结合径向基函数和KPCA的食用油太赫兹光谱特征提取方法

王卓薇[1]，罗鉴鹏[1]，李学识[2*]，程良伦[2]

1. 广东工业大学计算机学院，广东 广州 510006
2. 广东工业大学自动化学院，广东 广州 510006

摘 要 针对太赫兹光谱线性不可分的情况，提出结合径向基函数和核主成分分析(KPCA)的方法进行食用油太赫兹光谱特征提取。该方法所提取到的特征类内距离小，类间距离大，在大多数支持向量机(SVM)分类器可以建立准确的分类模型。太赫兹光谱是检测食用油种类和品质的一种重要手段，研究针对食用油太赫兹光谱的特征提取技术对于食用油种类和品质快速检测具有重要意义。虽然利用太赫兹光谱检测食用油种类和品质已经具备理论基础，但是如何准确提取食用油太赫兹光谱的特征，从而建立更加准确的分类模型依然是一个难点。目前研究人员常常采用化学计量学中的主成分分析法(PCA)提取特征，结合机器学习的方法建立物质分类模型。然而，食用油的太赫兹光谱的线性可分情况在不同频段有不同的特性。当食用油的太赫兹光谱线性可分时，使用PCA提取特征是可行的，容易建立准确的分类模型。但是，当食用油的太赫兹光谱线性不可分时，使用PCA提取到的特征往往不够准确，需要选择合适的分类器去建立准确的分类模型。结合径向基函数和KPCA的特征提取方法通过径向基函数将线性空间不可分的太赫兹光谱数据映射到径向基空间，然后使用KPCA提取特征，最终实现特征线性可分，从而可以建立更加准确的分类模型。实验首先使用滑动窗口平均滤波算法对3种食用油太赫兹光谱数据进行滤波处理，接着使用径向基函数对太赫兹光谱进行非线性映射，然后采用KPCA进行数据降维，最后用支持向量机对食用油建立分类模型，验证特征提取效果。类间可分性计算结果表明，该方法所提取的特征类内距离更小，类间距离更大，整体上特征提取效果优于PCA和KPCA。基于不同内核的SVM模型上进行分类验证的实验结果表明，在PCA和KPCA提取的特征在一些分类模型上无法准确区分食用油种类的情况下，该工作特征提取方法在各种内核的SVM模型上均能准确区分食用油种类。所提出的方法用于食用油太赫兹光谱特征提取有更好的效果，在食用油品质检测与分析方面具有良好的应用价值。

关键词 太赫兹光谱；径向基函数；核主成分分析；支持向量机

中图分类号：O433 **文献标识码**：A **DOI**：10.3964/j.issn.1000-0593(2020)02-0391-06

引 言

太赫兹光谱在食品检测方面的应用研究越来越多，食用油检测是其中的一个重要部分。Liu等[1]比较了偏最小二乘－支持向量机(LS-SVM)、BP神经网络(BPNN)、随机森林(RF)、主成分分析(PCA)这些不同的化学计量学方法判断橄榄油产地的效果，验证了化学计量学在太赫兹光谱定性分析中的重要作用。聂美彤等[2]使用衰减全反射式太赫兹光谱研究了大豆油、核桃油和葡萄籽油的光谱特性，证明了太赫兹光谱在食用油定性分析方面具备理论基础。李利龙等[3]使用太赫兹光谱对7种植物油和2种调和油进行研究，结果表明：脂类有机大分子对THz辐射具有差异性吸收，具备在THz波段的识别基础，可通过THz技术进行鉴别和定性分析。Yin等提出了一种通过使用太赫兹(THz)光谱结合遗传算法(GA)和偏最小二乘判别分析(PLS-DA)来区分食用油的方法。结果表明，GA-PLS-DA模型具有较小的预测均方根误差(RESEP)，较大的预测相关系数(R_p)，以及比其他模型

收稿日期：2018-12-27，**修订日期**：2019-04-08

基金项目：国家自然科学基金项目(61505035，61672168，61672172)，国家重点研发计划(2016YFB1200402-019)，国家自然科学基金广东联合基金项目(U1801263)，广东省信息物理融合系统重点实验室项目(2016B030301008)，智能制造信息物理融合系统集成技术国家地方联合工程研究中心，高分辨率对地观测系统重大专项(83-Y40G33-9001-18/20)资助

