

多元线性模型和数据降维用于评价复杂体系中各因素的影响程度

吴晶晶¹, 黄善松¹, 周芸¹, 蒋光辉¹, 邵利民^{2*}

(1. 广西中烟工业有限责任公司技术中心, 广西省南宁市北湖南路 28 号, 530001;

2. 中国科学技术大学化学系, 安徽省合肥市金寨路 96 号, 230026)

摘要: 评价复杂体系中各因素对特定对象的影响程度是一类常见问题。通过影响程度, 可以发现主要影响因素, 从而实现相应的科研和生产目标。本文提出了一种影响程度因子, 用于评价复杂体系中各种因素的影响程度。该影响程度因子基于多元线性模型, 通过偏导数得出, 并且消除了量纲, 以对比不同类型因素的影响程度。在建模样本数量有限的情况下, 通过主成分分析对原始建模数据进行降维, 然后在主成分和特定对象之间建立多元线性模型, 最终获得各因素的影响程度因子。将影响程度因子用于研究卷烟产品 9 项物理指标对于 4 种烟气成分含量的影响, 得到了各烟气成分含量的主要影响因素, 为产品研发提供数据支持。

关键词: 影响程度; 多因素; 主成分分析; 复杂体系

中图分类号: TS41+1

文献标识码: A

文章编号: 1001-4160(2017)02-167-171

1 引言

复杂体系广泛见于科学研究和生产实践, 其中考察对象受到多个因素的影响。在复杂体系的研究中, 评价各因素对考察对象的影响程度是一类常见问题。根据影响程度的大小, 可以发现主要影响因素; 通过控制这些因素, 可以实现相应的科研和生产目的。

卷烟烟气成分含量受多种物理指标的影响, 包括重量、吸阻、滤嘴通风率、总通风率等。某种烟气成分的含量与这些物理指标构成了一个复杂多因素体系。通过研究各种物理指标的影响程度, 可以确定该烟气成分的主要影响因素, 进而为产品设计和生产工艺提供参考, 以实现烟气有害成分含量的高效调控。

卷烟烟气的成分非常复杂, 含有超过 5000 种化学物质^[1]。2003 年, Rodgman 和 Green 总结了已经被报道过的烟气有害成分, 认为在烟气中存在有 149 种有害成分^[2]。2009 年, 谢剑平等筛选出 7 种最具有代表性的卷烟烟气有害成分, 并建立了一种新的烟气危害性指数^[3]。关于烟气成分的研究多数集中于这 7 种成分。

烟草行业多采用数理统计方法研究各种因素的

影响程度。有些研究通过相关系数或者 P -值的大小来评价各种因素的影响程度^[4-6]。有的研究采用广义线性模型对结果进行方差分析^[7]。还有一些研究是建立多因素与主流烟气有害成分释放量的回归方程, 然后根据回归系数评价影响程度^[8]。

关于因素的影响程度, 目前尚无统一评价方法。另外, 以相关系数或者 P -值作为评价指标似乎欠妥, 因为从数理统计的角度看, 相关系数超出临界值(或者 P -值小于显著性水平, 二者等价)只是表明了统计显著性, 与各因素的影响程度并没有直接关系。本文基于考察对象和各种影响因素之间的多元线性模型, 在偏导数的基础上, 进行无量纲处理, 定义了影响程度因子。该因子与偏导数具有同样涵义, 因此可以评价各因素的影响程度; 该因子是一个无量纲参数, 因此可以用于比较不同因素的影响程度; 该因子的物理意义是“自变量 X 相对于其平均值发生了 100% 的改变时, 导致因变量 Y 相对于其平均值的改变”, 因此可以定量研究因素影响程度。在统计分析中, 标准化回归系数也可以比较不同自变量 X 对于因变量 Y 的影响程度, 但是其物理含义不如本文提出的“影响程度因子”明确; 另外提供标准化回归系数的模型通常不能用于预测, 而用于计算“影响程度因子”的模型无此限制。但是获得标准化回归系数需

