

卷烟工业

领导信箱

ldxx@tobacco.gov.cn

烟草论坛

留言板

电子邮件定制

短信互动

国家烟草专卖局总机

010-63605000

新闻投稿热线:

010-63606303

010-63605947

010-63605142

cx-out@tobacco.gov.cn

多变量分析法结合近红外光谱表征卷烟配方的过程质量

2006-11-10

长期以来,在我国卷烟制造业中,对卷烟配方工艺过程以及产品质量的分析与控制主要以物理质量为主体,而对于内在主要化学成分方面的信息则因常规化学分析周期长、时间滞后等原因难以实现现场分析,以致于在配方过程中,应用多参量综合信息对卷烟配方的均匀性和产品内在质量的稳定性进行分析与控制还处于探索阶段。现代化学计量学诞生至今已有 20 多年,将化学计量学中的多变量分析方法(如多元校正、模式识别)与近红外光谱分析技术相结合,通过建立相应的校正模型和分类模型,对复杂的多组分体系进行分析,不但可预测未知样品的多个组分或性质,同时还可以对其品质属性进行分类,因该项分析技术具有样品前处理简便、分析速度快、精度高等优点,已被广泛应用于石化、农业、食品和医药等领域^[1-2],近年来也被应用于国内烟草行业的原料质量控制和卷烟品质检测等方面^[3-5]。随着现代近红外光谱分析技术的日趋普及,面向复杂化学体系的分析,无论是离线分析还是在线分析,化学计量学的方法将不断渗透。为此,采用多变量分析方法中的偏最小二乘(PLS)法、主成分分析-马氏距离(PCA - Mahalanobis)法结合近红外光谱(NIRs),对复烤分组加工以及卷烟配方等过程质量进行分析探讨。

1 基本原理

在分组加工或卷烟配方过程中,按一定采样周期,应用 FT - NIR 分析仪,采集不同品质特征配方模块的近红外光谱数据,即可构成样本的光谱数据集。对所采集的任意一个样本 x_i 均可用一组参量来表征,这些参量其实就是原始的光谱数据,可用向量表示为

数据,可用大童表示为:

$$x_i = [x_{i1} \quad x_{i2} \quad \cdots \quad x_{in}]$$

那么,对 m 个样本构成的样本集,就可表达为 $m \times n$ 阶的光谱数据矩阵:

$$X_{m \times n} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix}$$

在矩阵 $X_{m \times n}$ 中,既包含样本性质的信息,也包含样本特征或变量的信息。从数学角度来看,样本集可由特征坐标轴构成的 n 维特征空间中的点集来表示。一个样本对应一个点, n 个特征包含了样本性质的全部信息,样本与样本之间的关系、变量与变量之间的关系以及样本与变量之间的关系等,就是多变量分析要处理的 3 个基本问题^[6]。

在多变量分析方法中,主成分分析-马氏距离法^[7-8]是一种较为实用的方法。采用该方法建立不同品质特征配方模块的类模型,便可表征在样本集中样本与样本之间、样本与类之间的隐含关系,即对矩阵 $X_{m \times n}$ 进行主成分分析,得到在主成分 f 下的得分矩阵 $T_{m \times f}$ 和载荷矩阵 $P_{f \times n}$ 。其中,得分矩阵反映了样本与样本之间的关系。通过样本集的得分矩阵计算马氏距离,主要步骤为:

设得分矩阵 $T_{m \times f}$ 中任意样本 i 的得分向量为 $t_{i(1 \times f)}$,那么,所有样本得分向量的平均向量为 $\bar{t}_{(1 \times f)} = \frac{1}{m} \sum_{i=1}^m t_i$;经中心化处理后,任意样本 i 的得分向量为 $t_{i(1 \times f)}^{cent} = t_{i(1 \times f)} - \bar{t}_{(1 \times f)}$;则样本的马氏距离为:

$$D_i^2 = t_i^{cent} M^{-1} t_i^{cent'}$$

式中: M^{-1} ——Mahalanobis 矩阵 M 的逆矩阵, $M_{(f \times f)} = \frac{T_{(m \times f)}^{cent' } T_{(m \times f)}^{cent}}{m-1}$; $T_{(m \times f)}^{cent}$ ——中心化得分矩阵, $T_{(m \times f)}^{cent} = \{t_i^{cent}\}$ 。

