

基于人工神经网络的傅里叶变换中红外光谱法 对食用油油烟种类识别研究

叶树彬^{1,2}, 徐亮^{1*}, 李亚凯¹, 刘建国¹, 刘文清¹

1. 中国科学院安徽光学精密机械研究所, 中国科学院环境光学与技术重点实验室, 安徽 合肥 230031
2. 中国科学技术大学, 安徽 合肥 230026

摘要 随着餐饮业的发展, 餐饮烟气已经成为某些城市三大空气污染源之一。由于餐饮烟气对人体健康威胁很大, 近年来对餐饮烟气的研究愈来愈热。餐饮烟气中包含有大量食用油加热过程中裂解而产生的不饱和烃类, 危害着人类健康。不同食用油裂解出来的成分以及含量有所不同, 通过构建一定的分类识别数学模型, 从而实现食用油分类识别。采用自主研发的傅里叶变换红外光谱仪, 采集了不同食用油油烟烟气红外光谱数据。同时构建了主成分分析(PCA)分别结合概率神经网络(PNN)以及误差反向传播神经网络(BPANN)的分类识别算法。将两种分类识别算法对不同食用油油烟烟气的傅里叶变换红外光谱数据进行分析。通过样本数据对数学模型进行训练, 将训练好的数学模型对未知光谱数据进行分析, 来确定产生油烟烟气的食用油种类。实验结果表明, 两种算法都能对不同的油烟种类进行较好地分类识别。在全波段识别时, 识别率分别达到 90.25% 和 97.0%。通过对烟气光谱数据的吸收波段进行分析, 提取大气窗口并且具有较强挥发性有机物(VOCs)吸收特征的波段(1 300~700 cm^{-1} 以及 3 000~2 600 cm^{-1}); 将吸光度数据分成两个分离的吸收波段, 两种算法在 3 000~2 600 cm^{-1} 波段都有较好的识别效果, PCA-PNN 算法识别率为 90.25%, PCA-BPANN 算法识别率为 92.25%。可见, 两种神经网络算法都能有效对食用油种类进行识别。

关键词 FTIR; 人工神经网络; 油烟识别; 主成分分析

中图分类号: O433.4 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2017)03-0749-06

引言

随着餐饮行业的发展, 餐饮烟气已经成为某些城市三大空气污染源之一^[1]。研究表明, 在中国地区, 餐饮烟气是肺癌的重要原因之一^[2], 尤其是中国女性受到的健康威胁更大^[3]。因此近年来, 对餐饮烟气的研究^[4-5], 尤其是餐饮烟气对环境以及人体的危害的研究愈来愈热^[3, 6-9]。同时, 人们对食用油种类的关注, 尤其是对食用转基因油的争议越来越多^[10]。

最近几年, 傅里叶变换红外光谱法(FTIR)得到了飞速发展^[11]并在大气、以及组分定性^[12]、定量分析^[13]等方面得到了很广泛的应用研究。因此, 本工作试图通过结合近年来

发展的傅里叶变换红外光谱(FTIR)技术与相关的分类识别算法, 对食用油进行识别。

2012年, Aderval S. Lunna 等采用近红外光谱, 通过化学计量学(SVM-DA, PLS-DA)模型对转基因与非转基因大豆油进行快速识别^[14]; 2015年, Zhou 等采用 SRC、GS3VM 方法对食用油以及地沟油进行分类识别^[15-16]。传统方法直接采取不同种类油的红外光谱, 通过建模进行分类分析。由于采样过程操作复杂, 效率较低以及复杂的预处理操作, 不能进行在线实时分析。其次由于餐饮烟气组分多而复杂^[17], 仅低分子量的非甲烷总烃类(NMHC)就多达几十种^[5], 无法采用传统线性及非线性拟合算法对其定性、定量分析。

