

应用近红外光谱判别烟叶等级模型的可靠性及化学成分特征分析

刘艺琳¹, 张海燕², 彭海根³, 赵龙莲¹, 陶晓秋^{2*}, 李军会^{1*}

1. 中国农业大学信息与电气工程学院, 北京 100083

2. 四川省烟草质量监督检测站, 四川 成都 610041

3. 四川威斯派克科技有限公司, 四川 成都 610041

摘要 应用近红外光谱技术实现农产品产地、质量等级等属性特征的快速、准确鉴别, 在农产品的收购、加工中可发挥重要作用。目前, 应用近红外技术实现上述目标虽有一些文献报道, 但已实际应用的实例却很少, 其主要原因在于所建模型的可靠性尚存在一定问题。以来自四川省内不同部位(不同等级)的烟叶样品为例, 基于主要化学成分, 近红外光谱和定性判别结果评价了烟叶等级识别模型的可靠性; 并通过探究主要化学成分和光谱特征分析了四川省烟叶的等级特征。研究表明: 在一定生态产区范围内, 可建立可靠性较好的烟叶等级识别模型, 物质信息基础和模型识别结果的一致性可验证模型的可靠性。通过探究化学成分和光谱特征, 分析了四川省不同等级烟叶具有的化学成分特征: 上部烟叶具有低总糖, 高烟碱, 高总氮, 高纤维素, 高酰胺的等级特征; 中部烟叶具有高总糖, 中烟碱, 中总氮, 中纤维素, 中酰胺的等级特征; 下部烟叶具有高总糖, 低烟碱, 低总氮, 低纤维素, 低酰胺的等级特征。本文得出的依据物质信息基础判定定性模型可靠性的方法, 以及基于化学成分和近红外光谱特征分析烟叶等级特征的方法, 对其他农产品的建模和化学成分特征分析具有参考价值。

关键词 近红外; 烟叶; 定性判别; 可靠性; 等级特征

中图分类号: O657.33 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2020)10-3260-05

引言

近红外光谱主要体现对含氢基团振动的倍频和合频信息的吸收, 包含了大多数有机化合物的组成信息。基于近红外光谱的分析技术包括了定量分析和定性分析两大方面, 两者已日渐成熟, 能够实现快速, 无损以及实时的物质检测和分[1]。通过近红外定性分析, 能够确定样品的产地, 等级, 真假, 种类, 贮藏时间等。目前, 使用近红外光谱技术来对农产品进行定性判别已经得到了广泛的研究, 如竹子[2], 药用木瓜[3], 圆枣[4], 植物纤维[5], 小麦等。常用的判别农产品属性特征的方法, 除近红外分析技术外, 还有电子鼻技术[7], 矿物元素指纹图谱分析技术等。

对农产品进行定性判别, 不仅有利于鉴别农产品的真假, 品质和等级, 方便市场的分类和分级, 品种的选育等, 而且在收购, 加工等方面, 有利于实现稳定的大规模工业生产。然而, 对于定性判别模型的实际应用仍存在一些问题。

在实际生产中, 对于建立的定性模型, 一般存在能够识别的属性类型不是很多; 识别准确率不高; 定性识别模型的可靠性较低等问题。因此, 应用近红外技术实现上述目标虽有一些文献报道, 但已实际应用的实例却很少。在应用近红外定性模型时, 需对模型的可靠性及样品的物质信息基础进行分析研究。

研究以四川省烟草为例, 在部位判别模型可行的基础上, 结合主要化学成分, 近红外光谱及定性模型判别结果, 通过物质信息基础的分析, 进一步对模型的可靠性进行了验证; 对不同部位烟叶的光谱特征进行分析, 得到了四川省烟叶的部位(等级)特征。

1 实验部分

1.1 仪器

MPA型傅里叶变换近红外光谱仪(德国BRUKER公司)。工作参数: 光谱采集范围12 000~4 000 cm^{-1} ; 光谱分

收稿日期: 2019-07-24, 修订日期: 2019-11-08

基金项目: 国家重点研发计划子课题(2016YFD0700304)资助

作者简介: 刘艺琳, 女, 1996年生, 中国农业大学信息与电气工程学院硕士研究生 e-mail: lyl3350liu@163.com

* 通讯联系人 e-mail: caunir@cau.edu.cn

分辨率: 8 cm^{-1} ; 扫描次数: 64 次。

1.2 样品及数据

选择 2014 年—2016 年收集的四川省不同产区不同部位的烟叶烤烟样品为实验样品。实验样品共 550 份, 如表 1 所示。采用四川省质量监督检测站对样品进行化学成分检测获得的样品各项常规化学成分数据。

