

基于原烟外观图像和近红外光谱的 烟叶感官质量模型研究

陈思昂¹, 王敏², 杜薇¹, 高旭¹, 李海涛², 邓羽翔¹, 王俊稀¹, 郭彦旭¹

(1. 四川中烟工业有限责任公司, 四川 成都 610066;

2. 上海创和亿电子科技发展有限公司, 上海 200082)

摘要: 为了探讨不同产地、不同等级以及不同品种来料烟外观指标、物理特性及近红外光谱对其感官质量的影响, 找到一种可以客观反映烟叶感官质量的方法。选取了四川3个产地、3个部位、10个等级和2个品种的单料烟, 以组为单位进行采集, 每组10片, 共采集烟叶7897片。使用变量标准化(SNV)方法对原始光谱数据进行校正, 用Standard Scaler方法对烟叶外观指标、物理特性、校正后的光谱数据进行标准化。采用主成分分析(PCA)法对数据进行降维。分别运用二次判别分析算法(QDA)、K近邻算法(KNN)、支持向量机(SVC)构建训练模型, 用Grid Search CV算法进行超参优化, 以平衡准确率作为模型评价指标。结果表明, 3个训练模型中, SVC的泛化能力最优。其中, 香气风格彰显程度、香气质、香气量、杂气、浓度、劲头、刺激性、余味及甜感的预测平衡准确率分别为0.747、0.751、0.715、0.720、0.712、0.774、0.685、0.725、0.700。外观指标、物理特性及近红外光谱共同影响着烟叶的感官质量。

关键词: 烟叶综合评价; 外观指标; 近红外光谱; 感官质量

中图分类号: S-3

文献标志码: A

文章编号: 2097-2172(2023)03-0260-10

doi: 10.3969/j.issn.2097-2172.2023.03.013

Study of Sensory Quality Classification Model Based on Appearance Image and Near-infrared Spectroscopy of Primary Flue-cured Raw Tobacco

CHEN Siang¹, WANG Min², DU Wei¹, GAO Xu¹, LI Haitao², DENG Yuxiang¹, WANG Junxi¹, GUO Yanxu¹

(1. China Tobacco Sichuan Industrial Co., Ltd., Chengdu Sichuan 610066, China; 2. Shanghai Micro Vision

Technology Ltd., Shanghai 200082, China)

Abstract: To investigate the influence of the appearance index, physical properties and near-infrared spectroscopy of cigarettes from different origins, different grades and different varieties on the sensory quality and to find a way that could objectively reflect the sensory quality of tobacco leaf, in this study, single tobaccos including 3 parts, 10 grades and 2 varieties were selected from 3 origins of Sichuan Province, and 10 pieces of tobacco leaves in each group were adopted for the collection, i.e. a total of 7897 tobacco leaves were collected. The original spectral data was corrected by standard normalized variate (SNV), and the appearance index, physical properties and standardised spectral data of tobacco leaf were standardized with the Standard Scaler method subsequently. Principal Component Analysis (PCA) was used to reduce the dimensionality of the data. The quadratic Discriminant Analysis Algorithm (QDA), K-Nearest Neighbors (KNN), and Support Vector Machine (SVC) were performed to construct the training model, and the Grid Search CV algorithm was executed for hyperparameter optimization, and the balance accuracy was used as the model evaluation. The results showed that among the three training models, SVC had the best generalization ability. The prediction balance accuracy of aroma style highlighting, aromatic, aroma volume, miscellaneous gas, concentration, strength, irritation, aftertaste and sweetness were 0.747, 0.751, 0.715, 0.720, 0.712, 0.774, 0.685, 0.725, 0.700, respectively. The appearance index, physical properties and near-infrared spectra jointly affect the sensory quality of tobacco leaf.

Key words: Comprehensive evaluation of tobacco leaf; Appearance index; Near-infrared spectroscopy; Sensory quality

收稿日期: 2022-11-18

基金项目: 四川中烟工业有限责任公司项目(KJSB202104150020)。

作者简介: 陈思昂(1995—), 男, 四川自贡人, 助理工程师, 硕士, 主要从事烟叶质量评价及模块配方研究工作。Email: 229780733@qq.com。

通信作者: 郭彦旭(1980—), 男, 河南新乡人, 工程师, 主要从事卷烟配方研究工作。Email: 290974009@qq.com。

烟草是我国重要的经济作物, 而烟叶是卷烟工业最重要的原料基础。烟叶的质量直接影响着烟叶的可用性, 影响着烟草经济的发展。烟叶质量是一个综合概念, 包含了烟叶的外观指标、物理特性、化学成分以及感官质量, 它们从不同方面反映了烟叶的品质特征。外观特征中的烟叶颜色、物理特性中的长度、叶片结构等是烟叶分级的重要依据。化学成分和感官质量反映了烟叶的内在质量, 化学成分的含量可通过近红外光谱数据建模测量。外观特征、物理特性在某种程度上和烟叶的内在质量存在着必然的联系^[1-2], 而感官质量是目前鉴定烟叶内在质量的重要手段, 是烟叶产品质量的基础和核心^[3]。目前, 烟叶的感官质量主要依靠专家的评吸, 较大程度上受专家的知识经验水平、心理、情绪等的影响, 指标较主观、片面。因此亟须寻找一种可以客观反映烟叶感官质量的方法。

近年来的研究表明, 通过一些指标可以一定程度上客观反映烟叶的感官质量。通过对烟叶的厚度、颜色和弹性与感官舒适性指标中的甜润、清晰、清爽度进行相关性分析, 发现这些外观指标可以反映烟叶的感官舒适性^[4]。对烟叶化学成分与感官质量进行偏最小二乘回归分析表明, 烟叶的糖碱比、糖氮比、总植物碱、挥发碱、总氮、蛋白质、水溶性总糖^[5]、还原糖、K、Cl、挥发碱等分别与感官刺激性、杂气、柔细度、余味、香气质、劲头等感官质量指标有不同程度的相关性^[6]。上述研究仅通过烟叶的某一类质量, 比如外观质量/内在化学成分来评价烟叶的质量, 这种评价是片面的, 因为烟叶的质量是各个指标综合表征的结果。同时, 相关研究主要集中在简单的相关、偏回归分析方法, 这些分析方法只能反映变量内各项指标间的简单相关性, 难以客观反映变量整体间的内在联系。有研究者对云南普洱烟区初烤烟叶的外观质量、常规化学成分进行主成分分析或因于分析, 之后对烟叶外观质量、常规化学成分与感官质量指标进行典型相关分析, 同样表明外观质量、化学成分与感官质量存在显著的相关性^[7], 但这同样不能反映烟叶的这些指标对感官质量的影响。除此之外, 上述研究采用的是化学成分数据, 是经过模型计算出的数据, 存在一定