针对上述问题, 提出直接采集食用油经过加热过后油烟的 FTIR 光谱对食用油种类进行实时在线识别。近年来, 在

收稿日期: 2016-03-04, 修订日期: 2016-07-18

基金项目: 中国科学院战略性先导科技专项(XDB05050300), 国家重大科学仪器设备开发专项(2013YQ22064302), 中国科学院战略性先导科技专项(XDB05040500), 工业区 VOCs 排放通量遥测方法研究项目(41405029)资助

作者简介: 叶树彬, 1990年生, 中国科学院安徽光学精密机械研究所博士研究生 e-mail: yeshubin9589@163.com

* 通讯联系人 e-mail: xuliang@aiofm.ac.cn

光谱学领域发展了多种分类、识别方法^[18-19],而人工神经网络算法应用于模式识别较多^[19]。2014年,Vassilis等将人工神经网络算法应用于傅里叶变换红外光谱,实现了对肉质损坏的分类识别^[19],取得了很好的识别效果。为了比较不同识别算法模型对食用油油烟的识别效果,结合PCA算法以及人工神经网络(ANN)算法,分别构建了主成分分析-误差反向传播神经网络分类识别算法(PCA-BPANN)以及主成分分析概率人工神经网络算法(PCA-PANN)。通过对餐饮烟气红外光谱吸收特征分析,从全波段吸光度光谱数据中分离出两个主要影响波段,对食用油进行分类识别。

1 实验部分

1.1 仪器及参数

在实验室条件下搭建了一个模拟厨房。实验仪器为课题组研发的傅里叶变换红外光谱仪。如图1,装置主要由傅里叶变换红外光谱仪,经过改装的多次反射池,温控系统,抽气泵,气体导管,计算机以及配套的软件系统组成。在多次反射池中安装了一个温度传感器,将加热保温层以及温度传感器与一台高精度温控系统连接。测量时将温度设置为180℃。



图1 实验室
Fig. 1 Laboratory

为了保证数据不受人为操作的影响,整个测量过程严格按照表1所示的操作过程进行测量。室内温度保持恒温(24℃)。每次实验测量的样本食用油量为100 mL。通过固定液化气出气口气压和加热过程中油的温度,以保持数据测量条件的一致性,实验时严格控制液化气出气口气压,以保证液化气参与燃烧的气体体积。测量一条干涉图、光谱图的时间在10 s左右,即整个系统包括换气的时间内需要超过160 s的时间来测量一条光谱。为了区分食用油加热过程中所产生油烟的情况,实验采集了每种食用油在加热0/30/60/90/120/150 s时所产生的油烟的情况,单次测量10条光谱。

1.2 数据获取

实验一共测了4组不同食用油油烟的数据,如表2为四种食用油的基本信息,每组油烟数据在不同燃烧时间下测量了10条以上光谱,单种食用油油烟光谱超过50条。实际测

量光谱有效范围为4 000~500 cm⁻¹的中红外光谱部分,光谱分辨率为1 cm⁻¹,单条光谱的数据点数达到7 255个。如图2为测量的一条油烟1的典型光谱。

表1 实验操作过程表

Table 1 Experimental procedures

步骤	抽气泵	温度控制 (180℃)	抽油烟机	时间
初始状态	关	开	开	—
量取100 mL样品油	关	开	开	—
将样品放入锅中	关	开	开	—
加热样品油/s	关	开	开	0/30/60/90/120/150
抽取油烟	开	开	开	90 s
稳定油烟	关	开	开	大于60 s
测量干涉图、光谱图	关	开	开	测量耗时