表 1 样品及其部位
Table 1 Sample and parts

部位	2014 年	2015 年	2016 年	总计
上部	85	65	50	200
中部	85	65	50	200
下部	85	65	0	150

1.3 方法

使用 CAUNIRS 近红外光谱分析软件, 结合判别偏最小二乘(DPLS)方法建立基于近红外光谱的烟草不同部位(等级)的相似性分析模型。使用二进制系统表示样本所属的类, 设置标准, 其中“1”表示属于该类, “0”表示不属于该类, 类别编码表如表 2。

表 2 类别编码
Table 2 Category code

部位类别	编码		
	A	B	C
3 类	100	010	001

预测值通常接近 0 或 1, 而不是精确地为 0 或 1, 通过比较预测值和阈值之间的关系将其设置为阈值以区分类别。

应用主成分及 Fisher 准则投影分析方法(PPF)方法建立基于近红外光谱的投影分析模型。PPF 是基于光谱的主成分数据与 Fisher 准则对样品进行相似性分析的投影方法。PPF 所实现的最终目标是得到各类中各个样品的低维投影值, 使类内投影值能反映类内的离散度与相邻类间的交叉度, 使各类投影均值能客观表达类间的远近关系。通过投影图得出各类样品的类间距离, 可对各类样品的相似程度进行直观评价; 通过投影图得出各类内离散度, 可对各类内样品的一致性进行直观评价^[7]。

使用 MATLAB 软件及编程语言对光谱文件进行标准正态变量(SNV)、平均等预处理后, 对不同部位的烟叶光谱进行特征分析。根据不同波数范围内的近红外光谱吸收的不同化学成分信息, 对烟叶的等级特征进行分析。

2 结果与讨论

2.1 基于近红外光谱数据建立全部烟叶部位判别模型的结果分析

将四川省全部样品共 550 份烟叶的近红外光谱数据导入 CAUNIRS 软件, 根据部位(上部, 中部, 下部)分为 3 类并分别设置编号。选择分析谱区范围为 $8\ 000\sim4\ 000\text{ cm}^{-1}$, 对光谱进行一阶导数和 15 点平滑预处理后, 利用 DPLS 方法建

模。判别结果如表 3 所示。不同部位(等级)的烟叶之间存在连续性和相似性, 因此当判别一个类别与其相邻类别时, 上部判别为中部, 中部判别为上部或下部, 下部判别为中部都是合理的; 而将一个类别区分为其非相邻类别, 如将上部判别为下部或下部判别为上部是错误的^[8]。

表 3 基于近红外数据的部位判别结果(内部交叉)
Table 3 Part discrimination results based on near-infrared data (internal cross)

识别合理率(%): 97.80		识别错误率(%): 2.20	
烟叶部位	上部	中部	下部
上部	92.86	7.14	0.00
中部	14.37	74.25	11.38
下部	2.20	26.37	71.43

由表 3 可以看出: (1)全部样品部位判别模型总的识别合理率为 97.80%, 达到了实际应用要求。该判别结果说明在一定产区范围内, 产地因素对部位判别结果的影响较小, 此模型具有可行性, 但仍需进一步验证模型的可靠性。(2)上部和下部之间误判较少; 中部与上部、下部之间存在部分误判。即不相邻部位之间存在识别错误较少, 相邻部位之间存在部分样品识别错误。该判别结果可能是由于烟叶采集时的误差及烟叶本身的连续性和相似性造成的, 此种判别结果合理。

根据判别结果分析可知, 在同一生态产区内, 烟叶的部位(等级)识别结果合理率较高, 模型具有可行性。然而此模型的可靠性仍需进一步分析。本文基于主要化学成分, 光谱特征及定性判别结果, 通过讨论不同等级烟叶的物质信息基础, 从而验证部位(等级)识别模型的可靠性, 以及通过探究主要化学成分和光谱特征从而分析烟叶的等级特征。

