

利用径向基神经网络对毒品太赫兹光谱的识别

梁美彦 赵树森 沈京玲

(北京市太赫兹波谱与成像重点实验室, 太赫兹光电子学教育部重点实验室, 首都师范大学物理系, 北京 100048)

摘要 利用太赫兹时域光谱技术(THz-TDS)对 11 种常见毒品进行了实验探测并得到它们在 0.2~2.6 THz 范围内的吸收谱, 然后利用径向基(RBF)神经网络对 11 种毒品的太赫兹吸收光谱进行了训练和识别。训练和识别的光谱在输入网络之前都经过归一化处理, 识别率达到了 96%。RBF 神经网络模型是用 Matlab 语言编制。识别结果表明, 利用 RBF 神经网络可以实现对不同种类毒品的识别, 而且训练和识别的速度非常快, 几乎是实时的。这也是 RBF 神经网络优于误差逆传播(BP)神经网络和自组织特征映射(SOM)网络的特点。因此, RBF 神经网络为太赫兹技术用于毒品的检测和识别提供了一种有效而快捷的方法。

关键词 光谱学; 毒品识别; 太赫兹吸收光谱; 径向基神经网络

中图分类号 O433.4; O235 **文献标识码** A **doi:** 10.3788/AOS200929s1.0226

Identification of Terahertz Absorption Spectra of Illicit Drugs Using Radial Basis Function Neural Networks

Liang Meiyang Zhao Shusen Shen Jingling

(Beijing Key Laboratory for Terahertz Spectroscopy and Imaging, Key Laboratory of Terahertz Optoelectronics, Ministry of Education, Department of Physics, Capital Normal University, Beijing 100048, China)

Abstract Absorption spectra of 11 illicit drugs were measured by using the terahertz time-domain spectroscopy technique (THz-TDS) in the range 0.2~2.6 THz, and then the absorption spectra of the 11 different illicit drugs were identified successfully by well trained radial basis function (RBF) neural networks. The spectra used for training and identification were normalized before input to the network and the positive identification rate is 96%. The model of RBF neural network was processed in Matlab. The results indicate that it is feasible to apply RBF neural network to the identification of illicit drugs, and the speed of training and identification is fast. The character makes RBF network superior to back propagation neural network (BP) and self-organizing feature map (SOM). Therefore, RBF neural network provides a fast and effective method in the secure inspection and identification for illicit drugs using THz-TDS technique.

Key words spectroscopy; illicit drugs identification; terahertz absorption spectra; radial basis function neural networks

1 引 言

毒品不仅危害人们的身心健康, 而且容易引发一系列社会犯罪活动, 因此缉毒工作对于打击毒品走私和维护社会稳定两方面来说都具有十分重要的

意义。针对不同类型的毒品, 刑法中对毒贩的量刑也有所不同, 目前存在的检测方法中, 一些方法仅能确定毒品的形状而不是种类, 所以从对毒贩的量刑来说, 也迫切需要一种快速有效的识别手段。太赫

基金项目: 国家 973 计划(2007CB310408, 2006CB302901)、北京教育委员会科技发展计划(KM200910028005)和北京市人才强教项目资助课题。

作者简介: 梁美彦(1982-), 女, 硕士研究生, 主要从事太赫兹光谱检测神经网络识别等方面的研究。

E-mail: liangyan661982@163.com

导师简介: 沈京玲(1957-), 女, 教授, 主要从事太赫兹光谱和非线性光学等方面的研究。

E-mail: sjl-phy@mail.cnu.edu.cn (通信联系人)

兹波是介于毫米波和红外光之间的电磁辐射,它能够穿透衣服、信封以及塑料包裹等材料而且对人体没有伤害;其单光子能量特别低,可以做到无损检测;而且一些有机分子内部的振动和转动以及晶体晶格的低频振动均位于太赫兹波段,因此使太赫兹时域光谱技术 (THz-TDS) 在该波段对海洛因、吗啡、盐酸氯胺酮、可卡因等有机分子的探测成为可能^[1~4]。2003年 Kodo Kawase 等使用太赫兹参量振荡器应用空间图样成分分析的方法对苯丙胺类毒品甲基安非他明(MA)和4,5-亚甲基二氧苯丙胺(MDA)进行了成像研究^[5,6];B. Fischer 等报道了啡、可卡因的太赫兹吸收光谱^[7];在国内,首都师范大学物理系太赫兹实验室利用太赫兹时域光谱和成像技术对系列毒品进行了研究,并得到了一定的研究成果^[1~4,8~10]。