作者简介：王卓薇，女，1985年生，广东工业大学计算机学院副教授 e-mail：whuwzw@gdut.cn

*通讯联系人 e-mail：lixueshi@gdut.edu.cn

更高的分类精度。他们得到 THz 光谱与化学计量学相结合是区分各种食用油的有效方法的结论[4]。

在目前的材料定性研究中，研究人员主要是通过提取太赫兹光谱特征结合支持向量机、人工神经网络等机器学习方法进行定性识别。陈涛[5]提出了一种基于 PCA 和模糊识别方法的生物分子太赫兹光谱识别方法，并采用多种典型糖类和氨基酸生物分子的太赫兹光谱作为实验介质证明所提方法的可行性和有效性。胡晓华等[6]采用 PCA 对 3 个产地的咖啡进行太赫兹光谱分析，构造了基于粒子群参数寻优的支持向量机鉴别模型，模型对不同产地咖啡样品的综合识别率达到 95%。张文涛等[7]在采用太赫兹时域光谱技术对转基因大豆油光谱检测的基础上结合 PCA 及支持向量机，构建 PCA-SVM 模型对转基因大豆油进行鉴别。Liu 等[8]使用连续投影算法结合加权线性判别法实现了各种类型转基因油的区分。在上述研究中，首先采用 PCA 提取太赫兹光谱特征然后采用非线性的分类器进行分类。但是，PCA 这类线性降维方法不适合对太赫兹光谱数据进行特征提取。由于物质的太赫兹光谱数据各维度呈现非线性，尤其是当不同物质的太赫兹光谱曲线整体非常相似时，线性处理方法易产生较大误差。

核主成分分析(KPCA)是一种非线性研究方法，通过核函数完成非线性映射的过程，最终实现对非线性数据降维同时最大程度保留原始数据的信息。KPCA 在捕捉数据的非线性特征比较有效。KPCA 应用在故障检测等场合比较多。Hu 等[9]提出了一种基于加权极限学习机(WELM)的小波包分解(WPD)和 KPCA 的特征提取方法。Deng 等[10]改进 KPCA 用于工业过程多模态诊断。但是 KPCA 在光谱识别方面应用极少，本文尝试使用 KPCA 提取光谱数据特征。径向基函数是一类其值只依赖于变量距原点距离的函数。如果原始数据是线性不可分的，通过径向基函数映射可能变得线性可分。太赫兹光谱数据整体上是线性不可分的，通过径向基函数可以将光谱数据映射到新的空间，然后进行线性区分。但是径向基函数映射后得到的太赫兹光谱数据也未必都是线性可分的，因此采用 KPCA 这种方法进行特征提取更加合适。针对太赫兹光谱线性不可分、特征提取难的问题，提出了结合径向基函数和 KPCA 的方法进行特征提取。首先采用径向基函数对去噪后的光谱数据进行映射，再采用 KPCA 进行特征提取，最后采用支持向量机对太赫兹光谱进行分类，验证特征提取效果。

1 结合径向基函数和 KPCA 的太赫兹光谱特征提取方法

1.1 径向基函数和 KPCA 的理论基础

1.1.1 径向基函数

径向基函数满足：若 $\| x_1 \| = \| x_2 \|$，则 $\varphi(x_1)=\varphi(x_2)$。根据定义可以发现，径向基函数是某种沿径向对称的函数，通常表示成变量到原点之间的欧氏距离的单调函数。径向基函数可以将非线性数据映射到新的径向基空间中，原始的非线性数据在新的径向基空间就有可能变成线性数据。径向基函数空间定义为：给定一个一元函数 ϕ：$R_+ \to R$，在定义域 $x \in R^d$ 上，所有形如 $\phi(x-c)=\phi(\| x-c \|)$及其线性组合张成的函数空间称为由函数 ϕ 导出的径向基空间。

食用油的主要成分是脂肪，食用油的脂肪包含饱和脂肪、反式脂肪、单不饱和脂肪和多不饱和脂肪。不同的食用油成分上的主要差异表现在不同种类脂肪的含量。这种含量的微小差异在太赫兹光谱中表现为光谱吸收谱线的微小差异。通过径向基函数映射，可以将食用油的太赫兹光谱映射到可能线性可分的径向基空间中，更便于特征提取。

