为生物医学领域中关键物质的识别和测量提供了新的思路和方法。

参 考 文 献

- [1] Li T, Ma H Y, Peng Y, *et al.* Gaussian numerical analysis and terahertz spectroscopic measurement of homocysteine[J]. *Biomedical Optics Express*, 2018, 9(11): 5467-5476.
- [2] Zhang X N, Chen J, Zhou Z K. THz time-domain spectroscopy technology[J]. *Laser & Optoelectronics Progress*, 2005, 42(7): 35-38.
张兴宁, 陈稷, 周泽魁. 太赫兹时域光谱技术[J]. *激光与光电子学进展*, 2005, 42(7): 35-38.
- [3] Zhou L, Chen L G, Ren G H, *et al.* Monitoring cis-to-trans isomerization of azobenzene using terahertz time-domain spectroscopy [J]. *Physical Chemistry Chemical Physics*, 2018, 20(42): 27205-27213.
- [4] Chen T, Cai Z H. Terahertz time-domain spectroscopy of L-, D- and DL-arabinose[J]. *Laser & Optoelectronics Progress*, 2018, 55(6): 063001.
陈涛, 蔡治华. L-, D-和 DL-阿拉伯糖的太赫兹时域光谱研究[J]. *激光与光电子学进展*, 2018, 55(6): 063001.
- [5] Zhao S S, Chen S J, Shen J L. Identification of terahertz absorption spectra of illicit drugs using support vector machines [J]. *Chinese Journal of Lasers*, 2009, 36(3): 752-757.
赵树森, 陈思嘉, 沈京玲. 用支持向量机识别毒品的太赫兹吸收光谱[J]. *中国激光*, 2009, 36(3): 752-757.
- [6] Fischer B M, Walther M, Uhd Jepsen P. Far-infrared vibrational modes of DNA components studied by terahertz time-domain spectroscopy[J]. *Physics in Medicine & Biology*, 2002, 47(21): 3807-3814.
- [7] Kawase K, Ogawa Y, Watanabe Y, *et al.* Non-destructive terahertz imaging of illicit drugs using spectral fingerprints[J]. *Optics Express*, 2003, 11(20): 2549-2554.
- [8] Yue W W, Wang W N, Zhao G Z, *et al.* THz spectrum of aromatic amino acid[J]. *Acta Physica Sinica*, 2005, 54(7): 3094-3099.
岳伟伟, 王卫宁, 赵国忠, 等. 芳香族氨基酸的太赫兹光谱研究[J]. *物理学报*, 2005, 54(7): 3094-3099.
- [9] Chen W Q, Peng Y, Jiang X K, *et al.* Isomers identification of 2-hydroxyglutarate acid disodium salt

