

非均匀固体籽粒近红外光谱采集方法研究

李浩光^{1,2}, 于云华^{1,2}, 逢燕¹, 沈学锋^{1,2}

1. 山东石油化工学院机械与控制工程学院, 山东 东营 257061

2. 中国石油大学(华东)新能源学院, 山东 东营 257061

摘要 近红外光谱定性分析技术研究中实验对象主要包括液态物质、粉末状物质和非均匀固体颗粒物质。液态物质与粉末状物质属于均匀分布被测对象, 光谱采集与分析相对容易。而非均匀固体样品大小不一、形态各异、内部物质分布不均匀, 近红外光谱中既包含定性分析中需要提取的种类信息, 又包含需要消除的个体差异信息, 因此相对于均匀分布的液体或粉末状物质, 非均匀固体分析难度更大。目前国内外尚未建立对于大小不一、形状各异的非均匀固体颗粒物质的有效近红外定性分析方法。以非均匀固体谷物玉米籽粒为主要研究对象, 在研究多种光谱采集方式的基础上, 分析了非均匀固体籽粒光谱采集的特点, 设计了针对非均匀固体籽粒的光谱采集装置, 为确保实验结果的客观性, 采用了朴素贝叶斯分类器、K近邻、支持向量机、BP神经网络、仿生模式识别等5种模式识别方法, 建立在漫反射及漫透射方式下单籽粒玉米近红外光谱多品种定性分析模型, 并对漫反射与漫透射方式下所建立的定性分析模型进行对比实验, 分析胚面朝向对玉米单籽粒的漫反射与漫透射光谱定性分析模型鉴别效果影响, 测试集与建模集时间间隔长短对漫反射及漫透射方式下玉米单籽粒品种鉴别精度的影响。实验结果表明采用漫透射方式建立的模型不易受到非均匀固体籽粒摆放方式的影响, 且模型具有更好的泛化能力, 为后续研究提供可行光谱采集方法。以非均匀固体玉米籽粒为主要实验对象, 对其采集方法与定性分析模型进行研究, 可为类似形态被测物体的近红外光谱定性分析提供有益借鉴, 具有重要研究意义。

关键词 非均匀固体籽粒; 漫反射; 漫透射; 模式识别

中图分类号: O657.3 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2021)09-2748-06

引言

近红外光谱定性分析研究对象有液体、粉末、非均匀固体籽粒等形式, 本工作以大小不一、形状各异的非均匀固体谷物颗粒为主要研究对象。非均匀固体籽粒因个体之间存在差异, 其光谱中既包含定性分析需要提取的种类信息, 又包含需要消除的个体差异信息, 因此, 非均匀固体籽粒光谱采集与分析相对于常规的均匀分布液体样本或者粉末状样本具有更大难度。

针对这一难题, 以玉米籽粒为研究对象, 在分析非均匀固体籽粒特点的基础上, 搭建了漫反射与漫透射采集装置, 采用漫反射及漫透射两种方式采集玉米籽粒的光谱, 并使用支持向量机、仿生模式识别等模式分类方法建立了非均匀固体籽粒的近红外光谱定性鉴别模型, 对漫反射与漫透射两种

方式的定性分析模型进行了对比实验, 确定了采用漫透射方式作为玉米这种非均匀固体籽粒的近红外光谱采集方式, 为该领域后续研究提供了参考。

1 光谱采集方法

近红外光谱分析技术中通常分为透射、漫反射、漫透射3种采集方法。

(1)透射采集方法: 适合于分析均匀分布的液体样品或者透明样品, 测量时, 被测样品置于近红外光源与近红外探测器之间, 探测器所采集的近红外光需透过样品并与样品内部分子相互作用, 若将所采光谱用于定量分析, 被测样品浓度与对光的吸收关系符合朗伯-比尔定律^[1-2]。

(2)漫反射采集方法: 适合于固体或粉末类不透明物质, 该测量方式下近红外光源与探测器位于被测样品同侧, 近红

收稿日期: 2021-01-20, 修订日期: 2021-04-11

基金项目: 国家重大科学仪器设备开发专项(2014YQ470377), 山东省教育厅科技计划项目(J18KA329), 东营市科技发展基金项目(DJ2020032)资助

