



赵振杰,李石头,田雨农,等. 基于非参数核密度估计的卷烟配方相似性评价方法研究[J]. 轻工学报, 2025, 40(6): 107-117.  
ZHAO Z J, LI S T, TIAN Y N, et al. Research on cigarette blend similarity evaluation method based on non-parametric kernel density estimation[J]. Journal of Light Industry, 2025, 40(6): 107-117.  
DOI: 10.12187/2025.06.011

# 基于非参数核密度估计的卷烟配方相似性评价方法研究

赵振杰<sup>1</sup>, 李石头<sup>1</sup>, 田雨农<sup>1</sup>, 何文苗<sup>1</sup>, 张洪非<sup>2</sup>, 庞永强<sup>2</sup>, 廖付<sup>1</sup>

1. 浙江中烟工业有限责任公司,浙江 杭州 310008;

2. 国家烟草质量监督检验中心,河南 郑州 450001

**摘要:**【目的】提出一种基于非参数核密度估计(KDE)的配方相似性评价方法,以提高卷烟叶组配方维护效率。【方法】通过历史配方数据构建单项化学成分的KDE分布模型,计算单项化学成分相似性;采用不同权重分配策略,比较实验配方与历史配方的相似度;基于相关性分析评估模型的性能,分析叶组配方整体相似性与感官质量之间的规律。【结果】KDE拟合的概率密度估计函数,能较好地反映历史维护配方的化学成分数据分布;通过评审的叶组配方与实验叶组相比,糖碱比、总氮含量、还原糖含量3项指标与历史配方的整体相似性均值分别提升了30.57%,18.97%,13.06%;虽然4种赋权方法的相关性相近,但它们在单项化学成分相似性相关程度的倾向性上表现不同,其中博弈论组合,赋权法在多数感官指标上的相关性介于主观、客观赋权之间,可降低主客观赋权间的综合权重偏差。【结论】通过历史配方数据构建KDE分布模型,以评估实验配方与历史配方的相似性,为评估和调控配方维护过程中叶组的相似性与稳定性提供参考。

**关键词:**卷烟叶组配方维护;非参数核密度估计;博弈组合赋权;相似性

**中图分类号:**TS452    **文献标识码:**A    **文章编号:**2096-1553(2025)06-0107-11

## 0 引言

卷烟叶组配方设计是形成卷烟产品风格特点的基础,而配方稳定性则是保障产品质量的基本要求<sup>[1-2]</sup>。在实际的叶组配方维护设计中,配方设计需兼顾成本、库存、产量、化学成分等多种因素,烟叶等级可能经历多次替代,配方结构亦需不断调整,而这些变化均需最大程度保持配方的相似性<sup>[3-4]</sup>。

长期以来,许多科研人员致力于深入研究烟叶替代、配方叶组维护及与感官质量之间的关系。在叶组设计和维护方面,已引入一些辅助设计方法,例如热解差异分析<sup>[3]</sup>、近红外光谱相似度<sup>[5-6]</sup>、烟叶挥发性成分分析<sup>[7]</sup>、烟叶原料配方功能分类<sup>[8]</sup>等。杨超等<sup>[9]</sup>采用多元统计方法,研究了因子综合得分聚类分析,用于判断维护配方与初始配方之间的相似程度,该方法已投入实际应用并取得了一定成果。然而,其评价仅依靠与单一初始配方化学成分

收稿日期:2024-09-10;修回日期:2024-12-09;出版日期:2025-12-15

基金项目:中国烟草总公司“揭榜挂帅”项目(中烟办(2021)150号-110202103003);浙江中烟科技项目(ZJZY2022B012)

作者简介:赵振杰(1992—),男,河南省项城市人,浙江中烟有限责任公司工程师,主要研究方向为卷烟配方研究与数理统计。

E-mail: zhaozhenjieaa@163.com

通信作者:廖付(1981—),男,广西壮族自治区博白县人,浙江中烟有限责任公司高级工程师,主要研究方向为卷烟产品研发与烟草原料。E-mail: liaofu@zjtobacco.com

的相似性,存在明显局限性。在卷烟设计过程中,历史配方数据包含了丰富的卷烟设计风格和品质特征信息,是该产品特征的集中体现。王楠等<sup>[10]</sup>结合某烟草企业3年的卷烟历史配方数据,挖掘单等級片烟配伍规律,并给出替换方案,提出了基于非负矩阵分解的卷烟配方维护方法。冯润泽等<sup>[2]</sup>则利用卷烟历史配方数据,计算配方中单料片烟的共现频次,提出基于互信息的卷烟配方维护方法。然而,现有研究主要侧重于烟叶相似性替代或烟叶配伍性挖掘,而基于历史配方化学成分数据分布,评估卷烟维护过程中实验配方与该规格卷烟历史配方相似性的研究鲜有报道。若根据实验配方化学成分在历史配方数据中的分布特征进行相似性度量,相比仅依靠实验配方与单一初始配方的相似性评价,可克服现有模型仅依赖单一初始配方,无法全面表征叶组信息的局限性。

传统参数估计方法通常要求数据满足特定分布假设,当样本量较小或数据分布不符合常见假设时,估计结果易出现偏差。相比之下,非参数核密度估计(KDE)直接利用样本数据,无需对数据分布做出假设,能更准确地反映数据的真实分布<sup>[11]</sup>。卷烟叶组历史配方数据集通常有限,KDE 因其不依赖数据分布假设的优势,在有限数据集情境下仍具备稳健可靠的特点,并能灵活适应各种数据类型和分布形态<sup>[12]</sup>。

据此,本文拟以产品 A 为研究对象,对其烟叶常规化学成分、衍生指标及结构性大分子物质(如淀粉)分别进行对其历史配方的单项化学成分数据分布的非参数核密度估计。随后,根据实验配方中化学成分在历史配方数据中的分布情况,进行相似性度量,以评估实验配方与历史配方的相似性。进一步结合不同赋权方法,分析叶组配方整体相似性与感官质量之间的规律,以期为配方感官质量评价、配方维护过程中叶组相似性与稳定性的评估和调控提供参考。

## 1 材料与方法

### 1.1 材料与数据获取方法

某公司 2021—2022 年卷烟产品 A,共计约 200

万箱,包含 27 个历史配方(已经执行生产投料)和 24 个实验配方。历史配方时间为 2021 年 1 月—2022 年 4 月。实验配方(N1—N24)时间为 2022 年 5 月—2023 年 3 月的。实验配方等级与执行配方相比,变动比例均在 30% 以上。烟叶中常规化学成分及衍生指标按照文献[13]进行测定和计算,烟叶中淀粉含量采用流动分析法,按照《烟草及烟草制品 淀粉的测定连续流动法》(YC/T 216—2013)<sup>[14]</sup>进行测定。由具有省级评吸资质的 9 名评价专家组成评价小组,参照《烤烟 烟叶质量风格特色感官评价方法》(YC/T530—2015)<sup>[15]</sup>进行感官质量评价,并结合企业自身需求制定评价指标,评价小组采用定量打分的方式,结果取均值。

