

# 基于近红外光谱的卷烟质量投影模型

吴海云<sup>1</sup>, 束茹欣<sup>2</sup>, 陈德莉<sup>1</sup>, 陈勇<sup>2</sup>

1 上海烟草集团有限责任公司信息中心, 上海市杨浦区长阳路 717 号 200082;

2 上海烟草集团有限责任公司技术中心, 上海市杨浦区长阳路 717 号 200082

**摘要:** 为提高卷烟配方的稳定性, 科学量化卷烟的质量水平, 建立基于卷烟产品近红外光谱信息的质量投影模型。通过主成分分析对光谱数据进行降维, 利用主成分建立 Fisher 判别函数对各卷烟产品进行投影。实验结果表明, 模型得到的二维投影图能直观反映各卷烟牌号质量的高低关系, 为卷烟质量的评价提供了客观的科学依据。

**关键词:** 烟叶; 质量评价; 近红外光谱

**引用本文:** 吴海云, 束茹欣, 陈德莉, 等. 基于近红外光谱的卷烟质量投影模型 [J]. 中国烟草学报, 2015, 21 (1)

卷烟质量的稳定是消费者的基本诉求, 也是衡量一个品牌优劣的重要因素。“维稳”因此成为配方设计人员首先要考虑的问题。传统的分析监控手段需要耗费大量的时间和精力, 难以掌控卷烟的整体品质质量, 无法直观体现各品牌卷烟质量间的高低关系。通过采用新的分析技术和方法, 建立卷烟质量评价模型, 在配方设计及工艺生产中对产品进行监控, 对提高卷烟质量的稳定性将具有重要作用。

在各化工领域, 近红外光谱分析由于其快速、无损、绿色等特点<sup>[1-4]</sup>, 得到了广泛的应用, 也成为烟草企业在进行生产过程质量控制和配方辅助设计时的理想技术。目前, 烟草企业近红外分析的应用主要集中在两个方面: 一是光谱的定量分析, 即使用 PLS (偏最小二乘法)、PCR (主成分回归法) 等方法对卷烟的各种化学成分进行预测和监控等<sup>[5-6]</sup>; 二是定性分析, 普遍采用 PCA (主成分分析法)、ANN (人工神经网络)、SVM (支持向量机)、Fisher (费歇尔判别法) 等方法进行研究, 集中于对卷烟类别和风格质量的识别及评价<sup>[7-11]</sup>。将多种方法结合研究的投影分析结果并不多, 最近的有温亚东等利用 PTF (Projection Method Based on Target Principal Component and Fisher Criterion) 方法对近红外光谱进行投影分析, 使用了两次有监督的方式对光谱数据进行了特征提取<sup>[12]</sup>。陶帅等通过使用一维投影值讨论了 PPF-PCA (Projection of basing on Principal

Component and Fisher Criterion, PPF) 和 PPF-Eig 两种投影矢量求解方法的区别<sup>[13]</sup>。张雅娟等通过 PPF 投影验证了近红外光谱线性加和的可行性<sup>[14]</sup>。

本文通过分析近红外光谱, 利用基于主成分分析的 Fisher 判别法, 采用无监督与有监督模式相结合的思想, 建立卷烟质量二维投影模型, 为卷烟产品质量控制与配方辅助设计提供科学直观的依据和使用参考。

## 1 材料与方法

### 1.1 实验材料

建模试验样品为 2012 年至 2013 年收集的烤烟型卷烟样品共 3375 个, 分别来自上海烟草集团公司所产的 12 个品牌, 以 A ~ L 表示, 其中 A ~ J 参与建模与验证, K、L 用做外部验证。各品牌样品数量及价类见表 1, 表 2。

表 1 样品数量

Tab. 1 Number of samples

| 品牌 | A  | B  | C  | D   | E | F   | G   | H  | I    | J  | K  | L |
|----|----|----|----|-----|---|-----|-----|----|------|----|----|---|
| 数量 | 87 | 18 | 44 | 605 | 8 | 495 | 114 | 31 | 1853 | 64 | 48 | 8 |

表 2 样品价类

Tab. 2 Prices of samples

| 价类 | 品牌        |
|----|-----------|
| 一类 | E、D、I、C、K |
| 二类 | A、B       |
| 三类 | G、F、J、H、L |