1.1.2 核主成分分析 KPCA

目前，数据降维的方法主要分为两大类：线性降维和非线性降维。主成分分析(PCA)因为其概念简单、计算方便、线性重构误差最优等优良性能，成为数据处理中应用最广泛的线性降维方法之一，而 KPCA 作为 PCA 在处理非线性问题的扩展，得到快速发展。Xia 等[11]使用 KPCA 方法提取高光谱图像特征，使用随机森林方法对图像进行分类，获得良好的分类性能。Gan 等[12]将 KPCA 集成到基于多特征的内核稀疏表示分类中，提取高光谱图像特征并分类。针对 PCA 提取非线性特征的不足，本文提出使用 KPCA 提取太赫兹光谱的非线性特征。KPCA 的流程示意图如图 1 所示。

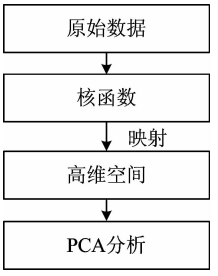

图 1 KPCA 流程图

Fig. 1 Flow chart of KPCA

对于给定的 n 维 N 个经过中心化的太赫兹时域光谱数据集 $X=\{x_1, x_2, x_3, \cdots, x_N\}$，$x_i \in R^n (i=1, 2, 3, 4, \cdots, N)$，首先将其映射到特征空间，得到 $\phi(x)$，则在特征空间中的的协方差矩阵表示为式(1)

$$C = \phi(X)\phi(X)^{\mathrm{T}} \tag{1}$$

在特征空间中进行 PCA 降维，可得到

$$\phi(X)\phi(X)^{\mathrm{T}} = \lambda w \tag{2}$$

式(2)中，w 是特征空间中的特征向量，λ 是特征向量对应的特征值。

对于任意第 j 个特征向量 $w_j (j=1, 2, 3, \cdots, n)$，$\lambda_i$ 是对应的特征值，由式(2)得到式(4)

$$\sum(\phi(x_i)\phi(x_i)^{\mathrm{T}})w_j = \lambda_i w_j \tag{3}$$

化简式(3)，

$$w_j = \frac{1}{\lambda_i}\sum(\phi(x_i)\phi(x_i)^{\mathrm{T}})w_j = \frac{1}{\lambda_i}\sum(\phi(x_i)^{\mathrm{T}}w_j)\phi(x_i) \tag{4}$$

令 $a = \frac{1}{\lambda_i}\phi(x_i)^{\mathrm{T}}w_j$，则有

$$w_j = \sum a\phi(x_i) \tag{5}$$

将式(5)代入式(4)，可得

$$\phi(X)\phi(X)^{\mathrm{T}}\phi(X)a=\lambda_i\phi(X)a \tag{6}$$

将式(6)两边同时乘 $\phi(X)^{\mathrm{T}}$，得

$$\phi(X)^{\mathrm{T}}\phi(X)\phi(X)^{\mathrm{T}}\phi(X)a=\lambda_i\phi(X)^{\mathrm{T}}\phi(X)a \tag{7}$$

令核方法 $K=\phi(X)^{\mathrm{T}}\phi(X)$，则式(7)可变为式(8)

$$K^2=\lambda_iKa \tag{8}$$

进一步化简，为式(9)

$$K=\lambda_ia \tag{9}$$

KPCA 常用的核函数有：线性核函数、多项式核函数、高斯核函数、指数核函数和拉普拉斯算子核函数。核函数的选择是核方法研究及应用的核心内容，选择的准则和方法目前并没有成型的理论方法，通过实际数据的验证结果来指导核函数的选择是常用的方法之一[13]。经过多次迭代实验比较，高斯核函数函数作为 KPCA 的核函数在本实验中是有效的。

1.2　太赫兹光谱特征提取与分类验证

1.2.1　太赫兹光谱特征提取

在实际问题中，原始数据经常包含一些多余的或者重复的信息，为了减少整个识别系统获取测量数据的代价和相应的计算工作量以及改善识别系统的性能，有必要通过特征提取把数据变换到低维数的特征空间中。太赫兹光谱通过 KPCA 可将有效信息降维到低维空间，作为识别特征。传统的线性投影方法，不能有效地将太赫兹光谱投影到一个可以线性区分的平面上。因此，采用结合径向基函数和 KPCA 这种非线性降维方法提取太赫兹光谱的特征。具体步骤如下：