### 1.2 权重构建

**1.2.1 客观赋权** 主成分分析的核心是通过降维将相关性较强的多个变量转化为少数线性无关的向量,从而解释原始数据集。主成分赋权法强调变量之间的相关性,以减少信息重叠<sup>[16]</sup>。参考文献[17],结合历史配方各项评价化学成分数据集,采用主成分分析法进行权重确定。

**1.2.2 主观赋权** 层次分析法(Aalytic Hierarchy Process, AHP)是一种用于多准则决策问题的定量分析技术。通过层次分析法将决策问题分解为层次结构,并通过专家判断两两比较的方式使用成对比较矩阵和特征向量法计算主观权重<sup>[13,18]</sup>。为确保比较矩阵的客观性,本研究邀请 9 位卷烟配方设计专家,参考文献[13]提出的三标度 AHP 法,对常规化学成分及衍生指标进行评估,并构建比较矩阵。通过一致性检验,得到各因素的主观权重  $w_2$ ,并确保专家逻辑一致、矩阵合理。

**1.2.3 博弈论组合赋权** 博弈论是以纳什均衡为协调目标,寻找研究决策者之间相互作用的数学模型和分析方法,旨在不同赋权方法下寻求一致性或均衡性,减小综合权重与主客观赋权间的偏差<sup>[19]</sup>。本文参考文献[18]和[20]的方法,将主成分分析法确定的客观权重和 AHP 法确定的主观权重作为博弈双方,求解最优权重 Z,具体步骤如下:

1) 权重线性组合。若针对评价指标赋权采用

了  $l$  种方法(在本研究中,  $l=2$ )。假设权重向量集为  $w_k = \{w_{k1}, w_{k2}, w_{k3}, \dots, w_{kj}\}$ ,  $k=(1, 2, 3, \dots, l)$ ; 线性组合权重系数为  $\alpha_l = \{\alpha_1, \alpha_2, \alpha_3, \dots, \alpha_l\}$ 。以上向量任意组合:

$$Z^* = \sum_{k=1}^l \alpha_k w_k^T$$

式中,  $Z^*$  为组合权重向量,  $\alpha_k$  为线性组合系数。

2) 优化组合。根据博弈论思想, 以离差极小化为目标建立目标函数, 对组合系数  $\alpha_l$  进行寻优。其对策模型如下:

$$\min \left\| \sum_{k=1}^l \alpha_k w_k^T - \omega_k^T \right\|_2$$

3) 解优化组合系数。根据矩阵微分性质, 可得到满足目标函数最优一阶导数条件如下:

$$\sum_{k=1}^l \alpha_k w_k w_k^T = w_k w_k^T$$

4) 计算博弈论组合权重向量。利用 R 软件求解上式的解  $\alpha_k$ , 并进行归一化处理:

$$\alpha_k^* = \frac{\alpha_k}{\sum_{k=1}^n \alpha_k}$$

博弈论组合权重为:

$$Z = \sum_{k=1}^n \alpha_k^* \omega_k^T, k=1, 2, \dots, l$$

### 1.3 KDE 模型的构建及配方相似性计算

对于有限数据集的卷烟叶组配方, KDE 是一种可靠且稳健的分布推断工具, 其灵活性和适应性使其无需对数据分布作假设, 即能够处理各种数据类型和分布形态, 从而更准确反映真实分布情况<sup>[11-12]</sup>。

**1.3.1 基于历史叶组配方的非参数核密度估计** 对于卷烟规格  $A$  的历史配方表征数据集, 第  $j$  项化学成分有  $n$  个样本点构成的数据集  $A_j = [x_1, x_2, \dots, x_i, \dots, x_n]$ , 该项化学成分非参数核密度估计的概率密度函数为:

$$f(x) = \frac{1}{nh_s} \sum_{i=1}^n K_i(\cdot)$$

其中  $K_i(\cdot)$  为核密度估计的核函数,  $n$  为样本数量,  $h_s$  为设定窗宽。

高斯核函数具有良好的平滑性, 且已被广泛验证适用于多种场景, 因此本文选择将其用于

KDE<sup>[21-22]</sup>。KDE 的概率密度函数表达式为:

$$P_n(x) = \frac{1}{\sqrt{2\pi} nh_s} \sum_{i=1}^n e^{-\frac{(x-x_i)^2}{2h_s^2}} \quad ①$$

其中,  $n$  为样本数,  $x_i$  为总体化学指标数据  $x$  独立同分布第  $i$  个的样本。

窗宽  $h_s$  的选择直接影响到 KDE 的准确性。窗宽过大或过小都会导致估计结果失真, 窗宽过大会导致过度平滑, 而窗宽过小则容易产生噪声波动。因此, 合适的窗宽可以使估计结果更加逼近真实分布<sup>[23]</sup>。Silverman 准则是一种常用的窗宽选择规则, 能有效平衡估计的偏差和方差, 从而提高估计结果的准确性和稳健性<sup>[24-25]</sup>。最优窗宽  $h_s$  的计算方法如下:

$$h_s = \left( \frac{4\sigma^5}{3n} \right)^{\frac{1}{5}} \quad ②$$

为衡量模型拟合效果, 本研究采用平均绝对误差 (Mean Absolute Error, MAE) 和均方根误差 (Root Mean Square Error, RMSE) 评估预测模型适应性。MAE 反映了模型预测误差的平均幅值, 反映模型在不同数据点的平均偏差, 而 RMSE 则反映误差的分散度。2 种误差值越小, 说明模型与原始概率密度分布越接近, 模型的拟合精度就越高<sup>[26]</sup>。