的误差, 不同机器测量出的结果也会略有不同。为探讨影响烟叶感官质量各指标的因素, 我们从数据源头出发, 采用最原始的近红外光谱数据, 同时加上烟叶的外观指标、物理特性数据, 把这些数据综合起来训练机器学习模型, 从而建立一种全面、客观表征烟叶的感官质量的方法。

1 材料与方法

1.1 材料

1.1.1 供试品种 指示单料烟品种为云 87、红大, 产自四川凉山州会东县、会理县、德昌县。

1.1.2 实验设备 烟叶综合测试台(型号为 GTM-600), 由上海创和亿电子科技有限公司自主研发, 主要用于拍摄原烟图像以及检测原烟的物理特性指标。手持式近红外光谱仪(型号为 AURA), 由 Carl Zeiss 公司生产, 主要用于采集 950~1 650 nm 波长的近红外光谱数据。

1.2 样品采集

于 2021 年 9 月至 2022 年 2 月收集四川 3 个采样点、2 个单料烟品种的上中下 3 个部位、10 个等级的烟叶, 以组为单位进行采集, 每组数量为 10 片, 共采集烟叶 7 897 片。以上单料烟由烟叶分级专家按照烤烟国家标准(GB 2635—1992)进行等级分选, 共包含 10 个等级, 分别为上部橘黄二级烟(B2F)、下部橘黄二级烟(X2F)、上部橘黄三级烟(B3F)、中部橘黄一级烟(C1F)、中部橘黄三级烟(C3F)、上部橘黄一级烟(B1F)、中部橘黄二级烟(C2F)、中部橘黄四级烟(C4F)、中部柠檬黄三级烟(C3L)、中部微带青三级烟(C3V)。采集烟叶时每 10 片烟叶作为一个整体, 使用手持近红外光谱仪采集光谱。每种单料烟采集的光谱数量为 25~50, 采集 5~10 组。

1.3 方法

1.3.1 原烟外观指标检测 RGB 图像是最常见的一种图像表示方式, 由 R(红色)、G(绿色)、B(蓝色)3 个通道组成。本研究通过对原烟图像分别提取 R、G、B 通道的颜色值, 并计算平均值和标准偏差, 可以获取图像的 R-均值、R-标准偏差、G-均值、G-标准偏差、B-均值、B-标准偏差等外观指标。HSV 图像是另外一种图像表示方式, 由 H(色调)、S(饱和度)、V(亮度)3 个通道组成, 本研究同时将图像转换到 HSV 颜色空间, 分别提取

H、S、V 通道的颜色值，并计算各通道的平均值和标准偏差。

1.3.2 原烟物理特性的检测 采用烟叶综合测试台称取原烟重量，并获取原烟的厚度。采用图像预处理的方法从烟叶综合测试台拍摄的原烟图像中提取长度、宽度、周长、面积、结构等物理特性指标。

1.3.3 原烟近红外光谱的检测 采用手持式近红外光谱仪检测范围为 950 ~ 1 650 nm 的原烟近红外光谱数据。

1.3.4 原烟感官质量评价 烟叶感官质量指标包括香气风格彰显程度、香气质、香气量、杂气、浓度、劲头、刺激性、余味及甜感，按照 GB5606.4—2005 进行评吸，利用香味轮廓法采用 9 分制判定。

1.4 数据挖掘

1.4.1 数据划分 以组为单位进行训练集和测试集的划分。先采集的组数作为训练集，后采集的组数作为测试集。每个光谱号下的单料烟均会采集 8 ~ 10 组，用每个光谱号下单料烟的前 8 组来训练模型，剩下的组用来预测。

1.4.2 数据探索 采用 matplotlib 库绘制外观指标、物理特性及近红外光谱的频数直方图查看数据的整体分布情况。采用 seaborn 库绘制相关系数矩阵查看外观指标、物理特性及近红外光谱各指标之间的相关性，探索数据之间是否存在多重共线性。

1.4.3 管道模型训练及优化 考虑到测试集数据需要和训练集数据进行相同的预处理，我们把样本平衡、数据标准化、数据降维以及模型训练写入一个管道模型。

由于感官质量各指标中各类别的样本数存在明显的样本不平衡，因此样本需要进行平衡处理。本研究采用 imbalanced-learn 库对样本数少的类别进行上采样。

烟叶的外观指标、物理特性、近红外光谱数据之间的量纲不同，可能会影响后续模型的精度，因此这些数据在建模前需要进行标准化处理。我们采用 Standard Scaler 方法对数据进行标准化。原始近红外光谱数据在采集时可能会受到散射带来的光谱误差，需要先进行变量标准化(standard normalized variate, SNV)处理。

本研究的数据涵盖了来自不同维度的特征，特征之间可能会存在着多重共线性，这会导致解的空间不稳定，从而导致模型的泛化能力弱；同时，高维空间样本具有稀疏性，导致模型比较难找到数据特征。我们采用主成分分析方法(Principal Component Analysis, PCA)进行数据降维。

QDA(Quadratic Discriminant Analysis Algorithm)为二次判别分析算法，是一种生成模型，它假设每个类别服从高斯分布，同线性判别分析一样，求在输入为 x 的情况下分类为 k 的概率最大的分类。KNN(K-Nearest Neighbors)为 K 近邻算法，是一种分类算法，其思想是一个样本与数据集中的 k 个样本最相似，如果这 k 个样本中的大多数属于某一个类别，则该样本也属于这个类别。SVC(Support Vector Machine)是支持向量机中的一种分类算法，其原理是寻找一个能够区分两类的超平面，使得边际最大。我们选取 QDA、KNN、SVC 3 种模型作为数据的训练模型，采用网格搜索 + 交叉验证(Grid Search CV)作为超参数优化算法选取最优的超参数及模型。