**作者简介:** 吴海云 (1971—), 本科, 经济师, 主要从事经济方面研究, Email: wuh@sh.tobacco.com.cn

**通讯作者:** 束茹欣 (1974—), 工程硕士, 高级工程师, 主要从事烟草质量控制研究, Email: shurx@sh.tobacco.com.cn

收稿日期: 2014-03-19

### 1.2 仪器、设备与测试方法

Antaris 型傅立叶变换近红外光谱仪。工作参数：光谱采集范围  $10000 \sim 4000\text{cm}^{-1}$ ；光谱分辨率： $8\text{cm}^{-1}$ ；扫描次数：64 次。

将烟丝磨成烟末<sup>[15]</sup>，在低于  $50^\circ\text{C}$  下烘 0.5 h，使含水率达到  $10\% \sim 12\%$ 。过 60 目筛，取适量烟末放入样品杯中，取一固定重量砝码放置在样品上方，使其自然压实后进行近红外光谱扫描。

### 1.3 建模方法

模型所有步骤通过 matlab R2012b 编程实现。谱段选择范围采用  $8000 \sim 4000\text{cm}^{-1}$ ，光谱预处理方法采用 SG 一阶导 17 点三次多项式平滑处理，最后有 1021 个变量进入模型。利用所有样本数据的协方差矩阵进行主成分分析，提取前 16 个主成分，将所有主成分作为变量建立 Fisher 判别函数，取前两个判别函数的得分投影作图，并由此生成各类的类中心点和置信椭圆。

## 2 结果与讨论

### 2.1 模型投影结果

挑选 10 类不同质量品级的卷烟牌号，将建模光谱数据按 2:1 的比例进行随机划分，多的部分进行建模投影，剩余部分进行分类的正确率验证。分别取不同的主成分个数分别建立模型，得到各模型分类的正确率，见图 1（横轴为主成分数，纵轴为正确率）。

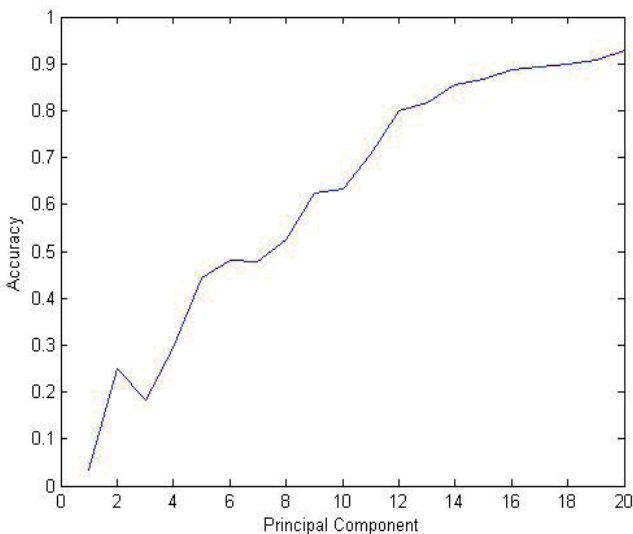


图 1 不同主成分模型的正确率

Fig. 1 Accuracy of different principal component models

以各主成分模型的正确率和投影结果，结合实际经验，最终选取 16 个主成分进行质量模型的建立。

选取最能解释光谱差异的前两个判别函数进行做图，见图 2。各类的类中心及置信椭圆长短轴数据见表 3 和表 4。其中类中心点 = (类内均值 - 总体均值) \* 非标准化判别函数系数矩阵，椭圆的中心即为类中心，长短轴为各维度上判别函数得分的类内 2 倍标准差。类中心点表示类的中心位置，置信椭圆表示样本的可能投影位置。通过这两者可以直观评价各品牌类内的一致性和稳定性。若样本投影落在某类椭圆外，则代表其不属于这个类或与这个类的其余样本差异较大，为异常样本。表明该样本在配方设计或者生产时出现问题，应该给予重视。

由图 2 可以直观看出各品牌间的相互位置关系，横轴上各品牌从左至右的位置顺序与其真实质量分级位置基本一致（见表 2，由价类反映品牌的质量等级），纵轴则代表了一部分的风格差异。