收稿日期: 2016-01-09; 修回日期: 2016-11-29

基金项目: 广西中烟 2015 年度技术创新项目 (1212014043); 中央高校基本科研业务费专项资金 (wk2060190040)

作者简介: 吴晶晶(1983—), 女, 硕士, Email: wujingjing3421@163.com

联系人: 邵利民(1972—), 男, 副教授, Email: lshao@ustc.edu.cn

要。在建立多元线性模型的过程中, 由于实际情况的限制, 有时无法获得足够多的建模样本。为了解决这一问题, 本文通过主成分分析 (principal component analysis, PCA) 对建模数据进行降维, 以主成分代替原始变量, 然后在考察对象和主成分之间建立多元线性模型, 最终获得各因素的影响程度因子。

2 材料与方法

2.1 材料

研究了两类产品。1 号产品的烟支规格为 (54+30) 24.5 mm; 同心圆嘴棒; 共有 18 个建模样本。2 号产品的烟支规格为 (54+30) 24.5 mm; 三元嘴棒, 共有 13 个建模样本。

RM-200A 转盘型吸烟机 (博瓦特-凯西公司, 德国), 7890A 气相色谱 (Aglient, 美国), FED240 烘箱 (宾得, 德国), 电子天平 (梅特勒, 瑞士), OM-IIS 综合测试台 (欧美利华公司, 北京)。

对烟气化学成分焦油、烟碱和 CO 进行了测量, 方法分别是 GB/T19609-2004、GB/T23355-2009 和 GB/T23356-2009。对烟支的物理指标进行了测量, 包括重量、吸阻、滤嘴通风率、总通风率、圆周、圆度、硬度和长度, 测量方法分别是 GB22838.4Y-2009、GB22838.5Y-2009、GB22838.15Y-2009、GB22838.3Y-2009、GB22838.13Y-2009、GB22838.6Y-2009 和 GB22838.2Y-2009。

2.2 方法

在原理介绍中, 大写字母表示变量, 相应的小写字母表示该变量的观测值。大写粗体字母表示矩阵, 下标表示矩阵大小, 上标 T 表示矩阵转置, 上标-1 表示矩阵求逆。

2.2.1 影响程度因子的定义

以变量 Y 表示考察对象, 考察对象受到 n 个因素的影响, 这些影响因素以变量 X_i ($i = 1, 2, \dots, n$) 表示。设 X_i 与 Y 之间具有线性或者近似线性关系, 那么存在以下线性模型

$$Y = k_1 X_1 + k_2 X_2 + \dots + k_n X_n + b \quad (1)$$

其中, k_i ($i = 1, 2, \dots, n$) 为系数, b 为常数项。

令 $X_i = \bar{x}_i \Delta_{xi}$, $Y_i = \bar{y} \Delta_y$, 其中 \bar{x}_i 和 \bar{y} 分别表示变量 X_i 和 Y 的观测值的平均值; Δ_{xi} 和 Δ_y 是新变量, 其含义分别是变量 X_i 和 Y 相对于其观测值的平均值的变化率, 与原始变量相比, 这些新变量没有量纲。通过变量代换, 式(1)可变为

$$\bar{y} \Delta_y = k_1 \bar{x}_1 \Delta_{x1} + k_2 \bar{x}_2 \Delta_{x2} + \dots + k_n \bar{x}_n \Delta_{xn} + b \quad (2)$$

整理式(2), 得到

$$\Delta_y = \frac{k_1 \bar{x}_1}{\bar{y}} \Delta_{x1} + \frac{k_2 \bar{x}_2}{\bar{y}} \Delta_{x2} + \dots + \frac{k_n \bar{x}_n}{\bar{y}} \Delta_{xn} + \frac{b}{\bar{y}} \quad (3)$$

通过式(3)求 Δ_y 关于 Δ_{xi} 的偏导数, 得到

$$\frac{\partial \Delta_y}{\partial \Delta_{xi}} = \frac{k_i \bar{x}_i}{\bar{y}} \quad (4)$$