**1.3.2 实验配方与历史配方的相似性** 对于历史配方的第  $j$  项化学成分含量  $A_j$ , 将积分区间设定为  $[\min_j A_j, \max_j A_j]$ 。对于第  $i$  个待选配方的第  $j$  项化学成分含量  $X_{ij}$ , 分别设定相应积分区间为  $[\min_j A_j, X_{ij}], [X_{ij}, \max_j A_j]$ 。结合式①计算积分区间数值  $S_T, S_1$  和  $S_2$ 。此时, 对于待选配方中, 第  $i$  个待选配方的第  $j$  项化学成分相似性定义为:

$$p_j = 1 - \left| \frac{S_1 - S_2}{S_T} \right|$$

对于第  $i$  个待选配方, 计算整体相似性  $P$ :

$$P = \sum_{j=1}^m p_j Z_j$$

其中,  $p_j, Z_j$  分别为第  $j$  项化学成分对应的单项化学成分相似性及在评价中的权重,  $m$  为参与计算的单项化学成分指标总数。

### 1.4 数据处理

采用 R 编程语言中 Psych 软件包进行描述性统

计分析和相关性分析。组合赋权、相似性计算均采用 R version 4.3.2 自行编写代码实现。

## 2 结果与讨论

### 2.1 描述性统计

**2.1.1 历史叶组配方** 卷烟 A 的 27 个已执行历史叶组配方的常规化学成分描述性统计分析结果见表 1。由表 1 可知,历史叶组配方烟叶总糖含量为 24.57%~27.16%,总植物碱含量为 2.24%~2.41%,还原糖含量为 20.31%~23.05%,氯含量为 0.46%~0.52%,钾含量为 2.28%~2.55%,总氮含量为 1.94%~2.03%,淀粉含量为 4.27%~4.79%,糖碱比为 8.71~9.65,氮碱比为 0.82~0.91,钾氯比为 4.48~5.33,两糖比为 0.81~0.85。变异系数分析显示,总氮、两糖比、总植物碱变异程度较低,而钾氯比、氯、还原糖变异程度较高。历史叶组配方化学成分变异系数均小于 5%,表明叶组之间化学成分较为一致。

**2.1.2 实验叶组配方** 卷烟实验叶组配方的化学成分析结果见表 2。由表 2 可知,实验配方和历史配方的化学成分含量平均值较为接近。实验配方变异系数较高的依次是氯含量、钾氯比和糖碱比,分别为 5.06%、3.80%、3.45%,而两糖比、总氮含量和氮碱比的变异系数较低,分别为 0.80%、0.82% 和 1.37%。

### 2.2 基于博弈论计算组合权重

根据主成分分析法和层次分析法得到化学成分客观权重  $w_1$  和主观权重  $w_2$ 。通过 1.2.3 中博弈论法计算出组合系数,求得调整系数  $\alpha, \beta$  分别为

0.265 和 0.735。经组合赋权后,得到基于博弈论的组合权重  $Z_1$  和常规组合权重  $Z_2$  见表 3。比较表 3 中  $Z_1$  和  $Z_2$  的排序结果发现,基于博弈论的组合权重与常规组合权重结果基本一致。

基于不同赋权方法的权重对比结果见图 1。由图 1 可知,  $w_1$  与  $w_2$  相差较大;在不同组合赋权法下,组合权向量  $Z_1$  和  $Z_2$  与  $w_1, w_2$  的偏离程度不同。具体而言,  $Z_1$  的偏离幅度较小,而  $Z_2$  的偏离幅度较大。相对常规赋权法,博弈论组合赋权算法在主、客观赋之间选择了相对妥协的结果,在不同赋权情况下维持相对一致的权重分配。

### 2.3 叶组配方相似性评价分析

**2.3.1 基于 KDE 的历史叶组配方化学指标分布** 将卷烟 A 的历史叶组配方化学数据按照式②计算带宽,并构建 KDE 函数。历史叶组配方化学成分数据的直方图及基于 KDE 概率密度曲线见图 2 所示。由图 2 可知,KDE 能反映历史维护配方单项指标在不同含量范围下的概率密度分布,且能适应不同类型数据样本分布。与正态分布拟合效果相比,KDE 拟合后的概率密度估计函数,直方图基本契合曲线的形状,并在峰值与波谷之间进行平滑处理。对于历史叶组配方化学数据而言,并非所有的化学成分都精准契合正态分布。本研究中提出的 KDE 拟合模型,可直接根据样本数据拟合,不作预先分布假设,能够更灵活地适应各种类型数据,并且能够有效地处理非线性关系。因此,对于有限历史维护叶组配方化学数据,KDE 估计模型具备较好的适应性。

表 1 历史叶组配方的常规化学成分描述性统计分析结果

Table 1 Descriptive statistical analysis results of chemical composition in historical tobacco blends

化学成分	均值	最小值	最大值	中位数	标准差	偏度	峰度	变异系数/%
总糖含量/%	26.01	24.57	27.16	26.12	0.69	-0.38	-0.69	2.65
总植物碱含量/%	2.33	2.24	2.41	2.33	0.05	-0.09	-1.44	2.15
还原糖含量/%	21.72	20.31	23.05	21.74	0.84	-0.09	-1.37	3.87
氯含量/%	0.49	0.46	0.52	0.49	0.02	-0.02	-0.16	4.08
钾含量/%	2.40	2.28	2.55	2.39	0.08	0.33	-1.17	3.33
总氮含量/%	2.00	1.94	2.03	2.01	0.02	-0.62	-0.36	1.00
淀粉含量/%	4.51	4.27	4.79	4.49	0.15	0.10	-1.20	3.33
糖碱比	9.31	8.71	9.65	9.36	0.23	-0.78	0.20	2.47
氮碱比	0.86	0.82	0.91	0.86	0.02	0.02	-0.98	2.33
钾氯比	4.87	4.48	5.33	4.87	0.21	0.21	-0.59	4.31
两糖比	0.83	0.81	0.85	0.84	0.01	-0.22	-1.17	1.20

表2 卷烟实验叶组配方的化学成分分析结果  
Table 2 Chemical composition analysis of experimental tobacco blends

