

文章编号:1672-6987(2015)05-0482-05; DOI: 10.16351/j.1672-6987.2015.05.002

# 基于支持向量机和太赫兹光谱技术的农药分类

黄翼虎<sup>1</sup>, 栾浩杰<sup>1</sup>, 刘淑泉<sup>1</sup>, 张硕<sup>1</sup>, 马吉祥<sup>2</sup>, 林其安<sup>1</sup>

(1. 青岛科技大学 自动化与电子工程学院, 山东 青岛 266042; 2. 山东泰开自动化有限公司, 山东 泰安 271000)

**摘要:** 基于农药的太赫兹吸收光谱和支持向量机对4种农药进行了分类。利用太赫兹时域光谱系统对4种农药, 即三氯杀螨砜、敌百虫、亚胺硫磷和六氯苯进行检测, 获得样品的太赫兹时域信号, 然后计算获得其在0.3~2.2 THz范围内的吸收光谱。用主成分分析法对太赫兹吸收光谱进行分析, 提取光谱特征, 用支持向量机分别对太赫兹吸收光谱和光谱特征进行分类。实验结果表明, 这两种方案均能有效地对农药进行分类, 分类正确率为100%。

**关键词:** 太赫兹吸收光谱; 农药; 主成分分析; 支持向量机

中图分类号: TS 255 文献标志码: A

## Pesticide Classification Based on Support Vector Machine and Terahertz Spectroscopy

HUANG Yi-hu<sup>1</sup>, LUAN Hao-jie<sup>1</sup>, LIU Shu-quan<sup>1</sup>, ZHANG Shuo<sup>1</sup>, MA Ji-xiang<sup>2</sup>, LIN Qi-an<sup>1</sup>

(1. College of Automation and Electronic Engineering, Qingdao University of Science and Technology, Qingdao 266042, China;

2. Shandong Taikai Automation Co. Ltd, Taian 271000, China)

**Abstract:** Based on pesticide terahertz absorption spectroscopy and support vector machines for four pesticides are classified. Use of terahertz time-domain spectroscopy system for four pesticides that tetradifon, trichlorfon, phosmet and hexachlorobenzene were detected to obtain a sample of terahertz time-domain signal, and then calculating gain its 0.3—2.2 THz absorption spectrum within the range. Principal component analysis of the THz absorption spectra were analyzed to extract the spectral characteristics of terahertz absorption spectra and spectral characteristics were using support vector machine for classification, the experimental results show that the two programs can effectively classify pesticides classification the correct rate of 100%.

**Key words:** terahertz absorption spectroscopy; pesticide; principal component analysis; SVM

在农业生产中,为了保障、促进植物及农作物的成长,需要施用杀虫、杀菌、杀灭有害动物或杂草的农药。我国农药产业发展非常迅速,已经成为世界第二大农药生产国,2012年我国所生产

的化学农药原药(折合有效成分100%)已达354.9万t,为保障和促进我国农业生产的顺利进行提供了重要支持。然而,农药的大量使用在发挥积极作用的同时,也对生态环境造成污

收稿日期: 2014-06-17

基金项目: 国家自然科学基金项目(61340038); 山东省自然科学基金项目(2013ZRB01254); 济南出入境检验检疫局项目(2012IK182).

作者简介: 黄翼虎(1971—), 男, 副教授。\*通信联系人。

染及破坏,同时农药在农产品及食品中的残留日益成为一个影响人类安全的社会问题,越来越受到人们的关注。目前,我国有600多种化合农药,每年均有新的农药研制成功,要区分这些种类繁多的农药,单凭感观和经验是不可行的,此时就需要有一种快速简单准确的方法对农药种类进行鉴别。