2.2 部位判别模型的可靠性分析

将烟叶样品的光谱文件按照不同部位, 即上部, 中部和下部, 分别用符号 B, C, X 命名后, 设置分析谱区范围为 $4\ 000\sim8\ 000\text{ cm}^{-1}$ 并进行预处理, 选择主成分为 8 后进行基于 PPF 方法的投影分析, 获得投影图如图 1 所示。

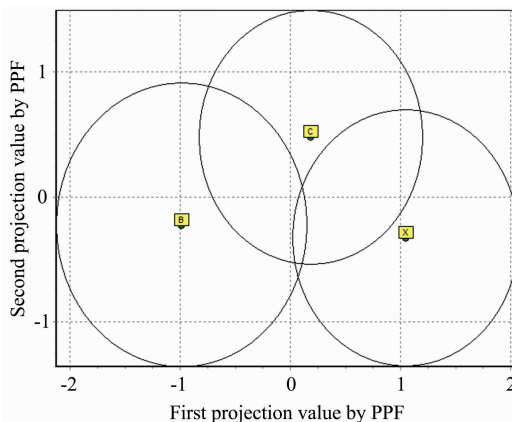


图 1 基于 PPF 的主成分为 8 的投影图

Fig. 1 PPF projection with 8 principal components

由图 1 可以看出: 上部和下部烟叶的光谱投影图交叉部分较小, 表明两者相似性较低; 中部烟叶的光谱投影图与上部和下部的投影图交叉部分都较大, 表明中部烟叶与上部、下部烟叶均具有一定的相似性。此结果符合烟叶的连续性和相似性, 与定性判别结果中上部和下部之间误判较少; 中部和上部、下部之间存在部分误判的结果一致。

根据光谱投影图分析可知, 烟叶不同部位光谱之间的相似性分析结果与定性判别结果一致。因此, 存在基于近红外光谱建立部位(等级)识别模型的光谱信息基础, 此部位(等级)判别模型具有可靠性。

2.3 不同等级烟叶的化学成分特征分析

2.3.1 基于化学成分的烟叶等级特征分析

通过近红外定量分析技术建模得到烟叶样品的各项主要化学成分含量后, 计算不同部位烟叶样品的主要化学成分含量均值: 上部烟叶具有低总糖、高烟碱、高总氮; 中部烟叶具有高总糖、烟碱适中、总氮适中; 下部烟叶具有高总糖、低烟碱、低总氮。

根据化学成分规律分析可知, 不同部位(等级)烟叶的化学成分特征具有明显差异, 与定性判别模型结果一致, 且烟叶化学成分规律具有一定的等级特征。因此, 存在基于光谱建立部位(等级)识别模型的化学成分信息基础, 此部位(等级)判别模型具有可靠性。不同等级烟叶的化学成分特征为: 上部烟叶具有低总糖、高烟碱、高总氮特征; 中部烟叶具有高总糖、中烟碱、中总氮特征; 下部烟叶具有高总糖、低烟碱、低总氮特征。

2.3.2 通过光谱特征的烟叶等级特征分析

为进一步分析烟叶的等级特征, 对不同频率范围内的光谱特征进行分析。将烟叶光谱按照不同部位, 即上部, 中部, 下部分别用符号 B, C, X 命名并设置分析谱区范围为 $4\ 000\sim 8\ 000\ \text{cm}^{-1}$ 后, 导入 MATLAB 软件进行分析。由于样品较多且不同年份采集的光谱存在系统误差, 根据不同部位将光谱进行平均后, 进行 SNV 预处理, 放大光谱图的 $4\ 200\sim 5\ 200\ \text{cm}^{-1}$ 波数范围, 其光谱特征如图 2 所示。不同频率范围内吸收的有机基团信息如表 4 所示。