在利用太赫兹光谱和成像技术对毒品识别和检测的基础上,用神经网络识别技术已经取得了一些的成果。例如,多层感知器(MLP)神经网络已经用于识别炸药的光谱^[11,12],误差逆传播(BP)神经网络和自组织特征映射(SOM)神经网络已经用于识别毒品太赫兹吸收光谱^[9,10],这些神经网络共同的缺点是速度太慢、适用性差。因此需要一种方法来弥补这些缺陷,本文提出利用径向基(RBF)神经网络对毒品太赫兹光谱的识别。这种识别方法在一定程度上可以弥补这两种方法的缺陷,提高了光谱的识别速度。

RBF神经网络是一种有教师的神经网络,它具有结构自适应确定、输出与初始值无关等特点。而且这种网络是局部逼近的,训练速度快,对光谱的训练和识别几乎是实时的。把实验测到的11种毒品的太赫兹吸收光谱归一化后,用110个光谱训练RBF网络,训练完毕后让RBF网络识别不同时间测得的27个光谱,RBF网络对这些光谱进行了成功的识别。

2 实验

用RBF神经网络对11种常见毒品在不同时间测得的太赫兹光谱(每个样品各10次)进行了训练,并用不同时间测得的光谱来识别,识别率达到96%,这11种样品分别为盐酸安眠酮、盐酸巴比妥、盐酸苯巴比妥、盐酸杜冷丁、盐酸二氢埃托菲、盐酸甲基麻黄碱、盐酸可卡因、盐酸麻黄碱、盐酸美沙酮、盐酸伪麻黄碱和盐酸罂粟碱。11种毒品的纯度均为

99%以上,先烘干后再压成厚度为0.6~1.0 mm,直径为10.0 mm的圆片。然后把压好的片置于太赫兹装置中,分别在不同时间进行测量,得到这11种样品在0.2~2.6 THz范围内的光谱。每个光谱的获取时间为3 min。每一种样品测量了12~13次,一共得到了137个太赫兹光谱。

实验采用透射式太赫兹产生装置如图1所示,所用激光器是光谱物理公司的自锁模可调谐式钛蓝宝石激光器,激光中心波长为810 nm,脉宽为100 fs,重复频率为82 MHz,激光输出功率为980 mW。激光器产生的飞秒激光脉冲经过分光镜(BS)分为两束:透过BS的一束为抽运光,经过斩波器和延迟器(由反射镜RM₂、RM₃组成)后,再经反射并经准直后通过凸透镜L₁会聚于发射晶体InAs上,从而激发出太赫兹电磁波,太赫兹波经过离轴抛物面镜PM₁~PM₄准直聚焦于电光探测晶体上;经BS反射的一束作为探测光,经过一系列反射镜RM₆~RM₁₁和凸透镜L₂后经偏振片P打在高阻硅片(Si)上,经反射打在电光探测晶体ZnTe上,在电光晶体上与载有样品信息的太赫兹波相遇,再通过λ/4波片,由渥拉斯顿棱镜(PBS)分成偏振方向相互垂直的两束光,最后由差分探测器通过测量两偏振分量的差异而解调出太赫兹信号,通过计算机进行数据采集,进而得到样品信息。实验时,把样品放在PM₂焦点上;同时为了减少空气中水份对太赫兹的吸收,提高实验精度,将有太赫兹存在的区域置于充有氮气的有机玻璃罩中,如图1虚线所示;湿度小于4%,实验时温度为21℃。图2是氮气中测得的时域光谱和傅里叶变换后的相应的频域谱,太赫兹脉冲直接通过氮气得到的光谱作为参考光,通过毒品后得到的光谱作为样品的光谱。从图2可以看出样品的太