结束测量,返回初始状态
返回初始测量,如果是新样品,抽气10 min以上,并从初始状态开始测量

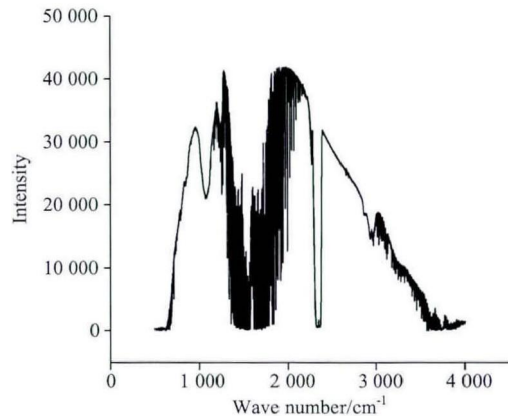


图2 典型光谱

Fig. 2 Typical spectra

表2 四种食用油信息

Table 2 Information of four kinds of edible oil

油烟种类	名称	加工原料
种类1	金龙鱼营养强化菜籽油	转基因油菜籽
种类2	福临门食用调和油	转基因大豆及转基因油菜籽
种类3	福临门葵花籽油	非转基因葵花籽
种类4	福临门一级大豆油	转基因大豆

中红外部分可用来反演的大气窗口有三个,分别为3 000~2 400, 2 250~2 000以及1 300~700 cm⁻¹三个波段。在实验室条件下可以采用这三个波段来计算光谱的100%线,从而计算信噪比。表3给出了采用2 200~2 100以及2 600~2 500 cm⁻¹两个波段计算四种油烟组分信噪比剔除异常值后的平均结果。

由表3可见,油烟1的信噪比在波段1、波段2都远远高于油烟2,3和4。

表3 不同波段下不同油烟组分的平均信噪比

Table 3 The average SNR of different fume components in different wavebands

信噪比(SNR)	油烟1	油烟2	油烟3	油烟4
波段1(2 200~2 100)	2 503.08	1 015.66	908.292	1 010.1
波段2(2 600~2 500)	4 903.18	2 901.19	2 706.57	3 145.03

2 分类识别算法

2.1 主成分提取

直接将光谱数据作为神经网络识别算法的输入会造成光谱数据冗余比较大。同时,对于4组分识别而言,数据维度远远超过所需的识别因子,这会增加神经网络算法的学习时间。因此,首先对获取的样本数据进行主成分提取(PCA)。提取有效的主成分,能显著降低数据维数,剔除冗余信息,而又不丢失有效信息。典型的提取结果能将7 255维的光谱数据降到15维。表4给出了前面几个特征值、贡献率以及前面几个主成分的累积贡献率。

从表4可以看出,第一个主成分对原始光谱数据的解释贡献率为67.623 7%。累积贡献率大于90%的主成分个数为15个,也即,能将7 255维数据降成15维用于识别的有效数据。

表4 前15个主成分特征值、贡献率以及累积贡献率

Table 4 The first 15 principal component characteristic value, the contribution rate and cumulative contribution rate

主成分	特征值	贡献率/%	累积贡献率/%
1	232.077 9	67.623 7	67.623 7
2	22.004 6	6.411 8	74.035 4
3	15.075 4	4.392 7	78.428 2
4	6.032 0	1.757 6	80.185 8
5	5.380 4	1.567 8	81.753 6
6	4.551 5	1.326 2	83.079 8
7	4.213 5	1.227 8	84.307 5
8	3.595 6	1.047 7	85.355 2
9	3.303 7	0.962 6	86.317 9
10	3.017 0	0.879 1	87.197 0
11	2.509 2	0.731 1	87.928 1
12	2.415 0	0.703 7	88.631 8
13	2.185 8	0.636 9	89.268 7
14	1.467 6	0.427 6	89.696 4
15	1.428 8	0.416 3	90.112 7

2.2 神经网络识别算法

神经网络算法广泛应用于模式识别,信号处理,图像处理等领域,此处分别构建BP-ANN神经网络算法以及概率神经网络算法来对油烟种类进行识别。

2.2.1 误差反向传播神经网络算法

构建的BP-ANN神经网络算法共三层,输入层、输出层以及一个隐层。一般来说,输入层节点数取决于输入数据维数,为一个固定量。输出层的数目为分类数目,本工作

为4层(四种油烟组分)。而隐层的节点数基本上就决定了所构建神经网络的性能。隐层节点数过多会消耗更多的学习时间,同时可能出现过拟合现象,而隐层数目选择过少可能会出现欠拟合现象。一般来说,隐层节点数 Q 由公式(1)计算得到