Step 1 对实验测得的太赫兹时域光谱采用滑动平均滤波算法进行去噪预处理，获得实验样本集；

Step 2 对去噪后的太赫兹光谱采用径向基函数进行非线性映射。

Step 3 选择高斯核函数作为 KPCA 的核函数。高斯核函数的表达公式为

$$K(x_i,x_j)=\exp\left(-\frac{\|x_i-x_j\|^2}{2\sigma^2}\right) \tag{10}$$

式(10)中，x_i，x_j 分别表示非线性映射后的光谱样品，σ 表示一个常数。

Step 4 使用高斯核函数对样本集进行变换，计算核矩阵。

Step 5 计算核矩阵的特征值和特征向量，将特征值按照从大到小的顺序进行排列，将特征向量与特征值一一对应。

Step 6 将特征向量进行正交化处理，得到正交后的特征向量，获得降维后的数据。

1.2.2　分类验证

支持向量机(SVM)是一种监督式机器学习算法。支持向量机的原理是在两类样本间寻找一个最优的分类超平面，使得该超平面两侧与最近样本点的距离最大化。支持向量机方法建立在统计学习理论的 VC 维理论和结构风险最小原理基础上，根据有限的样本信息在模型的复杂性和学习能力之间折衷，希望获得最好的推广能力。本文使用支持向量机构建分类模型，最终实现太赫兹光谱识别，验证特征提取效果。图 2 为实验流程。

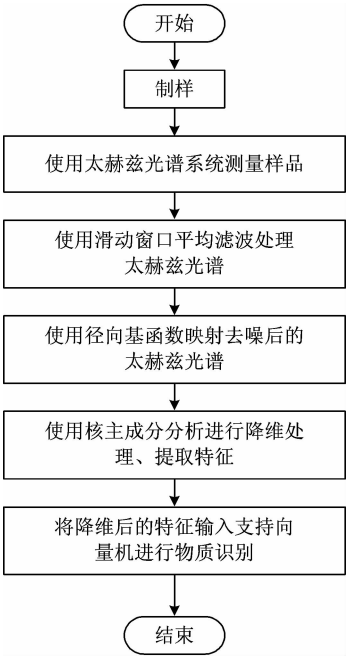

图 2　实验流程

Fig. 2　Flow chart of experiment

2　实验与结果

2.1　食用油太赫兹光谱测量

不同的太赫兹时域光谱系统的采样率、有效太赫兹光谱范围、采样频率等参数不完全相同。实验中采用爱德万公司生产的型号为 TAS7400TS GDU1 太赫兹时域光谱系统。

测量时，太赫兹光谱系统温度为 22 ℃，相对湿度维持在 5%以下。实验系统参数设置如表 1 所示。

表 1　实验参数设置

Table 1　Experimental parameter setting

参数名称	值
样品采样次数	1 024
背景采样次数	1 024
光谱分辨率/GHz	1.9
光谱起始频率/THz	0.5
光谱截止频率/THz	2
样品厚度/mm	6

实验测量了芝麻油、葡萄籽油、茶籽油的透射式太赫兹时域光谱，对测量所得的太赫兹光谱采用滑动窗口平均去噪算法进行去噪处理后，各样品太赫兹吸光度谱如图 3 所示。

2.2　径向基函数映射

对食用油光谱去噪后，存在重叠部分。对食用油光谱采用径向基函数映射，将光谱映射到不同的空间。采用的映射函数公式为

$$y=\mathrm{e}^{-\frac{x^2}{\sigma}} \tag{11}$$

式(11)中，y 为映射后的光谱数据，x 为去噪后的光谱数据，

σ 为常数。

采用的径向基函数能够将光谱中的吸收峰，变平缓，因此使用径向基函数映射变换后原来混叠的光谱曲线重叠部分大大减少，增加了可区分性。径向基函数映射后食用油光谱如图 4 所示。