作者简介: 李浩光, 1981年生, 山东石油化工学院副教授 e-mail: lihaoguang@upc.edu.cn

外光照射到样品表面后, 经过多次反射、折射、散射及吸收后返回样品表面^[3-5]。

(3)漫透射采集方法: 适合于采集不透明或固体样品, 采集时被测样品与光源及探测器相对位置与透射采集方式相同。由于样品中含有光散射物质, 近红外光在穿透被测样品时, 除吸收作用外还存在多次散射作用, 因此朗伯-比尔定律^[3]并不适用于漫透射采集方法。

漫反射和漫透射方式是非均匀固体近红外光谱定性分析中常用的两种检测手段。漫反射采集方式通常可以穿透被测样品较浅的深度, 该方式下漫反射光携带样品的浅层信息。而漫透射方式能够贯穿被测样品, 深达样品内部, 因此漫透射光谱能够反映样品内部较深层信息^[3]。

近红外光谱定性分析中的光谱采集方法针对不同应用可分为群体和单籽粒个体采集两种方式。

(1)群体测量方式: 群体样本光谱采集常采用整杯测量方式, 测量杯中装有一定深度、一定数量的固体籽粒, 测量所得近红外光谱包含了多颗固体籽粒光谱平均信息, 因此该方式在采集过程中可基本消除个体籽粒差异对光谱的影响。一般对大批量非均匀固体籽粒进行分类时, 可选用群体测量方式。

(2)单粒测量方式: 由于单粒固体样品体积小, 受光面积小、形态各异、大小不一, 每幅光谱仅反映当前所测籽粒的信息, 单粒方式采集的光谱既包含了品种信息, 也包含了每颗籽粒之间的差异信息, 在建立定性鉴别模型时, 应尽可能扩大单籽粒光谱之间的品种差异信息, 缩小或消除同类样本之间的个体差异信息。在需要对每颗籽粒进行分类的检测中则需选用单粒测量方式。

综上所述, 单粒个体定性鉴别技术相比于群体定性鉴别技术难度更大。

2 非均匀固体籽粒光谱特点分析

针对非均匀固体单籽粒鉴别, 在使用漫反射或漫透射采集方式时, 有以下 5 个特点:

(1)易漏光: 采集时需保证探测器窗口采集所得为经样品反射或透射的近红外光, 并排除外界杂散自然光及近红外光源直接照射探测器窗口造成的干扰。

(2)固体籽粒形态各异、大小不一: 漫反射或漫透射采集装置应对各种不同形状、不同大小的籽粒具有较好适应性。

(3)部分籽粒具有活性: 对于不透明非均匀固体籽粒, 如玉米籽粒等, 漫反射或漫透射方式所使用近红外光源应限定在一定强度范围内。采用漫透射方式时, 光源首先应达到穿透不透明固体籽粒的强度, 此外, 考虑到光源强度对籽粒活性的影响, 光源强度也不宜过高, 过高则会灼伤籽粒, 破坏籽粒活性。

(4)部分固体为不规则多面体: 非均匀固体籽粒摆放位置与近红外光入射及出射角度密切相关, 采集近红外光谱时应消除摆放位置对所采光谱的影响。

(5)非均匀固体内部物质分布非均匀: 非均匀固体不同

位置物质组成不同, 采集近红外光谱时应寻找最有利于籽粒品种分类的部位, 或在分析时消除籽粒部位对分类性能的影响。

综上, 对于不规则、非透明、非均匀固体小颗粒可使用漫反射或漫透射方式采集其近红外光谱, 本工作在综合分析上述五点影响因素基础上, 进一步明确适合于该类固体的光谱采集装置。

3 采集装置设计

在分析漫反射与漫透射采集方法原理的基础上, 采用 MicroNIR1700 型微型近红外光谱仪分别搭建了非均匀固体籽粒漫反射与漫透射光谱采集装置, 光谱采集装置如图 1 所示。