太赫兹时域光谱技术是近些年发展起来的一种无损光谱检测技术。许多有机物分子的低频振动频率落在太赫兹波段,这些振动对应分子内低频集体振动模式、大分子的骨架振动模式、分子间弱相互作用(氢键与范德华力)引起的振动模式以及晶格的声子振动模式<sup>[1-3]</sup>;太赫兹波能够透射由非极性分子构成的材料<sup>[4]</sup>,如纸制品、塑料制品、织物、陶瓷等材料,可以对包装内的待检品进行检测。太赫兹光子能量比紫外光低,频率为1 THz的光子的能量大约为4.1 meV,这样低的能量不会对待测物产生光致电离<sup>[5]</sup>,对人体和待测物没有危害;太赫兹技术应用相干探测技术<sup>[6]</sup>,能够同时测得太赫兹辐射电场的瞬时幅值与相位(而其他传统的光谱技术只能直接测得幅值信息),进而计算样品的光学参数。该技术的信噪比非常高,可以达到80 dB。鉴于太赫兹时域光谱技术所具有的特点,可以利用该技术对物质进行分类与鉴别。梁美彦等<sup>[7]</sup>利用自组织特征映射(SOM)神经网络对6种毒品的太赫兹吸收光谱进行了聚类识别;李茜<sup>[8]</sup>以4种炸药太赫兹吸收光谱的特征吸收峰为属性特征,对4种炸药进行模糊聚类分析;马治浩<sup>[9]</sup>等用小波包变换来对4种杀菌剂太赫兹吸收光谱进行特征提取,然后用欧式距离法对4种农药进行分类。

本工作利用太赫兹时域光谱系统(THz-TDS)对4种农药(三氯杀螨砜、敌百虫、亚胺硫磷和六氯苯)进行检测,获得样品的太赫兹时域信号,然后计算获得其太赫兹吸收光谱。用主成分分析法对太赫兹吸收光谱进行分析,提取光谱特征,用支持向量机分别对太赫兹吸收光谱和光谱特征进行分类,实验结果表明,这两种方案均能有效地对农药进行分类。

## 1 实验部分

### 1.1 试验与仪器

三氯杀螨砜、敌百虫、亚胺硫磷六氯苯,购于百灵威科技有限公司,纯度均大于99%,实验前

均进一步提纯。

钛蓝宝石飞秒激光器,Vitesse-800-5型,美国Coherent公司;太赫兹光谱仪,Model Z-3型,美国Zomega公司。

### 1.2 实验方法

实验前将4种药品经干燥后,在红外灯照射下用玛瑙研钵充分研磨。然后分别与聚乙烯粉末以质量的1:1混合,用漩涡混合器混合均匀,放入压片机中以16 MPa压力压成直径为13 mm的圆形薄片。各压片结构均匀,表面光滑无破损,且两表面平行。

钛蓝宝石飞秒激光器和太赫兹光谱仪构成,装置详细说明见参考文献[10]。整个实验在室温下进行,且太赫兹光谱仪置于充满高纯氮气的箱体内,箱内相对湿度约为1%。利用THz-TDS系统对4种农药进行检查,得到太赫兹时域信号,每种药品17个样本。

## 2 4种农药的太赫兹吸收光谱分析

4种农药样本的太赫兹吸收光谱见图1。

从图1吸收系数光谱曲线可以看出,在0.4~2.2 THz频率范围内,三氯杀螨砜分别在0.65,0.96,1.30,1.57,1.84和2.18 THz处有6个吸收峰;敌百虫分别在1.17,1.56和2.06 THz处有3个吸收峰;亚胺硫磷分别在0.90,1.23,1.68,1.84,1.98和2.11 THz处有6个吸收峰;六氯苯对太赫兹波的吸收非常弱,分别在0.64,0.87,1.06,1.23,1.54,1.74和1.93 THz处有7个吸收峰。