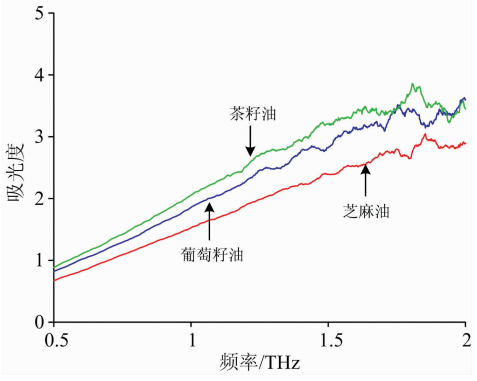

图 3 去噪处理后的食用油吸光度谱

Fig. 3 Terahertz absorption spectra of edible oil

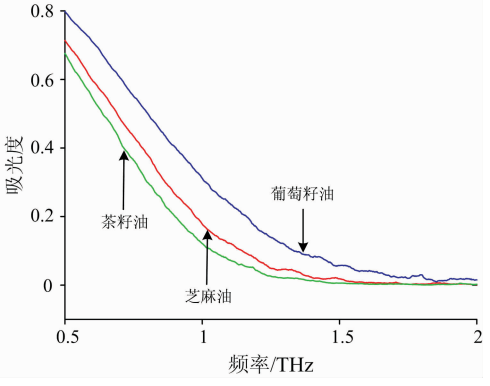

图 4 径向基函数映射后太赫兹吸光度谱图

Fig. 4 Terahertz absorption spectra of edible oil after radial basis function mapping

2.3 特征提取

对去噪后的 90 例食用油样本(30 例芝麻油，30 例葡萄籽油，30 例茶籽油)光谱分别采用本方法、PCA 和 KPCA 进行特征提取，特征提取的维数为 12 维。食用油样本光谱数据除了使用滑动窗口平均去噪算法预处理外，没有经过归一化、中心化等其他预处理过程。为了观察方便，选取贡献率最大的 3 个主因子绘制散点图，结果分别如图 5(a,b,c)所示。从图 5(a,b)可以看到，PCA 和 KPCA 提取出来的前 3 个主成分占光谱的变化不到 50%，因此需要更多特征光谱信息。而这 3 种方法提取出来的前 12 个主因子的累积贡献率均超过了 90%，因此前 12 个主因子可以作为食用油光谱的特征。

采用类内距离和类间距离来评价特征提取效果，类内距离越小，类间距离越大，表示特征的紧密程度越大，不同类之间的可分离程度越好，特征提取效果越好。反之，则特征提取效果不好。实验使用欧氏距离计算类内距离和类间距离。类内距离是指类内所有点两两之间距离的平均。类间距离采用的是中间距离法，计算类中心之间的距离。为了解决不同特征提取方法映射范围不一的问题，把 3 种特征提取方法计算得到的 12 维特征映射到各坐标轴范围均为[−1，1]的高维坐标系中，然后进行类内距离和类间距离计算。所提取的特征类内距离计算结果如表 2 所示，类间距离计算结果如表 3 所示。

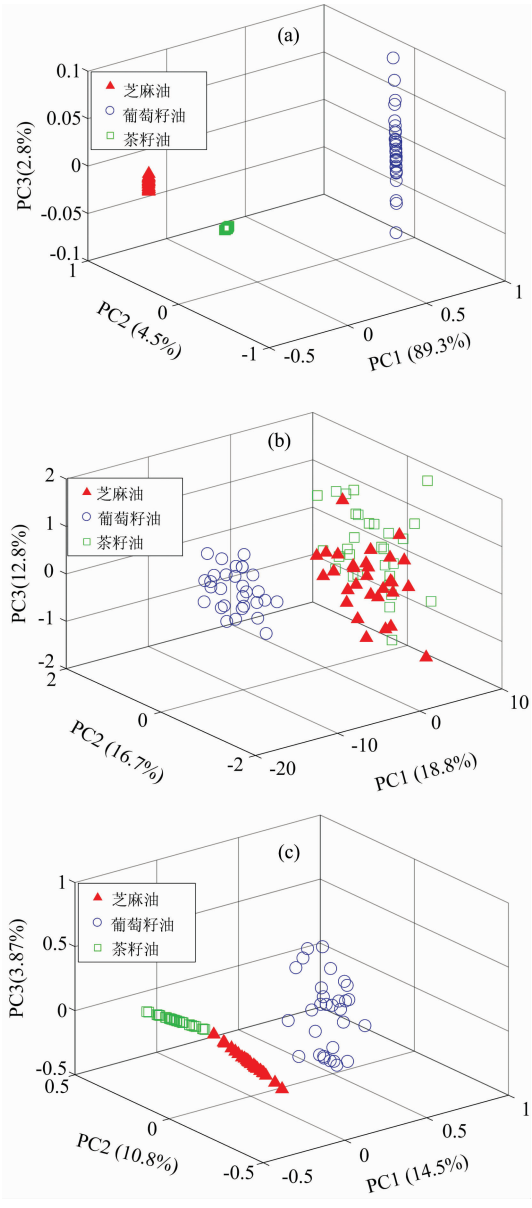

图 5 (a) 结合径向基函数和 KPCA 特征提取结果；(b) PCA 特征提取结果；(c) KPCA 特征提取结果

Fig. 5 (a) The feature extraction results of combining radial basis function and KPCA; (b) The feature extraction result of PCA; (c) The feature extraction result of KPCA