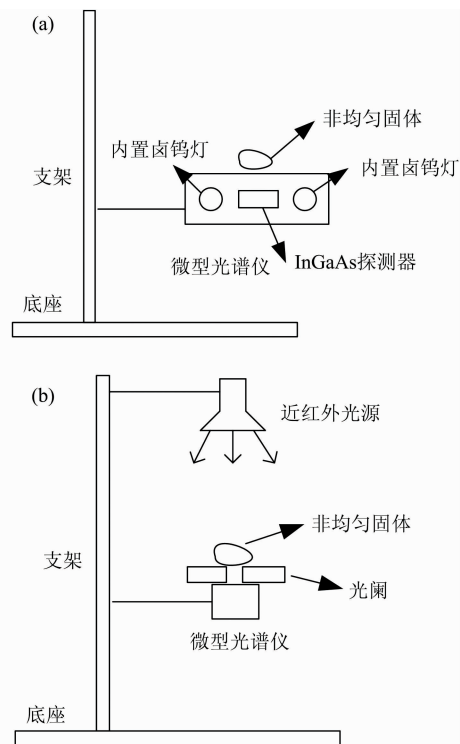


图 1 自制光谱采集装置结构示意图

(a): 漫反射实验装置; (b): 漫透射实验装置

Fig. 1 Structure diagram of self-made spectrum acquisition device

(a): Diffuse reflection experimental device;

(b): Diffuse transmission experimental device

(1)漫反射实验装置: 如图 1(a)所示, 为便于采集非均匀固体近红外光谱, 漫反射采集装置采用底座及支架固定微型光谱仪, 使用光谱仪内置卤钨灯作为近红外光源, 采集光谱时将玉米籽粒置于光谱仪检测窗口正上方。该装置特点为近红外光源与探测器均位于被测样品下侧。光谱仪的检测窗口如图 2 所示, 窗口中心矩形为探测器铟镓砷(InGaAs)线性阵列, 铟镓砷线性阵列两侧各有一个内置卤钨灯, 为避免外界杂散自然光干扰, 使用漫反射测量方式采集光谱时, 需使

用黑色遮光罩置于被测物体外侧。

(2)漫透射实验装置:如图 1(b)所示,测量时需关闭微型光谱仪内置卤钨灯光源,并使用安装在支架上方的外置近红外光源,该装置特点为被测物位于近红外光源与检测器的中间位置。

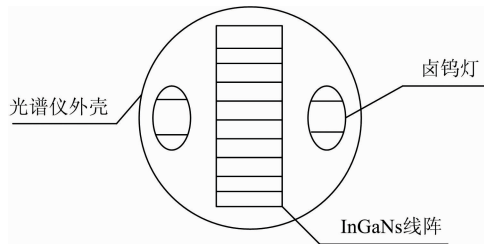


图 2 MicroNIR1700 型微型光谱仪俯视图

Fig. 2 Top view of MicroNIR 1700 micro spectrometer

据前期研究工作^[6-8],为保证最佳测量效果,设置近红外光源光源下端与光谱仪测量面距离为 5 cm,光源电压为 5 V,光阑孔径为 5.2 mm,为保证测量可靠性,采集光谱时每隔一定时间对光谱仪进行一次白板校正与黑板校正。

4 结果与讨论

4.1 实验数据

以玉米种子品种真实性鉴别任务为例,对漫透射和漫反射两种采集方法下所建定性分析模型的鉴别效果进行分析。

针对形状各异、大小不一的非均匀玉米籽粒,为确定适合于单籽粒玉米近红外光谱采集方法,首先采用图 1 所示的两种自制光谱采集装置分别采集多个品种玉米籽粒漫透射以及漫反射近红外光谱。

实验数据为 2016 年 10 月份连续 5 日采集的农华 032、农大 108、京玉 16、洛单 248 四个品种玉米籽粒的近红外光谱。

采集时,各品种分别挑选 100 颗作为待测样品,分别以漫反射方式与漫透射方式采集每玉米胚面和非胚面光谱各一次,每次采集十条光谱并取平均。采用交替采集方法,即按照农华 032 一粒—农大 108 一粒—京玉 16 一粒—洛单 248 一粒……的方式采集单颗籽粒光谱。

将连续 5 日采集数据分别命名为 T1, T2, T3, T4 和 T5 数据集,各数据集中均包含胚面(germ surface)光谱和非胚面(non-germ surface)光谱。

