

近红外光谱的投影分析方法在工业分级与复烤模块配方中的应用

温亚东^{1,2}, 王毅¹, 王能如², 马翔¹, 李军会³

1 红塔集团技术中心, 云南省玉溪市红塔大道 118 号 653100;

2 中国科技大学, 安徽省合肥市美菱大道 121 号 230052;

3 中国农业大学信息与电气工程学院, 北京市海淀区清华东路 17 号 100083

摘要: 为提高工业分级的稳定性, 使复烤配方更加科学合理, 建立了基于近红外光谱信息目标主成分及 Fisher 准则的投影方法 (PTF), 并通过软件程序来实现结果分析。研究表明: 应用 PTF 方法得到的投影值能客观表达不同类间的远近关系, 以及同一类内的离散度与相邻类间的交叉度, 使类间离散度与类内离散度之比极大化, 为工业分级提供了客观科学的依据, 对复烤模块配方设计有重要的指导作用。

关键词: 近红外光谱; 投影分析; 目标主成分; Fisher 准则; 工业分级; 复烤模块配方

doi: 10.3969/j.issn.1004-5708.2009.05.002

中图分类号: TS44

文献标识码: A

文章编号: 1004-5708(2009)05-0006-05

Application of NIR spectra projection method in tobacco leaf grading and blending in redrying module

WEN Ya-dong^{1,2}, WANG Yi¹, WANG Neng-ru², MA Xiang¹, LI Jun-hui³

1 R&D Center, Hongta Group, Yuxi 653100, China;

2 University of Science & Technology of China, Hefei 230052, China;

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

Abstract: Projection method based on target principal component of NIR spectra and Fisher criterion was established to improve leaf grading stability and redrying module credibility. Result showed that the application of PTF method can exactly describe classification of tobacco samples, and the projection values indicated the distance of different categories, the deviation of one category, and the intersection of two nearby categories, and hence the maximum of the ratio between the deviation of different categories and the deviation of one category was realized. The application of this method can facilitate tobacco leaf grading and redrying module designing.

Key words: NIR spectra; projection method; target principal component; Fisher criterion; tobacco leaf grading; redrying module

卷烟产品风格特色的稳定是消费者的需求,也是配方设计人员追求的目标,其中原料烟叶质量特色的稳定起关键作用。通过对烟叶的工业分级,并根据其风格质量特征掺配出具有一定特色和品质一致的烟叶

复烤配方模块,在复烤工艺生产中对烟叶的主要物理指标与化学指标进行检测和控制^[1-2],保证稳定、优质的烟叶原料供应,可以提高卷烟产品的内在质量。

近红外光谱分析技术为烟草企业提供了快速、高效并能实现在线应用的分析手段^[3-4],已成为许多烟草企业进行生产过程质量控制及辅助配方设计的重要依据。应用近红外定量分析烟草中的总糖、还原糖、总氮、烟碱等化学指标,国内外许多烟草企业均已做了大

作者简介: 温亚东,男,在读硕士,主要从事烟草近红外技术应用与原料研究。Tel: 0877-2965139, E-mail: wendy@hongta.com

收稿日期: 2008-11-13

量的研究和应用^[4-6],其中红塔集团于2005年率先将该技术成功应用到复烤在线质量监控中,对复烤成品进行主要化学指标的标识,为叶组配方设计和执行提供参考。

总糖、还原糖、总氮、烟碱等化学成分的检测只利用了近红外光谱中的部分信息,全信息的定性分析方法越来越受到人们的重视,特别是一些经典和新型的有监督模式识别方法如^[5-8]:PCA-NN(主成分-最近距离判别法)、DPLS(定性偏最小二乘)、DANN(定性人工神经网络)、PCA-Fisher(主成分-费歇尔判别法)、SIMCA(Soft Independent Modelling of Class Analogy)等。

在烟草近红外定性分析中,应用主成分—马氏距离判别法于烟叶风格及产地的识别^[9-10]与应用SIMCA方法建立烟丝配方模块的类模型,并应用验证样品与类模型的“类距离”和“置信度”2个参数对烟丝配方模块的宏观稳定性进行监测和评价^[2]已有成功的范例。而基于目标主成分及Fisher准则的投影方法(Projection Method Based on Target Principal Component and Fisher Criterion—简称为PTF),是红塔烟草集团和中国农业大学联合开发的“烟草配方设计及在线质量监控软件”中的一种新方法。通过PTF方法在烟叶复烤配方中的实际应用,得到的投影值能客观表达不同类间的远近关系,以及同一类内的离散度与相邻类间的交叉度,使类间离散度与类内离散度之比极大化,为工业分级提供了客观科学的依据,对复烤配方设计有重要的指导作用。