$$Q = \lfloor \sqrt{M+N} + k \rfloor \quad (1)$$

式中, M 为输入节点数, N 为输出节点数, k 为一个0~10的常数。此处 k 取10。输入层、输出层激发函数分别设计为S型激发函数。

得到训练结果的神经网络时,最后添加一层竞争输出层,即可以得到想要的人工神经网络。为了能够快速学习,并使得神经网络的训练结果不对初值过于敏感,采用Levenberg-Marquardt学习算法,对神经网络进行学习。它具有收敛速度快,消耗内存低的特点。

2.2.2 概率神经网络算法

为了实现对比,同时构建了另外一种模式识别神经网络算法,即概率神经网络算法(PNN)。概率神经网络算法的主要思想是利用贝叶斯决策规则,在多维输入空间内分离决策空间。它是以Parzen窗口函数为激发函数的一种前馈网络模型,同时吸收了径向基神经网络与经典的概率密度估计原理的优点,与传统的前馈网络模型相比,在分类方面尤其具有较为显著的优势。

PNN概率神经网络一共有4层,包括输入层,样本层,求和层以及竞争输出层。在输入层中,网络计算输入向量与所有训练样本向量之间的距离,输入层神经元的个数即为特征向量维数,也即2.1节提到的主成分个数。样本层的激发函数为高斯函数:

$$\theta_i = \exp\left(-\sum \|x - c_i\|^2 / 2\sigma_i^2\right) \quad (2)$$

式中, c_i 为径向基函数的中心, σ_i 表示特性函数第 i 个分类对弈的开关参数,样本层神经元个数为训练样本的个数;求和层将样本层的输出按类相加,相当于 N 个加法器;竞争层输出分类结果,输出结果中只有一个为1,其余都为0,其中,概率值最大的哪一类输出结果为1,因此,神经元个数设计为1个,是一个4维的输出向量,与BP-ANN输出层经过竞争输出后的结构相同。

3 结果与讨论

3.1 识别结果

首先将不同加热时间的油烟数据,作为样本全集,每个组分中有50条以上的光谱数据。通过一定的算法,每种组分随机提取10个样本作为验证算法的样本集合,其余的作为算法学习样本集。在对学习样本进行PCA之前,首先对样本数据进行一定的预处理,得到样本光谱的吸光度数据。将得到的吸光度谱样本数据提取主成分,降低数据的维度,构建适合于ANN算法学习的样本集合。如图3(a)和(b)分别为BP-ANN和PNN学习算法学习的结果,可见,两种分类算法都能对训练样本进行准确无误地分类仿真。