4.2 分析方法设置

使用朴素贝叶斯等 5 种模式分类方法^[9-10]分别对所采集的漫反射与漫透射光谱建立单籽粒玉米定性鉴别模型,5 种模式分类方法简介如下:

(1)朴素贝叶斯分类器(Naive Bayes classifier, NBC)是一种基于贝叶斯决策理论的模式分类方法。

(2)K 近邻(K-nearest neighbor, KNN)方法是一种基于实例的机器学习方法,其特点是在训练过程中存储已知类别的训练数据,在预测新的未知样本时,从已存储数据中抽取相似样本,作为分类的重要辅助信息。

(3)支持向量机(support vector machine, SVM)方法是一种二分类机器学习方法,支持向量机的理论基础是 VC 维理论和结构风险最小化原理,依据少量样本并折中考虑模型的复杂性和学习能力,以增强模型的泛化能力^[10],SVM 分类器的目标就是寻找最优超平面,将两类样本正确分开,且分类间隔最大。

(4)BP 神经网络(back propagation artificial neuron network, BPANN)算法通过正向传播与反向传播交替运行,使用梯度最速下降法与动态迭代搜索一组权重向量,使网络误差函数达到最小值^[10]。

(5)仿生模式识别(biomimetic pattern recognition, BPR)方法是一种基于同源连续性原理的模式识别方法,其特点是基于识而不识别,通过构建在高维空间中能够对同类样本具有最优覆盖的超几何体为目标^[11]。

根据分类器工作原理的特点,实验中使用两种不同分类器前处理策略,分类器前处理策略即光谱的预处理方法以及特征提取组合的具体样式。

分类前处理策略 1 仅采用平滑、一阶导、归一化,分类前处理策略 2 除使用策略 1 中的预处理外还需使用偏最小二乘及正交线性鉴别分析方法进行特征提取^[8]。

使用 NBC, SVM, KNN, BPANN 和 BPR 共 5 种模式分类方法建立单籽粒玉米定性鉴别模型,其中 NBC, KNN, SVM 和 BPR 使用分类前处理策略 1, BPANN 使用分类前处理策略 2。

针对每种分类方法的特点对其参数进行如下设置:

NBC: 贝叶斯分类器的分类先验概率可设定为 0.25,各个特征类条件概率可利用 T1 数据集数据计算得到。

KNN: 分类器使用欧氏距离,以鉴别准确率为选择标准,通过 5 折交叉验证的方法确定适合该分类任务的最优近邻数为 5。

SVM: 实验中使用 LIBSVM 工具箱,设置 SVM 分类器类型为 Nu-SVC,多分类类型,选用径向基核函数,通过工具箱中的网格寻参函数交叉验证来确定参数 Nu 为 0.6,核函数中的 gamma 值为 2.8。

BPANN: 采用分类器前处理策略 1,将预处理方法得到的输出值直接作为 BP 神经网络的输入值,通过 5 折交叉验证的方法确定最优隐藏层数目为 9。

BPR: 使用 KS 方法选择构网样本点,超香肠作为基本覆盖单元,通过交叉验证确定超香肠神经元的半径。

为评价模式分类器性能,对于多分类问题,以类别代号为 i 的样本为例:

(1)正确识别率(correct acceptance rate, CAR)

$$CAR = \frac{N_{ca}}{N_{thiskind}} \quad (1)$$

式(1)中, N_{ca} 为正确识别的 i 类光谱数; $N_{thiskind}$ 为 i 类总光谱数。

(2)正确拒识率(correct rejection rate, CRR)

$$CRR = \frac{N_{cr}}{N_{otherkind}} \quad (2)$$

式(2)中, N_{cr} 为正确拒识的非 i 类光谱数; $N_{otherkind}$ 为非 i 类

总光谱数。

(3) 鉴别准确率或平均识别率(correct discriminate rate, CDR)

$$CDR = (CAR + CRR) / 2 \quad (3)$$

实验 1

以 T1 作为实验数据集, 分析胚面朝向对玉米品种单籽粒的漫反射与漫透射光谱定性分析模型鉴别效果影响。

实验过程中, 每次从 T1 数据集中随机抽取二分之一光谱样本作为训练集, 其余二分之一光谱样本作为测试集对模型性能进行测试, 重复 50 次实验后取鉴别准确率均值。