1 PTF方法原理

PTF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。与SIMCA方法具有相似相逆性,相似性指该2种方法均是有监督-无监督特征提取相结合的模式;SIMCA的方法是首先对每个独立的类进行无监督方式的特征提取,然后利用样品在主成分空间中不同类别的样品之间的类距离来对类的归属进行判别^[5],从而实现有监督方式所能实现的判别分析能力;而PTF方法则是首先对类的群体进行2次有监督方式(分别为目标主成分提取和Fisher准则提取)的特征提取,然后利用提取的特征向量逐步计算得到各类中各个样品的投影值,而投影值所反映的类与类之间的远近关系,属于无监督模式(如系统聚类)所能实现的功能。

1.1 目标主成分提取

对得到的光谱数据矩阵 A_{nm} (n :样本数, m :光谱数据点数),传统主成分提取一般是通过主成分迭代等算法^[5]将其分解为 T_{nk} 、 W_{km} 和 E_{nm} (k :主成分数),并且满足 $T=AW+E$,其中 T 为求得的主成分矩阵, W 为得到 T 所需的载荷矩阵或权重矩阵, E 为残差矩阵。对得到的样本光谱数据矩阵 A_{nm} 中如包含 q 类,每一类平均光谱构成的平均光谱矩阵记为 \bar{A}_{qm} (q :类个数),由 \bar{A}_{qm} 经主成分迭代算法得到的载荷矩阵记为 \bar{W}_{km} ,本文中称由 \bar{W}_{km} 和 A_{nm} 主计算得到的主成分矩阵为目标主成分 T_{nk} ,每一类平均主成分构成的平均主成分矩阵称为目标平均主成分 \bar{T}_{qk} 。该方法的实质是使提取的目标主成分信息体现的类间差异极大化。

1.2 Fisher准则提取

基于Fisher准则的线性鉴别已被公认为特征抽取的最好方法之一,特别是在人脸识别和图像识别等领域得到了广泛的应用^[11]。设 S_b 和 S_w 分别表示训练样本的类间散布矩阵和类内散布矩阵,Fisher准则函数定义如下:

$$Jf(x) = x^T S_b x / x^T S_w x, x \neq 0$$

Fisher准则函数能够将样本在投影矢量上的类间离散度和类内离散度结合在一起,为确定最优投影方向提供了一个准则。取极大化目标函数 $Jf(x)$ 的矢量 x 作为投影方向,其物理意义是:特征样本在 x 上投影后,类间离散度和类内离散度之比达到最大^[11],投影矢量 x 的求解方法参见相关文献^[12-15]。

1.3 PTF算法程序的实现步骤

1)对包含 q 类样本光谱数据矩阵 A_{nm} 计算得到平均光谱矩阵 \bar{A}_{qm} 。

2)对 \bar{A}_{qm} 进行预处理后,应用主成分迭代算法计算得到目标载荷矩阵 \bar{W}_{km} 。

3)对 A_{nm} 进行2)中相同的预处理后,应用 \bar{W}_{km} 计算得到目标主成分矩阵 T_{nk} 和目标平均主成分矩阵 \bar{T}_{qk} 。

4)对包含 q 类样本的目标主成分矩阵 T_{nk} ,分别计算得到类间散布矩阵 S_b 和类内散布矩阵 S_w 。

5)依据Fisher准则,确定目标主成分矩阵的最优投影矢量 x 。

6)应用投影矢量 x 分别计算得到各个样本的投影值及各类样本的投影均值。

7)各类样本投影均值及类内样本投影值离散度的图形和数据报告的输出。

2 材料与方法

2.1 材料

采用红塔集团 2006 年工业分级抽检样品 8 mm 烟片近红外光谱,烟叶品种为 K326,样品数大于 25 的作为测试样品,共有 33 个工业分级等级,相同等级取样个数在 25~100 个之间,全部样品数共计 1465 个。