1.4.4 模型测试和评估 对测试集数据采用和训练集相同的处理方法，之后采用最优模型对测试集数据进行预测，并用平衡准确率作为模型的评估指标。

2 结果与分析

2.1 数据探索结果

2.1.1 数据分布情况 自变量的数据分布对后面的数据预处理以及建模有着非常重要的作用。本研究分别绘制了外观指标、物理特性及近红外光谱数据的直方图。从烟叶外观指标和物理特性的直方图(图 1)可以看出，重量、宽度、周长、面积、结构、R-均值、R-标偏、G-均值、G-标偏、B-均值、B-标偏、H-均值、S-均值、S-标偏、V-均值和 V-标偏数据近似服从正态分布；烟叶的厚度存在明显的异常值，烟叶厚度为 0.005 ~ 0.020 的数据量极少；长度数据为左偏态分布；H-标偏数据为右偏态分布。剔除烟叶厚度超过 0.005 的异常值后，再次探索烟叶厚度的数据分布情况，表明数据近似服从正态分布。

近红外光谱数据维度比较高，因此每间隔 50 nm 绘制该波长的分布。从烟叶近红外光谱数据的

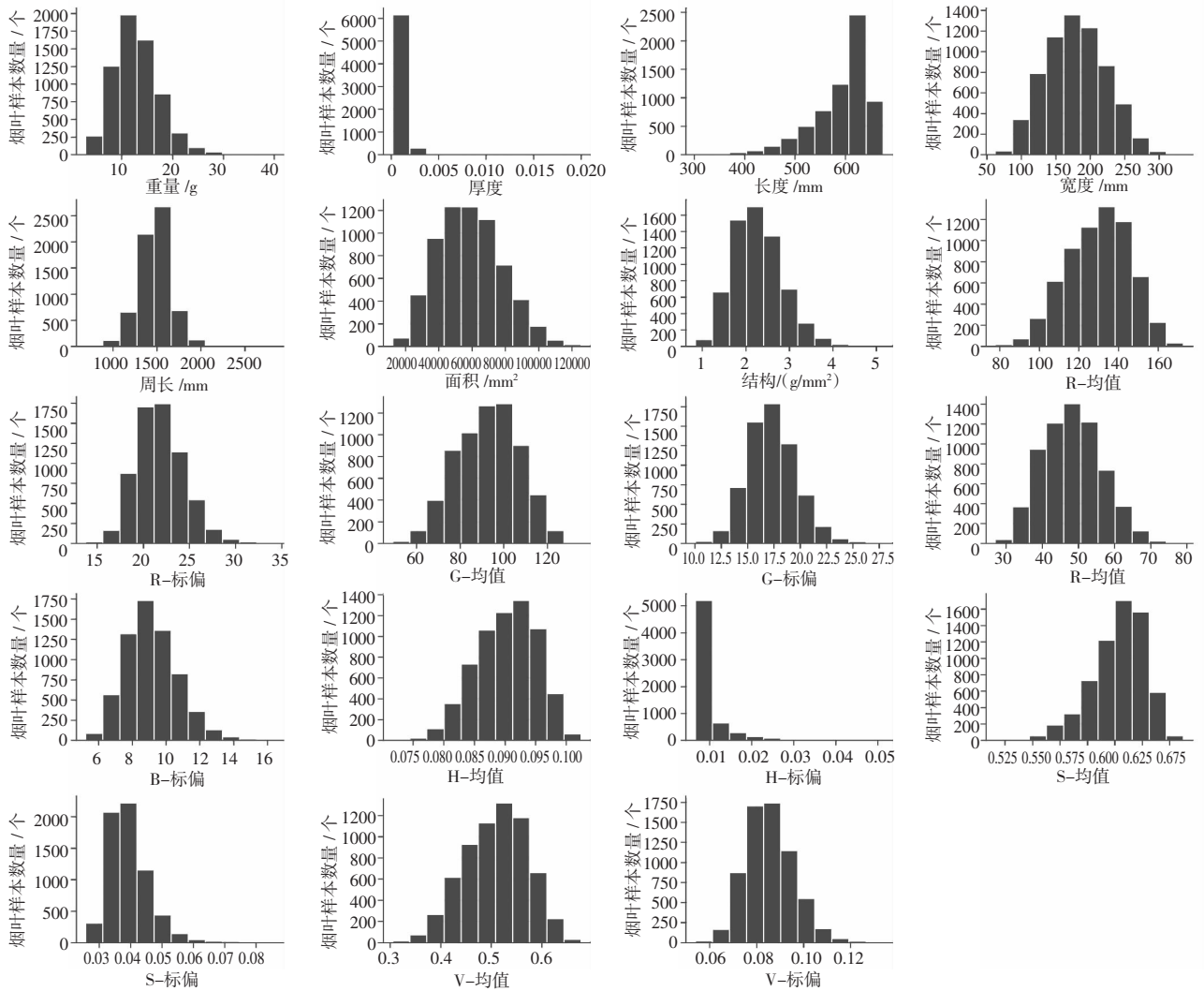


图 1 烟叶外观指标和物理特性的直方图

直方图(图 2)可以看出, 数据近似服从正态分布, 没有明显的异常值。综上可知, 除了烟叶长度和 H- 标偏数据服从偏态分布外, 其他烟叶外观指标、物理特性及近红外光谱数据均近似服从正态分布。

进一步探讨感官质量指标作为因变量的各类别数据情况, 以查看各类别数据之间是否存在不平衡。如图 3 所示, 各个感官质量指标均出现中间类别数量多、两边类别数量少的情况。各个类别的数量极度不平衡, 有些类别的数量所占总数比甚至小于 5%, 这种情况下建立的模型会比较偏重于数量占比多的类别, 导致模型的泛化能力差, 因此需要对占比小于 5 的类别进行相邻类别合并。本研究将香气风格彰显程度中类别 10 合并到类别 9 中, 香气量中类别 9 合并到类别 8 中; 将杂气中

类别 10 合并到类别 9 中, 杂气中类别 4 合并到类别 5 中; 将浓度中类别 9 合并到类别 8 中; 将劲头中类别 8 合并到类别 7 中; 将刺激性中类别 10 合并到类别 9 中; 将余味中类别 9 与 10 归并为类别 8, 余味中类别 4 归并为类别 5; 将甜感中类别 9 归并为类别 8。经过合并后, 极少数占比的类别不存在, 但各类别之间的数据仍然是不平衡的。综上所述, 感官质量各指标中的各类别之间不平衡, 后续建模时需要对数据进行平衡处理。