表 1 S1 情形胚面识别胚面鉴别准确率 (%)

Table 1 Identification accuracy of case S1 (%)

方法	NBC	KNN	SVM	BPANN	BPR
漫反射	75.0	82.0	81.5	83.0	85.5
漫透射	80.0	85.6	90.6	91.1	95.0

表 2 S2 情形非胚面识别非胚面鉴别准确率 (%)

Table 2 Identification accuracy of case S2 (%)

方法	NBC	KNN	SVM	BPANN	BPR
漫反射	74.0	81.5	82.5	83.5	85.5
漫透射	80.5	85.5	91.2	91.0	95.0

表 3 S3 情形非胚面识别胚面鉴别准确率 (%)

Table 3 Identification accuracy of case S3 (%)

方法	NBC	KNN	SVM	BPANN	BPR
漫反射	69.0	70.0	75.3	75.5	80.2
漫透射	81.0	85.0	90.8	90.5	95.0

表 4 S4 情形胚面识别非胚面鉴别准确率 (%)

Table 4 Identification accuracy of case S4 (%)

方法	NBC	KNN	SVM	BPANN	BPR
漫反射	65.0	70.3	73.5	75.6	80.6
漫透射	81.2	85.3	91.5	91.5	96.5

根据训练集与测试集选择玉米胚面或非胚面光谱不同排列组合分为以下 4 种情形 S1—S4:

S1: 训练集选择胚面光谱, 测试集选择胚面光谱;

S2: 训练集选择非胚面光谱, 测试集选择非胚面光谱;

S3: 训练集选择非胚面光谱, 测试集选择胚面光谱;

S4: 训练集选择胚面光谱, 测试集选择非胚面光谱。

表 1—表 4 是从 T1 数据集的胚面或非胚面光谱数据中随机抽取二分之一作为训练集, 其余二分之一作为测试集进行测试时, 4 种可能排列组合情形下漫反射与漫透射方式实验结果对比。

由 S1—S4 四种情形对比结果可见:

(1) 漫透射方式: S1—S4 四种情形识别率差异较小。以 BPR 方法为例, 四种情形所得识别率均在 95% 左右。

(2) 漫反射方式: S1—S4 四种情形识别率并不完全一致。S1 和 S2 两种情形下, 定性分析模型所得识别率均相对

较高, 以 BPR 方法为例, 其识别率达到 85% 左右; S3 和 S4 两种情形下, 定性分析模型所得识别率均相对较低, 以 BPR 方法为例, 其识别率为 80% 左右。

产生上述现象的原因分析如下:

(1) 采用漫反射方式时, 近红外光源未能完全穿透玉米籽粒, 近红外探测器接收到的漫反射光只携带了玉米籽粒某一面或者表皮的信息, 因此在 S3, S4 两种情形下, 玉米品种鉴别准确率出现明显下降。

(2) 采用漫透射方式时, 近红外探测器采集所得近红外光一定为贯穿玉米籽粒的部分, 其光谱中既包含了玉米籽粒胚面信息, 又包含了玉米籽粒非胚面信息, 同时还包含了种子内部物质信息, 因此, 在 S3 和 S4 两种情形下, 正确鉴别率并未出现明显下降。

采用 5 种模式分类方法建立定性分析模型性能实验结果表明, 在单日数据集上, 漫透射定性分析模型不受玉米胚面摆放方式影响, 而漫反射定性分析模型易受玉米籽粒摆放方式影响, S1—S4 四种情形下漫透射定性分析模型性能均高于漫反射定性分析模型性能。

实验 2

以 T1 数据集作为训练集, T2—T5 作为测试集对模型性能进行测试。训练集可选择胚面光谱或非胚面光谱, 而测试集也可以选择胚面或者非胚面光谱, 由此产生 S1—S4 四种情形, 使用 BPR 方法建立定性分析模型, 并对四种情形下的模型预测性能进行比较。

表 5 为使用 BPR 方法时, 4 种可能排列组合情形下, 漫反射与漫透射方式下鉴别准确率对比。

表 5 BPR 方法不同测试集的鉴别准确率 (%)

Table 5 Identification accuracy of different test sets with BPR method (%)