在构建类模型时,样本集中点的分布规律是“物以类聚”,即不同特征的样本点位于模式空间中不同的区域,具有相似性质或是相似特征的样本在模式空间中处于相近的位置,聚集为一类,特征的集合就构成了模式。不同品质特征的配方模块可形成不同的类,在每一类中,通过马氏距离可表征样本点的聚集程度,马氏距离变异越小,样本性质或特征的相似程度就越高;反之变异越大,其相似程度就越低,或是归属其它类别。因此,应用

相似性及灵敏度,或是归属其它类别。因此,应用马氏距离可以表征在分组加工过程或配方过程中模块质量的均匀性和稳定性;应用建立的类模型也可进行辅助模块的替代性研究,以寻求品质特征相似的模块互换。与此同时,通过校正模型还可快速定量监测配方模块的多组分内在化学成分信息。因近红外光谱谱带较宽,相互重叠严重,谱带归属困难,对近红外光谱进行定量分析时,为了获得理想的模型,通常应用多变量分析方法中的偏最小二乘(PLS)法^[9],并用交互校验(Cross-validation)方法^[7]确定最优主因子数(Optimum Factor)来建立样本性质矩阵与光谱矩阵之间的定量关系——校正模型。

2 材料与方 法

2.1 主要实验设备及工作参数设置

Nicolet Antaris FT-NIR 光谱仪,包括 InGaAs 检测器的漫反射积分球、5cm 石英采样杯、旋转台等附件和 TQ Analyst 6.2 数据处理软件(美国 THERMO NICOLET 公司)。

仪器的主要工作参数:光谱扫描范围 10000 ~ 4000 cm^{-1} ;分辨率 8 cm^{-1} ;扫描次数 72。

2.2 方 法

采用多变量分析方法结合近红外光谱表征过程质量,基本实验方法见图 1。

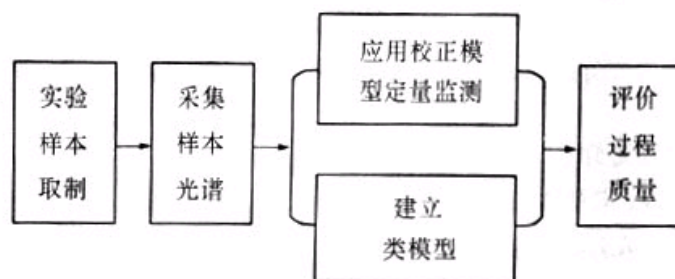


图 1 基本实验方法

2.2.1 建立校正模型

选择 1200 个具有代表性的样品,在 60 ~ 70 $^{\circ}\text{C}$ 下烘至含水率 12% 左右,粉碎,过 60 目筛,常温下避光密封贮存,即为建模样本。在常温下开机预热光谱仪 2h,将样本装入 5cm 石英杯中,容量过半,轻压至平,置于旋转台,即可采集建模样本的近红外漫反射光谱数据。同时,按文献[10-12]的方法分别测定建模样本的总氮、烟碱和总糖等主要化学成分的含量。然后,应用 PLS 法构建建模样

要化学成分的含量。然后,应用 PLS 法把建模样本的光谱数据与相应的化学成分含量数据结合“剔一”(Leave-one-out)的交互校验方法进行拟合建模,逐步优化模型,使其达到最佳状态。最后,采用独立校验集进一步验证模型,以保证校正模型的预测准确性。

2.2.2 采集实验样本的光谱数据

按 30min 1 次的采样频率,现场采集传统配方 A、三级配方 B 两种模式下不同模块的复烤片烟(F1 与 F2)和配方烟丝(P1 与 P2)作为实验样本,样本容量均为 F1 和 F2 各 110 个,P1 和 P2 各 40 个。将样本粉碎,过 60 目筛,即可采集相应的近红外光谱数据。与此同时,调用已建好的校正模型,进行过程内在质量监测,获取相应的品质信息。