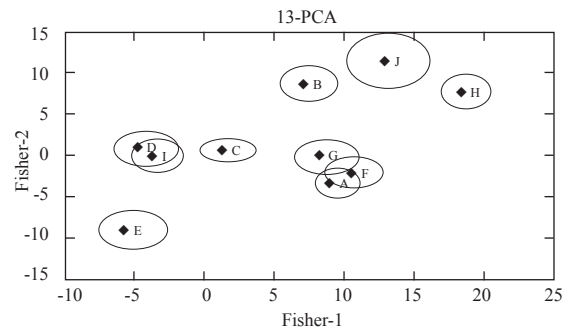


图 2 选取 16 个主成分建模的投影图

Fig. 2 Projection of 16 principal components model

表 3 建模各品牌的类中心坐标

Tab. 3 Center coordinates of different brands in model

| 牌号       | A     | B    | C    | D     | E     | F     | G     | H     | I     | J     |
|----------|-------|------|------|-------|-------|-------|-------|-------|-------|-------|
| Fisher-1 | 9.53  | 7.45 | 1.68 | -4.14 | -5.09 | 10.66 | 8.79  | 18.68 | -3.35 | 13.12 |
| Fisher-2 | -3.36 | 8.63 | 0.64 | 0.78  | -8.99 | -2.05 | -0.27 | 7.70  | -0.07 | 11.41 |

表 4 建模各品牌的置信椭圆的长短轴值

Tab. 4 Confidence ellipses of different brands in model

| 牌号       | A    | B    | C    | D    | E    | F    | G    | H    | I    | J    |
|----------|------|------|------|------|------|------|------|------|------|------|
| Fisher-1 | 1.59 | 2.01 | 2.01 | 2.27 | 2.45 | 2.09 | 2.34 | 1.77 | 1.84 | 3.02 |
| Fisher-2 | 1.80 | 2.20 | 1.41 | 2.09 | 2.33 | 1.84 | 2.02 | 2.04 | 1.97 | 3.36 |

### 2.2 模型外部验证

分别取建模品牌的外部样本（除 E 取 1 个样本外，

其余品牌均为 10 个样本) 作为验证, 分别做投影。见图 3, 模型能很好的区分识别和区分各个产品。

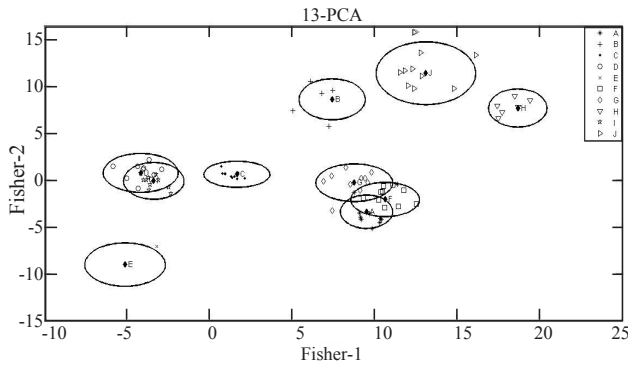


图 3 建模品牌的外部样本验证图

Fig. 3 Projection for validation samples

选取两类不参与建模的 K 和 L 品牌 (分别有 48 条和 8 条光谱数据) 进行投影, 见图 4。投影显示 K 的质量介于 C 与 I 之间, L 介于 H 与 J 之间, 与真实情况相符。