序号	总糖含量/%	总植物碱含量/%	还原糖含量/%	氯含量/%	钾含量/%	总氮含量/%	糖碱比	氮碱比	钾氯比	两糖比	淀粉含量/%
N1	27.17	2.35	22.51	0.46	2.30	2.01	9.58	0.86	4.96	0.83	4.59
N2	25.84	2.33	21.04	0.52	2.47	2.04	9.01	0.88	4.74	0.81	4.49
N3	26.29	2.38	21.74	0.51	2.34	2.02	9.14	0.85	4.59	0.83	4.59
N4	26.52	2.35	22.14	0.47	2.35	2.00	9.43	0.85	5.00	0.83	4.65
N5	25.63	2.43	21.32	0.56	2.41	2.03	8.78	0.84	4.32	0.83	4.46
N6	27.19	2.35	22.54	0.46	2.30	2.01	9.60	0.85	5.02	0.83	4.62
N7	25.22	2.36	21.01	0.51	2.42	2.03	8.91	0.86	4.79	0.83	4.36
N8	25.73	2.34	21.14	0.51	2.47	2.03	9.02	0.87	4.84	0.82	4.55
N9	26.33	2.37	21.90	0.48	2.36	2.01	9.24	0.85	4.91	0.83	4.56
N10	25.81	2.41	21.39	0.53	2.41	2.03	8.89	0.84	4.58	0.83	4.62
N11	26.82	2.36	22.15	0.48	2.37	2.00	9.40	0.85	4.96	0.83	4.68
N12	26.30	2.37	21.70	0.50	2.35	2.01	9.17	0.85	4.66	0.83	4.54
N13	26.75	2.29	22.59	0.49	2.36	1.99	9.87	0.87	4.80	0.84	4.77
N14	26.82	2.35	22.20	0.48	2.35	2.01	9.47	0.86	4.91	0.83	4.63
N15	25.57	2.35	21.06	0.51	2.47	2.03	8.98	0.87	4.86	0.82	4.50
N16	25.80	2.38	21.28	0.53	2.48	2.04	8.96	0.86	4.68	0.82	4.47
N17	26.70	2.28	22.42	0.46	2.31	2.00	9.86	0.88	5.05	0.84	4.83
N18	25.76	2.43	21.29	0.51	2.39	2.04	8.75	0.84	4.70	0.83	4.53
N19	25.76	2.35	21.08	0.53	2.46	2.03	8.98	0.87	4.67	0.82	4.49
N20	25.65	2.35	20.98	0.51	2.47	2.04	8.91	0.87	4.88	0.82	4.51
N21	25.26	2.30	21.05	0.50	2.43	2.01	9.16	0.88	4.83	0.83	4.34
N22	26.27	2.37	21.68	0.51	2.35	2.00	9.15	0.85	4.60	0.83	4.53
N23	26.32	2.36	21.88	0.51	2.34	2.01	9.28	0.85	4.64	0.83	4.53
N24	26.49	2.31	21.88	0.47	2.38	2.01	9.49	0.87	5.07	0.83	4.46
均值	26.17	2.35	21.67	0.50	2.39	2.02	9.21	0.86	4.79	0.83	4.55
标准差	0.57	0.04	0.55	0.03	0.06	0.02	0.32	0.01	0.18	0.01	0.11
变异系数/%	2.16	1.60	2.54	5.06	2.44	0.82	3.45	1.37	3.80	0.80	2.45

表3 各化学成分的权重  
Table 3 Weighting of individual chemical components

化学成分	客观权重 $w_1$	主观权重 $w_2$	组合权重 $Z_1$	排序	常规组合权重 $Z_2$	排序
总糖含量/%	0.105	0.107	0.107	3	0.122	3
总植物碱含量/%	0.087	0.126	0.116	2	0.120	4
还原糖含量/%	0.107	0.081	0.088	5	0.094	5
氯含量/%	0.059	0.070	0.067	10	0.045	11
钾含量/%	0.095	0.078	0.082	8	0.080	8
总氮含量/%	0.066	0.079	0.076	9	0.057	9
糖碱比	0.095	0.137	0.126	1	0.141	1
氮碱比	0.093	0.080	0.084	7	0.081	6
钾氯比	0.117	0.098	0.103	4	0.125	2
两糖比	0.089	0.057	0.066	11	0.055	10
淀粉含量/%	0.085	0.087	0.086	6	0.080	7

注: $Z_2$ 是将  $w_1, w_2$  中对应元素相乘,归一化所得的组合权向量,采用的是一种最常见的组合赋权法<sup>[18]</sup>。

正态分布与非参数核密度估计的误差分析结果见表4。由表4可知,对于大多数化学成分,非参

数核密度估计的MAE和RMSE均低于正态分布,这表明前者的拟合精度更高。因此,非参数核密度

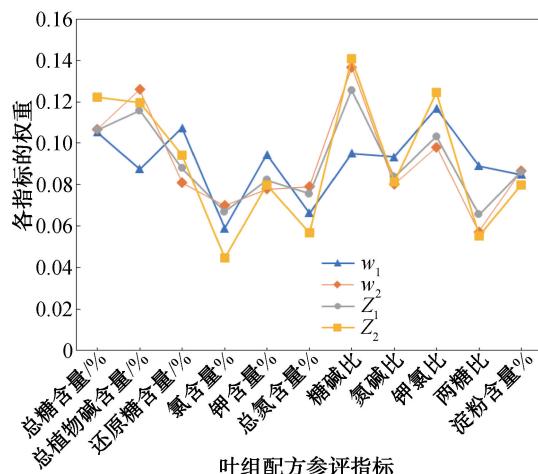


图 1 基于不同赋权方法的权重对比结果

Fig. 1 Comparison of weighting results using different evaluation-index weighting methods

估计具有较强的适应性,且模拟性能更优。

**2.3.2 基于 KDE 的配方相似性** 通过数值积分的方式计算实验叶组配方单项化学成分相似性结果见表 5,其中较高相似性值表示待选配方在历史配方数据分布中更居中,数据在密度分布中更接近分布中心。由表 5 可知,实验叶组配方单项化学成分相似性均值较大的依次是氮碱比、淀粉含量、两糖比分别为 0.707、0.638 和 0.618。单项化学成分相似性均值较小的依次是氯含量、糖碱比、总氮,分别为 0.219、0.312 和 0.436。变异系数较大的依次是氯含量、总氮含量、糖碱比,分别为 101.68%、84.40% 和 82.12%,而变异系数较小的依次是氮碱比、还原糖含量、两糖比分别为 22.78%、34.62% 和