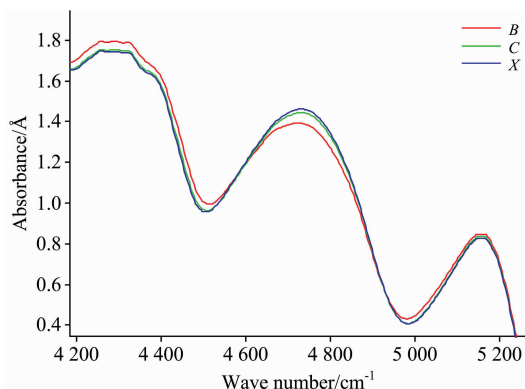


图 2 $4\ 200\sim 5\ 200\ \text{cm}^{-1}$ 范围光谱

Fig. 2 Spectrum during $4\ 200\sim 5\ 200\ \text{cm}^{-1}$

由图 2 可以看出, 上部、中部、下部烟叶的平均光谱存

在不同特征。在不同的频率范围内, 不同部位烟叶光谱的位置特征不同。由图 2, 表 4 可知: 在 $4\ 250\sim 4\ 350\ \text{cm}^{-1}$ 频率范围内, 光谱主要表征吸收含 C—H 基团物质的振动信息, 推测主要为纤维素的含量信息, 即上部烟叶可能具有高纤维素, 中部烟叶可能具有适中纤维素, 下部烟叶可能具有低纤维素; 在 $4\ 700\sim 4\ 800\ \text{cm}^{-1}$ 频率范围内, 光谱主要表征吸收含 O—H 基团物质的振动信息, 推测主要为含糖物质的含量信息, 即上部烟叶具有低总糖, 中部和下部烟叶具有高总糖; 在 $5\ 100\sim 5\ 200\ \text{cm}^{-1}$ 频率范围内, 光谱主要表征吸收含 C=O 与 O—H 基团物质的振动信息, 推测主要为酰胺类物质的含量信息, 即上部烟叶可能具有高酰胺, 中部烟叶可能具有适中酰胺, 下部烟叶可能具有低酰胺。

表 4 不同频率范围吸收的有机基团

Table 4 Organic groups absorbed in different frequency ranges

波数/ cm^{-1}	振动基团	主要结构
$4\ 250\sim 4\ 350$	C—H 等	纤维素
$4\ 700\sim 4\ 800$	O—H 等	淀粉等糖类物质
$5\ 100\sim 5\ 200$	N—H, C=O 等	酰胺

根据光谱特征分析可知: 不同频率范围内烟叶的部位特征不同, 即存在使用数学方法对烟叶进行部位(等级)分类的光谱信息基础, 此部位(等级)判别模型具有可靠性。对不同波数范围内的光谱特征及光谱吸收的化学成分信息进行分析可知: 对于纤维素, 总糖和酰胺, 不同部位(等级)烟叶具有不同的等级特征。

对不同等级烟叶的化学成分特征规律进行总结, 结果如表 5 所示。

表 5 四川省不同等级烟叶的化学成分特征

Table 5 Chemical composition characteristics of different grades of tobacco in Sichuan Province

	总糖	烟碱	总氮	纤维素	酰胺
上部	低	高	高	高	高
中部	高	中	中	中	中
下部	高	低	低	低	低

根据化学成分信息基础和光谱信息基础分析可知, 部位定性判别结果、主要化学成分及光谱相似性具有一致性, 此定性判别模型具有可靠性。由表 5 可知, 通过化学成分规律分析及预处理后光谱的特征分析, 可以得到一定生态产区范围内(四川省)烟叶的部位(等级)特征: 上部烟叶具有低总糖, 高烟碱, 高总氮, 高纤维素, 高酰胺的等级特征; 中部烟叶具有高总糖, 中烟碱, 中总氮, 中纤维素, 中酰胺的等级特征; 下部烟叶具有高总糖, 低烟碱, 低总氮, 低纤维素, 低酰胺的等级特征。此方法对验证其他农产品质量等级等属性特征判别模型的可靠性和分析农产品的等级特征具有参考价值。

3 结 论

通过基于光谱的四川省烟叶部位(等级)判别模型结果,说明了在一定生态产区范围(四川省)内,基于光谱数据建立的部位(等级)定性判别模型识别合理率较高,即产地因素对部位判别结果的影响较小,模型具有可行性。根据主要化学成分、光谱特征及定性判别结果三个方面,进一步验证了模型的可靠性及分析了烟叶样品的等级特征。

对模型判别结果进行分析,发现上部和下部烟叶之间误判较少,上部和中部、中部和下部烟叶之间存在部分误判。对不同部位烟叶的化学成分特征进行分析,发现上部烟叶具有低总糖、高烟碱、高总氮的特征;中部烟叶具有高总糖、烟碱适中、低总氮的特征;下部烟叶具有高总糖、低烟碱、低总氮的特征;即不同部位烟叶的化学成分特征存在明显差异。对光谱的相似性投影结果进行分析,发现上部和下部烟