2.2 仪器与设备

MPA 型傅立叶变换近红外光谱仪(德国 BRUKER 公司)。工作参数:光谱采集范围 12000~4000 cm^{-1} ;光谱分辨率:8 cm^{-1} ;扫描次数:64 次。

2.3 方法

采用红塔烟草集团和中国农业大学联合研制的“烟草配方设计及在线质量监控软件”进行数据处理。数据处理各环节中采用的数学方法和参数分别为:

- 1) 谱区选择范围采用 7000~4000 cm^{-1} 。
- 2) 光谱预处理方法采用 1 阶导数 17 点平滑处理。
- 3) 目标主成分个数分别采用 5 个(9 类样品处理)和 10 个(33 类样品处理)。
- 4) 目标主成分矩阵投影矢量 x 的计算方法采用迭代算法求解。
- 5) 类内样本投影值离散度采用方差计算公式。

3 结果与讨论

3.1 工业分级样品的 PTF 投影分析结果

从 33 类工业分级样品中依据外观品质特性挑选出整体差异性明显、部分分级等级间相似程度较高的 9 类具代表性样品进行 PTF 投影分析。类内投影值离散度和第 1 维投影均值见表 1,各类投影均值之间距

离见表 2,各类投影均值之间距离(注:第 i 类与第 j 类之间的距离在一维投影情况下为:第 i 类投影均值减去第 j 类投影均值所得数值的绝对值;在多维投影值情况下为其两两投影值之间的欧氏距离)与类内投影值离散度之间的比值见表 3(注:第 i 类与第 j 类之间的比值 = |第 i 类投影均值 - 第 j 类投影均值| * 2 / (第 i 类投影值离散度 + 第 j 类投影值离散度))。其投影图分析结果见图 1,图 1 中的圆圈从左到右分别以各类投影均值为圆心,半径表示类内投影值的离散度。

根据表 1 中类内投影值的离散度数据来评价工业分级样品类内的一致性与稳定性;表 2 中各类投影均值之间的距离可评价工业分级样品类间的差异和趋势,表 3 中各类投影均值之间距离与类内投影值离散度的比值,描述其差异和趋势的一个相对量指标;图 1 中将类内投影值的离散度用不同半径的圆来表示,同时按照类间距离的差异和趋势排序,可直观地评价和描述工业分级样品类内一致性与稳定性以及工业分级样品类间的差异程度。

表 1 类内投影值的离散度与第 1 维投影均值

分级名	离散度	第 1 维投影均值
BBL1	0.50	0.182
BBL2	0.36	0.396
BB01	0.35	-0.479
BB02	0.38	-0.364
BOA1	0.39	-1.159
BOA2	0.34	-1.297
CO1	0.37	0.519
CO3	0.50	1.043
MOA1	0.44	1.468

表 2 各类投影均值之间的距离

分级名	BBL1	BBL2	BB01	BB02	BOA1	BOA2	CO1	CO3
BBL2	0.21							
BB01	0.67	0.88						
BB02	0.55	0.76	0.12					
BOA1	1.33	1.55	0.67	0.80				
BOA2	1.48	1.70	0.82	0.94	0.15			
CO1	0.34	0.13	1.00	0.88	1.29	1.82		
CO3	0.86	0.65	1.53	1.46	2.30	2.35	0.51	
MOA1	1.29	1.08	1.95	1.83	2.62	2.77	0.95	0.43

表 3 各类投影均值之间距离与类内投影值离散度之间的比值

分级名	BBL1	BBL2	BBO1	BBO2	BOA1	BOA2	CO1	CO3
BBL2	0.49							
BBO1	1.58	2.48						
BBO2	1.25	2.05	0.33					
BOA1	2.99	4.13	1.81	2.08				
BOA2	3.52	4.86	2.38	2.61	0.41			
CO1	0.78	0.36	2.78	2.35	3.39	5.13		
CO3	1.72	1.51	3.60	3.32	5.17	5.60	1.17	
MOA1	2.74	2.70	4.94	4.46	6.31	7.10	2.35	0.91