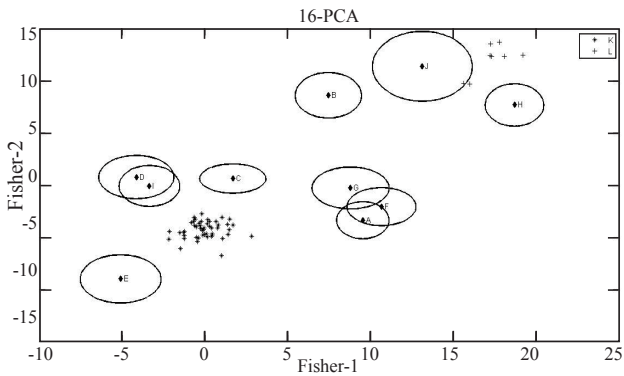


图 4 不参与建模的品牌 K、L 投影图

Fig. 4 Projection for brand K and brand L

2.3 模型对比

为与模型算法做比较, 不对原光谱数据进行主成分降维, 而在数据做相同预处理后直接进行 Fisher 判别分析, 将得到的前两个判别函数进行作图及做相同的验证, 见图 5、6、7。两种建模方法运行对比结果见表 4。通过对比可以看出, 虽然直接使用 Fisher 判别法做分类时的正确率较高, 但所做的投影无法很好表征各品牌间的质量关系 (见图 5, E 位于 I、D 之后, 与实际不符), 且由于进入模型的变量过多, 导致出现过拟合现象, 使得验证光谱与建模样本间的投影值差异较大。

通过综合比较, 基于主成分分析的 Fisher 判别法能更真实地反映品牌间的质量关系, 更适用于对卷烟产品的实时质量监控中。

表 5 两种建模方法运行结果对比

Tab. 5 Comparison between the results of two models

| 建模方法       | 建模和验证时间 /s | 正确率   |
|------------|------------|-------|
| PCA+Fisher | 10.4       | 0.92  |
| Fisher     | 23         | 0.966 |

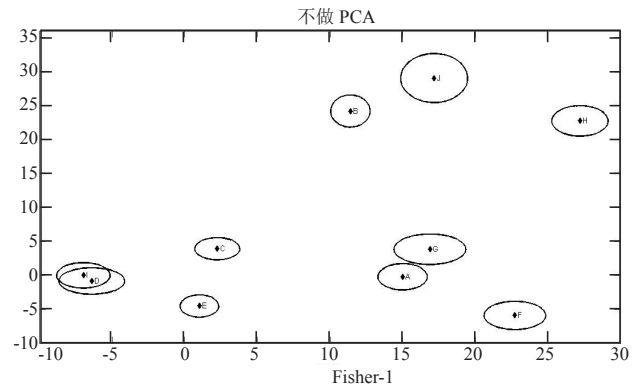


图 5 Fisher 判别法投影图

Fig. 5 Projection of Fisher Criterion

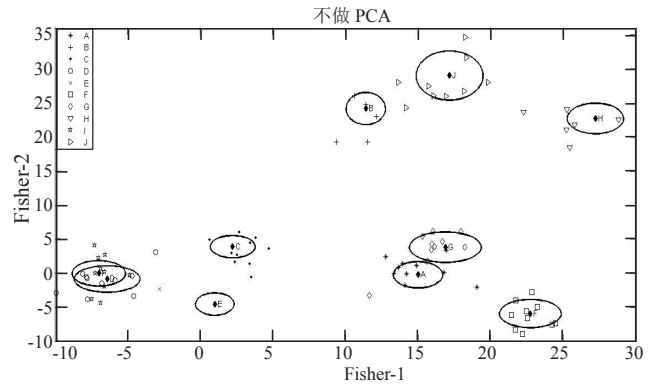


图 6 Fisher 判别法同类品牌外部样本验证图

Fig. 6 Projection of Fisher Criterion for validation samples

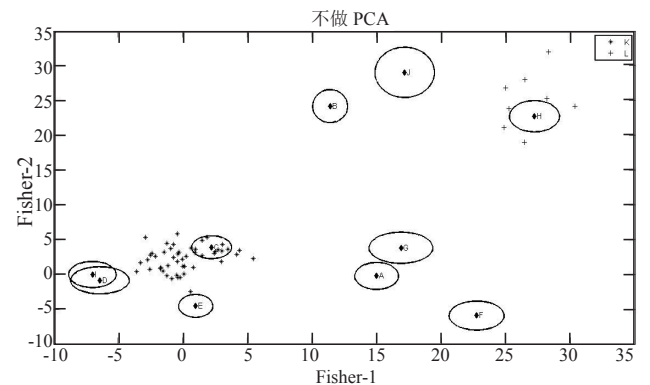


图 7 Fisher 判别法不参与建模的品牌投影图

Fig. 7 Projection of Fisher Criterion for K and L

### 3 结论

卷烟质量投影模型的建立, 为卷烟产品质量控制与配方辅助设计提供科学直观的依据和使用参考。通过投影的二维图, 可以直观地发现生产过程中产生的异常情况, 及时预警。配方设计人员可以根据投影二维图对配方进行及时的调整, 减少评吸人员的负担, 进而节省大量的人力与物力。

