

Multi-Domain Feature Fusion Based Chinese Liquor Recognition Using a Hand-Held Electronic Nose^{*}

WANG Qian, MENG Qinghao, JIN Licheng^{*}

(Institute of Robotics and Autonomous Systems, Tianjin Key Laboratory of Process Detection and Control; School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China)

Abstract: In order to improve the accuracy of the hand-held electronic nose (e-nose) for distinguishing Chinese liquor, a recognition method based on multi-domain feature fusion is proposed. First, the time domain, time-frequency domain and spatial domain features of the signals are extracted based on statistical analysis, wavelet packet analysis and one versus rest common spatial pattern (OVR-CSP), respectively. Second, the time domain, time-frequency domain and spatial domain features are fused by a weighting method. A self-made hand-held e-nose was used to collect 6 kinds of Chinese liquor samples and perform recognition experiments. Taking the support vector machine (SVM) as the classifier, and the results show that the recognition accuracy of the proposed multi-domain feature fusion method has been increased by 9.83%, 8% and 1.5%, compared with time domain, time-frequency domain and spatial domain features alone, respectively. The recognition accuracy and running time of the system based on SVM, K-nearest neighbour (KNN), and back propagation artificial neural network (BP-ANN) were compared, and the results show that the KNN algorithm takes the shortest time and has a relatively high recognition accuracy.

Key words: hand-held electronic nose; liquor recognition; wavelet package; OVR-CSP; feature extraction

EEACC: 7230

doi: 10.3969/j.issn.1004-1699.2021.02.001

基于多域特征融合的手持式电子鼻白酒识别^{*}

王 茜, 孟庆浩, 靳荔成^{*}

(天津大学机器人与自主系统研究所, 天津市过程检测与控制重点实验室, 天津大学电气自动化与信息工程学院, 天津 300072)

摘 要: 为提升手持式电子鼻辨别白酒的准确率, 本文提出了一种基于多域特征融合的识别方法。首先, 分别基于统计学分析、小波包分析和一对多共同空间模式 (One Versus Rest Common Spatial Pattern, OVR-CSP) 提取信号的时域、时频域和空域特征。其次, 基于特征加权的方法将时域、时频域和空域特征进行融合。利用自制的手持式电子鼻采集了 6 种白酒样本并进行识别实验, 结果显示, 采用支持向量机 (Support Vector Machine, SVM) 作为分类器时, 与单独使用时域、时频域和空域特征相比, 所提多域融合特征方法的识别准确率分别提高了 9.83%、8% 和 1.5%。进一步比较了 SVM、K 近邻 (K-Nearest Neighbour, KNN) 和 BP 神经网络 (Back Propagation Artificial Neural Network, BP-ANN) 三种分类识别方法识别准确率和运行时间, 结果表明, KNN 算法的用时最短, 且识别率较高。

关键词: 手持式电子鼻; 白酒识别; 小波包; OVR-CSP; 特征提取

中图分类号: TP391.4

文献标识码: A

文章编号: 1004-1699(2021)02-0143-07

我国是白酒生产和消费的大国。根据香型、品牌及制作工艺的不同, 我国白酒可有几千种之多。近些年, 白酒市场中以次充好、以假乱真的现象时有发生, 如何快速鉴别白酒的品质困扰着普通消费者。现有用于白酒鉴别的方法包括感官法、色谱法、光谱法和传感器阵列等方法。感官法主要通过有经验的

品酒师完成, 具有较强的主观性, 受影响因素较多。色谱法和光谱法依赖昂贵的分析仪器, 成本高、耗时长, 且仪器的体积普遍偏大, 不便于日常使用^[1-2]。电子鼻^[3]作为一种新兴检测技术, 已在食品鉴别、生物检验、医疗诊断、环境监测、烟草检测等^[4-8]多个领域有成功应用的案例。