2.2.3 建立类模型

应用主成分分析-马氏距离法,对所采集的 A、B 两种配方模式下的复烤片烟和配方烟丝的光谱数据分别建立相应的类模型。

2.2.4 评价过程质量

对定量监测结果以及类模型的有关马氏距离进行数理统计分析,以评价和表征过程质量。

3 结果与讨论

3.1 校正模型的建立与应用

为使所建立的校正模型在应用时具有较好的适应能力和预测准确性,选择建模样本时应充分考虑样本组分不同含量分布的均匀性、结构的代表性和样本基体的一致性,要尽可能避免组分含量形成正态分布,同时还要注意不同等级、不同品种烤烟样本数量的均衡性。对于初烤烤烟与陈化烤烟,因二者之间存在明显差异,应分开建模,以保证校正模型能获得较为准确的预测结果。

在应用 PLS 结合交互校验法的建模过程中,剔除异常值是优化模型的一个关键步骤。异常值通常有两类,一类由来自基础数据的测定误差较大所致,对回归分析影响明显,一般通过残差分析,比较容易剔除,或是再次测定样本进行修正;在建模过程中还常存在另一类异常值(也称为强影响点),其光谱不具代表性,为远离模型整体的平均光谱,常表现出残差值不一定很大,甚至很小,但杠杆值(Leverage)较大,且具有较强的相互掩蔽效应,对模型的稳健性有强烈的扰动作用,一般可采用 Maha-

lanobis 距离、杠杆值等指标联合诊断^[6],并在交互校验过程中逐步剔除,进而确定建模的最优主因子数。图 2 为初烤烤烟烟碱的测定值(Actual)与模型预测值(Calculated)的散点图和相应的主因子数与全交互校验均方残差的变化图,其中最优化主因子数为 12,所对应的 RMSECV 最小。

同理,其余初烤烤烟的总氮、总糖等组分的建模也如此,用最优化主因子数建立的校正模型的数理指标见表 1。通常,应用全交互校验法建立的模型,根据以上这些主要数理指标基本上可衡量模型的预测能力。但在建模过程中,难免要剔除一些样本,校正集的代表性就可能因此受到影响,因此当模型建好后,有必要采用独立校验集进一步对其进行验证。本实验另行随机抽样(不低于 30 个)作为独立校验集,用建好的模型进行预测,将其结果与相应的化学测定值通过配对 t -检验

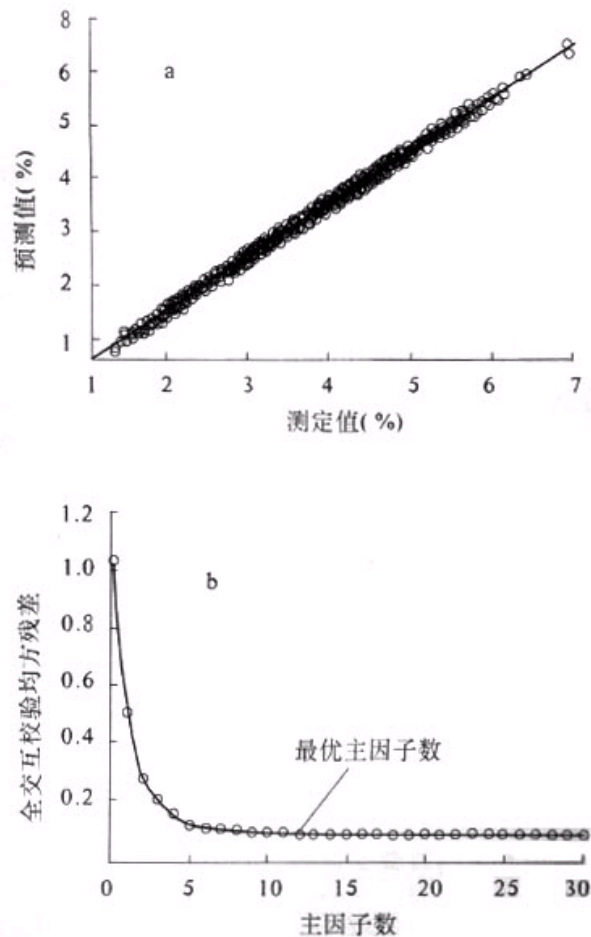


图 2 烟碱测定值与模型预测值的散点图(a)