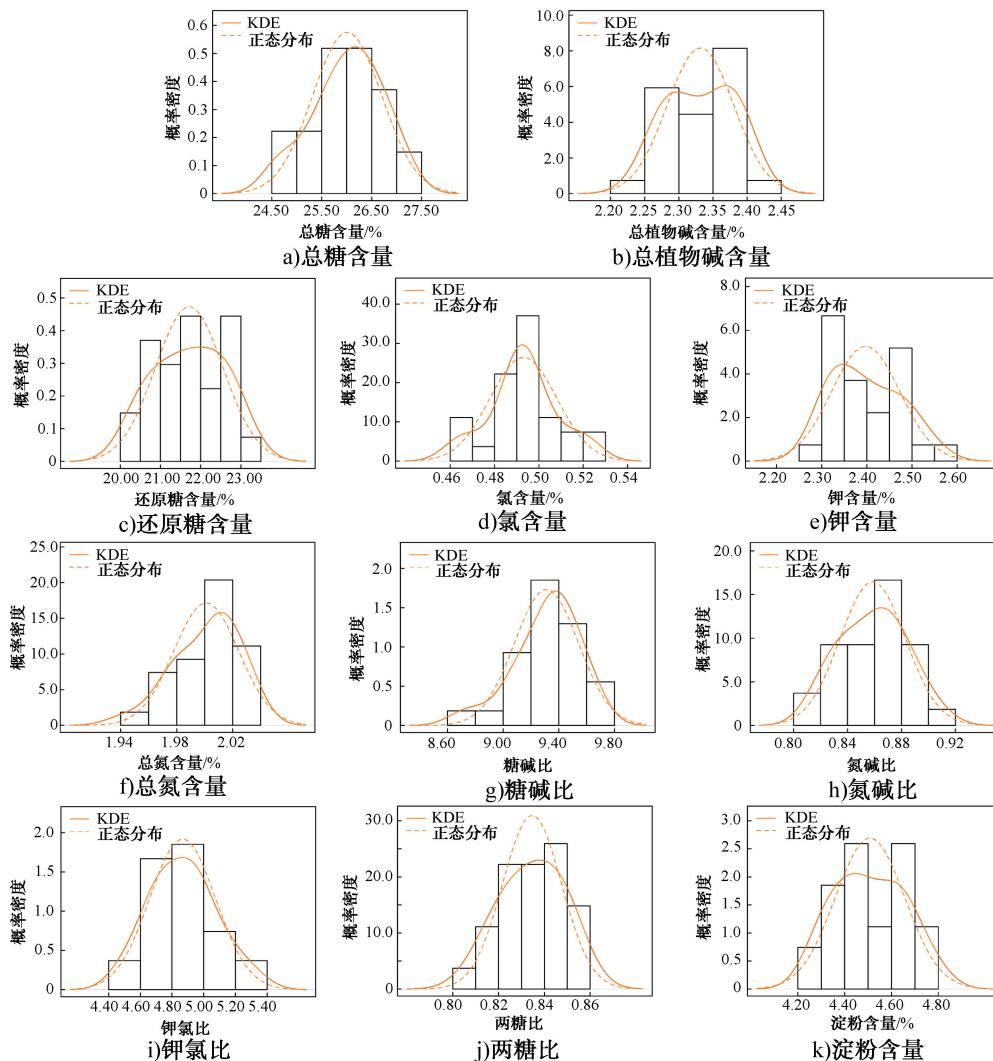


图 2 历史配方叶组化学成分数据的直方图及基于 KDE 概率密度曲线

Fig. 2 Histogram and KDE-based probability density curve of chemical composition data for historical tobacco blends

表4 正态分布与非参数核密度估计的误差分析结果

Table 4 Error analysis of normal distribution versus nonparametric kernel density estimation

化学成分	非参数核密度估计		正态分布	
	MAE	RMSE	MAE	RMSE
总糖含量/%	0.031 2	0.042 5	0.058 0	0.067 2
总植物碱含量/%	1.171 8	1.300 5	1.749 2	2.221 8
还原糖含量/%	0.085 5	0.097 1	0.104 3	0.126 1
氯含量/%	3.665 7	4.520 7	5.969 3	6.865 8
钾含量/%	1.271 7	1.492 9	1.595 5	1.925 9
总氮含量/%	1.706 2	2.329 6	2.929 7	3.691 2
糖碱比	0.107 5	0.136 5	0.132 8	0.142 4
氮碱比	1.391 7	1.918 2	2.226 8	2.982 1
钾氯比	0.162 7	0.180 7	0.161 3	0.191 4
两糖比	2.068 2	2.419 5	3.920 6	4.883 7
淀粉含量/%	0.423 0	0.530 5	0.566 3	0.755 1

36.46%。这表明在多次配方烟叶替代过程中,实验叶组配方中氮碱比、还原糖含量、两糖比与历史配方整体相似性较高,波动幅度相对较小,而氯含量、糖碱比、总氮含量的相似性较低,波动幅度较大。

此外,N3、N7、N9、N14、N17、N19、N23、N24为经二次评审后通过的实验叶组配方,用于生产执行,叶组配方化学成分对比分析结果见表6。由表6可知,单项化学成分中执行配方的糖碱比、总氮含量、还原糖含量与历史配方整体相似度均值提升幅度较大,分别为30.57%,18.97%,13.06%。而钾氯比、淀粉含量、氯含量的相似度均值均有所降低,分别为-8.09%,-4.87%,-2.10%。配方维护过程中,在于最大限度地保持配方的相似性和稳定性,

表5 实验叶组配方的相似性分析结果

Table 4 Similarity analysis results of experimental tobacco blends

序号	总糖含量/%	总植物碱含量/%	还原糖含量/%	氯含量/%	钾含量/%	总氮含量/%	糖碱比	氮碱比	钾氯比	两糖比	淀粉含量%
N1	0.000	0.767	0.363	0.015	0.157	0.624	0.131	0.884	0.682	0.694	0.649
N2	0.772	0.951	0.456	0.012	0.399	0.000	0.245	0.545	0.561	0.126	0.880
N3	0.720	0.366	0.999	0.244	0.485	0.380	0.468	0.676	0.173	0.615	0.667
N4	0.477	0.781	0.664	0.106	0.612	0.943	0.594	0.784	0.543	0.966	0.383
N5	0.575	0.000	0.663	0.000	0.893	0.064	0.037	0.367	0.000	0.857	0.743
N6	0.000	0.787	0.334	0.000	0.120	0.858	0.083	0.819	0.489	0.734	0.516
N7	0.271	0.648	0.436	0.360	0.785	0.153	0.127	0.971	0.721	0.940	0.328
N8	0.666	0.841	0.530	0.233	0.425	0.064	0.257	0.818	0.916	0.383	0.813
N9	0.672	0.489	0.863	0.373	0.714	0.734	0.739	0.646	0.838	0.860	0.775
N10	0.741	0.030	0.720	0.000	0.898	0.027	0.106	0.557	0.139	0.715	0.534
N11	0.205	0.671	0.652	0.251	0.760	0.890	0.696	0.641	0.670	0.583	0.295
N12	0.710	0.548	0.971	0.445	0.552	0.769	0.549	0.673	0.332	0.537	0.874
N13	0.257	0.458	0.302	0.952	0.719	0.625	0.000	0.746	0.774	0.465	0.026
N14	0.206	0.818	0.614	0.299	0.601	0.641	0.463	0.921	0.846	0.665	0.490
N15	0.522	0.800	0.472	0.270	0.403	0.002	0.196	0.815	0.983	0.469	0.936
N16	0.735	0.410	0.636	0.000	0.336	0.000	0.177	0.910	0.399	0.524	0.793
N17	0.303	0.300	0.431	0.000	0.202	0.855	0.000	0.501	0.406	0.701	0.000
N18	0.693	0.000	0.646	0.271	0.982	0.000	0.024	0.463	0.457	0.611	0.896
N19	0.694	0.785	0.483	0.000	0.489	0.046	0.199	0.856	0.351	0.253	0.898
N20	0.588	0.696	0.411	0.318	0.390	0.000	0.126	0.767	0.952	0.240	0.983
N21	0.295	0.586	0.459	0.456	0.707	0.541	0.517	0.517	0.876	0.936	0.229
N22	0.745	0.487	0.954	0.210	0.587	0.906	0.489	0.590	0.191	0.546	0.904
N23	0.684	0.667	0.880	0.364	0.539	0.588	0.873	0.818	0.273	0.837	0.920
N24	0.505	0.699	0.877	0.073	0.838	0.752	0.390	0.686	0.351	0.585	0.777
均值	0.502	0.566	0.617	0.219	0.566	0.436	0.312	0.707	0.538	0.618	0.638
标准差	0.245	0.270	0.214	0.222	0.237	0.368	0.256	0.161	0.285	0.225	0.294
变异系数/%	48.95	47.67	34.62	101.68	41.81	84.40	82.12	22.78	52.96	36.46	46.06