项目来源: 国家自然科学基金项目 (61573252); 国家重点研发计划项目 (2017YFC0306200)

收稿日期: 2020-09-30 **修改日期:** 2020-12-19

电子鼻在酒类识别方面也有了一些研究案例。Pornpanomchai 等^[9]将电子鼻技术和神经网络技术相结合实现了 10 种啤酒品牌的分类。Qi 等^[10]研制了一种由采样模块、气室和 ARM (Advanced RISC Machines) 处理器组成的便携式电子鼻,采用仿生分段呼吸采样方法实现快速采样,并用优化后的支持向量机 (Support Vector Machine, SVM) 算法对六种白酒进行识别,准确率达 90.8%。商业化便携式电子鼻 Cyranose320^[11]已被广泛应用在多个领域。Li 等^[12]研制了用于检测猪肉新鲜度的手持式电子鼻系统。Li 等^[13]利用手持式电子鼻研究了检测白酒方法,能够很好的实现白酒分类。目前,针对于酒类检测的电子鼻系统提取的特征^[14-15]可分为静态特征和动态特征。静态特征包括基线值、峰值、均值、稳态响应时间等,动态特征包括响应曲线在某点的积分值、微分值、瞬态响应值等^[16-21]。

为提升白酒识别的准确率,除常用的时域特征提取方法外,本文将基于小波包分析的时频域特征及基于一对多共同空间模式 (One Versus Rest Common Spatial Pattern, OVR-CSP) 分析的空域特征^[22]引入到手持式电子鼻的白酒识别研究,并通过可分度加权的方式对时域、时频域和空域三种特征进行融合。利用实验室自制的手持式电子鼻对 6 种白酒进行检测与识别,实验结果表明本文所提的多域特征融合提取方法提高了分类的平均准确性。在此基础上,比较了 SVM、K 近邻 (K-Nearest Neighbour, KNN) 和 BP 神经网络 (Back Propagation Artificial Neural Network, BP-ANN) 三种分类器的实时性和分类效果。

1 实验设计

1.1 白酒用手持式电子鼻平台

本文所采用的实验室自制手持式电子鼻如图 1 所示,左侧为电子鼻展开结构图,主要包括 Arm 系统板、系统扩展板、气室、气路、LCD 触摸屏等部件,右侧为电子鼻整体结构图,该电子鼻系统总重 283 g,长 11 cm,宽 9 cm,高 3 cm,平均功率小于 5 W。

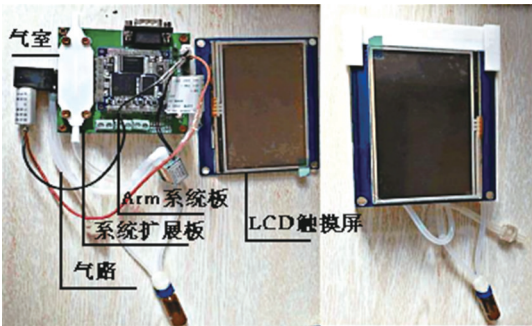


图 1 用于白酒检测的手持式电子鼻实物图

传感器阵列是电子鼻系统的气味感知部件,传感器阵列的选择是根据不同传感器对白酒气味物质的响应特点,并以较少的传感器数量获取更多的气味信息,以及基于传感器之间对不同气味信息交叉敏感为原则进行阵列优化。本文所用的手持式电子鼻的传感器阵列包含 6 种金属氧化物半导体 (Metal Oxide Semiconductor, MOS) 型传感器和一个温湿度传感器, MOS 型气体传感器型号分别为 CCS801、CCS803、TGS8100、MiCS-5914、MiCS-5524、MiCS-4514,其中 MiCS-4514 输出两路信号,传感器具体信息如表 1 所示。

