

〈太赫兹技术〉

基于层次分析法的太赫兹时域光谱定性识别

赵渺钰，燕芳，李文文，刘洋硕

(内蒙古科技大学 自动化与电气工程学院，内蒙古 包头 014010)

摘要：太赫兹时域光谱技术属于远红外光谱技术的一种，其光谱能够反映物质的内部特性，具有丰富的物理和化学信息，因此利用太赫兹波可对含氮元素的添加剂进行定性识别。层次分析法原用于解决评价类问题，本文将其引入太赫兹光谱定性分析领域，提出一种结合层次分析法的太赫兹时域光谱定性识别方法，并采集数据对其进行验证。文中以6种添加剂作为实验样品，首先对采集到的太赫兹时域光谱数据进行预处理，将其转化为由峰值、峰位、峰数和整体趋势组成的数据集；然后将数据划分为对比集和测试集，构建结合层次分析法的添加剂定性识别模型，并进行参数寻优。结果表明，基于单一因素（即整体趋势、峰值、峰位、峰数）的添加剂定性识别精度分别为80.23%、70.93%、67.44%、40.70%，而基于多因素的层次分析定性识别方法识别精度可提升至92.44%。此外，在数据预处理阶段对吸收谱数据进行二值化模糊表征，并作为整体趋势的数据集，将层次分析定性识别模型与此种数据表征法结合后，识别精度可提至94.19%。研究结果证明了结合层次分析法的定性识别算法的有效性，且该法步骤简单，无需训练，适用于小样本的快速定性检测。

关键词：层次分析；定性识别；太赫兹时域光谱技术；添加剂

中图分类号：O433.4; TS202.3 文献标志码：A 文章编号：1001-8891(2025)05-0656-06

Qualitative Identification of Terahertz Time-domain Spectra Based on Analytic Hierarchy Process

ZHAO Miaoyu, YAN Fang, LI Wenwen, LIU Yangshuo

(School of Automation and Electrical Engineering, Inner Mongolia University of Science and Technology, Baotou 014010, China)

Abstract: Terahertz time-domain spectroscopy (THz-TDS) is a type of far-infrared spectroscopy that reflects the internal characteristics of substances and provides rich physical and chemical information. Therefore, terahertz waves can be used to qualitatively identify food additives containing nitrogen. Hierarchical analysis, originally developed for solving evaluation-type problems, is introduced in this study to the field of qualitative analysis of terahertz spectra. This paper proposes and evaluates a qualitative identification method that combines THz-TDS with hierarchical analysis. In this study, six nitrogen-containing food additives were selected as experimental samples. First, the acquired terahertz time-domain spectral data were preprocessed and transformed into four datasets: peak values, peak positions, peak numbers, and overall spectral trends. Next, the data were divided into comparison and test sets to construct a qualitative identification model incorporating hierarchical analysis, followed by parameter optimization. The results indicated that the qualitative identification accuracy of additives based on single factors: overall trend, peak value, peak position, and peak number were 80.23%, 70.93%, 67.44%, and 40.70%, respectively. The multi-factor hierarchical analysis-based method improved the identification accuracy to 92.44%. In addition, by binarizing the fuzzy characterization of the absorption spectrum data during preprocessing and using it as the basis for assessing overall trends, the recognition accuracy increased to 94.19% when combined with the hierarchical analysis.

收稿日期：2023-08-29；修订日期：2024-04-11。

作者简介：赵渺钰（1999-），女，学士，研究方向为太赫兹技术应用。E-mail：13663540812@163.com。

通信作者：燕芳（1980-），女，教授，博士，研究方向为太赫兹技术应用。E-mail：0472yanfang@163.com。

基金项目：国家自然科学地区基金项目（62161042）；内蒙古自治区关键技术攻关计划项目（2021GG0361）；内蒙古自治区直属高校基本科研业务费项目。

model. These results demonstrate the effectiveness of the proposed qualitative identification algorithm. The method is straightforward, does not require training, and is well-suited for rapid qualitative detection of small sample sets.