式(4)中偏导数的意义是 Δ_{xi} 的变化引起的 Δ_y 的变化, 反映了 Δ_{xi} 对 Δ_y 的影响程度。 Δ_{xi} 与 X_i 之间以及 Δ_y 与 Y 之间存在线性关系, 所以该偏导数可以评价变量 X_i 对变量 Y 的影响程度。

通过式(4)定义变量 X_i 对 Y 的影响程度因子 (impact factor, IF)

$$IF_i = \frac{k_i \bar{x}_i}{\bar{y}} \quad (5)$$

IF_i 的物理意义是“变量 X_i 相对于其平均值发生了 100% 的改变时, 导致变量 Y 相对于其平均值的改变”。式(5)中, k_i 、 \bar{x}_i 和 \bar{y} 的量纲相互抵消, 所以, IF_i 不具有量纲, 因此可以用于比较不同变量 X 对于 Y 的影响程度。正值 IF_i 表明变量 X_i 和 Y 正相关, 即 X_i 增大 (或者减小) 导致 Y 增大 (或者减小); 负值 IF_i 表明变量 X_i 和 Y 负相关, 即 X_i 增大 (或者减小) 导致 Y 减小 (或者增大)。

2.2.2 影响程度因子的计算

设有 m 个样本, n 个自变量 X_i ($i = 1, 2, \dots, n$) 和因变量 Y 在这些样本上的观测值表示为列向量 $[x_{1i}, x_{2i}, \dots, x_{mi}]^T$ 和 $[y_1, y_2, \dots, y_m]^T$ 。变量 X_i 对 Y 的影响程度因子计算步骤如下。

(1) 将上述观测值整理为如下形式的数据矩阵 \mathbf{X} 和 \mathbf{Y} ; 线性模型表示为 $\mathbf{Y} = \mathbf{X}\mathbf{K}$, \mathbf{K} 为系数矩阵, 其中元素如下:

$$\mathbf{X} = \begin{bmatrix} x_{11} & \dots & x_{1n} & 1 \\ x_{21} & \dots & x_{2n} & 1 \\ \dots & \dots & \dots & \dots \\ x_{m1} & \dots & x_{mn} & 1 \end{bmatrix}_{m \times (n+1)} \quad \mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_m \end{bmatrix}_{m \times 1} \quad \mathbf{K} = \begin{bmatrix} k_1 \\ \dots \\ k_n \\ b \end{bmatrix}_{(n+1) \times 1}$$

(2) 通过多元线性拟合 $\mathbf{K} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}$ 获得线性模型的系数矩阵 \mathbf{K}

(3) 通过 $\hat{\mathbf{Y}} = \mathbf{X}\mathbf{K}$ 对变量 Y 进行预测, 获得预测值矩阵 $\hat{\mathbf{Y}}$, 其中元素如下:

$$\hat{\mathbf{Y}} = [\hat{y}_1, \hat{y}_2, \dots, \hat{y}_m]^T$$

(4) 通过预测值 \hat{y}_i 与测量值 y_i 的相关系数对多元线性模型 $\mathbf{Y} = \mathbf{X}\mathbf{K}$ 进行评价。如果相关系数大于临界值, 则认为线性模型有效。

(5) 如果线性模型有效, 那么通过式(5)计算 X_i 对 Y

的影响程度因子 IF_i 。

2.2.3 数据降维

在建立线性模型的过程中 (2.2.2 中的步骤 2), 样本数量应该足够大, 通常要求**样本数是变量数的 3 倍以上, 即 $m > 3n$** 。如果**不满足这一要求**, 模型容易出现过拟合现象, 严重时**无法建立线性模型**。为了解决这一问题, **需要增加样本或者减少变量**。但是, 实际中难以获得足够多的样本, 所以**减少变量几乎是唯一可行的方法**。本研究通过**主成分分析 (PCA)** 减少原始变量的数目, 即**降维**, 具体方法如下。