2.1.2 数据相关性分析 自变量之间较强的关联性说明变量之间存在多重共线性, 这种共线性会对后面的模型训练造成一定的影响, 因此建模之前需要探索烟叶外观指标、物理特性及近红外光谱数据之间的关联性。从烟叶外观指标、物理特

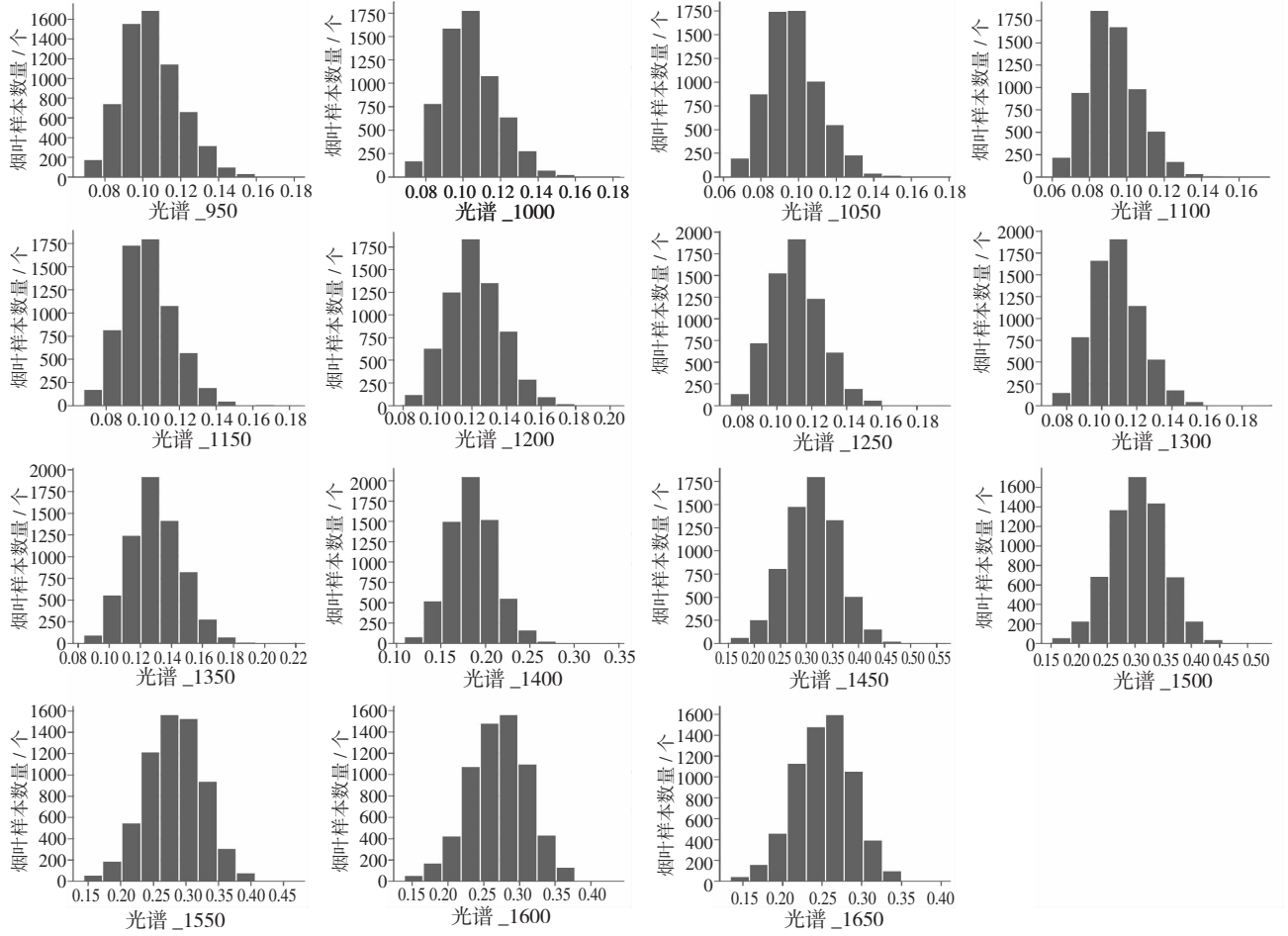


图 2 烟叶近红外光谱数据的直方图

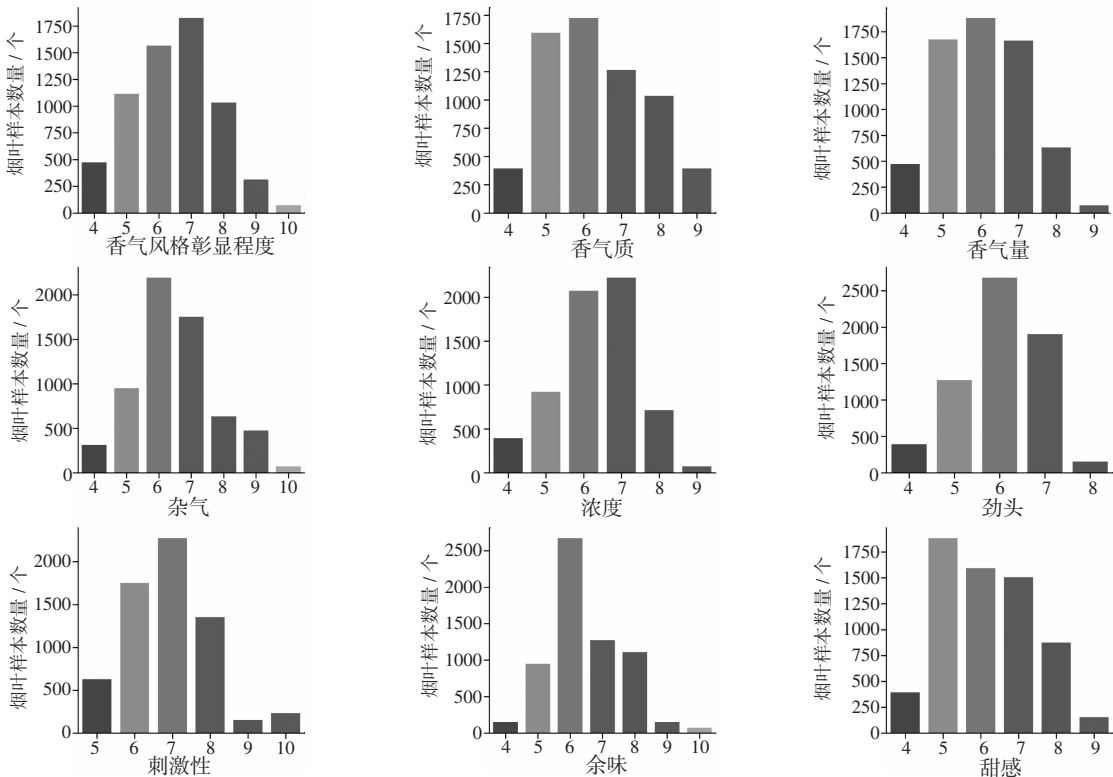


图 3 感官质量指标分类

性之间的相关性分析(图4)可以看出, 各物理特性之间, 重量与面积的相关系数为 0.7, 长度和周长的相关系数为 0.7, 宽度和面积的相关系数达到 0.9。各颜色外观指标之间, R-均值、G-均值、B-均值、H-均值、V-均值之间的相关性很强, 相关系数达到 0.9 和 1.0。R-标偏、G-标偏、B-标偏、V-标偏之间的相关性也很强, 相关系数也达到了