叶的光谱相似性较低,中部烟叶的光谱特征介于上部和下部之间。通过模型判别结果、化学成分特征及光谱相似性之间的一致性,证明了此模型的可靠性。

根据不同部位烟叶的化学成分特征分析,得到不同等级烟叶的部分化学成分特征。根据不同频率范围内的光谱特征及光谱吸收的物质信息,进一步分析了一定生态产区范围(四川省)内烟叶可能具有的部位(等级)特征。上部烟叶具有低总糖,高烟碱,高总氮,高纤维素,高酰胺的等级特征;中部烟叶具有高总糖,中烟碱,中总氮,中纤维素,中酰胺的等级特征;下部烟叶具有高总糖,低烟碱,低总氮,低纤维素,低酰胺的等级特征。

本文中的分析方法,可以验证烟叶部位(等级)识别模型的可靠性,以及分析烟叶的等级特征。此分析方法对其他农产品的质量等级等属性特征的建模和特征分析具有参考价值。

References

- [1] YAN Yan-lu, CHEN Bin, ZHU Da-zhou(严衍禄, 陈 斌, 朱大洲). Principle, Technology and Application of Near Infrared Spectroscopy (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), 2013.
- [2] Yang Z, Li K, Zhang M, et al. Biotechnology for Biofuels, 2016, 9(1): 35.
- [3] Han B, Peng H, Yan H. Pharmacognosy Magazine, 2016, 12(46): 93.
- [4] WANG Wan-jiao, HE Xiao-guang, YANG Xiao-chen, et al(王婉娇, 贺晓光, 杨晓忱, 等). Food Technology(食品科技), 2015, (6): 344.
- [5] Wei J, Chengfeng Z, Guangting H, et al. Frontiers in Plant Science, 2017, 7.
- [6] HUANG De-dong, HE Wei-wei, JIN Ling, et al(黄得栋, 何微微, 晋 玲, 等). China Journal of Experimental Traditional Medical Formulae(中国实验方剂学杂志), 2017, (23): 22.
- [7] WEN Ya-dong, WANG Yi, WANG Neng-ru, et al(温亚东, 王 毅, 王能如, 等). Acta Tabacaria Sinica(中国烟草学报), 2009, 15(5): 6.
- [8] Luan L, Wang Y, Li X, et al. Journal of Near Infrared Spectroscopy, 2016, 24(4): 363.

Reliability and Chemical Composition Analysis of Tobacco Leaf Grade Model by Near-Infrared Spectroscopy

LIU Yi-lin¹, ZHANG Hai-yan², PENG Hai-gen³, ZHAO Long-lian¹, TAO Xiao-qi^{2*}, LI Jun-hui^{1*}

1. College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China

2. Sichuan Tobacco Quality Supervision and Inspection Station, Chengdu 610041, China

3. Sichuan Vspec Technology Co., Ltd., Chengdu 610041, China

Abstract The application of near-infrared spectroscopy technology for the rapid and accurate identification of attributes such as the origin and quality grade of agricultural products can play an important role in the acquisition and processing of agricultural products. However, the practical application of near-infrared technology for qualitative identification is rare because of certain problems related to the reliable of the model. Taking tobacco leaf samples with different parts (grades) in Sichuan Province, China as an example, this paper discusses a method for determining whether it is reliable to a grade identification model according to the main chemical composition, near-infrared spectroscopy and qualitative discrimination results; and a method for analyzing tobacco leaves grade characteristics according to main chemical composition and near-infrared spectroscopy. Within a certain ecological production area, a reliable leaf grade recognition model can be established, and the consistency of the material information base and the model identification results can verify the reliability of the model. By exploring the chemical composition and spec-

tral characteristics, the possible grade characteristics of tobacco leaves in Sichuan Province are analyzed; the upper tobacco leaves have low total sugar, high nicotine, high total nitrogen, high cellulose and high amide content; the middle have high total sugar, middle nicotine, middle total nitrogen, middle cellulose and middle amide content; the lower has a high total sugar, low nicotine, low total nitrogen, low cellulose and low amide content. The methods used herein, such as the method for determining the reliability of qualitative models based on the material information basis and the method for analyzing the characteristics of tobacco grades based on the chemical composition and spectral characteristics, have reference value for modeling the attributes and chemical composition analysis of other agricultural products.

Keywords Near-infrared; Tobacco; Qualitative discrimination; Reliability; Grade feature

(Received Jul. 24, 2019; accepted Nov. 8, 2019)

* Corresponding authors