将训练好的BP-ANN与PNN网络对4组油烟共40条

样本进行识别, 识别结果如表 5。

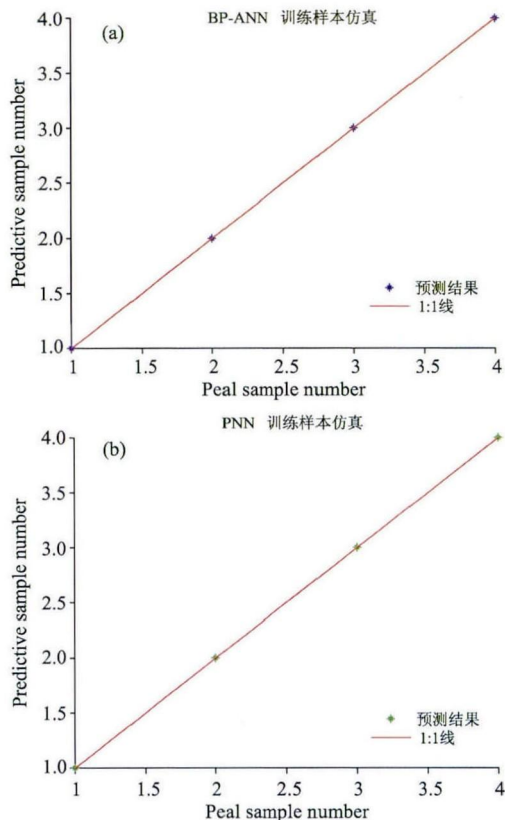


图 3 人工神经网络算法训练仿真结果

(a); BP-ANN 模型结果; (b); PNN 网络模型结果

Fig. 3 Simulation results of artificial neural network trainings

(a); Result of BP-ANN model; (b); Result of PNN model

表 5 两种人工神经网络算法的识别结果比较(1)

Table 5 Comparison of two artificial neural network algorithms (1)

算法结果	组分 1		组分 2		组分 3		组分 4	
	BPANN	PNN	BPANN	PNN	BPANN	PNN	BPANN	PNN
验证样本数目	10	10	10	10	10	10	10	10
正确识别数目	10	10	10	6	10	9	10	10
错误识别数目	0	0	0	4	0	1	10	10
识别率/%	100	100	100	60	100	90	100	100

由表 5 可知, 两种算法都能对四种组分的油烟烟气进行有效识别。其中 BPANN 算法能对四种组分进行完全识别, 识别率为 100%。而 PNN 算法能对组分 1 和组分 4 完全识别, 组分 2 的识别率较低, 只有 60%, 对组分 3 的识别正确率为 90%。

3.2 两种识别算法性能比较

上一节给出了识别结果, 其中 BPANN 算法的综合识别率为 100%, 而 PNN 的综合识别率则只有 87.5%。然而, 这并不能说明 BPANN 识别算法性能完全优于 PNN 识别算法。因为总的样本集不变, 而验证样本集是每种组分随机选择 10 个样本而构建的。同时, 两种算法在耗时上也具有很大差

异。表 6 给出了两种算法实验 10 次的实验结果。

表 6 两种人工神经网络算法的识别结果比较(2)

Table 6 Comparison of two artificial neural network algorithms (2)

实验次数	BPANN			PNN		
	正确识别	识别率/%	耗时/s	正确识别	识别率/%	耗时/s
1	39	97.5	1.308 489	37	92.5	0.794 781
2	38	95.0	1.269 070	37	92.5	0.803 191
3	40	100	1.212 789	35	87.5	0.838 228
4	39	97.5	1.379 311	36	90.0	0.857 740
5	39	97.5	1.169 148	35	87.5	0.824 773
6	40	100	1.126 398	38	95.0	0.873 853
7	38	95.0	1.230 982	37	92.5	0.833 884
8	39	97.5	1.140 576	34	85	0.769 117
9	38	95.0	1.438 556	37	92.5	0.892 013
10	38	95.0	1.227 166	35	87.5	0.839 343

表 6 清晰表明, 在识别率上 BPANN 算法平均正确识别率为 97.0%, 高于 PNN 算法 90.25% 的识别率, 而 BPANN 算法的平均学习耗时为 1.250 2 s, 要高于 PNN 算法的 0.832 7 s。BPANN 算法主要在迭代学习过程中耗时较长, 典型的学习次数为 39 次, 学习的均方根误差 (RMSE) 小于 1.0×10^{-20} 。PNN 神经网络以指数函数代替了神经网络中的 S 型激发函数, 从而, 构建能够计算非线性判别边界的神经网络, 该判定边界接近于贝叶斯最佳判定面。然而, 概率神经网络将贝叶斯分类决策理论引入到神经网络中时, 针对概率密度函数做了几个基本假设, 其中一条就是给分类的概率密度函数形态相同, 即为高斯分布。然而, 在试验中采样时, 分别采集不同油烟不同加热时间的光谱数据 10 条以上, 同一种加热时间下的光谱 PCA 特征值可能服从高斯分布, 而不同加热时间的光谱之间的 PCA 特征值并不一定服从高斯分布, 从而导致概率神经网络在识别率上显著差于 BPANN 识别算法, 当然基于贝叶斯分类决策理论的神经网络可能在样本数目上比 BPANN 具有更高的要求, 而本实验用来训练神经网络的单组份样本不足 40 个。如果在加热时间上也取随机值, 理论分析表明, PNN 算法在模式识别上要优于基于曲线拟合的 BPANN 构建的模式识别算法。