0.8 ~ 1.0。H-标偏与S- 标偏的相关系数为 0.7, 相关性也较强。总之物理特性与颜色指标之间的相关性比较弱或没有相关性。

对烟叶各近红外光谱数据之间的相关性进行分析(图 5)表明, 各近红外光谱数据之间的相关性很强, 说明变量之间存在很明显的多重共线性。

对烟叶的外观指标、物理特性和近红外光谱

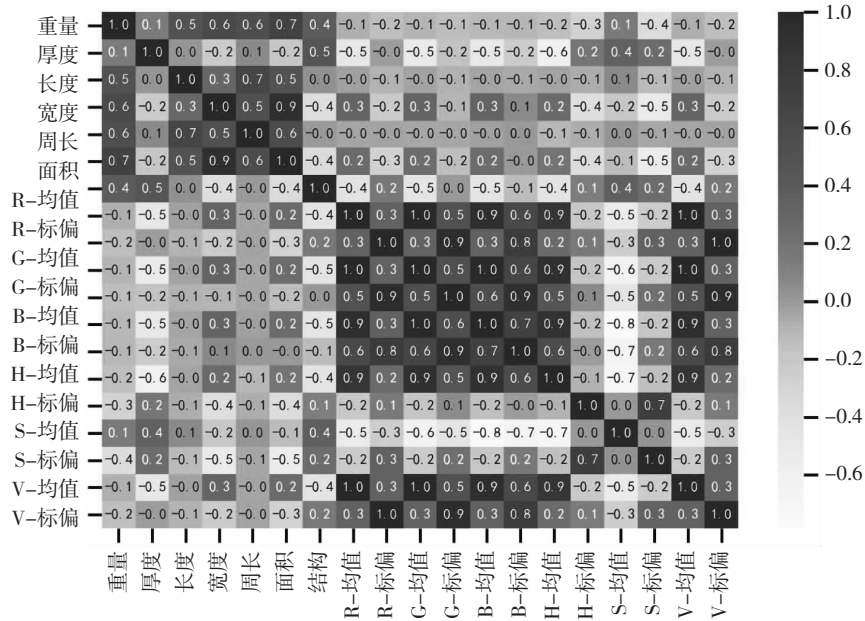


图 4 烟叶的外观指标、物理特性之间的相关性分析

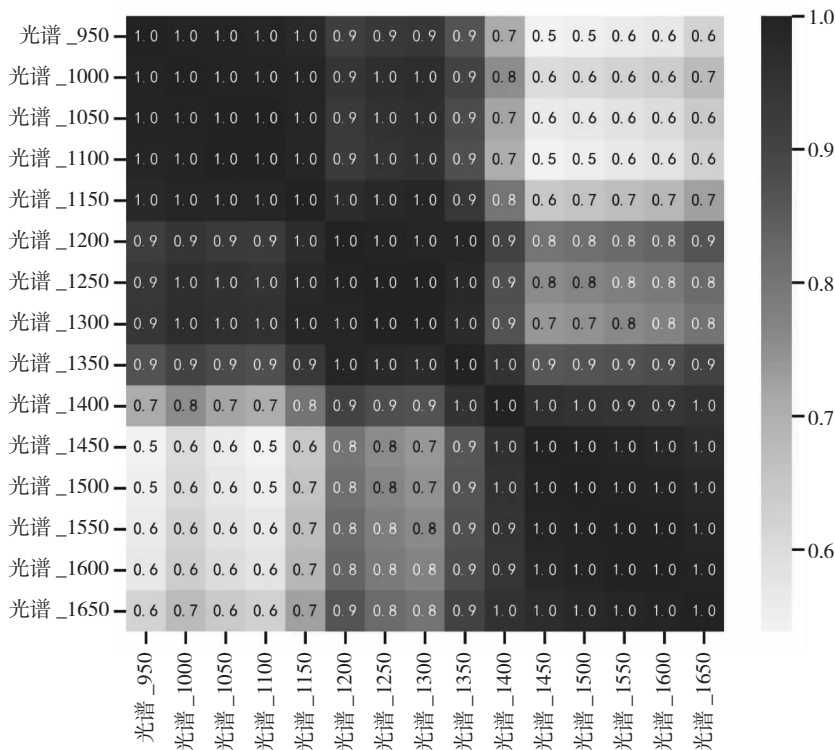


图 5 烟叶近红外光谱数据之间的相关性分析

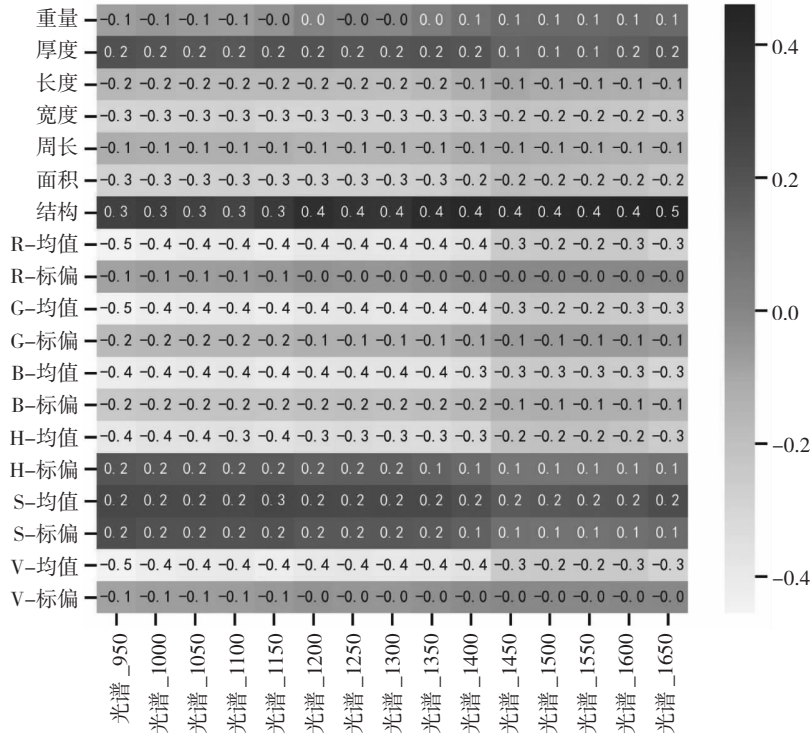


图 6 烟叶的外观指标、物理特性和近红外光谱数据的相关性分析

数据的相关性进行分析(图 6)可以看出, 烟叶的外观指标和物理特性与近红外光谱数据之间的相关性比较弱。综上所述, 自变量之间存在比较多的共线性, 后续建模时需要消除这种共线性。

2.2 模型训练与模型选择

对训练集数据进行样本平衡、数据标准化、数据降维(保留数据方差占比 0.999 5 的主成分)处理后, 采用 QDA、KNN、SVC 3 种模型对上述处理后的数据进行训练并优化。由表 1 可知, 烟叶感官质量各指标中, SVC 模型的平衡准确率均表现最高, 为 0.725~0.773。其中, 感官质量各指标模型中, 劲头的平衡准确率最高, 达到 0.773; 香气风格彰显程度的平衡准确率最低, 为 0.725。

表 1 验证集中感官质量各指标模型中各模型的平衡准确率^①