Key words: analytic hierarchy process, qualitative detection, terahertz time domain spectroscopy, food additives

0 引言

近年来, 太赫兹时域光谱技术(Terahertz Time -Domain Spectroscopy, THz-TDS)已被广泛应用于食品检测、光谱分析、安全监测、生物医学等领域^[1]。太赫兹光谱具有典型的高光谱数据特征, 从海量的光谱数据中提取特征信息和建立统一指纹谱库是 THz-TDS 技术应用所面临的挑战。

在基于太赫兹光谱的定性识别技术发展的早期, 定性多依赖于对光谱中特殊吸收峰峰位的指认。然而, 当面对种类繁多且结构相似的待测物和测量过程中环境的不可抗力因素时, 仅靠吸收峰峰位指认来鉴别物质种类已不再适用。后来, 随着机器学习的不断发展和完善, 基于太赫兹光谱数据的物质鉴别也多采用机器学习的方法。例如, 胡晓华等人采用粒子群优化算法(Particle Swarm Optimization, PSO)结合支持向量机(Support Vector Machine, SVM), 实现了咖啡豆产地的快速鉴别, 识别率达到 95%^[2]; 周博等人针对不同烟草配方组分, 利用其太赫兹时域光谱分别建立了 SVM、K 近邻(K-Nearest Neighbors, KNN)和袋装树分类模型, 模型准确率均在 90%以上, 为太赫兹技术在烟草物料无损检测的应用上提供了参考^[3]。但机器学习作为对数据依赖较强的方法, 通常需要大量的训练集和较长时间来训练模型才能得到较好结果, 且针对不同训练数据均需寻找特定的最优参数^[4]。在实际检测中, 所得数据大都为小样本数据, 用其来训练模型可能存在过拟合或欠拟合问题^[5], 这在一定程度上限制了太赫兹时域光谱技术在物质定性鉴别领域的应用。

层次分析法(Analytic Hierarchy Process, AHP)是美国运筹学家、匹茨堡大学教授萨蒂研究“根据各个工业部门对国家福利的贡献大小而进行电力分配”课题时, 提出的一种层次权重决策分析方法, 主要用于解决评价类或者风险评估的问题^[6]。在定性分析中, 可利用其思想, 综合光谱数据的整体趋势、峰位、峰值、峰数等多个吸收谱特征因素, 将主观评价转化为基于数据的客观识别。这样不仅能够有效避免“一锤定音”式的判定, 还可简化光谱数据的预处理过程, 对于不同物质的相似光谱也能够根据上述 4 个

因素有效区分。为此, 本文以添加剂吸收谱数据为例, 提出了一种综合多因素的层次分析定性识别方法, 并验证了其有效性, 为受多种因素影响的太赫兹吸收谱定性分析提供了新思路。

1 方法理论及实验

1.1 算法简介

层次分析法(AHP)的分析思路是: 将复杂分析对象分解为目标层、准则层和方案层, 对同一层元素两两之间的重要程度做出判断, 建立判断矩阵, 并计算出每一因素对高一层指标重要程度的权重, 得到判断矩阵对应的权重, 为最优选择提供依据^[7]。根据层次分析结构, 首先从准则层开始, 比较各个准则于目标层的相对重要程度。相对重要性尺度如表 1 所示。

表 1 AHP 相对重要性尺度

Table 1 AHP relative importance scale

Scale	Define	Clarification
1	Equally important	Elements a and b are equally important for an attribute
3	Slightly important	Element a is slightly more important than element b for an attribute
5	More important	Element a is more important than element b for an attribute
7	Clearly important	Element a is significantly more important than element b for an attribute
9	Special importance	Element a is particularly more important than element b for an attribute
2, 4, 6, 8.	Midpoint	A compromise between the two neighbouring scales above
The reciprocal of the above scales	Inverse comparison	For example, a compares to b as 3 and b compares to a as 1/3.