首先对原始数据 $\mathbf{X}_{m \times n}$ 进行**中心化处理**, 然后对中心化后的数据矩阵进行**奇异值分解 (singular value decomposition, SVD)**, 得到**特征向量矩阵 $\mathbf{V}_{n \times q}$** ($q = \min(m, n)$)。最后, 取矩阵 $\mathbf{V}_{n \times q}$ 的**前 p 个向量** ($p < m/3$), 通过下式完成降维

$$\tilde{\mathbf{X}}_{m \times p} = \mathbf{X}_{m \times n} \mathbf{V}_{n \times p} \tag{6}$$

式中, $\tilde{\mathbf{X}}_{m \times p}$ 是降维后的数据矩阵, 与原始数据矩阵 $\mathbf{X}_{m \times n}$ 相比, 变量数从 n 减小到 p , 从而满足建模的要求。

以降维数据 \tilde{x}_{ij} ($i = 1, 2, \dots, m; j = 1, 2, \dots, p$) 代替原始自变量数据 x_{ij} ($i = 1, 2, \dots, m; j = 1, 2, \dots, n$), 然后按照 1.2.2 中的步骤 1-4 进行建模。该模型的系数矩阵表示为 $\tilde{\mathbf{K}}_{(p+1) \times 1}$, 其中元素是降维后的新变量在线性模型中的系数, 并非原始自变量在线性模型中的系数, 后者通过下式计算得到

$$\mathbf{K}_{n \times 1} = \mathbf{V}_{n \times p} \tilde{\mathbf{K}}_{p \times 1} \tag{7}$$

其中, $\tilde{\mathbf{K}}_{p \times 1}$ 表示将 $\tilde{\mathbf{K}}_{(p+1) \times 1}$ 中最后一个元素 (即线性模型中的常数项) 删除后得到的矩阵。获得 $\mathbf{K}_{n \times 1}$ 后, 即可通过式 (5) 计算出各影响因素的影响程度因子。

需要指出的是, PCA 降维会导致信息损失, 那么基于降维后数据的分析结果不一定反映真实情况, 因此可能会产生某些数字假象。本研究在 3.2 中结合实际应用对这一问题进行了分析。

3 结果和讨论

3.1 线性模型

通过上述方法建立了产品物理指标与烟气成分含量的线性模型。首先通过主成分分析对产品的 9 个物理指标进行降维, 获得 3 个综合性指标。然后, 以这 3 个指标为自变量, 分别以 4 种烟气成分含量为因变量, 建立线性模型。最后, 计算烟气成分含量测量值和预测值的相关系数, 以此评价线性模型

的有效性。所有模型的相关系数以及 P -值列于表 1。

在显著性水平 $\alpha = 0.01$ 下检验, 模型预测值与实测值显著相关, 因此认为所建模型有效。

表 1 产品物理指标与 4 种烟气成分线性模型的相关系数^a
Table 1 Correlation coefficients of linear models built between factors of tobacco product and 4 harmful components in the mainstream smoke.

烟气成分 Mainstream component	1 号产品 Product #1	2 号产品 Product #2
总粒相物 Total particulate matter	0.93 (2.2×10 ⁻⁸)	0.74 (3.7×10 ⁻³)
焦油 Tar	0.93 (2.0×10 ⁻⁸)	0.77 (2.2×10 ⁻³)
烟碱 Nicotine	0.91 (2.0×10 ⁻⁷)	0.90 (2.5×10 ⁻⁵)
CO	0.90 (3.6×10 ⁻⁷)	0.95 (7.4×10 ⁻⁷)

a: 括号中为 P -值。

3.2 1 号产品物理指标对烟气成分含量的影响

通过式 (7), 计算出 1 号产品的 9 项物理指标对烟气成分含量的影响程度因子, 结果见图 1。

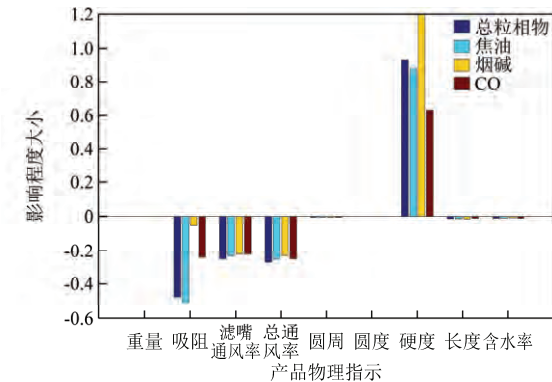


Fig.1 Impacts of 9 factors of tobacco product #1 on total particulate matter (TPM), tar, nicotine, and CO in the mainstream smoke.
图 1 1 号产品的 9 项物理指标对烟气总粒相物、焦油、烟碱和 CO 的影响程度因子