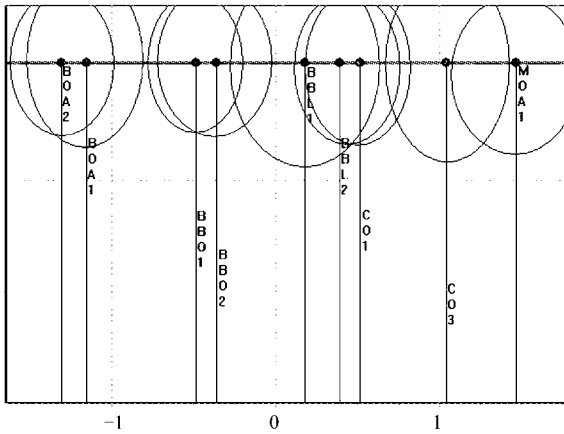


图 1 9类工业分级样品 PTF 投影
(横坐标为投影值,纵坐标无意义)

通过表 1、表 2、表 3 以及图 1 的数据表明：PTF 投影分析方法能够使类间离散度和类内离散度之比的最大值超过 7,在 36 个两类组合中其比值数据超过 1 的有 30 个,该水平远大于用无监督方式提取主成分,然后直接利用欧氏距离计算类间离散度和类内离散度达到的水平。用无监督主成分-欧氏距离法计算的数据分别见表 4、5、6。

表 4 无监督主成分-欧氏距离法计算得到的类内离散度

分级名	离散度	分级名	离散度
BBL1	1.2196	BOA2	0.6928
BBL2	0.7752	CO1	0.8421
BBO1	0.8129	CO3	1.0811
BBO2	0.741	MOA1	0.8627
BOA1	0.7465		

表 5 无监督主成分-欧氏距离法计算得到的各类均值之间的距离

分级名	BBL1	BBL2	BBO1	BBO2	BOA1	BOA2	CO1	CO3
BBL2	0.62							
BBO1	0.30	0.37						
BBO2	0.56	0.14	0.28					
BOA1	0.29	0.47	0.11	0.36				
BOA2	0.46	0.29	0.18	0.17	0.22			
CO1	0.08	0.58	0.24	0.51	0.23	0.41		
CO3	1.13	0.51	0.87	0.60	0.96	0.75	1.08	
MOA1	1.34	0.72	1.08	0.80	1.16	0.95	1.29	0.21

表 6 无监督主成分-欧氏距离法计算得到的类间类内比值

分级名	BBL1	BBL2	BBO1	BBO2	BOA1	BOA2	CO1	CO3
BBL2	0.62							
BBO1	0.30	0.47						
BBO2	0.57	0.18	0.36					
BOA1	0.29	0.61	0.15	0.50				
BOA2	0.48	0.40	0.24	0.22	0.30			
CO1	0.08	0.71	0.29	0.65	0.29	0.53		
CO3	0.98	0.55	0.92	0.66	1.05	0.85	1.12	
MOA1	1.29	0.88	1.29	1.00	1.44	1.22	1.51	0.22

通过对工业分级样品的近红外光谱进行 PTF 投影分析,更加充分全面地提取近红外光谱所包含的丰富的、有价值的信息,能够对工业分级的应用效果进行验证和评价,为工业分级的正确执行和适时调整提供了重要的参考依据。

3.2 PTF 在复烤配方中的应用与讨论

打叶复烤生产的工艺过程在卷烟生产中起着非常重要的基础性作用,不仅起着调整叶片含水量、除杂灭虫等作用,而且对参与卷烟生产的叶组配方的准确表达起着非常重要的作用^[1,16],复烤模块配方的设计就是为实现卷烟叶组质量稳定所作的基础工作。目前红塔集团已应用多种方法和技术手段进行烟叶复烤配方模块的设计,其中 PTF 对工业分级样品的投影分析已作为一种重要的参考依据。图 2 为 33 类工业分级样品 PTF 投影分析结果,在设计复烤配方模块时,依据 PTF 投影体现的工业分级各个等级间的距离关系,以相近可合,并以不同距离之间的连续性和过渡性为指导原则,进行复烤配方模块的重组设计,使复烤模块的配方更加科学、合理,保证和提高复烤产品的特色和稳定性。