感官质量指标	SVC	KNN	QDA
香气风格彰显程度	0.725	0.700	0.643
香气质	0.757	0.701	0.659
香气量	0.764	0.730	0.698
杂气	0.727	0.684	0.639
浓度	0.737	0.721	0.658
劲头	0.773	0.766	0.711
刺激性	0.727	0.688	0.634
余味	0.743	0.722	0.627
甜感	0.743	0.719	0.662

① SVC 为支持向量机中的一种分类算法; KNN 为 K 近邻算法; QDA 为二次判别分析算法。

2.3 模型预测与模型评估

采用最终选择的 SVC 模型对测试集数据进行预测的结果(表 2)表明, 感官质量各指标中, 劲头模型的平衡准确率最高, 达到 0.774, 这与验证集的表现一致; 刺激性的平衡准确率最低, 为 0.685。

表 2 测试集中感官质量各指标模型的平衡准确率

感官质量指标	SVC
香气风格彰显程度	0.747
香气质	0.751
香气量	0.715
杂气	0.720
浓度	0.712
劲头	0.774
刺激性	0.685
余味	0.725
甜感	0.700

测试集数据中感官质量各指标的混淆矩阵见图 7。对角线上的数据为预测值与真实值表现一致的数量。在香气风格彰显程度模型中, 类别 D 的检出率最高, 100 个样品中可以检测出 86 个; 类别 B 的检出率最低, 236 个样品中只检测出 160 个。在香气质模型中, 类别 C-的检出率最高, 359 个样品中检测出 304 个; 类别 B 的检出率最