从图 1 可以看出, 各项物理指标的影响程度因子呈现出较大差异; 然而对于每种烟气成分, 影响程度较大的物理指标均为“吸阻”、“滤嘴通风率”、“总通风率”和“硬度”。下面分别对各项物理指标进行讨论。

3.2.1 滤嘴通风率和总通风率的影响

对于该产品, 滤嘴通风率和总通风率对烟气焦油、烟碱、CO 含量产生负影响。通过影响程度因子计算得出, 当滤嘴通风率增大 20% 时, 焦油降低 4.6%, 烟碱降低 4.4%, CO 含量降低 4.4%。可见, 通过增大滤嘴通风率和总通风率, 可以降低焦油、烟碱、CO 含量。有研究表明^[9], 滤嘴通风率与总粒相物、焦油呈负相关, 与本研究的结论一致。

3.2.2 吸阻的影响

吸阻对于烟气总粒相物、焦油、烟碱和 CO 的影响是一致的, 均为负影响, 增大吸阻, 可以降低这四种烟气成分的含量; 另外, 吸阻对焦油的影响最大, 这些结论与文献[9]一致。从图 1 中可以看出, 吸阻对 CO 的影响大于对烟碱的影响; 而文献[9]表明吸阻对烟碱的影响大于对 CO 的影响。这可能与产品的工艺条件不同有关, 文献[9]选用的是醋酸纤维滤棒样品, 本研究选用的样品为同心圆嘴棒样品。由于不同的工艺结构, 对烟气化学成分吸附机理不同, 因此, 影响程度有所差异。

通过影响程度因子计算得出, 当吸阻提高 20% 时, 焦油降低 10.2%, 烟碱降低 1.02%, CO 降低 4.8%。所以, 提高吸阻可以有效地降低焦油的含量。

3.2.3 硬度的影响

对于该产品, 硬度是影响程度最大的物理指标, 随着硬度的增大, 焦油释放量、烟碱和 CO 释放量随之增加, 通过影响程度因子计算得出, 当硬度降低 20% 时, 焦油降低 17.6%, 烟碱降低 24%, CO 降低 12.6%。可见, 降低硬度有利于降低这 3 种烟气成分的含量; 对烟碱、焦油的降低作用尤其明显。

3.2.4 其他物理指标的影响

从图 1 还可以看出, “重量”、“圆周”、“圆度”、“长度”和“含水率”这些指标的影响程度因子相对较小, 原因可能是这些物理指标与烟气成分含量的相关程度很小。但是还存在另外一种可能, 是建模样品的相应数据变化太小而导致的数据假象。

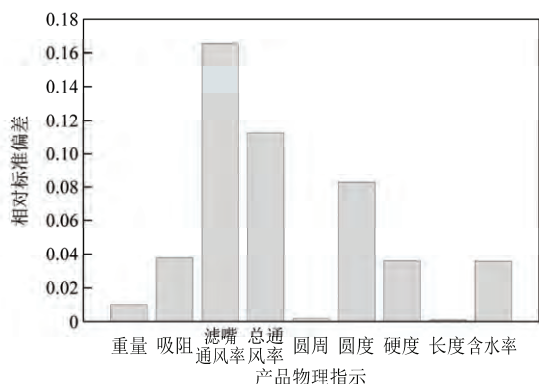


Fig.2 Relative standard deviations of 9 factors of the calibration data of tobacco product #1.