## 3 主成分分析与支持向量机理论

### 3.1 主成分分析理论

主成分分析(PCA)方法是一种特征提取方法,其中心思想是数据降维,以排除信息共存中相互重叠的部分<sup>[11-12]</sup>。它是将原变量进行转换,使少数几个新变量是原变量的线性组合,同时,这些新变量尽可能多的表征原变量的数据结构特征而不丢失信息。经转换得到的新变量相互正交互不相关,消除了众多共存信息中相互重叠的部分,即消除变量之间可能存在的多重共线性<sup>[13]</sup>。

假定对每个样本取n个特征,即

$$X = (x_1, x_2, \dots, x_n)^T. \quad (1)$$

要求构造个新特征 $p_1, p_2, \dots, p_n$ ,使它们满足以下条件:

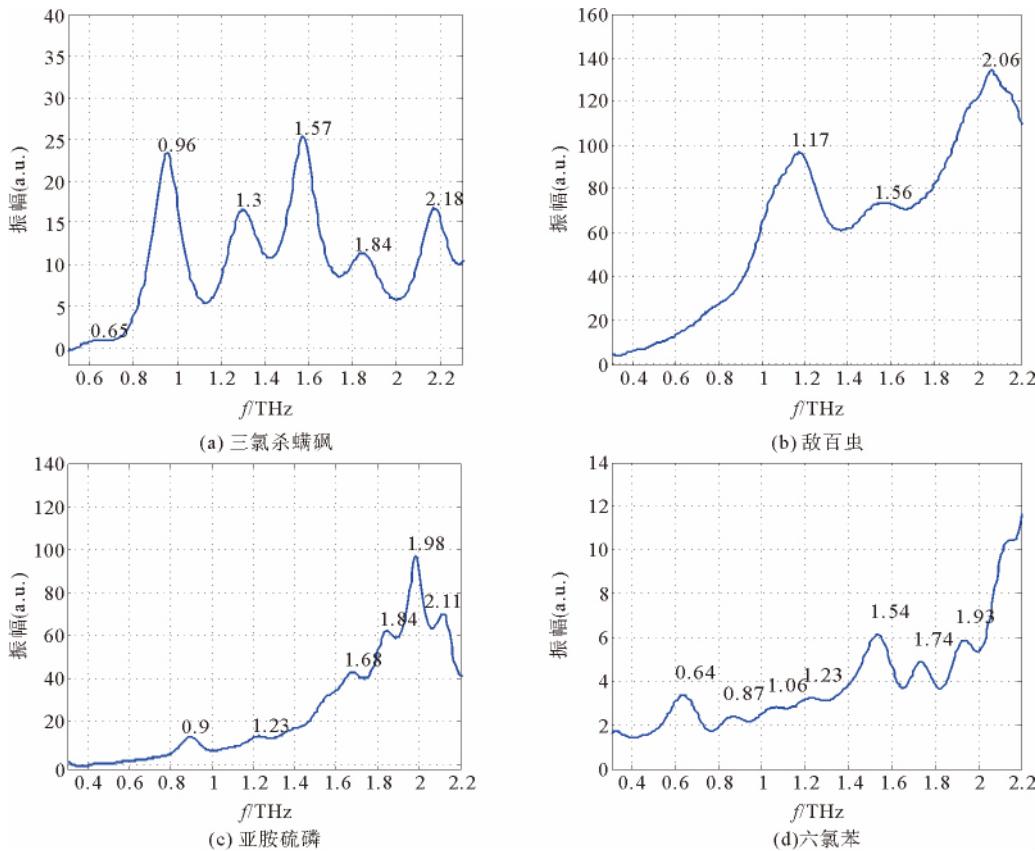


图1 4种农药的吸收系数光谱图

Fig. 1 Absorption coefficient spectrum of four pesticides

1) 每个新特征是原有特征的线性组合,即

$$p_i = u_{i1}x_1 + u_{i2}x_2 + \dots + u_{in}x_n, i = 1, 2, \dots, n \quad (2)$$

或

$$p_i = u_i^T X, u_i = (u_{i1}, u_{i2}, \dots, u_{in})^T, i = 1, 2, \dots, n. \quad (3)$$