表6 实验叶组配方化学成分对比分析结果

Table 6 Comparative analysis results of chemical composition in experimental tobacco blends

项目	相似性均值										
	总糖含量/%	总植物碱含量/%	还原糖含量/%	氯含量/%	钾含量/%	总氮含量/%	糖碱比	氮碱比	钾氯比	两糖比	淀粉含量/%
实验叶组配方	0.502	0.566	0.617	0.219	0.566	0.436	0.312	0.707	0.538	0.618	0.638
评审通过配方	0.507	0.597	0.698	0.214	0.581	0.519	0.407	0.760	0.495	0.682	0.607
相对变化比例/%	1.07	5.37	13.06	-2.10	2.68	18.97	30.57	7.40	-8.09	10.29	-4.87

以确保卷烟产品感官品质风格一致。杨超等<sup>[27]</sup>的研究也发现,在叶组配方中糖碱比与多等级组合后的叶组配方总体感官质量和风格密切相关。本文研究结果与之较一致。

## 2.4 不同赋权法下叶组配方相似度分析

**2.4.1 相似性值比较** 利用现有模型分析叶组配方的相似性时,通常采用等权重主成分因子替代原始化学成分直接进行聚类分析<sup>[9]</sup>。然而,这种方法会弱化高权重特征第一主成分因子的重要性,从而导致相似性评估失真<sup>[28]</sup>。本研究通过不同的赋权方法与单项化学成分相似性相结合,解决了现有模型的不足,提高了配方相似性评价的准确性。本文采用4种赋权方法,对叶组配方相似性值进行计算,结果见表7。由表7结合表6可知,4种赋权方法相似性均值大小依次为客观赋权法(0.532)>常规组

合赋权法(0.521)>博弈论组合赋权法(0.518)>主观赋权法(0.512),变异系数大小依次为主观赋权(19.94%)>常规组合赋权(19.48%)>博弈论组合赋权(19.03%)>客观赋权(17.30%)。可以看出,主观赋权法相似性均值最小,变异系数较大;客观赋权法相似性均值最大,变异系数最小;而常规组合赋权、博弈论组合赋权的相似性均值和变异系数介于两者之间,且在主客观赋权之间选择了一个妥协的结果。

**2.4.2 与感官质量的相关性分析** 对不同权重赋予方式计算的叶组相似度值与感官评价的各项指标(感官得分与设计分值的差值)进行相关分析,结果见表8。由表8可知,多个指标之间存在显著中等相关性,相关性大小依次为香气质>刺激性>余味>谐调>杂气>成团性。与感官总分相似性差值的相关程度依次是博弈论组合赋权法(-0.668)>主观赋权法(-0.665)>客观赋权法(-0.663)>常规组合赋权法(-0.654)。尽管4种赋权方法相关性接近,但在单项指标相关性倾向方面表现不同。客观赋权法在杂气、刺激和成团性方面的相关性较高,而主观赋权法则更偏重于香气质和协调性。常规组合赋权法相比其他3种方法,在香气质和协调方面显示出较高的相关系数,但在成团性、杂气及余味方面的相关性相对较小,与感官总分的相关性相对较

表7 不同赋权法的相似性值比较

Table 7 Comparison of similarity indices across different weighting methods

项目	客观赋权法	主观赋权法	常规组合赋权法	博弈论组合赋权法
最小值	0.337	0.308	0.308	0.317
最大值	0.716	0.695	0.709	0.702
极差	0.379	0.387	0.401	0.384
均值	0.532	0.512	0.521	0.518
变异系数/%	17.30	19.94	19.48	19.03

表8 叶组相似度值与感官指标的相关性分析

Table 8 Correlation analysis between sensory attributes and tobacco blends similarity values

相似性值	香气质	谐调	杂气	刺激性	成团性	余味	总分
客观赋权	-0.591 **	-0.466 *	-0.461 *	-0.602 **	-0.429 *	-0.523 **	-0.663 **
主观赋权	-0.627 **	-0.470 *	-0.427 *	-0.566 **	-0.386	-0.521 **	-0.665 **
常规组合赋权	-0.626 **	-0.472 *	-0.417 *	-0.573 **	-0.378	-0.503 *	-0.654 **
博弈论组合赋权	-0.618 **	-0.470 *	-0.438 *	-0.578 **	-0.400	-0.523 **	-0.668 **

注: \* 和 \*\* 分别表示在 0.01、0.05 水平(双尾)上显著相关。

低。博弈论组合赋权法在大多数感官指标上的相关性介于主、客观赋权之间,组合权重与主、客观赋权之间的偏差降低,并在感官评价总分及余味方面显示出一定的倾向性。

### 3 结论

本文提出了基于 KDE 的卷烟配方相似性评价方法。该方法利用历史配方数据构建了各组分化学成分的 KDE 分布模型,计算单项化学成分的相似性,并通过不同权重策略分析整体相似性。同时,结合相关性分析探讨叶组整体相似性与感官质量的关系,得到如下结论:KDE 拟合的概率密度估计函数能较好地表征历史配方维护的化学成分概率密度分布特征;与实验叶组相比,经评审的执行配方在糖碱比、总氮含量和还原糖含量 3 项指标上与历史配方的整体相似性均值提升幅度较大,分别为 30.57%、18.97% 和 13.06%,可用于指导配方设计。此外,4 种赋权方法的相似性值与感官指标相关性表现存在差异。博弈论组合赋权法与大多数感官指标的相关性介于主、客观赋权之间,有效减小了组合权重与主、客观赋权之间的偏差,并与感官评价总分相关性较好。该评价方法完善了卷烟产品叶组维护评价体系,适用于日常维护与指导配方调整,特别在高比例烟叶更替和结构调整时,有望辅助设计人员优化配方方案,确保产品稳定性,为配方优化提供科学依据。