图 2 1 号产品建模数据 9 项物理指标的相对标准偏差

图 2 是 1 号产品建模数据在 9 项物理指标上的相对标准偏差, 发现“重量”、“圆周”和“长度”值的变化非常小, 这样通过 PCA 降维后, 这 3 类指标在降维数据中所占的比重很低, 最终导致了非常小的影响程度因子, 并不一定说明其对烟气成分含

量的影响较小。对于“圆度”和“含水率”, 图 2 显示, 建模数据在这两项指标上的取值的变化程度较大, 因此 PCA 降维导致的数字假象并不严重。这种情况下, 可以认为其影响程度因子反映了实际情况。

综上所述, 对于 1 号产品, 对烟气成分含量影响最大的物理指标是“硬度”, 影响较大的物理指标有“吸阻”、“滤嘴通风率”和“总通风率”, 影响程度较小的指标是“圆度”和“含水率”, 尚不能确定影响程度的是“重量”、“圆周”和“长度”。

3.3 2 号产品物理指标对烟气成分含量的影响

对 2 号产品进行了研究。图 3 是该产品 9 项物理指标对烟气成分含量的影响程度因子。与 1 号产品的情况相似, 影响程度较大的物理指标也是“吸阻”、“滤嘴通风率”、“总通风率”和“硬度”。

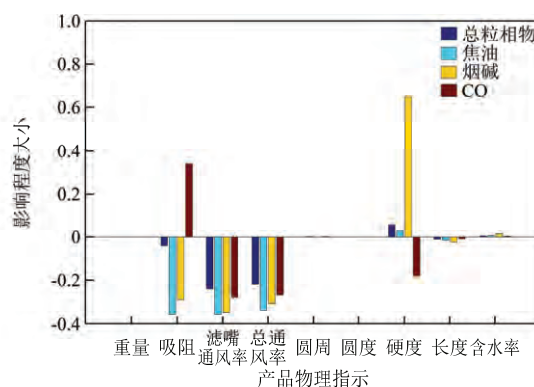


Fig.3 Impacts of 9 factors of tobacco product #2 on total particulate matter (TPM), tar, nicotine, and CO in the mainstream smoke.

图 3 2 号产品的 9 项物理指标对烟气总粒相物、焦油、烟碱和 CO 的影响程度因子

3.3.1 滤嘴通风率和总通风率的影响

滤嘴通风率和总通风率对于烟气总粒相物、焦油、烟碱和 CO 的影响是一致的, 均为负影响, 与 1 号产品的结论相同。可见, 对于 1、2 号产品, 提高滤嘴通风率和总通风率, 有助于降低烟气中的焦油、烟碱、CO 含量。通过影响程度因子计算得出, 当滤嘴通风率提高 20% 时, 焦油和烟碱分别降低 7.2% 和 7%。

3.3.2 吸阻的影响

吸阻对 CO 含量有显著的正影响, 与 1 号产品的结论相反。可能与该产品的三元嘴棒结构有关。通过影响程度因子计算得出, 当吸阻降低 20% 时, CO 含量降低 6.8%。吸阻与总粒相物、焦油和烟碱含量呈负相关。通过影响程度因子计算得出, 当吸阻增加 20% 时, 焦油相对降低 7.2%, 烟碱相对降低 5.8%。

由以上分析可知, 对于该产品, 增加吸阻有利

于降低焦油、烟碱的含量,同时却增加了烟气CO的含量。由于“吸阻”对焦油的影响大于对CO和烟碱的影响,所以增大吸阻对降低烟气有害成分是有益的。

3.3.3 硬度的影响

硬度与总粒相物、焦油、烟碱的含量呈正相关,通过影响程度因子计算得出,当硬度降低20%时,烟碱降低13%。硬度与CO含量呈负相关,这与1号产品的结论不同,可能与两种产品采用的不同工艺条件有关,1号产品采用的是同心圆嘴棒结构,而2号产品采用的是三元复合嘴棒结构。不过硬度的影响程度不大,当其降低20%时,CO增加3.6%。

3.3.4 其他物理指标的影响

对于影响程度因子较小的其他5项物理指标,同样研究了建模数据的变化情况,得出了与1号产品相同的结论,即“圆度”和“含水率”对烟气成分含量的影响较小,尚不能确定“重量”、“圆周”和“长度”的影响程度。