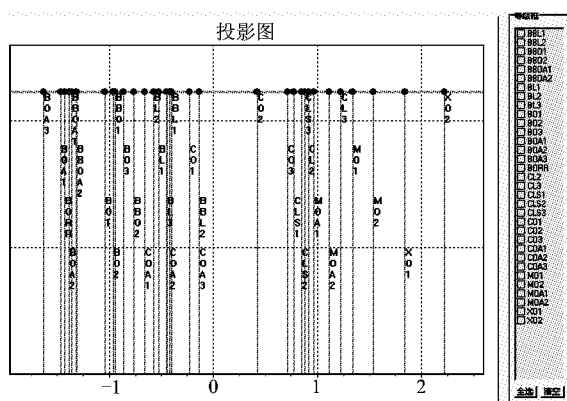


图 2 33 类工业分级样品 PTF 投影
(横坐标为投影值,纵坐标无意义)

针对不同批次、不同年份的工业分级烟叶样品,依据 PTF 投影分析得到的稳定、客观的投影结果,进行复烤配方的调整,从而实现烟叶复烤配方模块内在质量的稳定,同时也是该分析方法的技术关键。从图 2 中可以看出,33 类样品中挑选出的 9 类样品投影值从小到大的顺序依次为: BOA1-BOA2-BBO1-BBO2-BBL1-BBL2-CO1-CO3-MOA1;在图 1 中 9 类样品单独进行投影分析时的投影值从小到大的顺序依次为: BOA2-BOA1-BBO1-BBO2-BBL1-BBL2-CO1-CO3-MOA1;在类别数不同的情况下得到的投影结果,BOA1 和 BOA2 的次序发生了互换,相邻相近的类间发生次序互换属于该方法的正常统计波动,总体认为投影关系基本一致。

其他因素如近红外光谱谱区范围选择、光谱预处理方法、主成分个数的变化等投影参数以及样品数等的不同会对投影结果产生影响,得到稳定、客观的投影图以应用于指导生产实践,应用更具量化指标的数学参数(例如采用多维投影)等更加全面地表达投影结果,这些工作目前已在开展,会有后续的研究论文报道。

参考文献

- [1] 王寅生.新编烟草收购预检预验与卷烟分级、加工工艺、技术标准规范及质量检验标准化实务手册[M].北京:轻工业出版社,2005.6.
- [2] 王家俊,李娟.基于 FT-NIR 分析技术的 SIMCA 建模及其在卷烟配方过程质量监测中的应用[J].烟草科技,2008(3): 5-9.
- [3] 马翔,温亚东,王毅,等.傅立叶变换近红外光谱仪在烟草制丝线上的应用[J].烟草科技,2007(3): 16-18.
- [4] 江苏,马翔,陈永福,等.近红外光谱分析技术及其在烟草行业中的应用[J].光谱实验室,2006(5): 633-637.
- [5] 严衍禄.近红外光谱分析基础与应用[M].北京:中国轻工业出版社,2005.1.
- [6] 王东丹,李天飞,吴玉萍,等.近红外光谱分析技术在烟草化学分析上的应用研究[J].云南大学学报,2001, 23(2): 135-137.
- [7] 陆婉珍.现代近红外光谱分析技术[M].北京:中国石化出版社,2000.4.
- [8] 赵龙莲,张录达,李军会,等.小波包嫡和 Fisher 判别在近红外光谱法鉴别中药大黄真伪中的应用[J].光谱学与光谱分析,2008(4): 817-820.
- [9] 张建平,陈江华,束茹欣,等.近红外信息用于烟叶风格识别及卷烟配方研究的初步探索[J].中国烟草学报,2007,13(10): 1-5.
- [10] 束茹欣,王国东,张建平,等.国产烤烟烟叶的 NIRS 模式识别[J].烟草科技,2006(8): 12-15.
- [11] 杨健,杨静宇,叶晖. Fisher 线性鉴别分析的理论研究及其应用[J].自动化学报,2003(7): 481-493.
- [12] Wilks S S. Mathematical Statistics[M]. New York: Wiley Press, 1962.
- [13] Duda R, Hart P. Pattern Classification and Scene Analysis[M]. New York: Wiley Press, 1973.
- [14] Tian Q. Image classification by the Foley Sammon transform. Optical Engineering[M], 1986, 25(7): 834-839.
- [15] Jin Zhong, Yang Jingyu, Hu Zhongshan et al. Face Recognition based on uncorrelated discriminant transformation[J]. Pattern Recognition, 2001, 33(7): 1405-1416.
- [16] 章涛.卷烟配方设计、生产工艺技术与感官鉴定及质量控制检验实务全书[M].北京:中国科技文化出版社,2005.6.