3.3 不同波段的识别性能比较

本文所构建的 PCA-BPANN 分类识别算法与 PCA-PNN 分类识别算法所处理的数据为吸光度 $A(\nu)$, 如式(3)

$$A(\nu) = \sum_{i=1}^N \sigma_i(\nu) c_i L \quad (3)$$

式中 $\sigma_i(\nu)$ 为第 i 种分子在波数 ν 处的吸光度, c_i 为该种组分分子的浓度, L 为吸收光程。如果对吸光度数据进行分析可以得知, 同种油烟的烟气光谱差别在 c_i , 也即油烟排放出不同组分的浓度差别, 而不同种类食用油烟的吸光度数据还有 $\sigma_i(\nu)$ 的差别。可见, PCA-ANN 与 PCA-PNN 算法主要是在提取不同种食用油烟吸光度光谱的 $\sigma_i(\nu)$ 特征差别, 以及忽略同种油烟在同种组分中 c_i 的差别。

实际上,餐饮油烟种类复杂,多达几百种。油烟中的VOCs的组分大致可以分为非甲烷碳氢化合物(NMHCs)、含氧挥发性有机物(OVOCs)以及卤代烃。从红外光谱学角度考察可知,NMHCs,OVOCs以及卤代烃的红外吸收特征主要集中在 $3\ 400\sim 2\ 600\ \text{cm}^{-1}$ 的官能团振动吸收对应区、 $2\ 000\sim 500\ \text{cm}^{-1}$ 的指纹吸收区,以及 $2\ 400\sim 2\ 200\ \text{cm}^{-1}$ 与—CN等相关的三键和累积双键伸缩振动区。考虑到中红外部分的三个大气窗口,采用 $1\ 300\sim 700$ 、 $3\ 000\sim 2\ 600\ \text{cm}^{-1}$ 以及以上两个波段联合波段进行识别分析。表7给出了两种算法在两个分离波段分别识别的结果以及两个波段数据联合识别的结果。从联合波段分析知,PCA-PNN以及PCA-BPANN平均识别率分别为81.5%和84.75%,比两种波段在全波段时的识别率稍低,全波段识别时两种算法的分别识别率为90.25%和97.0%。可见,两个波段的数据,占据了全波段数据用来分类识别时的大部分信息。

从表7可知,两个波段数据单独分析时,PCA-PNN分类识别算法与PCA-BPANN分类识别算法识别率相当。通过联合波段数据分析可知,PCA-BPANN算法识别率总体高于PCA-PNN算法识别率3个百分点。

表7 两种识别算法在不同波段下的识别率比较

Table 7 Comparison of recognition rate of two recognition algorithms in different

实验次数	联合波段		$1\ 300\sim 700\ \text{cm}^{-1}$		$3\ 000\sim 2\ 600\ \text{cm}^{-1}$	
	PNN	BPANN	PNN	BPANN	PNN	BPANN
1	80.0	85.0	75.0	80.0	90.0	92.5
2	82.5	85.0	75.0	77.5	90.0	90.0
3	80.0	85.0	77.5	77.5	92.5	92.5
4	82.5	82.5	77.5	80.0	87.5	95.0
5	80.0	82.5	72.5	80.0	87.5	90.0
6	85.0	92.5	80.0	85.0	90.0	90.0
7	80.0	82.5	72.5	82.5	87.5	90.0
8	85.0	87.5	75.0	82.5	92.5	95.0
9	80.0	82.5	75.0	77.5	90.0	92.5
10	80.0	82.5	82.5	77.5	95.0	95.0
均值	81.5	84.75	76.25	80.0	90.25	92.25