1.2 综合多因素的层次分析法定性识别

1.2.1 评价指标设计

定义 1: 设矩阵 $\mathbf{W}=(w_{ij})_{(n+1)\times m}$, 其中 n 为准则层个数, m 为方案层个数。若有 $0 \leq w_{ij} \leq 1$, 则矩阵 \mathbf{W} 为最终权重矩阵。

定义 2: 设矩阵 $\mathbf{A}=(a_{ij})_{n\times n}$, 若有 $0 \leq a_{ij} \leq 1$, 则矩阵 \mathbf{A} 为准则层到目标层的判断矩阵, 其中 a_{ij} 表示指

标 i 相对于目标层的重要程度/指标 j 相对于目标层的重要程度。

1.2.2 模型设计

综合多因素的层次分析定性识别模型流程如图 1 所示, 具体步骤如下:

1) 由判断矩阵 A 可得 4 个准则于目标层的初始权重 $A_A = [0.6545, 0.2045, 0.0955, 0.0456]$, 将其填入权重矩阵的第一列。

2) 设单条测试数据矩阵为 $X = (x_1, x_2, x_3, \dots, x_j)$, $j=1, 2, 3, \dots, z$; 对比库数据矩阵为 $Y = (y_{ij})_{t \times z}$, 其中 z 表示数据的维度; t 表示对比库中光谱的数量。由于对比库中每类数据随机取 5 条后取平均, 故此处 $t=6$ 。导入测试数据后, 首先计算测试数据 X 与对比库数据 Y 关于每个准则的相似度, 其中相似度用欧式距离表征, 且与欧式距离成反比, 如式(1)所示; 其次, 设固定权重向量 $R = [0.29, 0.24, 0.19, 0.14, 0.092, 0.048]$, 依据测试数据在每个准则上与 6 类物质的相似度对 R 中元素进行排序, 由此构成方案层关于各个准则层的权重矩阵 B ; 最后将权重矩阵 B 填入最终权重矩阵 W 中。

$$\text{dist} = \sqrt{\sum_{j=1}^z (x_j - y_{ij})^2}, \quad j=1, 2, 3, \dots, z \quad (1)$$

3) 根据权重矩阵 W 得出单条测试数据的综合得分, 结果为 $C = A_A^T \times B$, 即 W 的第一列按列与其余列相乘求和。 C 数据存储形式为 $[c_1, \dots, c_6]$, 向量 C 中

最大值对应的列坐标即为最终识别结果。以此类推, 得出所有测试数据的识别结果。

1.3 实验

1.3.1 实验设备及样品制备

本文中所采集的实验数据均来自于北京市工业波谱成像工程技术研究中心的透射式 THz-TDS 实验系统^[8], 实验环境湿度始终保持在 5% 以下。所用实验样品均购置于阿拉丁试剂网, 其纯度 ≥99%。实验分别采集了 L-丙氨酸、苯甲酸、三聚氰胺、山梨酸、苏丹红 I 号、木糖醇这 6 种添加剂的单质太赫兹时域光谱, 其中补偿样品为聚乙烯 (PE)。为方便表述, 下文使用 $S_1 \sim S_6$ 来指代这 6 类添加剂, 每种添加剂样品有 3 种浓度, 每种浓度各 3 个样片, 共制备 9 个样片 (见表 2), 每个样片测量 4 次, 共 216 条数据。

1.3.2 数据预处理

根据 T. D. Dorney 与 L. Duvillaret 提出的光学参数基本模型^[9-10], 计算各样片在太赫兹频段的吸收系数 $\alpha(\omega)$, 公式如下:

$$n(\omega) = \frac{\rho(\omega)c}{\omega d} + 1 \quad (2)$$

$$\alpha(\omega) = \frac{2}{d} \ln \left\{ \frac{4n(\omega)}{\rho(\omega)[n(\omega)+1]^2} \right\} \quad (3)$$