和全交互校验均方残差与随主因子数的变化图(b)

($\alpha = 5\%$)进行确认,只有在两种测定方法不存在显著性差异的条件下,提交的模型才能付诸应用。

应用已验证的校正模型监测 A、B 两种配方模式下的过程中在质量,以复烤片烟 C2 模式为

模式下的过程内在质量,以及烤片烟 F2 模块为例,所获取的品质信息统计结果见表 2。

3.2 类模型的建立

应用 A、B 两种配方模式下复烤片烟和配方烟丝的光谱数据建模。为使类模型既获得理想的分类效果,同时又方便于 A、B 两种配方模式下不同模块过程质量的分析评价,通过反复寻优,本实验在相对统一的条件设置、预处理方法下建模,见表 3。所建立的类模型以复烤片烟 F2 模块和配方烟丝 P2 模块为例,分类图见图 3。在图 3 中,图(a)和图(b)分别表示 A、B 两种配方模式下, F2 模块和 P2 模块在模式空间中的类与类之间(A - F2 与 B - F2 之间、A - P2 与 B - P2 之间)类间样本点的聚集和离散程度,如图(a)的横坐标表示 B - F2 模块到 A - F2 模块的距离。在各模块中样本点的马氏距离的有关数理统计结果见表 4。

3.3 过程质量分析与评价

从以上结果可知,与传统配方(A)模式相比较,在三级配方(B)模式下,无论是复烤片烟模块

表 1 初烤烟总糖、烟碱和总氮等组分校正模型的有关数理指标

模型名称	校正集样本数量	相关系数	最优主因子数	交互校验均方残差	预测范围 (%)
总糖	1133	0.9955	12	0.7102	10.00 ~ 42.00
烟碱	1163	0.9982	12	0.0820	1.00 ~ 6.00
总氮	1112	0.9859	7	0.0824	1.00 ~ 3.00

表 2 两种配方模式下复烤片烟 F2 模块总糖、烟碱和总氮等组分的监测^①结果统计

组分名称	配方模式	平均值 (%)	标准偏差 (%)	变异系数 (%)
总糖	A	23.32	1.0009	4.29
	B	22.57	0.8221	3.64
烟碱	A	3.13	0.1846	5.90
	B	3.53	0.1805	5.11
总氮	A	2.14	0.0732	3.42
	B	2.28	0.0760	3.33

注:监测样本容量 107。

(F1, F2), 还是配方烟丝模块(P1, P2), 其相应的马氏距离变异系数均有不同程度减小(表 4), 即类间样本点的聚集程度提高, 说明在三级配方模式下, 混配的均匀性和模块质量的稳定性均优于传统的配方模式。当然, 也可通过内在化学组分变

传统的配方模式。当然,也可通过内在化学组分变异系数来衡量模块质量的稳定性(表2)。但值得注意的是,若选取过少的或变异小的组分来评价过程质量的稳定性,就可能存在潜在的风险。如在两种配方模式下,复烤片烟 F2 模块的总氮含量变异系数相差很小,就难以说明两种配方模式下模块质量稳定性的差异。因校正模型提取的是光谱中较为单一的信息,只有综合更多的变异明显的特征组分(如总糖、烟碱等主要组分),才能更好地反映两种配方模式过程质量稳定性的差异。与之相比,应用主成分分析-马氏距离法,可充分地提取光谱所包含的特征信息(已知的或隐含的),较为全面地表征模块中样本点聚集的离散程度。用马氏距离对时序作图,便可反映出模块在加工过程中的质量波动情况。若要定量监测过程质量,获取必要的多组分定量品质信息,就必须建立相应的校正模型。

红河卷烟厂三级配方模式的运行结果表明,无论是打叶复烤还是制丝等生产过程,配方模块的质量稳定性和均匀性与传统配方模式相比,均获得了提升,整个三级配方的应用效果在产品内在质量的稳定性方面得到了体现。由表5可知,“红河”一类卷烟产品评吸得分的变异系数均小于传统的配方模式。