从两个波段两种算法识别率的比较发现,两种分类识别

算法识别率相差不大。同时,在 $3\ 000\sim 2\ 600\ \text{cm}^{-1}$ 波段,两种算法都表现出较好的识别性能。这种情况可能有两方面的原因。首先,与两个波段的吸光度数据特性有关。波段 $1\ 300\sim 700\ \text{cm}^{-1}$ 中包含有各种VOCs指纹吸收峰,更具有特异性,但具有较弱的吸收峰面积。而波段 $3\ 000\sim 2\ 600\ \text{cm}^{-1}$ 包含有VOCs官能团振动以及三键和累积双键的伸缩振动信息,具有较强的吸收峰面积。其次,与两种ANN算法性能有关。前文提到,基于贝叶斯分类决策理论的概率神经网络在处理具有高斯分布的概率密度函数的特征数据上具有优势,而试验中,油烟的光谱数据在时间维度上并不随机,因此PCA-PNN分类识别算法的识别性能总体比BPANN算法识别率低3个左右百分点。

4 结论

分别构建了PCA-PNN以及PCA-BPANN分类识别算法,实验结果表明,两种分类识别算法在全波段以及联合波段的光谱数据上都具有很好分类识别性能。

通过对两种算法性能的比较,在采用时间维度不随机的样本光谱数据库中建模时,PCA-BPANN分类识别算法相比PCA-PNN算法具有更高的识别率。通过对油烟组分红外光谱特性的分析,将全波段数据分为两个离散波段。将两种算法分别应用于不同波段进行分类识别,结果表明,PCA-BPANN算法总体识别率优于PCA-PNN算法识别率3个百分点。并且两种算法在 $3\ 000\sim 2\ 600\ \text{cm}^{-1}$ 波段,都有较高的识别率,分别达到90.25%(PCA-PNN)、92.25%(PCA-BPANN)。通过对两种算法的比较以及不同波段的识别效果的分析,表明该算法能有效对食用油烟种类进行分类识别。

结合傅里叶变换红外光谱仪,构建主成分分析分别结合两种神经网络算法的分类识别算法,可以直接将红外光谱吸光度数据用于油烟食用油分类识别,拓展了傅里叶变换红外光谱仪以及红外光谱的应用领域。下一步,将通过使用这种方法,在基于大规模样本条件下,用于地沟油等油烟的建模,从而可以通过油烟,在线识别食用油的种类。

References

- [1] CHEN Xue-zhang, YOU Jin(陈学章, 尤 今). Standard Science(标准科学), 2012, 1: 243.
- [2] Sophie Sun, Joan H, Schiller, et al. Nature Reviews Cancer, 2007, 7(10): 778.
- [3] Yin Zhihua, Cui Zhigang, Ren Yangwu, et al. Medical Oncology, 2014, 31: 114.
- [4] Rikke Bramming Jorgensen, Bo Strandberg, Ann Kristin Sjaastad, et al. Journal of Occupational and Environmental Hygiene, 2013, 10(3): 122.
- [5] ZHANG Chun-yang, MA Yong-ling(张春洋, 马永亮). Acta Scientiae Circumstantiae(环境科学学报), 2011, 38(8): 1768.
- [6] Young Shun-Chieh, Chang L W, Lee Hui-Ling, et al. Environmental and Molecular Mutagenesis, 2010, 51(4): 315.
- [7] Cao Jiyu, Ding Rui, Wang Yong, et al. Environmental Toxicology and Pharmacology, 2013, 36(2): 320.
- [8] Lai C H, Jaakkola J J K, Chuang C Y, et al. Journal of Exposure Science and Environmental Epidemiology, 2013, 23(1): 94.
- [9] LIANG Bin, WANG Yin-er, JIN Jia-jia, et al(梁 斌, 王寅儿, 金嘉佳, 等). Anhui Chemical Industry(安徽化工), 2011, 37(3): 59.
- [10] ZHONG Fu-ning, CHEN Xi, YE Xi-jun(钟甫宁, 陈 希, 叶锡君). China Economic Quarterly(经济学季刊), 2006, 5(4): 1311.