综合1号产品和2号产品的分析结果,可以发现多数结论是一致的,包括影响程度较大的物理指标、“滤嘴通风率”和“总通风率”对烟气成分的影响、“吸阻”对焦油和烟碱的影响、“硬度”对烟碱的影响等。明显不一致的结论有,“吸阻”对CO的影响、“硬度”对总粒相物、焦油和CO的影响。

4 结论

复杂体系中特定对象受到多种因素的影响,考察这些因素的影响程度是科研和生产中的常见问题。

本文提出影响程度因子用于此目的。该影响程度因子基于多元线性模型,通过偏导数得出,并且消除了量纲,以对比不同类型因素的影响程度。在建模样本数量有限的情况下,首先通过主成分分析对建模数据进行降维,以主成分代替原始变量,然后在特定对象和主成分之间建立多元线性模型,最终获得各因素的影响程度因子。将影响程度因子用于研究卷烟产品9项物理指标对于4种烟气成分含量的影响,得到了不同烟气成分含量的主要影响因素,为产品研发提供数据支持。

References:

- 1 Rodgman A, Perfetti T A. The chemical components of tobacco and tobacco smoke[M]. 2nd ed. Boca Raton: CRC Press, 2013.
- 2 Rodgman A, Green C R. Toxic chemicals in cigarette main stream smoke—Hazard and Hoopla[J]. Beitr Tabakfor Int, 2003, 20: 481-545.
- 3 谢剑平, 刘惠民, 朱茂祥, 等. 卷烟烟气危害性指数研究[J]. 烟草科技, 2009, 2: 5-15.
- 4 李国栋, 于建军, 董顺德, 等. 河南烤烟化学成分与烟气成分的相关性分析[J]. 烟草科技, 2001, 8: 28-30.
- 5 黄朝章, 李桂珍, 连芬燕, 等. 卷烟纸特性对卷烟主流烟气7种有害成分释放量的影响[J]. 烟草科技, 2011, 4: 29-32.
- 6 谢国勇, 李斌, 银董红, 等. 卷烟燃吸温度分布与主流烟气中7种有害成分释放量的关系[J]. 烟草科技, 2013, 11: 67-72.
- 7 张霞, 刘巍, 张涛, 等. 烟叶原料对卷烟主流烟气7项有害成分释放量的影响研究[J]. 中国烟草学报, 2013, 6: 1-8.
- 8 罗彦波, 庞永强, 姜兴益, 等. PLS回归法分析多因素对卷烟燃烧温度及主流烟气有害成分释放量的影响[J]. 烟草科技, 2014, 10: 56-60.
- 9 彭斌, 孙学辉, 尚平平, 等. 辅助材料设计参数对烤烟型卷烟烟气焦油、烟碱和CO释放量的影响[J]. 烟草科技, 2012, 2: 61-65.

Using multivariate models and dimensionality reduction to assess the impacts of various factors in a complex system

Wu Jingjing¹, Huang Shansong¹, Zhou Yun¹, Jiang Ganghui¹ and Shao Limin^{2*}

(1. Technology Center, China Tobacco Guangxi Industrial Co., Ltd., Nanning 530001, China)

(2. Department of Chemistry, University of Science and Technology of China, Hefei 230026, China)

Abstract: A common task in academic and industrial research is to assess the impacts of various factors on an interested target. Such information helps to identify the factors that influence the target the most. An impact index was proposed in this paper based on multivariate linear model and partial derivative. This index has a unit dimension, so it can be used to compare the impacts of different factors. When the number of samples is limited to build an effective multivariate linear model, principal component analysis was used for dimensionality reduction, then the linear model was built on principal components and the interested target, and finally impacts were calculated. The proposed impact index was applied to investigate impacts of 9 factors of tobacco products on 4 harmful components in the mainstream smoke, factors that have large impacts were identified, which are pressure drop, filter ventilation, overall ventilation and hardness. The conclusion serves supportive information for research and development of tobacco industry.

Keywords: impact; factors; principal component analysis; complex system

(Received: 2016-01-09; Revised: 2016-11-29)