式中: c 为真空光速; ω 为振动频率; d 为样品厚度; $\rho(\omega)$ 为样品信号与参考信号的相位差; $\rho(\omega)$ 为样品信号与参考信号的振幅比; $n(\omega)$ 为样品的复折射率。

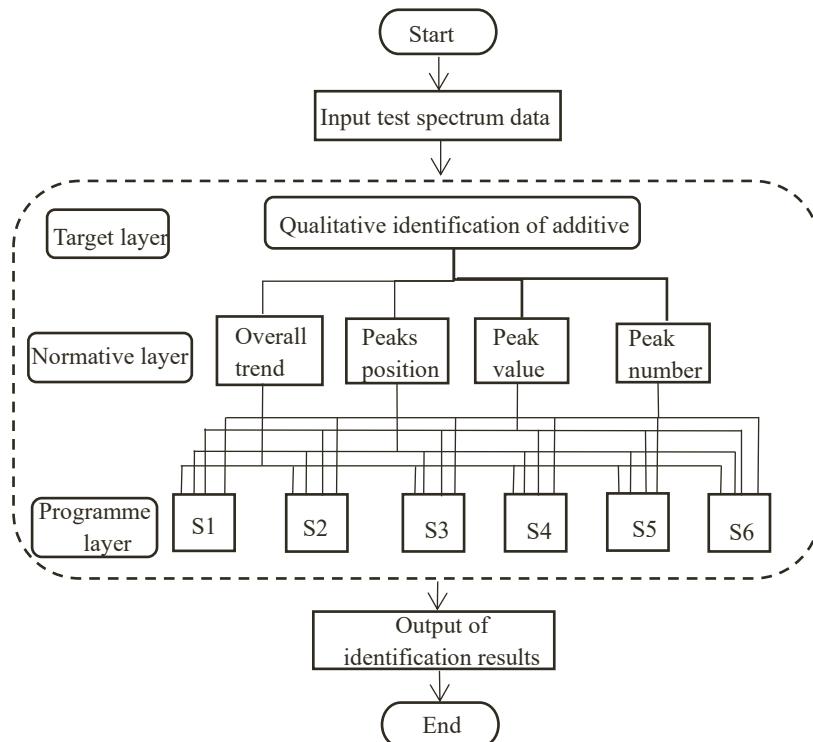


图 1 综合多因素的层次分析定性识别模型的流程图

Fig.1 Flow chart of qualitative identification model with AHP based on multiple factors

表2 添加剂样片信息统计

Table 2 Statistical of additives samples

Sample name	Category number	Sample concentration /%	Sample quantity
L-Alanine	S_{1-1} S_{1-2} S_{1-3}	10%	3
	S_{1-4} S_{1-5} S_{1-6}	25%	3
	S_{1-7} S_{1-8} S_{1-9}	35%	3
Benzoic acid	S_{2-1} S_{2-2} S_{2-3}	10%	3
	S_{2-4} S_{2-5} S_{2-6}	25%	3
	S_{2-7} S_{2-8} S_{2-9}	35%	3
Melamine	S_{3-1} S_{3-2} S_{3-3}	10%	3
	S_{3-4} S_{3-5} S_{3-6}	25%	3
	S_{3-7} S_{3-8} S_{3-9}	35%	3
2,4-Hexadienoic acid	S_{4-1} S_{4-2} S_{4-3}	10%	3
	S_{4-4} S_{4-5} S_{4-6}	25%	3
	S_{4-7} S_{4-8} S_{4-9}	35%	3
Sudan-I	S_{5-1} S_{5-2} S_{5-3}	10%	3
	S_{5-4} S_{5-5} S_{5-6}	25%	3
	S_{5-7} S_{5-8} S_{5-9}	35%	3
Xylitol	S_{6-1} S_{6-2} S_{6-3}	10%	3
	S_{6-4} S_{6-5} S_{6-6}	25%	3
	S_{6-7} S_{6-8} S_{6-9}	35%	3