- [11] Granada E, Eguia P, Vilan J A. *Renewable Energy*, 2012, 41: 416.
- [12] Nicholas W Turner, Michael Cauchi, Elena V Piletska. *Biosensors and Bioelectronics*, 2009, 24(11): 3322.
- [13] Smith T E L, Wooster M J, Tattaris M, et al. *Atmos. Meas. Tech.*, 2011, 4(1): 97.
- [14] Aderval S Lunaa, Arnaldo P da Silva, Jéssica S A Pinhoa, et al. *Analytical Molecular and Biomolecular Spectroscopy*, 2013, 100(1): 115.
- [15] Zhou Yang, Liu Tiebing, Li Jinrong. *Analytical Methods*, 2015, 7(6): 2367.
- [16] Zhou Yang, Liu Tiebing, Li Jinrong. *Chemometrics and Intelligent Laboratory*, 2015, 143: 1.
- [17] Zhu X D, Wang K X, Zhu J L. *Journal of Agricultural and Food Chemistry*, 2001, 49(10): 4790.
- [18] Vassilis S Kodogiannisa, Eva Kontogiannib, John N Lygourasc. *Journal of Food Engineering*, 2014, 142: 118.
- [19] Liu F, Wang J. *Journal of Environmental Science and Health Part a-Toxic/Hazardous Substances & Environmental Engineering*, 2004, 39(6): 1525.

Study on Recognition of Cooking Oil Fume by Fourier Transform Infrared Spectroscopy Based on Artificial Neural Network

YE Shu-bin^{1,2}, XU Liang^{1*}, LI Ya-kai¹, LIU Jian-guo¹, LIU Wen-qing¹

1. Key Laboratory of Environmental Optics and Technology, Anhui Institute of Optics and Fine Mechanics, Chinese Academy of Sciences, Hefei 230031, China
2. University of Science and Technology of China, Hefei 230026, China

Abstract With the developing of catering trade, cooking oil fume has become one of the three major air pollution sources in some cities. In recent years, a lot of research on the cooking oil fume have been done for its high threaten to human health. The cooking oil fume contains a large amount of unsaturated hydrocarbons produced by pyrolysis of edible oil, which are harmful to human health. The characteristics of the composition and content of edible oil fumes produced by pyrolysis of different edible oil are different. For classification and identification of edible oil, two kinds of classification and identification mathematical model are constructed. The spectrum data of different edible oil fume are collected by Fourier transform infrared spectrometer which is independent research and development. At the same time, different classification algorithms of the principal component analysis (PCA) combining probabilistic neural network (PNN) and the error back propagation artificial neural network (BPANN) are constructed respectively. Two kinds of classification algorithms are used to analyze the Fourier transform infrared spectrum data of different cooking fume gas. The mathematical models are trained by the sample data, and the trained mathematical model are used to analyze the unknown spectral data to determine the type of edible oil. The experimental results show that the two algorithms can classify and identify different types of oil fume. In the whole band recognition, the recognition rate is 90.25% and 97% respectively. By analyzed spectral data of flue gas absorption band, spectrums of atmospheric window and the strong absorption feature bands of volatile organic compounds (VOCs) (from 1 300 to 700 cm^{-1} and from 3 000 to 2 600 cm^{-1}) were extracted. The absorbance data are divided into two parts with separated absorption band, and the two algorithms in 3 000~2 600 cm^{-1} band have better recognition rate. PCA-PNN algorithm recognition rate is 90.25% and PCA-BPANN algorithm recognition rate is 92.25%. Obviously, two kinds of artificial neural network algorithm combining principle component analysis respectively can effectively identify the types of edible oil fume.

Keywords FTIR; ANN; Oil identification; PCA

(Received Mar. 4, 2016; accepted Jul. 18, 2016)

* Corresponding author