### 参考文献:

- [1] 朱亚昆,梅吉帆,郭文孟,等. 基于 PPF 投影算法和高光谱技术的卷烟牌号识别模型 [J]. 轻工学报,2024,39(4): 118–126.  
ZHU Y K, MEI J F, GUO W M, et al. Cigarette brand identification model based on PPF projection algorithm and hyperspectral technology [J]. Journal of Light Industry, 2024, 39(4): 118–126.
- [2] 冯润泽,雒兴刚,张忠良,等. 基于单料片烟互信息的卷烟配方维护方法 [J]. 烟草科技,2021,54(3): 65–71.  
FENGR Z, LUO X G, ZHANG Z L, et al. Cigarette blend maintenance based on mutual information from single-gradetobaccostrips [J]. Tobacco Science & Technology, 2021, 54(3): 65–71.
- [3] 李巧灵,陈昆焱,邓小华,等. 基于烟草热解差异度分析的烟叶替代方法 [J]. 烟草科技,2018,51(8): 77–84.  
LI Q L, CHEN K Y, DENG X H, et al. Method of tobacco substitution based on differential analysis of tobacco pyrolysis [J]. Tobacco Science & Technology, 2018, 51(8): 77–84.
- [4] 李晓,张浩博,郭朋玮,等. 不同规格卷烟配方叶丝柔软性与物理质量的关联性研究 [J]. 中国烟草学报,2024,30(4): 33–43.  
LI X, ZHANG H B, GUO P W, et al. Study on the correlation between softness and physical quality of formula tobacco in cigarettes of different specifications [J]. Acta Tabacaria Sinica, 2024, 30(4): 33–43.
- [5] 郝贤伟,帖金鑫,何文苗,等. 基于近红外光谱-感官评价的巴西烟叶风格模拟及替代 [J]. 烟草科技,2018,51(10): 83–89.  
HAO X W, TIE J X, HE W M, et al. Style simulation and substitution of Brazilian tobacco based on near infrared spectrum and sensory evaluation [J]. Tobacco Science & Technology, 2018, 51(10): 83–89.
- [6] 李石头,廖付,何文苗,等. 基于近红外光谱相似的烟叶替代与卷烟配方维护 [J]. 烟草科技,2020,53(2): 88–93.  
LI S T, LIAO F, HE W M, et al. Tobacco substitution and cigarette blend maintenance based on near infrared spectral similarity [J]. Tobacco Science & Technology, 2020, 53(2): 88–93.
- [7] 龙明海,李长昱,向虎,等. 支持向量机在卷烟配方原料分组中的应用 [J]. 云南农业大学学报(自然科学),2017,32(6): 1045–1051.  
LONG M H, LI C Y, XIANG H, et al. Applications of support vector machine in tobacco blend grouping of chinese type cigarette [J]. Journal of Yunnan Agricultural University (Natural Science), 2017, 32(6): 1045–1051.
- [8] 田震,许强,苗晨琳,等. 基于配方功能的烟叶原料动态平衡预测及分配方法探索 [J]. 中国烟草学报,2023,29(1): 1–10.  
TIAN Z, XU Q, MIAO C L, et al. Exploration of dynamic equilibrium prediction and distribution method of tobacco raw materials based on formula function [J]. Acta Tabacaria Sinica, 2023, 29(1): 1–10.
- [9] 杨超,万里兴,杨盛刚,等. 卷烟叶组化学成分分析及与配方维护的关系 [J]. 烟草科技,2019,52(4): 76–84.  
YANG C, WAN L X, YANG S G, et al. Analysis of chemical component in blended tobacco and relationship between chemical component and cigarette blend maintenance [J]. Tobacco Science & Technology, 2019, 52(4): 76–84.

- [10] 王楠,雒兴刚,张忠良,等.基于非负矩阵分解的卷烟配方维护方法[J].烟草科技,2019,52(8):67-76.  
WANG N, LUO X G, ZHANG Z L, et al. Cigarette blend maintenance based on non-negative matrix factorization [J]. Tobacco Science & Technology, 2019, 52(8): 67-76.
- [11] 吴喜之,赵博娟.非参数统计[M].北京:中国统计出版社,2019.  
WU X Z, ZHAO B J. Nonparametric Statistics [M]. Beijing: China Statistics Press, 2019.
- [12] 杨瑞君,张楚,杨评,等.非参数核密度估计模型预测双酚A的物种敏感度分布规律[J].生态毒理学报,2024,19(4):120-130.  
YANG R J, ZHANG C, YANG P, et al. Application of non-parametric kernel density estimation model in prediction of species sensitivity distribution of bisphenol A [J]. Asian Journal of Ecotoxicology, 2024, 19(4): 120-130.
- [13] 蒋佳磊,陆扬,苏燕,等.我国主要烟叶产区烤烟化学成分特征与可用性评价[J].中国烟草学报,2017,23(2):13-27.  
JIANG J L, LU Y, SU Y, et al. Chemical characteristics and usability of flue-cured tobacco growing in main tobacco producing areas in China [J]. Acta Tabacaria Sinica, 2017, 23(2): 13-27.
- [14] 国家烟草专卖局.烟草及烟草制品 淀粉的测定 连续流动法:YC/T 216—2013[S].北京:中国标准出版社,2013.  
State Tobacco Monopoly Administration. Tobacco and tobacco products-Determination of starch-Continuous flow method: YC/T 216—2013 [S]. Beijing: Standards Press of China, 2013.
- [15] 国家烟草专卖局.烤烟 烟叶质量风格特色感官评价方法:YC/T 530—2015[S].北京:中国标准出版社,2015.  
State Tobacco Monopoly Administration. Flue-cured tobacco-Sensory evaluation methods for quality and style characteristics of tobacco leaves: YC/T530—2015 [S]. Beijing: China StandardsPress, 2015.
- [16] XU S G, CUI Y X, YANG C X, et al. The fuzzy comprehensive evaluation (FCE) and the principal component analysis (PCA) model simulation and its applications in water quality assessment of Nansi Lake Basin, China [J]. Environmental Engineering Research, 2021, 26(2): 200022.
- [17] 王晓宾,王军,刘兰,等.广东烟区烤烟主要化学成分可用性评价[J].中国烟草科学,2019,40(2):64-72.  
WANG X B, WANG J, LIU L, et al. Chemical component usability of flue-cured tobacco leaves in Guangdong tobacco-growing area [J]. Chinese Tobacco Science, 2019, 40(2): 64-72.
- [18] 汪璇,卞建锋,杨民烽,等.基于综合赋权法的四川省烤烟化学成分可用性时空差异评价[J].中国农学通报,2021,37(22): 131-138.  
WANG X, BIAN J F, YANG M F, et al. Evaluation of temporal and spatial differences of flue-cured tobacco chemical composition usability in Sichuan province based on comprehensive weighting method [J]. Chinese Agricultural Science Bulletin, 2021, 37 (22): 131-138.
- [19] 樊艳翔,雷社平,解建仓.广东省河流水体富营养化综合评价及分异特征:基于博弈论组合赋权法与VIKOR模型[J].生态环境学报,2023,32(10): 1811-1821.  
FAN Y X, LEI S P, XIE J C. Comprehensive evaluation and differentiation characteristics of eutrophication in river waters of Guangdong province: Based on game theory combined empowerment method and VIKOR model [J]. Ecology and Environmental Sciences, 2023, 32 (10): 1811-1821.
- [20] 吴瑞英,孙怀卫,严冬,等.基于格网数据和博弈论组合赋权的中巴经济走廊洪灾风险评价[J].农业工程学报,2021,37(14): 145-154.  
WU R Y, SUN H W, YAN D, et al. Evaluation of flood disaster risk in China-Pakistan Economic Corridor by combination weighting based on improved game theory and grid data [J]. Transactions of the Chinese Society of Agricultural Engineering, 2021, 37 (14): 145-154.
- [21] 王颖,冯承莲,穆云松,等.非参数核密度估计在铜、银物种敏感度分布中的应用[J].中国环境科学,2017,37(4): 1548-1555.  
WANG Y, FENG C L, MU Y S, et al. Application of non-parametric kernel density estimation for developing species sensitivity distributions of copper and silver [J]. China Environmental Science, 2017, 37(4): 1548-1555.
- [22] 李智冈,吕莉,谭德坤,等.基于加权核密度估计与微簇合并的密度峰值聚类算法[J].信息与控制,2024,53(3):302-314.  
LI Z G, LYU L, TAN D K, et al. Density peaks clustering algorithm based on weighted kernel density estimation and micro-cluster merging [J]. Information and Control, 2024, 53 (3): 302-314.
- [23] 韩忠成,林金官,汪红霞.基于局部多项式展开的多元非参数模型贝叶斯带宽选择[J].数理统计与管理,2020,39(1): 93-103.  
HAN Z C, LIN J G, WANG H X. Bayesian bandwidth selection for multivariate nonparametric regression models based on local polynomial expansion [J]. Journal of Applied Statistics and Management, 2020, 39(1): 93-103.
- [24] 兰博,田喆,牛纪德,等.基于TRNSYS仿真数据的平板型太阳能集热器净得热非参数核密度估计概率模型[J].太阳能学报,2021,42(2): 1-7.  
LAN B, TIAN Z, NIU J D, et al. Probabilistic model of nonparametric kernel density estimation for useful energy gain of flat-plate solar collector based on trnsys simulation data [J]. Acta Energiae Solaris Sinica, 2021, 42(2): 1-7.