模型产生的投影二维图是基于两个 Fisher 判别函数而产生, 其中的一维经过验证, 可以表述为卷烟产品的质量, 另一维则具有一定的风格特性。在今后的研究中, 将对模型进行优化, 通过调整模型中的光谱处理方法, 以求得到对第二维度的准确表征。

### 参考文献

- [1] 杨凯, 刘鹏, 王维妙, 等. 原烟在线近红外光谱转移模型研究 [J]. 中国烟草学报, 2012(18): 27-31.
- [2] 段焰青, 陶鹰, 者为, 等. 近红外光谱分析技术在烟叶产地鉴别中的应用 [J]. 云南大学学报 (自然科学版), 2011 (33) 1: 77-82.
- [3] 祝元元, 陈永宽, 刘志华, 等. 近红外光谱技术在烟草行业的应用进展 [J]. 应用化工, 2010(11): 1750-1753.
- [4] 葛炯, 王瑾, 王维妙, 等. 近红外技术在卷烟真伪鉴别中的应用 [J]. 烟草科技, 2007(4): 29-31.
- [5] 唐果, 田旷达, 李祖红, 等. 近红外光谱结合 PLS-DA 划分烟叶等级 [J]. 烟草科技, 2013(4): 60-62.
- [6] 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.
- [7] 章平泉, 龚珍林, 杜秀敏, 等. 主成分分析-支持向量机在卷烟主流烟气中的应用 [J]. 中国烟草学报, 2010(12): 21-24.
- [8] 王德吉, 李广才, 粟卫军. 基于信息几何的卷烟焦油 SVM 预测 [J]. 中国烟草学报, 2009(8): 22-25.
- [9] 陈清, 刘巍, 钟科军. 基于烟草中致香成分的人工神经网络分类模型 [J]. 湖南大学学报 (自然科学版), 2006(4): 103-105.
- [10] 束茹欣, 孙平, 杨凯, 等. 基于 NIR-PCA-SVM 联用技术的烤烟烟叶产地模式识别 [J]. 烟草科技, 2011(11): 50-52.
- [11] 汪显国, 申晓锋, 刘泽, 等. 基于主成分分析法的烟丝加香均匀性评价 [J]. 中国烟草学报, 2013 (10): 33-42.
- [12] 温亚东, 王毅, 王能如, 等. 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 [J]. 中国烟草学报, 2009(5): 6-10.
- [13] 陶帅, 马翔, 李军会, 等. 基于烟叶近红外光谱主成分数据的投影方法研究及其在复烤配方中的应用 [J]. 光谱学与光谱分析, 2009(11): 2970-2974.
- [14] 张雅娟, 马翔, 张晔晔, 等. 近红外漫反射线性加和光谱在烟叶复烤配方中的应用 [J]. 光谱学与光谱分析, 2011(2): 390-393.
- [15] 张建平, 谢雯燕, 束茹欣, 等. 烟草化学成分的近红外快速定量分析研究 [J]. 烟草科技, 1999(3):37-38.

## NIR spectra-based projection model of cigarette quality

WU Haiyun<sup>1</sup>, SHU Ruxin<sup>2</sup>, CHEN Deli<sup>1</sup>, CHEN Yong<sup>2</sup>

<sup>1</sup> Information Center, Shanghai Tobacco Group CO. LTD., Shanghai, 20082, China;

<sup>2</sup> Technology Center, Shanghai Tobacco Group CO. LTD., Shanghai, 20082, China

**Abstract:** Projection model of cigarette quality based on NIR spectra was established to improve stability of cigarette formula and quantification of cigarette quality. Principal component analysis was used to reduce dimensionality of NIR data, based on which, Fisher discriminant functions were established for projection of cigarette products. Results showed that projective images could exactly describe the quality of different cigarette brands. Application of this model can facilitate the evaluation of cigarette quality.

**Keywords:** cigarette; quality evaluation; near-infrared

**Citation:** WU Haiyun, SHU Ruxin, CHEN Deli, et al. NIR spectra-based projection model of cigarette quality [J]. Acta Tabacaria Sinica, 2015,21(1)