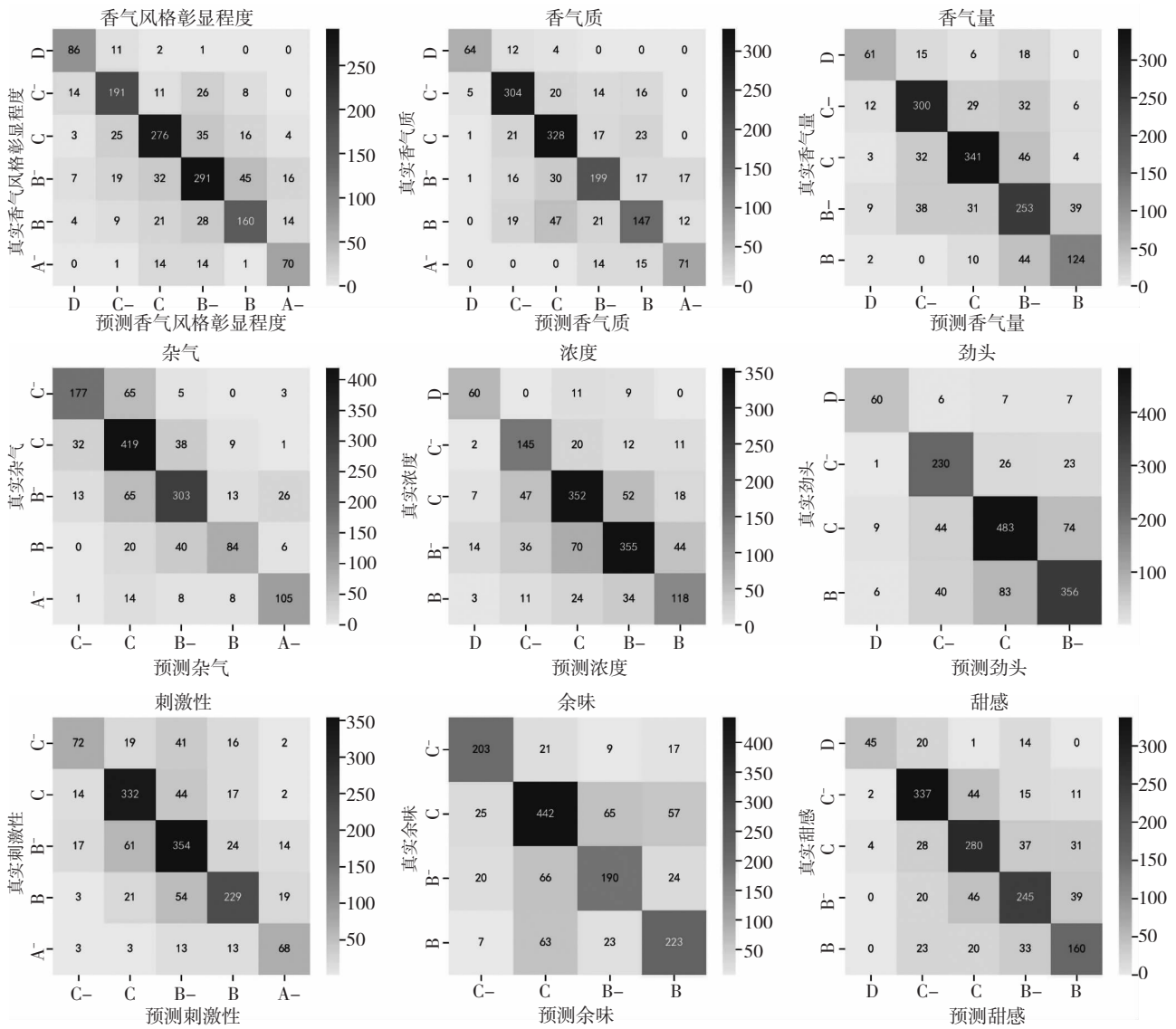


图 7 测试集数据中感官质量各指标混淆矩阵

低, 246 个样品中只检测出 147 个。在香气量模型中, 类别 C 的检出率最高, 426 个样品中检测出 341 个; 类别 D 的检出率最低, 100 个样品中只检测出 61 个。在杂气模型中, 类别 C 的检出率最高, 499 个样品中检测出 419 个; 类别 B 的检出率最低, 150 个样品中只检测出 84 个。在浓度模型中, 类别 C- 的检出率最高, 190 个样品中检测出 145 个; 类别 B 的检出率最低, 190 个样品中只检测出 118 个。在劲头模型中, 类别 C- 的检出率最高, 280 个样品中检测出 230 个; 类别 B- 的检出率最低, 485 个样品中只检测出 356 个。在刺激性模型中, 类别 C 的检出率最高, 409 个样品中检测出 332 个; 类别 C- 的检出率最低, 150 个样品中只检测出 72 个。在余味模型中, 类别 C- 的检出

率最高, 250 个样品中检测出 203 个; 类别 B- 的检出率最低, 300 个样品中只检测出 190 个。在甜感模型中, 类别 C- 的检出率最高, 409 个样品中检测出 337 个; 类别 D 的检出率最低, 80 个样品中只检测出 45 个。

3 讨论与结论

探索数据的分布情况对后续的建模起着至关重要的作用, 是一个不容忽视的环节。通过研究自变量数据的分布情况, 可以观察样本中是否存在异常点, 因为有些模型比如 QDA 等对于异常点会非常敏感^[8], 模型会试图拟合这部分数据, 导致测试集数据的预测效能比较差。因此, 本研究通过观察数据的分布去除了一些厚度异常的数据。同时, 观察数据的分布也可以帮助研究选择合适

的模型, 比如生成式模型对数据的分布要求很高。QDA 作为一种生成式模型, 不要求每个类别的协方差相同, 普适性更高, 但其假设条件就是样本必须服从正态分布。本研究对数据的分布进行探索, 去除异常值后, 发现数据基本服从正态分布, 因此可以构建 QDA 模型。从烟叶的感官质量各类别的数据分布来看, 各类别之间均存在不平衡的情况, 主要表现在中间类别多、两边类别少的情况, 这说明工业生产中的烟叶质量大部分处于一般水平, 质量很高和质量较低的烟叶均较少, 这跟工业生产中烟叶质量的真实情况一致。这种不平衡数据挖掘在其他很多领域也是普遍存在的, 基于精度的传统分类算法和 SVM 在进行分类时^[9], 模型会偏重于训练类别多的样本, 会忽视类别少的样本, 造成类别少的数目被错误预测的概率增大; 而通过增加类别少的数据的数量, 可以纠正模型预测的“有偏性”; 最后采用有效的评估指标进行模型评估, 而不能再继续采用准确率来进行评价。本研究通过对类别少的样本进行上采样, 以达到各类别之间样本数目的平衡, 从而避免模型预测的“有偏性”, 同时采用平衡准确率对模型效能进行评估。

对自变量间的相关性分析可以探讨数据之间是否存在多重共线性。多重共线性问题在很多模型的实际应用中普遍存在, 比如多元线性回归分析^[10], SVM 分类算法^[11]。而多重共线性产生的原因可能来自两个方面, 一是自变量之间客观存在共线性关系, 二是样本数据不是足够多。本研究应该属于第一种情况, 烟叶的颜色值比如 R, G, B 分别代表红色, 绿色和蓝色通道的值, 其不仅包含了颜色信息, 还包含了亮度等信息。而 H 代表色调, 包含了所有颜色值的信息, 因此, R, G, B, H 通道之间必然相互关联, 这与本研究中这 4 个通道之间相关性很高的结果是一致的。主成分分析为这种共线性问题的解决提供了一种有效的方法。通过主成分分析可以改进多重共线性的问题, 将多个指标综合为几个主成分进而减少信息的冗余^[12-13], 本研究也是通过 PCA 实现数据降维, 从而消除变量间的多重共线性。

本研究同时采用烟叶的外观物理特性及近红外光谱数据, 各类型之间的数据由于量纲的不同,

其数值差异也很大, 比如近红外光谱数据为小于 1 的值, 而烟叶的颜色值可达到 200 多。基于距离度量的模型如 KNN、SVM 模型以及线性回归类的模型如逻辑回归、岭回归、拉索回归等对于特征之间不同取值范围非常敏感, 因此建模前必须要进行标准化, 以消除数据之间的差异。而近红外光谱数据, 由于受到采集环境、仪器的运行状态或者技术人员操作等各种因素的影响, 在采集时会存在基线偏移、干扰噪声等问题^[14], 这会导致后期建立的模型出现偏倚, 因此必须对光谱数据进行校正。本研究采用 SNV 对光谱数据进行预处理, 提高 KNN、SVM、QDA 模型的精度。