首先, 截取吸收光谱的 0.6~2.6 THz 作为特征频

段, 经 S-G 平滑和剔除含坏点数据后, 得到 203 条有效光谱数据, 每条数据 632 个数据点; 其次, 对光谱数据进行寻峰, 并建立含 4 个因素的数据集; 最后划分对比集和测试集。

2 分析及优化

2.1 识别结果分析

基于太赫兹光谱中不同因素的添加剂定性识别结果如图 2 所示。由图 2 可知, 基于单因素的定性识别结果并不理想(图(a)~(d))。原因可以归结为由于实验环境和样品厚度的非绝对一致, 导致同类数据在某特征区间有差异, 且不同类样品之间存在相似特征。图(e)中, 由于基于多因素的层次分析算法按特定权重考虑了各因素对识别的影响, 故识别准确率相较于仅考虑单一因素的识别结果有了较大的提升。

2.2 识别结果优化

2.2.1 数据模糊化表征

对比图 2 的识别结果可知, 吸收谱中谱线整体趋势在各个因素中重要程度较高。因此, 本文提出将光谱整体趋势模糊化表征的方法, 具体如下:

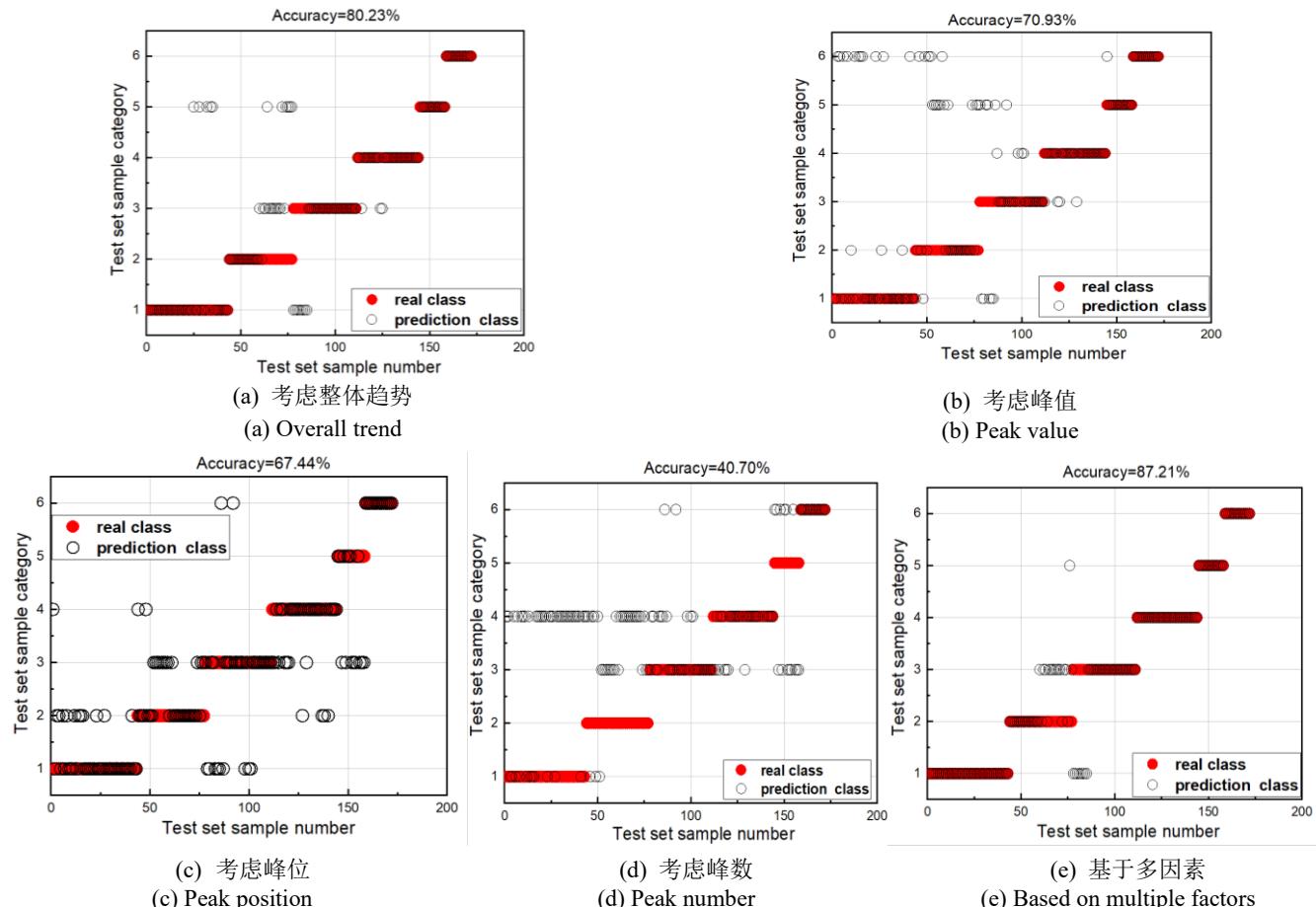


图 2 考虑不同因素对 6 类添加剂样品的识别结果

Fig.2 The recognition results of 6 types of additive samples based on different factors

首先, 记单条原始光谱为 $\mathbf{H}=(h_1, h_2, h_3, \dots, h_j)$, $j=1, 2, 3, \dots, g$, 其中 g 为原始光谱数据的维度; 其次, \mathbf{H} 中相邻数据的差记为 $\mathbf{E}=(e_j)_{1 \times j}=(h_2-h_1, h_3-h_2, \dots, h_j-h_{j-1})$, 当 $e_j \geq \lambda$ 或 $-\lambda < e_j < 0$ 时, 令 $e_j=1$, 代表上升; 当 $e_j < -\lambda$ 或 $0 < e_j < \lambda$ 时, 令 $e_j=-1$, 代表下降。其中 λ 代表数据差的平均值, 这样 \mathbf{H} 就会转化成只有 -1 和 1 的数据集。

样品 S_5 的部分原始吸收光谱数据与模糊表征后数据的对比如图 3 所示。从图 3 可知, S_5 在 1.46 THz 为一个特征点。原始数据中 S_{5-1} 在 1.46 THz 右侧数据呈现上升趋势, 与 S_{5-2} 和 S_{5-3} 在 1.46 THz 右侧的趋势并不匹配, 这可能导致识别精度的降低。而模糊表征使 3 条数据在 1.46 THz 处均体现出了其特征, 并在特征不明显的区域也有一定的特征表现。故该方法可突出光谱数据的整体趋势特征, 且整数运算相比于小数运算节省了程序的运算时间。此外, 在后续方案层对于整体趋势的相似度上, 将不再依据欧式距离来排序 R , 而用数据之间的重复率来做 R 排序的依据。

2.2.2 参数优化

结合层次分析法的定性识别算法的优化主要是针对整体趋势、峰值、峰位、峰数的权重指标比例进行参数寻优。从初始参数出发, 寻参范围设为 $0.2 \sim 0.7$, 步长为 0.05 。采用枚举法对整体趋势分别用原始光谱数据表征和模糊表征后的层次分析定性识别模型进行参数寻优, 寻优结果如图 4 所示。