- (2HG) by terahertz time-domain spectroscopy [J]. *Scientific Reports*, 2017, 7: 12166.
- [10] Peng Y, Yuan X R, Zou X, *et al.* Terahertz identification and quantification of neurotransmitter and neurotrophin mixture [J]. *Biomedical Optics Express*, 2016, 7(11): 4472-4479.
- [11] Shen Y C, Upadhyay P C, Linfield E H, *et al.* Temperature-dependent low-frequency vibrational spectra of purine and adenine[J]. *Applied Physics Letters*, 2003, 82(14): 2350-2352.
- [12] Peng Y, Shi C J, Xu M Q, *et al.* Qualitative and quantitative identification of components in mixture by terahertz spectroscopy[J]. *IEEE Transactions on Terahertz Science and Technology*, 2018, 8(6): 696-701.
- [13] Walther M, Fischer B, Schall M, *et al.* Far-infrared vibrational spectra of all-trans, 9-cis and 13-cis retinal measured by THz time-domain spectroscopy [J]. *Chemical Physics Letters*, 2000, 332(3/4): 389-395.
- [14] Liu G F, Zhang Z Y, Ma S H, *et al.* Quantitative measurement of mixtures by terahertz time-domain spectroscopy [J]. *Journal of Chemical Sciences*, 2009, 121(4): 515-520.
- [15] Chen T, Li Z, Hu F R, *et al.* Quantitative analysis of mixtures using terahertz time-domain spectroscopy and different PLS algorithms[J]. *Advanced Materials Research*, 2013, 804: 23-28.
- [16] Naftaly M, Miles R E. Terahertz time-domain spectroscopy for material characterization [J]. *Proceedings of the IEEE*, 2007, 95(8): 1658-1665.
- [17] Ji T, Zhao H W, Zhang Z Y, *et al.* Terahertz time-domain spectroscopy of D-, L-, and DL-penicillamines [J]. *Acta Physico-Chimica Sinica*, 2006, 22(9): 1159-1162.
吉特, 赵红卫, 张增艳, 等. D-, L-和 DL-青霉胺的太赫兹时域光谱[J]. *物理化学学报*, 2006, 22(9): 1159-1162.
- [18] Ueno Y, Rungsawang R, Tomita I, *et al.* Quantitative measurements of amino acids by terahertz time-domain transmission spectroscopy[J]. *Analytical Chemistry*, 2006, 78(15): 5424-5428.
- [19] Zhang Z Y, Yu X H, Zhao H W, *et al.* Component analysis to isomer mixture with THz-TDS[J]. *Optics Communications*, 2007, 277(2): 273-276.
- [20] Hua Y F, Zhang H J. Qualitative and quantitative detection of pesticides with terahertz time-domain spectroscopy[J]. *IEEE Transactions on Microwave Theory and Techniques*, 2010, 58(7): 2064-2070.
- [21] El Haddad J, de Miollis F, Bou Sleiman J, *et al.* Chemometrics applied to quantitative analysis of ternary mixtures by terahertz spectroscopy [J].

- Analytical Chemistry, 2014, 86(10): 4927-4933.
- [22] Truong B C Q, Fitzgerald A J, Fan S T, *et al.* Concentration analysis of breast tissue phantoms with terahertz spectroscopy [J]. Biomedical Optics Express, 2018, 9(3): 1334-1349.
- [23] Palermo R, Cogdill R P, Short S M, *et al.* Density mapping and chemical component calibration development of four-component compacts via terahertz pulsed imaging [J]. Journal of Pharmaceutical and Biomedical Analysis, 2008, 46(1): 36-44.
- [24] Savitzky A, Golay M J E. Smoothing and differentiation of data by simplified least squares procedures[J]. Analytical Chemistry, 1964, 36(8): 1627-1639.
- [25] Lu J. Parallelizing Mallat algorithm for 2-D wavelet transforms [J]. Information Processing Letters, 1993, 45(5): 255-259.
- [26] Server A, Josefsen R, Kulle B, *et al.* Proton magnetic resonance spectroscopy in the distinction of high-grade cerebral gliomas from single metastatic brain tumors[J]. Acta Radiologica, 2010, 51(3): 316-325.
- [27] Sarkar B K, Chakraborty C, Sharma A R, *et al.* Novel biomarker for prostate cancer diagnosis by MRS[J]. Frontiers in Bioscience, 2014, 19(7): 1186-1201.
- [28] Chang C C, Lin C J. LIBSVM: a library for support vector machines[J]. ACM Transactions on Intelligent Systems and Technology(TIST), 2011, 2(3): 27.
- [29] Alam S, Kang M, Pyun J Y, *et al.* Performance of classification based on PCA, linear SVM, and multi-kernel SVM [C] // 2016 Eighth International Conference on Ubiquitous and Future Networks (ICUFN), July 5-8, 2016, Vienna, Austria. New York: IEEE, 2016: 987-989.
- [30] Zhang L. The relationship between kernel functions based SVM and three-layer feedforward neural networks[J]. Chinese Journal of Computers, 2002, 25(7): 696-700.
张铃. 基于核函数的 SVM 机与三层前向神经网络的关系[J]. 计算机学报, 2002, 25(7): 696-700.
- [31] Feng G H. Parameter optimizing for support vector machines classification [J]. Computer Engineering and Applications, 2011, 47(3): 123-124, 128.
奉国和. SVM 分类核函数及参数选择比较[J]. 计算机工程与应用, 2011, 47(3): 123-124, 128.