SVC 由于其在高维空间中非常有效, 采用了结构风险最小化使模型泛化能力提高, 因此在烟草等多个领域广泛应用。有研究对中红外和近红外光谱数据进行融合后, 利用支持向量机可以建立烟叶浓香型、中间香型及清香型的分类判别模型, 准确率均大于 85%^[15]。通过烟叶的化学成分可以较好地预测烟叶感官品质中的香气量^[16]。本研究选用 QDA、KNN、SVC 算法构建烟叶感官质量各个指标的多分类模型, 结果表明 SVC 的预测效能最优, 烟叶感官质量各指标的平衡准确率 0.685 ~ 0.774, 这进一步说明 SVC 模型在预测烟叶感官质量各指标方面有着较大的优势。

近年来, 计算机图像处理技术在生物、医学、农业、建筑等方面都得到广泛应用。如通过对花牛苹果叶片图像进行处理, 提取图像中 R(红色)、G(绿色)、B(蓝色)各参数值以及各组合值, 建立模型, 可以快速有效无损检测苹果叶片的营养状况^[17]。近红外光谱技术作为一种快速、无损检测物质的方法, 在农业、工业、食品等领域也得到广泛应用^[18]。本研究也是通过图像处理技术提取烟叶图像各颜色信息(RGB、HSV 颜色空间的颜色信息)和物理特性, 结合近红外光谱技术检测烟叶化学成分的源头数据(近红外光谱数据), 进而预测烟叶的感官质量各指标, 其数据种类更全面, 更能代表烟叶的整体质量水平, 这为工业生产中判定烟叶的整体质量水平带来了便利, 可以快速检测烟叶的整体质量, 提高工业检测效率。

烟叶的外观指标、物理特性及近红外光谱共同影响着烟叶的感官质量。基于烟叶的外观物理

特性以及近红外光谱数据, 可以预测烟叶的香气风格彰显程度、香气质、香气量、杂气、浓度、劲头、刺激性、余味和甜感等感官质量指标。在这些感官质量指标中, 基本上类别 C 和类别 C- 的检出率最高, 类别 B、类别 B- 和类别 D 的检出率最低, 这可能与工业生产中类别 C 和类别 C- 的烟叶样品比较普遍, 而类别 B、类别 B- 和类别 D 的样品数量比较稀少有关。同时, 由于本研究是基于 2 个品种的烟叶构建的感官质量模型, 而这 2 个品种之间的感官评吸标准可能会有些差异, 这也可能是模型平衡准确率未超过 0.8 的原因。未来可能需要把两个品种分别构建烟叶的感官质量模型。3 个训练模型中, SVC 的泛化能力最优。其中, 香气风格彰显程度、香气质、香气量、杂气、浓度、劲头、刺激性、余味及甜感的预测平衡准确率分别为 0.747, 0.751, 0.715, 0.720, 0.712, 0.774, 0.685, 0.725, 0.700。

参考文献:

- [1] 梁洪波, 李念胜, 元 建, 等. 烤烟烟叶颜色与内在品质的关系[J]. 中国烟草科学, 2002, 23(1): 9-11.
- [2] 王 欣. 湖北烟区烤烟质量综合评价及与国内外优质烤烟的差异分析[D]. 郑州: 河南农业大学, 2008.
- [3] 张勇刚, 宋朝鹏, 李常军. 烟叶感官质量评价研究进展[J]. 湖北农业科学, 2010, 49(9): 2271-2274.
- [4] 闫铁军, 马俊桃, 刘文锋, 等. 烟叶外观质量与感官舒适性的相关性分析[J]. 湖北农业科学, 2021, 60(23): 109-113.
- [5] 陈剑明, 杨式华, 冯洪涛, 等. 云南主产烟区烟叶水溶性糖与感官质量相关及通径分析[J]. 西南农业学报, 2017, 30(11): 2506-2511.
- [6] 林顺顺, 张晓鸣. 基于 PLSR 分析烟叶化学成分与感官质量的相关性[J]. 中国烟草科学, 2016, 37(1): 78-82.
- [7] 潘义宏, 李佳佳, 蒋美红. 烟叶外观质量、常规化学成分与其感官质量的典型相关分析[J]. 江苏农业科学, 2015, 43(10): 384-388.
- [8] 王秀梅. 基于二次判别模型(QDA)的土壤镉安全阈值判定[D]. 北京: 中国农业科学院, 2018.
- [9] 郑恩辉, 李 平, 宋执环. 不平衡数据知识挖掘_类分布对支持向量机分类的影响[J]. 信息与控制, 2005, 34(6): 703-708.
- [10] 鲁 茂, 贺昌政. 对多重共线性问题的探讨[J]. 统计与决策, 2007(8): 6-9.
- [11] 冼广铭, 齐德昱, 方 群. 改进 SVM 分类算法中多重共线性问题研究[J]. 计算机工程与应用, 2010, 46(26): 142-144.
- [12] 陶 然. Logistic 模型多重共线性问题的诊断及改进[J]. 统计与决策, 2008(15): 22-24.
- [13] 姜有虎, 李玉梅, 李旭林, 等. 基于主成分分析的嘉峪关产区马瑟兰葡萄最佳采收期确定[J]. 甘肃农业科技, 2022, 53(1): 94-98.
- [14] 李尚科, 李 跑, 杜国荣, 等. 基于近红外光谱技术和优化预处理方法的不同品牌燕麦无损鉴别分析[J]. 食品安全质量检测学报, 2019, 10(24): 8204-8210.
- [15] 沙云菲, 黄 雯, 王 亮, 等. 中红外和近红外数据融合的香型风格判别[J]. 光谱学与光谱分析, 2021, 41(2): 473-476.
- [16] 赵青松, 李兴兵, 唐小松. 基于支持向量机的烟叶感官品质评价[J]. 计算机工程与应用, 2007, 43(10): 236-238.
- [17] 杨焕昱, 李帼英, 马建芳, 等. 基于图像数字处理技术的元帅系苹果叶片营养水平诊断研究[J]. 甘肃农业科技, 2022, 53(3): 59-63.
- [18] 张华瑜, 潘永东, 柳小宁, 等. 利用近红外谷物分析仪快速检测青稞粗蛋白质含量研究[J]. 甘肃农业科技, 2020(1): 33-36.