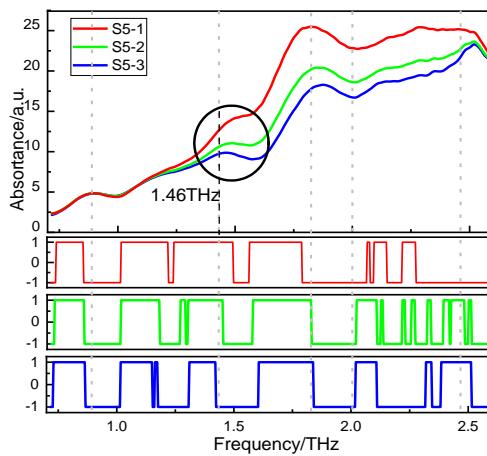


图 3 样品 S_5 原始吸收谱和模糊表征谱对比图

Fig.3 Comparison of original absorption spectra and fuzzy characterization spectra of sample S_5

表 3 展示了在结合层次分析法的定性识别模型下, 整体趋势用模糊表征(记为 MH)和原始数据表征(记为 YS)各自的识别精度与对应参数。由表 3 可知: YS+AHP 的最优识别率为 92.44%; MH+AHP 的最优识别率为 94.19%; 仅考虑整体趋势时, 使用 YS

法的识别率为 80.23%, 而使用 MH 法的识别率为 93.60%。

对于 YS 法, 依据欧式距离来判定相似度。由于每个坐标对欧式距离的贡献是同等的, 当两组数值某个坐标有较大偏差时, 会造成最终欧式距离偏大, 而影响判别结果; 而对于 MH 法, 在突出表示光谱数据整体趋势的同时, 也会受阈值 λ 取值大小的影响。

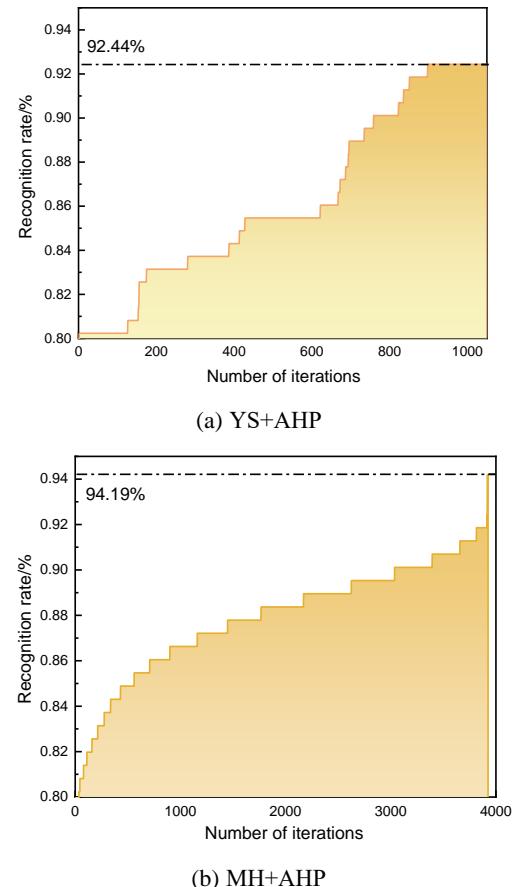


图 4 层次分析法权重指标寻优结果图

Fig.4 Optimization results of AHP weight index

表 3 识别精度及其对应参数统计表 (AHP 方法)

Table 3 Identification accuracy and corresponding parameters statistical table

Data representation	Parameter weighting				Identification accuracy/%
	Overall trend	Peak value	Peak position	Peak number	
YS	1	0	0	0	80.23
YS	0.6	0.1	0.2	0.1	92.44
MH	1	0	0	0	93.60
MH	0.8	0.1	0.1	0.0	94.19

3 结论

随着太赫兹光谱数据库物质种类的不断充盈, “同物异谱”和“异物相似谱”的现象普遍存在。仅

依靠峰位或峰值来定性已不能满足于识别精度的需求。本文提出的结合层次分析法的定性识别算法,综合考虑吸收光谱的整体趋势、峰值、峰位、峰数等多种影响因素及其对应的客观权重,对光谱进行定性识别。此方法无需复杂的光谱预处理和特征提取过程。