其中,  $u_{ij}$  是实常数。

2) 各个新变量之间是不相关的,即相关系数为零:

$$r(y_i, y_j) = 0, i, j = 1, 2, \dots, n, i \neq j. \quad (4)$$

$u_1$  使  $p_1$  的方差达到极大,  $u_2$  使  $p_2$  的方差达到次大, 以此类推。

满足以上 2 个条件的新特征  $p_1, p_2, \dots, p_n$  分别称为样本点的第  $1, 2, \dots, n$  个主成分。

主成分计算步骤<sup>[14]</sup>:

首先求出全体样本点的协方差矩阵:

$$S_x = \begin{bmatrix} S_{11} & S_{12} & \cdots & S_{1n} \\ S_{21} & S_{22} & \cdots & S_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ S_{n1} & S_{n2} & \cdots & S_{nn} \end{bmatrix}. \quad (5)$$

然后求出  $S_x$  的  $n$  个特征值  $\lambda_1, \lambda_2, \dots, \lambda_n$  (各

特征值按从大到小的顺序排列, 即  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ ) 和与之对应的正交归一化的特征向量  $u_1, u_2, \dots, u_n$ 。此时, 就得到主成分分析的正交变换矩阵  $U$ :

$$U = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}. \quad (6)$$

### 3.2 支持向量机理论

支持向量机(SVM)着重研究在小样本情况下的统计学习问题, 在解决小样本、非线性及高维模式识别中表现出许多特有的优势。它克服了其它机器鉴别技术无法解决的过学习、非线性、维数灾难、局部最小点等问题<sup>[15]</sup>。

支持向量机分类的基本思想是<sup>[16]</sup>: 通过非线性映射  $\Phi$  将线性不可分类样本  $(x_i, y_i)$ ,  $i = 1, 2, \dots, N, x_i \in \mathbf{R}^n, y_i \in \{-1, +1\}$  映射到高维特征空间  $H$  中进行线性分类。即在新空间建立最优分类超平面:

$$[\omega, \Phi(x)] + b = 0. \quad (7)$$

在条件中增加松弛项  $\xi_i \geq 0$  及错误惩罚因子

C。此时目标函数变为:

$$\text{最小化: } V(w, \xi) = \frac{1}{2} [w, w] + C \sum_{i=1}^N \xi_i. \quad (8)$$

约束条件为  $y_i([w, \Phi(x_i)] + b) \geq 1 - \xi_i$ , 其中  $\xi_i \geq 0$ ,  $i = 1, 2, \dots, N$ 。  $(9)$

只要一种核函数  $K(x_i, x_j)$  满足 Mercer 条件, 它就对应某一变换空间中的内积, 而不必知道低维空间是怎样向高维空间变换的。选择好核函数  $K(x, y) = [\Phi(x), \Phi(y)]$  后, 非线性支持向量机的最优化问题的对偶问题就变为最大化:

$$W(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j K(x_i, x_j), \quad (10)$$

使满足条件:

$$\sum_{i=1}^N \alpha_i y_i = 0, (0 \leq \alpha_i \leq C; i = 1, 2, \dots, N). \quad (11)$$

解得其最优解

$$\alpha^* = (\alpha_1^*, \dots, \alpha_l^*)^\top, \quad (12)$$

$$b^* = y_i - \sum_{i=1}^l y_i \alpha_i^* K(x_i, x_j). \quad (13)$$

此时最优分类函数为

$$f(x) = \text{sgn}\{[w^*, \Phi(x)] + b^*\} = \text{sgn}\left\{\sum_{i=1}^N \alpha_i^* y_i K(x_i, x_j) + b^*\right\}. \quad (14)$$