组合类型	采集方式	T2	T3	T4	T5
S1	漫反射	84.5	83.0	83.5	82.0
	漫透射	96.5	96.0	95.5	94.0
S2	漫反射	84.5	83.5	83.0	82.5
	漫透射	96.7	95.5	95.0	94.5
S3	漫反射	80.3	79.5	79.5	78.0
	漫透射	95.5	95.5	94.0	93.0
S4	漫反射	82.5	81.5	80.5	79.0
	漫透射	95.5	95.3	95.3	95.0

由表 5 可知, 在 S1—S4 情形下, 在 T2—T5 测试集中, 有如下结果:

(1) 漫反射方式: S3 及 S4 情形的识别率低于 S1、S2 情形的识别率。

(2) 漫透射方式: S3 及 S4 情形的识别率与 S1、S2 情形的识别率基本接近。

以 T3 测试集为例, 漫透射方式下, 四种情形的模型识别率均在 95% 左右。而漫反射方式下, S1、S2 情形的模型识别率为 83% 左右, S3、S4 情形下的识别率在 80% 左右。

由此可见, 漫透射方式 T2—T5 测试集的 S1—S4 四种情形鉴别准确率基本一致, 漫反射方式 T2—T5 测试集的

S1、S2 情形鉴别准确率高于 S3、S4 情形。总体而言,在多个测试集上,漫反射方式下所采光谱数据建立模型性能均低于漫透射方式下所采光谱数据建立的模型。实验 2 在多个实验数据集上证明:漫透射方式更适合于玉米单籽粒品种鉴别。

实验 3

以 T1 胚面与非胚面混合光谱作为训练集, T2—T5 的胚面与非胚面混合光谱作为测试集进行测试。选择训练集时,随机从 T1 胚面光谱及非胚面光谱中各抽取 50 条光谱样本,以模拟实际测量近红外光谱时玉米胚面朝向随机摆放的情形,即胚面和非胚面朝上概率各为 50%,选择测试集时使用相同方式。鉴别准确率为 50 次实验取均值得到。

表 6 是在漫透射方式下以 T1 数据集作为建模集建立定性分析模型, T2—T5 作为测试集进行测试时所得鉴别准确率。

表 6 漫透射方式下不同测试集的鉴别准确率 (%)

Table 6 Identification accuracy of different test sets under diffuse transmission mode (%)

方法	T2	T3	T4	T5
BPR	95.5	94.5	92.3	91.5

由表 6 可以看出,对于胚面与非胚面光谱混合的建模集

T1,使用不同测试集进行测试时,以漫透射光谱建立的定性分析模型具有较高识别率,在 T2—T5 四个测试集上的识别率均高于 91%,此外,漫透射采集方式下,建立定性模型对不同测试集数据进行鉴别时,随测试集与建模集时间间隔增大,模型性能下降较为平缓,因此漫透射方式定性分析模型具有较强泛化能力。

5 结 论

针对非均匀固体籽粒的近红外光谱采集问题,首先分析了不规则非均匀固体光谱分析的特点,研究了光谱采集方法,在此基础上,搭建了漫反射与漫透射光谱采集装置,然后使用仿生模式识别方法等 5 种模式识别方法建立了漫反射及漫透射方式下单籽粒玉米近红外光谱定性鉴别模型,并进行了漫反射与漫透射定性分析模型对比实验。实验结果表明,漫透射定性分析模型不受玉米胚面摆放方式影响,而漫反射定性分析模型易受玉米籽粒摆放方式影响;漫透射方式比漫反射方式更适合于玉米单籽粒近红外品种鉴别;漫透射采集方式下,建立定性模型对不同测试集数据进行鉴别时,随测试集与建模集时间间隔增大,模型性能下降较为平缓,因此漫透射方式定性分析模型具有较强泛化能力。该方式可作为玉米籽粒光谱采集的可靠方式。