从表 2 可以看到，本方法类内距离均小于 PCA 和 KPCA，这说明本方法聚类效果优于其余两种方法。

表 2　不同特征提取方法获得的食用油类内距离

Table 2　The intraclass distances of different feature extraction methods for various edible oils

食用油类型	类内距离		
	PCA	KPCA	结合径向基函数和 KPCA
芝麻油	1.492 0	2.544 2	0.128 7
葡萄籽油	0.273 6	1.973 4	0.234 7
茶籽油	0.578 9	1.902 2	0.212 4

表 3　不同特征提取方法获得的食用油类间距离

Table 3　The interclass distances of different feature extraction methods for various edible oils

组合食用油类型	类间距离		
	PCA	KPCA	结合径向基函数和 KPCA
芝麻油-葡萄籽油	4.821 9	2.251 6	3.854 0
芝麻油-茶籽油	1.475 2	1.645 5	3.812 6
葡萄籽油-茶籽油	6.421 3	2.367 4	6.193 3

从表 3 的结果可以看到，本方法类间距离均大于 KPCA，说明类间可分性优于 KPCA。本方法和使用 PCA 计算得到的芝麻油-葡萄籽油和葡萄籽油-茶籽油的类间距离差别不大，两种组合的类间可分性良好。但是，使用 PCA 计算得到的芝麻油-茶籽油组合类间距离偏小，容易出现错误分类的情况。而本方法各组合类间距离较大，可分性良好。整体上本方法类间可分性优于 PCA。

2.4　SVM 分类验证效果

为了更进一步验证上述特征提取方法的效果，采用支持向量机对提取后的特征进行建模分类。将上述 3 种食用油样本输入支持向量机中，采用 5 折交叉验证的方法，计算 6 种不同核函数的支持向量机分类的准确率，从而验证分类的效果。分类正确率结果如表 4 所示。

表 4　分类正确率结果对比

Table 4　Comparison of classification accuracy rate results

支持向量机类型	分类正确率/%		
	PCA	KPCA	结合径向基函数和 KPCA
Linear SVM	100	95.6	100
Quadratic SVM	100	95.6	100
Cubic SVM	100	96.7	100
Medium SVM	98.9	98.9	100
Coarse Gaussian SVM	95.6	94.4	100

从表 4 可以看出，本方法分类正确率高于 PCA 和 KPCA，说明本方法特征提取效果更好。

3　结　论

针对部分物质太赫兹吸收谱没有明显吸收峰特征，谱线整体相似难以识别的问题，提出了结合径向基函数和 KPCA 的特征提取方法。利用该方法对被测物质的太赫兹吸收谱进行非线性映射提取特征，使用支持向量机对其进行分类。本特征提取方法类内聚类效果好，类间可分性好，使用不同内核的支持向量机分类在本实验中正确率都能达到 100%。相比于 PCA，使用本方法提取出来的特征在支持向量机分类测试中正确率最大能提高约 4%。相比于 KPCA，使用本文提出的方法提取出来的特征在支持向量机分类测试中正确率最大能提高约 6%。因此所提出的特征提取方法效果良好，结合支持向量机能够对食用油进行分类，在食品安全检测领域有很好的应用价值。

References

[1] Liu W, Liu C, Yu J, et al. Food Chemistry, 2018, 251: 86.

[2] NIE Mei-tong, XU De-gang, WANG Yu-ye, et al(聂美彤, 徐德刚, 王与烨, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2018, 38(7): 2016.

[3] LI Li-long, XIANG Yang, WU Lei, et al(李利龙, 向　洋, 吴　磊, 等). High Power Laser and Particle Beams(强激光与粒子束), 2013, 25(6): 1566.

[4] Yin M, Tang S, Tong M. Analytical Methods, 2016, 8(13): 2794.

[5] CHEN Tao(陈　涛). Chinese Journal of Quantum Electronics(量子电子学报), 2016, 33(4): 392.