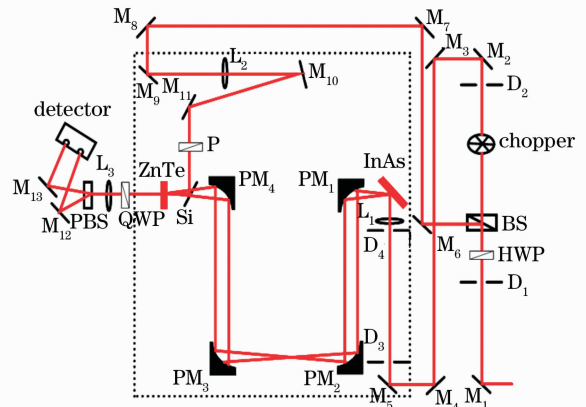


图1 太赫兹时域光谱系统示意图

Fig. 1 Schematic setup of the THz-TDS

赫兹光谱由于折射率而有一定延迟,而且振幅也由于样品的吸收而衰减。图3为盐酸罂粟碱的吸收光谱,RBF网络进行训练和识别的光谱为样品的太赫兹吸收光谱。

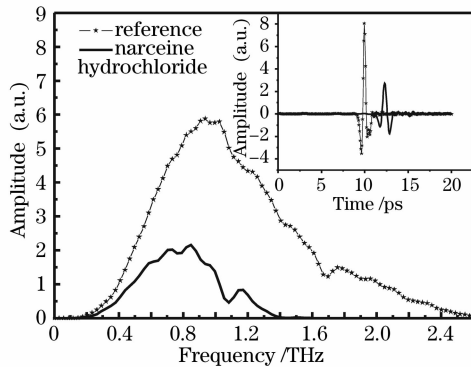


图2 参考光和盐酸罂粟碱的时域和频域谱

Fig.2 Time-domain spectra and the corresponding frequency-domain spectra of nitrogen and Narceine hydrochloride

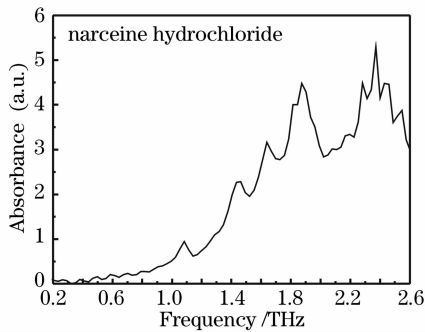


图3 盐酸罂粟碱的吸收光谱

Fig.3 Absorption spectra of narceine hydrochloride

3 人工神经网络

3.1 人工神经网络的介绍及分类

人工神经网络(ANN)由一些处理单元组成,这些处理单元叫神经元。人工神经网络是由大量神经元相互连接而成的网络,它模拟生物体中神经网络的某些结构和功能,如学习、控制和识别等。

人工神经网络可以分为两种:一种是无导师的,自组织神经网络是一种典型的无导师神经网络。这种网络没有预知的正确模式,通过对事物的反复观察、分析比较,自行揭示事物的内在规律,并对具有共同特征的事物进行正确归类,这种归类方法叫聚类。该方法可以对没有预期答案的问题做出归类并进行识别,但是这种方法的训练时间很长。另一种神经网络是有导师的。该方法采用有导师学习规则的神经网络要求对所学习的样本给出正确答案,以

便网络据此判断输出的误差,根据误差的大小改进自身的权值,以提高正确解决问题的能力。误差逆传播(BP)神经网络是一种有导师的神经网络,它的权值和阈值都是通过学习得到的。虽然BP网络有许多优点,但也存在不足。例如,BP网络隐含层神经元个数需要人为确定,而神经元个数的确定是一个比较复杂的问题,确定个数的合理与否,将直接影响到识别的结果;网络的学习和记忆具有不稳定性,使得计算结果不稳定;BP网络在训练过程中收敛速度慢、对初始条件敏感、容易陷入局部最小值等。

3.2 径向基神经网络

RBF神经网络是一种性能良好的前馈神经网络,它具有很强的生物背景和以任意精度逼近任意非线性函数的能力,是局部逼近神经网络,即对网络输入空间的某个局部区域只有少数几个连接权影响网络的输出,所以这种网络有学习速度快的优点,这一点对于实时性的要求来说至关重要。用于毒品光谱识别的神经网络是三层的RBF网络,即输入层、隐含层和输出层。输出层是线性函数,隐含层采用RBF函数作为激励函数,该RBF函数一般为高斯函数。隐含层每个神经元与输入层相连的权值向量为 w_{1j} 和输入向量 x^q (第 q 个输入向量)之间的距离乘上阈值 b_{1j} 作为本身的输入,如图4所示。