References

- [1] CHU Xiao-li, SHI Yun-ying, CHEN Bao, et al(褚小立,史云颖,陈 瀑,等). Journal of Instrumental Analysis(分析测试学报), 2019, 38(5): 603.
- [2] CHU Xiao-li(褚小立). User Manual of Near Infrared Spectroscopy(近红外光谱分析技术实用手册). Beijing: Machinery Industry Press(北京:机械工业出版社), 2016.
- [3] Jue T, Masuda K. Application of Near Infrared Spectroscopy in Biomedicine. Springer US, 2013.
- [4] LU Wan-zhen(陆婉珍). Near Infrared Spectrometer(近红外光谱仪器). Beijing: Chemical Industry Press(北京:化学工业出版社), 2010.
- [5] YAN Yan-lu, CHEN Bin, ZHU Da-zhou(严衍禄,陈 斌,朱大洲). Near Infrared Spectroscopy Analytical—Principles, Technology and Application(近红外光谱分析的原理、技术与应用). Beijing: China Light Industry Press(北京:中国轻工业出版社), 2007.
- [6] Qin H, Ma J Y, Chen S J, et al. Infrared Technology, 2015, 1(37): 78.
- [7] LI Hao-guang, LI Wei-jun, QIN Hong, et al(李浩光,李卫军,覃 鸿,等). Transactions of the Chinese Society of Agricultural Machinery(农业机械学报), 2016, 47(6): 259.
- [8] LI Hao-guang, LI Wei-jun, QIN Hong, et al(李浩光,李卫军,覃 鸿,等). Transactions of the Chinese Society of Agricultural Machinery(农业机械学报), 2017, (S1): 422.
- [9] QIN Hong, MA Jing-yi, CHEN Shao-jiang, et al(覃 鸿,马竞一,陈绍江,等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2015, 25(11): 1807.
- [10] Stellan Ohlsson. Deep Learning: How the Mind Overrides Experience. Cambridge: Cambridge University Press, 2011.
- [11] Wang S J, Lai J L. Neurocomputing, 2005, 67: 9.

Study on Near-Infrared Spectrum Acquisition Method of Non-Uniform Solid Particles

LI Hao-guang^{1,2}, YU Yun-hua^{1,2}, PANG Yan¹, SHEN Xue-feng^{1,2}

1. College of Mechanical and Contral Engineering, Shandong Institute of Petrochemical and Chemical Technology, Dongying 257061, China

2. New Energy College, China University of Petroleum(East China), Dongying 257061, China

Abstract In the research of near-infrared spectroscopy (NIRS) qualitative analysis technology, the experimental objects mainly include liquid matter, powder matter and non-uniform solid particles. The liquid and powder materials are evenly distributed, and the spectrum collection and analysis are relatively easy. The non-uniform solid particles are different in size, shape and internal material distribution. Near-infrared spectroscopy of these samples contain the type information to be extracted in qualitative analysis and the individual difference information to be eliminated. Therefore, the analysis of non-uniform solid particles is more difficult than that of liquid or powder materials with uniform distribution. At present, there is no effective NIRS qualitative analysis method for non-uniform solid particles with different sizes and shapes at home and abroad. In this paper, non-uniform solid grain maize is taken as the research object. Based on the study of various spectral acquisition methods, the characteristics of non-uniform solid grain spectral acquisition are analyzed, and the spectral acquisition device for non-uniform solid grain is designed. In order to ensure the objectivity of the experimental results, Five pattern recognition methods such as Naive Bayesian classifier, k -nearest neighbor, support vector machine, BP neural network and Biomimetic Pattern Recognition were used to establish the qualitative analysis model of single grain maize by near-infrared spectroscopy in diffuse reflection and diffuse transmission mode, and the qualitative analysis models established in diffuse reflection and diffuse transmission mode were compared to analyze the effect of embryo orientation on the identification of single grain maize. The effect of the time interval between train and test sets on the identification accuracy under diffuse reflection and diffuses transmission is studied too. The experimental results show that the diffuse transmission model is not easily affected by the lying style of non-uniform solid grains, and the model has better generalization ability, which provides a feasible spectral acquisition method for subsequent research. Taking the non-uniform solid maize grain as the main experimental object, the research on its collection method and qualitative analysis model can provide a useful reference for the qualitative analysis of near-infrared spectroscopy of similar objects, which has important research significance.

Keywords Non-uniform solid particle; Diffuse reflection; Diffuse transmission; Pattern recognition

(Received Jan. 20, 2021; accepted Apr. 11, 2021)