- [25] 李淑艳,翟友邦,王小龙,等. 基于核密度估计的拖拉机传动轴载荷外推方法 [J]. 中国农业大学学报, 2021, 26(10): 175–184.
- LI S Y, ZHAI Y B, WANG X L, et al. An extrapolation method of tractor propeller shaft load based on kernel density estimation [J]. Journal of China Agricultural University, 2021, 26(10): 175–184.
- [26] 张晓英,张晓敏,廖顺,等. 基于聚类与非参数核密度估计的风电功率预测误差分析 [J]. 太阳能学报, 2019, 40(12): 3594–3604.
- ZHANG X Y, ZHANG X M, LIAO S, et al. Prediction error analysis of wind power based on clustering and non-parametric kernel density estimation [J]. Acta Energiae Solaris Sinica, 2019, 40(12): 3594–3604.
- [27] 杨超,丁根胜,万里兴,等. 基于烟叶化学成分的卷烟叶组感官质量和风格研究 [J]. 烟草科技, 2018, 51(8): 85–91.
- YANG C, DING G S, WAN L X, et al. Relationships of main chemical components in tobacco blend of cigarette with its sensory quality and style [J]. Tobacco Science & Technology, 2018, 51(8): 85–91.
- [28] 褚旭,许强,殷瑜东,等. 基于不同聚类分析方法的卷烟叶组配方模块分类 [J]. 江苏农业科学, 2023, 51(15): 149–154.
- CHU X, XU Q, YIN Y D, et al. Classification of formula module of cigarette blend based on different clustering analysis methods [J]. Jiangsu Agricultural Sciences, 2023, 51(15): 149–154.

## Research on cigarette blend similarity evaluation method based on non-parametric kernel density estimation

ZHAO Zhenjie<sup>1</sup>, LI Shitou<sup>1</sup>, TIAN Yunong<sup>1</sup>, HE Wenmiao<sup>1</sup>, ZHANG Hongfei<sup>2</sup>, PANG Yongqiang<sup>2</sup>, LIAO Fu<sup>1</sup>

1. Technology Center, China Tobacco Zhejiang Industrial Co., Ltd., Hangzhou 310008, China;

2. China National Tobacco Quality Supervision & Test Center, Zhengzhou 450001, China

**Abstract:** **[Objective]** To enhance the efficiency of cigarette blend formulation maintenance, this study proposes a method based on non-parametric kernel density estimation (KDE) for evaluating the similarity of cigarette blend formulations. **[Methods]** A KDE distribution model for individual chemical components was constructed using historical formulation data to calculate the similarity of individual chemical components. Different weight allocation strategies were employed to compare the similarity between the experimental formulation and historical formulations. The model's performance was evaluated using correlation analysis, and the relationship between the overall similarity of leaf group formulations and sensory quality was analyzed. **[Results]** The probability density estimation function fitted using KDE effectively reflects the distribution of chemical component data from historical formulations. Compared with the experimental leaf groups, the average overall similarity of the leaf group formulations that passed evaluation to historical formulations in three indicators (sugar-alkaloid ratio, total nitrogen content, and reducing sugar content) increased by 30.57%, 18.97%, and 13.06%, respectively. Although the four weighting methods exhibited similar correlation levels, they differed in their tendency regarding the correlation degree of individual chemical component similarity. Specifically, the game theory-based combined weighting method showed correlations with most sensory indicators that were intermediate between those of subjective and objective weighting methods, thereby reducing the overall weight deviation between these two approaches. **[Conclusion]** Constructed a KDE distribution model using historical formulation data to evaluate the similarity between experimental formulations and historical formulations. The results provide a reference for assessing and controlling the similarity and stability of leaf groups during the formulation maintenance process.

**Key words:** cigarette blend maintenance; non-parametric kernel density estimation; game theory-based combinatorial weighting; similarity assessment