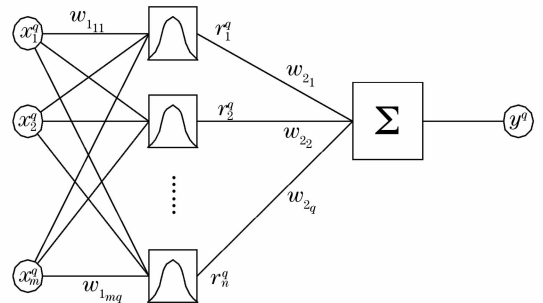


图4 径向基网络结构

Fig.4 Schematic representation of the three-layered radial basis function neural network

对应输入层第 q 个输入产生的隐含层第 i 个神经元的输入为 k_i^q

$$k_i^q = \sqrt{\sum_j (\omega_{1ji} - x_j^q)^2} \times b_{1i},$$

隐含层输出

$$r_i^q = \exp[-(k_i^q)^2] = \exp\left\{-\left[\sqrt{\sum_j (\omega_{1ji} - x_j^q)^2} \times b_{1i}\right]^2\right\} = \exp[-(\|\omega_{1j} - x^q\| \times b_{1i})^2],$$

输出为各隐含层神经元输出的加权求和,激励函数

采用纯线性函数,对应输入层第 q 个输入产生的输出层神经元输出 y^q 为

$$y^q = \sum_{i=1}^n r_i^q \times \omega_{2_i},$$

RBF 神经网络的训练过程分两步:第一步为无教师学习,确定训练输入层和隐含层之间的权值 ω_1 ;第二步为有教师学习,确定训练隐含层和输出层之间的权值 ω_2 。在训练前,需要提供输入向量 \mathbf{X} 、对应目标向量 \mathbf{T} ,与 RBF 函数分布密度 SPREAD。SPREAD 的大小实际上反映了输出对输入的响应宽度。在训练过程中需要用不同的 SPREAD 值进行尝试,以确定一个最优值。SPREAD 的大小不但影响网络的逼近精度,而且还影响着网络的预测精度训练的目的在于求取两层的最终权值 ω_1, ω_2 和最终阈值 b_1, b_2 (当隐含层神经元数等于输入矢量数的时候,取 $b_2=0$)。

RBF 网络在训练过程中,隐含层数量的确定是一个关键的问题,传统的做法是使其与输入向量的元素相等。但是,在输入矢量很多时,过多的隐含层神经元是很难让人接受的。为此,提出了一种改进的方法,基本原理是从 0 个神经元开始训练,通过检查输出误差使网络自动增加神经元,每次循环使用,使网络产生的最大误差所对应的输入向量作为权值向量 ω_{1_i} ,产生一个新的隐含层神经元,然后检查网络的误差,重复此过程直到达到误差要求或最大隐含层神经元数为止。由此可见,RBF 函数具有结构自适应确定、输出与初始权值无关等特点^[12]。

4 径向基网络的识别结果

把 11 种毒品的 110 个太赫兹吸收光谱(每种样品 10 个谱)经过归一化处理后,作为训练谱,输入到 RBF 神经网络,RBF 函数的分布密度 SPREAD 为 5,其训练次数是自适应确定的,通过自动改变神经元的数目来训练网络^[14]。网络训练好后,把不同时间测得的 27 个光谱经归一化后输入 RBF 网络进行识别,识别正确率达到了 96%,表 1 列出了网络识别正确数目占每种样品总数的比例,即样品的正确识别率。从表 1 可以看出,前 5 种样品用于网络识别的各有 3 个光谱,后 6 种用于识别的各有 2 个光谱,RBF 网络只对杜冷丁的 1 个光谱的识别是错误的。网络把杜冷丁识别成盐酸二氢埃托菲。这可能与样品本身的吸收特性有关系,也可能与网络的选择以及训练有一定的关系。结果表明,利用 RBF 人工神经网络对毒品太赫兹光谱的识别是有效的,利

用这种方法可以判定未知毒品的种类。

表 1 径向基网络的识别结果

Table 1 Result of identification using well-trained RBF neural network

Corresponding drug	Positive identification rate
Quaalude	3/3
Barbital	3/3
Phenobarbital	3/3
Pethidine	2/3
Dihydroetorphine hydrochloride	3/3
Methylephedrine	2/2
Cocaine	2/2
Ephedrine	2/2
Methadone	2/2
Pseudoephedrine	2/2
Narceine	2/2