[6] HU Xiao-hua, LIU Wei, LIU Chang-hong, et al(胡晓华, 刘　伟, 刘长虹, 等). Transactions of the Chinese Society of Agricultural Engineering(农业工程学报), 2017, 33(9): 302.

[7] ZHANG Wen-tao, LI Yue-wen, ZHAN Ping-ping, et al(张文涛, 李跃文, 占平平, 等). Infraredand Laser Engineering(红外与激光工程), 2017, 46(11): 159.

[8] Liu J, Kan J. Spectrochimica Acta Part A—Molecular and Biomolecular Spectroscopy, 2018, 194(5): 14.

[9] Hu Q, Qin A, Zhang Q, et al. IEEE Sensors Journal, 2018, 18(20): 8472.

[10] Deng X, Zhong N, Wang L. IEEE Access, 2017, 5: 23121.

[11] Xia J, Falco N, Benediktsson J A, et al. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2017, 10(4): 1601.

[12] Gan L, Xia J, Du P, et al. IEEE Transactions on Geoscience and Remote Sensing, 2018, 56(9): 5343.

[13] WANG Zhen-wu, HE Guan-yao(王振武, 何关瑶). Journal of Hunan University · Natural Sciences(湖南大学学报 · 自然科学版), 2018, (10): 155.

Edible Oil Terahertz Spectral Feature Extraction Method Combining Radial Basis Function and KPCA

WANG Zhuo-wei[1], LUO Jian-peng[1], LI Xue-shi[2*], CHENG Liang-lun[2]

1. School of Computers, Guangdong University of Technology, Guangzhou 510006, China
2. School of Automation, Guangdong University of Technology, Guangzhou 510006, China

Abstract In order to deal with the case where the terahertz spectrums are linearly inseparable, this paper proposes a method combining the radial basis function and the kernel principal component analysis (KPCA) to extract the terahertz spectral features of edible oils. By using this method, the extracted inner-class distance of features is small, meanwhile the extracted inter-class distance is large. An accurate classification model can be established in most support vector machine classifiers. Terahertz spectroscopy is an important method to detect the type and quality of edible oils. The research on the feature extraction technology of terahertz spectroscopy is of great significance for the rapid detection of edible oil types and quality. Although there have been a theoretical basis on how to use the terahertz spectroscopy to detect the type and quality of edible oils, it is still difficult to accurately extract the terahertz spectral features of edible oils and establish an accurate classification mode accordingly. Recently, researchers often use principal component analysis (PCA) in the field of chemometrics to extract features and use machine learning algorithms to establish a material classification model. However, the linear separability of the terahertz spectrum of edible oils has different characteristics in different frequency bands. When the terahertz spectrums of edible oils are linearly separable, it is feasible to extract features using PCA, and thus establish an accurate classification model. However, when the terahertz spectrums of edible oils are linearly inseparable, the features extracted using PCA are often not accurate enough, and an appropriate classifier is demanded to establish an accurate classification model. The method combining the radial basis function and KPCA feature extraction can be described as follows: the linear space-inseparable terahertz spectral data are mapped to the radial basis space by the radial basis function, then the features are extracted by KPCA which become linearly separable. As a result, a more accurate classification model can be established. For the experiment, firstly the sliding window average filtering algorithm is used to filter the terahertz spectral data of three edible oils. Then, the radial basis function is employed to nonlinearly map the terahertz spectrum. After that, KPCA is utilized for data dimensionality reduction. Finally, the support vector machine (SVM) is used to establish a classification model for edible oils and the feature extraction effect is verified. The calculated results of inter-class separability show that the inner-class distance of features extracted by the method is smaller, and the inter-class distance is larger. Thus, the overall feature extraction effect presented in this paper is better than those of PCA and KPCA. The experimental results of classification verification show that based on certain classification models the features extracted by PCA and KPCA cannot distinguish the type of edible oils very accurately. However, based on every classification model the feature extraction method proposed in this paper can distinguish the type of edible oils accurately. The method proposed in this paper has a better effect on the extraction of terahertz spectral features of edible oils, which makes it of great value in the detection and analysis of the quality of edible oils.

Keywords Terahertz spectroscopy; Radial basis function; Kernel principal component analysis; Support vector machine

(Received Dec. 27, 2018; accepted Apr. 8, 2019)

* Corresponding author