此外,本文验证了数据模糊表征法的有效性,有效避免了“同物异谱”现象造成整体趋势不一致问题。通过对6类添加剂的识别结果分析可知,结合层次分析法的太赫兹时域光谱定性识别方法可有效对物质进行定性识别,其识别准确率最高可达94.19%。

然而,该方法的识别精度受对比库中标准数据和数据预处理时寻峰算法参数的影响。因此,后续研究中还需要建立更加标准、统一的光谱数据库和检测识别的一体化系统,进一步实现小样本数据的快速准确识别。

参考文献:

- [1] 王芳,张春红,赵景峰,等.基于混合机器学习法的太赫兹波鉴别草种的研究[J].激光与光电子学进展,2021,58(3): 318-324.
WANG F, ZHANG C H, ZHAO J F, et al. Identification of a grass species using a terahertz wave based on hybrid machine learning method[J]. *Laser & Optoelectronics Progress*, 2021, 58(3): 318-324.
- [2] 胡晓华,刘伟,刘长虹,等.基于太赫兹光谱和支持向量机快速鉴别咖啡豆产地[J].农业工程学报,2017,33(9): 302-307.
HU X H, LIU W, LIU C H, et al. Rapid identification of producing area of coffee bean based on terahertz spectroscopy and support vector machine[J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2017, 33(9): 302-307
- [3] 周博,朱文魁,王赵改,等.基于太赫兹时域光谱技术的烟草组分识别[J].农业工程学报,2022,38(10): 310-316.
ZHOU B, ZHU W K, WANG Z G, et al. Identification of tobacco materials based on terahertz time-domain spectroscopy[J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2022, 38(10): 310-316.
- [4] 崔向伟,沈韬,刘英莉,等.小样本太赫兹光谱识别[J].激光与光电子学进展,2021,58(1): 321-328.
CUI X W, SHEN T, LIU Y L, et al. Recognition of small-sample terahertz spectrum[J]. *Laser & Optoelectronics Progress*, 2021, 58(1): 321-328.
- [5] 甘岚,沈鸿飞,王瑶,等.基于改进DCGAN的数据增强方法[J].计算机应用,2021,41(5): 1305-1313.
GAN L, SHEN H F, WANG Y, et al. Data augmentation method based on improved deep convolutional generative adversarial networks[J]. *Journal of Computer Applications*, 2021, 41(5): 1305-1313.
- [6] 吴建.以案例教学为核心的适应性学习系统的设计与实现[D].杭州:浙江工业大学,2009.
WU Jian. The design and implementation of case-based learning in adaptive learning[D]. Hangzhou: Zhejiang University of Technology, 2009.
- [7] 郑凤翥,宁飞,王惠林,等.光电系统伺服稳定平台可靠性定性分析[J].应用光学,2022,43(5): 853-858.
ZHEN F Z, NING F, WANG H L, et al. Qualitative analysis of reliability on servo stabilizationplatform of electro-optical system[J]. *Journal of Applied Optics*, 2022, 43(5): 853-858.
- [8] 燕芳,刘同华,张俊林.糖类同分异构体的太赫兹吸收峰形成机理研究[J].光学学报,2022,42(5): 228-234.
YAN F, LIU T H, ZHANG J L. Formation mechanism of terahertz absorption peaks of carbohydrate isomers[J]. *Acta Optica Sinica*, 2022, 42(5): 228-234.
- [9] Dorney T D, Baraniuk R G, Daniel M, et al. Material parameter estimation with terahertz time-domain spectroscopy[J]. *Journal of the Optical Society of America A*, 2001, 18(7): 1562.
- [10] Duvillare L, Garet F, Coutaz J L, et al. A reliable method for extraction of material parameters in terahertz time-domain spectroscopy[J]. *IEEE Journal of Selected Topics in Quantum Electronics*, 1996, 2(3): 39.