## 4 分类结果

### 4.1 4种农药的吸收光谱主成分提取

对4种农药在0.3~2.2 THz频率范围内的4×17组吸收光谱进行主成分分析,以提取特征。计算出各主成分,其中第一主成分贡献率为89.07%,第二主成分贡献率为9.85%,第三主成分贡献率为0.55%,前3个主成分的总贡献率大于99.47%,所以选择吸收光谱的前3个主成分作为数据特征。4种农药的吸收光谱在前两个主成分构成的体系内的分布图见图2。

每种农药测量17组数据,从中均匀的选择三分之一(每种农药选6组)作为预测集数据,其余11组作为训练集数据。

### 4.2 支持向量机对农药吸收光谱的主成分进行分类

选择C-支持向量机(C-SVM)类型,选择RBF函数作为核函数,经过十折交叉验证,确定参数C=0.5,γ=0.000 122 070 312 5。先以训练

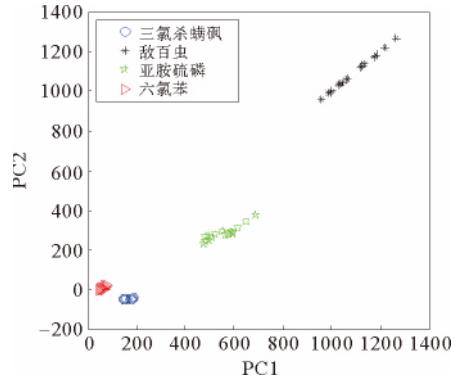


图2 4种农药的吸收光谱在前两个主成分

构成的体系内的分布图

Fig. 2 Distribution of the absorption spectrum of four kinds of pesticides in the system composed the first two principal components

集作为输入构造C-支持向量机,训练好的C-支持向量机模型参数如下:支持向量总个数为26,其中三氯杀螨砜5个,敌百虫9个,亚胺硫磷7个,六氯苯5个;应用训练好的支持向量机模型对预测集样本进行分类,识别结果正确率100%。

### 4.3 支持向量机对农药吸收光谱进行分类

鉴于上面的支持向量机分类方法需要先对各种农药的吸收系数谱进行主成分分析提取特征向量,试图采用更为简单的处理方法,直接对农药的吸收系数光谱作为分类处理的特征输入,在采用支持向量机方法进行4种农药的分类,训练集和预测集采用相同的样本。

以4类农药的太赫兹吸收系数光谱作为输入,选择C-支持向量机(C-SVM)类型,选择RBF函数作为核函数,经过10折交叉验证,确定最优参数C=0.5,γ=0.007 812 5,训练得到的模型参数为支持向量总数为28,其中,三氯杀螨砜5个,敌百虫10个,亚胺硫磷8个,六氯苯5个;然后用训练好的C-支持向量机模型对预测集样本进行分类,识别结果正确率100%。

## 5 结论

基于太赫兹时域光谱系统对4种农药进行检测,计算得到样品的太赫兹吸收光谱,分别以4种农药的吸收光谱和其前3个主成分作为特征向量,对其进行主成分分析进行分类;利用支持向量机对4种农药进行分类,2种分类方案识别结果正确率均为100%。因实验条件有限,对更多农药种类和样本数量的分类情况需进一步研究。