5 结 论

在对 11 种毒品的太赫兹光谱进行归一化处理,利用 RBF 神经网络对 11 种毒品在不同时间测得的太赫兹光谱进行训练,然后利用训练好的 RBF 神经网络成功的识别了另一次实验测得的太赫兹光谱,识别正确率达到了 96%。RBF 神经网络方法克服了 BP 网络的一些不足,具有结构自适应确定,输出与初始值无关等特点。在 Matlab 中计算时,它不需要人为确定隐含层神经元个数,计算结果也很稳定,而且训练和识别速度快,这些优点使得 RBF 网络在进行训练和识别时有一定的适用价值。研究表明,用 RBF 神经网络可以实现对不同种类毒品的识别和鉴定,为太赫兹时域光谱技术用于毒品的检测和识别提供了一种快速有效地方法。

参 考 文 献

- 1 Sun Jinhai, Shen Jingling, Liang Laishun *et al.*. Experimental investigation on terahertz spectra of amphetamine type stimulants [J]. *Chin. Phys. Lett.*, 2005, **22**(12): 3176~3178
- 2 Li Ning, Shen Jingling, Sun Jinghai *et al.*. Study on the THz spectrum of methamphetamine [J]. *Opt. Express*, 2005, **13**(18): 6750~6755
- 3 Wang Guangqin, Shen Jingling, Jia Yan. Vibrational spectra of ketamine hydrochloride and 3, 4-methylenedioxyamphetamine in terahertz range[J]. *J. Appl. Phys.*, 2007, **102**(1): 013106
- 4 Lu Meihong, Shen Jingling, Li Ning *et al.*. Detection and identification of illicit drugs using terahertz imaging [J]. *J. Appl. Phys.*, 2006, **100**(10): 103104
- 5 Kodo Kawase, Yuichi Ogawa, Yuuki Watanabe *et al.*. Non-destructive terahertz imaging of illicit drugs using spectral fingerprints[J]. *Opt. Express*, 2003, **11**(20): 2549~2554
- 6 Kodo Kawase. Terahertz imaging for drug detection and largescale integrated circuit inspection[J]. *Optics and Photonics News*, 2004, **15**(10): 35~39

- 7 B. Fischer, M. Hoffmann, H. Helm *et al.*. Chemical recognition in terahertz time-domain spectroscopy and imaging [J]. *Semicond. Sci. Technol.*, 2005, **20**(7): S246~S253
- 8 Sun Jinhai, Zhao Guozhong, Zhang Liangliang *et al.*. Effect of applied electric and magnetic field on THz radiation[J]. *Chinese J. Lasers*, 2005, **32**(2): 192~195
孙金海, 赵国忠, 张亮亮 等. 外加电场和磁场对太赫兹辐射产生的影响[J]. *中国激光*, 2005, **32**(2): 192~195
- 9 Jia Yan, Chen Sijia, Li Ning *et al.*. Identification of Terahertz absorption spectra of illicit drugs using back propagation neural networks[J]. *Chinese J. Lasers*, 2007, **34**(5): 719~722
贾 燕, 陈思嘉, 李 宁 等. 利用误差逆传播神经网络法识别几种毒品的太赫兹光谱[J]. *中国激光*, 2007, **34**(5): 719~722
- 10 Liang Meiyun, Shen Jingling, Wang Guangqin. Identification of illicit drugs by using SOM neural networks[J]. *J. Phys. D: Appl. Phys.*, 2008, **41**(13): 135306
- 11 Oliveira F, Barat R, Shulkin B *et al.*. Neural network analysis of Terahertz spectra of explosives and bio-agents[C]. *SPIE*, 2003, **5070**: 60
- 12 Oliveira F, Barat R, Shulkin B *et al.*. Analysis of Terahertz spectral images of explosives and bio-agents using trained neural networks[C]. *SPIE*, 2004, **5411**: 45
- 13 FEI SI Science and Technology REsearch Center. The Theory of Neural Network and Realization by Matlab 7 [M]. Beijing: Electronics Industry Press, 2006. 116~127
飞思科技产品研发中心. 神经网络理论与 MATLAB 7 的实现 [M]. 北京: 电子工业出版社, 2006. 116~127
- 14 Han Liqun. The Theory, Design and Application of Artificial Neural Networks[M]. Beijing: Chemical Industry Press, 2007. 164~177
韩力群. 人工神经网络理论、设计及应用[M]. 北京: 化学工业出版社, 2007. 164~177