表3 应用主成分分析-马氏距离法建立类模型的有关参数设置

条件	复烤片烟模块 (F1,F2)	配方烟丝模块 (P1,P2)
校正集样本容量	107	38
最大主成分数	10	10
谱区(cm^{-1})	7700 ~ 4120	7700 ~ 4120
数据预处理	中心化	中心化
信号降噪处理	二阶导数 + Norris 滤波器(13,5)	二阶导数 + Norris 滤波器 (13,3)

表4 在类模型各模块中样本点的马氏距离的有关数理统计结果

类模型	配方模式 及模块	平均值	标准 偏差	变异系数 (%)
复烤 片烟	A-F1	0.9668	0.2169	22.43
	B-F1	0.9747	0.1855	19.03

配方模型	A - F2	0.9510	0.1989	20.92
	B - F2	0.9820	0.1629	16.59
配方烟丝模型	A - P1	0.9640	0.2136	22.16
	B - P1	0.9706	0.1740	17.93
模型	A - P2	0.9694	0.1870	19.29
	B - P2	0.9513	0.1808	19.01

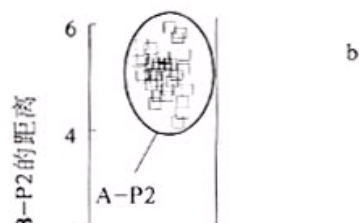
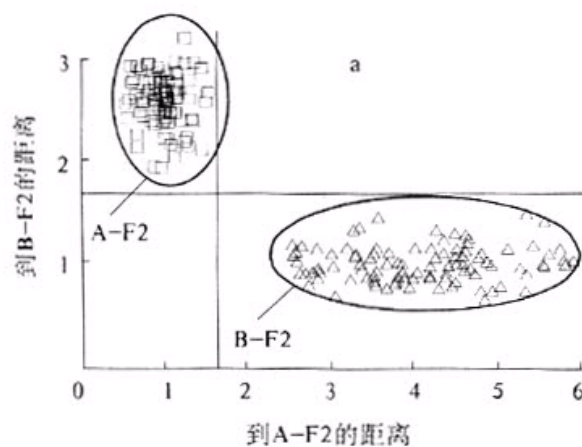
表5 A、B两种配方模式下“红河”一类卷烟评吸得分的变异系数比较

配方模式	香气	谐调	杂气	刺激	余味
A	2.30	1.56	2.57	2.29	2.23
B	1.81	1.46	1.66	2.06	2.10

注:数据来源于红河卷烟厂产品中心。

4 结论

实验结果表明,应用多变量分析方法(PLS 和 PCA - Mahalanobis)结合近红外光谱(NIRS)表征卷烟配方的过程质量是可行的,若有必要获取更为密集的信息,可选择在线近红外光谱仪进一步提高光谱数据的采样效率。值得深入研究的是,欲引入本方法进行统计过程控制(SPC - Statistical Process control),必须针对仪器性能和具体对象合理选取马氏距离的控制限度。基于该方法既能充分利用量测信息表征过程质量,同时又能快速获得过程多组分品质信息的实用性,因此将其应用于过程质量控制与分析以及产品质量监控等方面具有一定的实用价值。



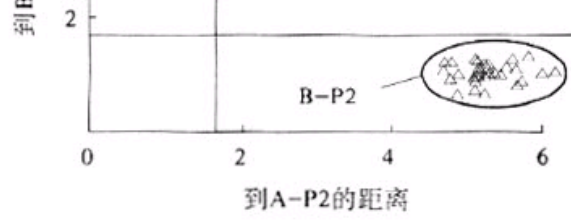


图3 A、B两种模式下复烤片烟 F2 模块(a)和配方烟丝 P2 模块(b)的分类示意图

(红河卷烟厂
北京化工大学材料学院)
王家俊 袁洪福 陈剑明 师建全
摘自《烟草科技》2006年第10期



主管：国家烟草专卖局办公室
地址：中国北京西城区月坛南街55号(100045)
建议使用：800*600分辨率以上，IE5.0以上浏览器
未经许可，本网站包括图像、图标、文字在内的所有数据不得转载

主办：国家烟草专卖局信息中心
备案序号：京ICP备05033420号