## 参 考 文 献

- [1] Walther M, Plochocka P, Fischer B, et al. Collective vibrational modes in biological molecules investigated by terahertz time-domain spectroscopy[J]. *Biopolymers*, 2002, 67(4/5): 310-313.
- [2] Tonouchi M. Cutting-edge terahertz technology[J]. *Nature Photonics*, 2007, 1(2): 97-105.
- [3] Pickwell E, Wallace V P. Biomedical applications of terahertz technology[J]. *Journal of Physics D: Applied Physics*, 2006, 39(17): R301.
- [4] Nuss M C. Chemistry is right for T-ray imaging [J]. *IEEE Circuits and Devices Margin*, 1996, 12: 25-30.
- [5] 王少宏, 许景周, 汪力, 等. THz 技术的应用及展望[J]. 物理, 2001, 30(10): 612-615.  
Wang Shaohong, Xu Jingzhou, Wang Li, et al. Applications and prospects of terahertz technology [J]. *Physics*, 2001, 30(10): 612-615.
- [6] Grischkowsky D, Keiding S, van Exter M, et al. Fattinger. Far-infrared time-domain spectroscopy with terahertz beams of dielectrics and semiconductors[J]. *Journal of Optical Society America B*, 1990, 7: 2006-2015.
- [7] Liang M, Shen J L, Wang G Q. Identification of illicit drugs by using SOM neural networks[J]. *Journal of Physics D: Applied Physics*, 2008, 41(13): 135306.
- [8] 李茜. 三种军用炸药的太赫兹光谱研究[D]. 太原: 中北大学, 2012.  
Li Qian. Research on three military explosives' terahertz spectrums[D]. Taiyuan: North University of China, 2012.
- [9] 马治浩, 王强, 王孝伟, 等. 杀菌剂类农药的太赫兹光谱分类方法[J]. 湖南科技大学学报: 自然科学版, 2012, 27(1): 107-111.  
Ma Yehao, Wang Qiang, Wang Xiaowei, et al. Terahertz spectroscopy cluster analysis of four kinds of bactericide[J]. *Journal of Hunan University of Science and Technology: Natural Science Edition*, 2012, 27(1): 107-111.
- [10] 侯迪波, 岳飞亨, 康旭升, 等. 持久性有机物硫丹的太赫兹光谱研究[J]. 光谱学与光谱分析, 2012, 32 (5): 1170-1174.  
Hou Dibo, Yue Feiheng, Kang Xusheng, et al. Terahertz time-domain spectroscopy of alpha endosulfan persistent organic pollutant [J]. *Spectroscopy and Spectral Analysis*, 2012, 32(5): 1170-1174.
- [11] 稚小立. 化学计量学方法与分子光谱分析技术[M]. 北京: 化学工业出版社, 2011: 55-57.  
Zhu Xiaoli. *Molecular Spectroscopy Analytical Technology Combined with Chemometrics and Its Applications* [M]. Beijing: Chemical Industry Press, 2011: 55-57.
- [12] 李弼程, 邵美珍, 黄浩. 模式识别原理与应用[M]. 西安: 西安电子科技大学出版社, 2008: 82-84.  
Li Bicheng, Shao Meizhen, Huang Hao. *Principle and application of pattern recognition* [J]. Xi'an: Xi'an Electronic and Science University Press, 2008: 82-84.
- [13] 许国根, 贾瑛. 模式识别与智能计算的 MATLAB 实现[M]. 北京: 北京航空航天大学出版社, 2012: 103-105.  
Xu Guogen, Jia Ying. *MATLAB implementation of pattern recognition and intelligent computation* [J]. Beijing: Beihang University Press, 2012: 103-105.
- [14] 盛立东. 模式识别导论[M]. 北京: 北京邮电大学出版社, 2010: 244-251.  
Sheng Lidong. *Introduction to Pattern Recognition* [M]. Beijing: Beijing University of Posts and Telecommunications Press, 2010: 244-251.
- [15] 李晶皎, 赵红丽, 王爱侠. 模式识别[M]. 北京: 电子工业出版社, 2010: 149-187.  
Li Jingjiao, Zhao Hongli, Wang Aixia. *Pattern Recognition* [M]. Beijing: Publishing House of Electronics Industry, 2010: 149-187.
- [16] 白鹏, 张喜斌, 张斌, 等. 支持向量机理论及工程应用实例 [M]. 西安: 西安电子科技大学出版社, 2008: 13-48.  
Bai Peng, Zhang Xibin, Zhang Bin, et al. *Support Vector Machine Theory and Engineering Application Examples* [M]. Xi'an: Xi'an Electronic and Science University Press, 2008: 13-48.

(责任编辑 林琳)