表 1 传感器阵列的传感器型号

型号	敏感气体	厂家
CCS801	甲醛、酒精、CO 等	英国 CCS
CCS803	甲醛、酒精、CO 等	英国 CCS
TGS8100	CO、烃类、酒精等	日本 Figaro
MiCS-5914	CO、烃类、酒精等	英国 E2V
MiCS-5524	CO、烃类、酒精等	英国 E2V
MiCS-4514	CO、烃类、酒精等	英国 E2V

1.2 数据采集与数据预处理

1.2.1 实验材料与数据采集

本文选取六种白酒进行实验,分别为国窖 1573、西凤、汾酒、剑南春、五粮液和飞天茅台。六种白酒的具体信息如表 2 所示。

表 2 六种白酒的具体信息

白酒名称	酒精含量	白酒香型	产地
国窖 1573	52%vol	浓香型	四川/泸州
西凤	52%vol	凤香型	陕西/宝鸡
汾酒	42%vol	清香型	山西/汾阳
剑南春	52%vol	浓香型	四川/绵竹
五粮液	52%vol	浓香型	四川/宜宾
飞天茅台	53%vol	酱香型	贵州/茅台

在实验室中进行白酒样本数据采集,室内温度为 28 ℃,空气相对湿度为 30%。手持式电子鼻实验采样参数设置:采样频率 400 Hz,采样时间 40 s,其中吸气时间 1 s,吸气等待时间 3 s,呼气时间 35 s,呼气等待时间为 1 s,气室清洗时间为 30 s。

1.2.2 数据预处理

采用气泵将洁净空气吹入白酒样本瓶,吹入空气携带挥发的白酒样本从气路系统进入气室中,与传感器阵列发生反应,其中吸气-吸气等待-呼气-呼气等待过程为采样周期,完成气体采样,采样结束后进行气室清洗。TGS8100 传感器原始响应曲线如图 2 所示。

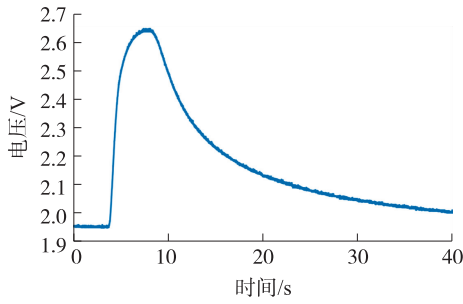


图2 TGS8100 传感器原始响应曲线

电子鼻获取的数据是经过 RC 电路消除工频干扰噪声后的结果,但此时响应曲线依旧存在一些噪声和波动,为便于后续数据处理,应进一步对数据进行滤波。本文先采用中值滤波去噪,中值滤波对去除硬件电路中的孤立脉冲噪声具有较好效果,同时也能使曲线更加平滑。中值滤波算法的思想是用某一点邻域窗口内所有响应值的中位值去代替这一点的真实值,如式(1)所示:

$$y(i) = \text{Median}\{x(i-n), \dots, x(i), \dots, x(i+n)\} \quad (1)$$

式中: $x(i)$ 为位于窗口中心的采样值,窗口中心采样值邻域窗口长度 $L=2n+1$,其中 n 为正整数,本文中 n 取值为51, $y(i)$ 为滤波后的响应值。

经过中值滤波后可以得到较为平滑响应曲线,但仍有一些波动,为减少这些波动对后续数据处理的不利影响,继续使用局部回归加权算法进行平滑滤波,得到波动更小、更平滑的响应曲线。局部回归加权滤波算法的具体方法如下:

在采样点处建立一个 $L=2n+1$ 的滤波窗口, n 为窗口半径,本文中 n 取31,设各个测量点为 $\mathbf{X} = [-n \dots n]^T$,每点对应的响应值 $\mathbf{Y} = [x(-n) \dots x(n)]^T$ 。对位于窗口内的数据用权值函数进行加权线性回归,取回归线的中心值作为该采样点滤波后的值。因为权值公式选取并不固定,满足一定规则即可,本文所采用的局部回归权值公式如式(2)所示,并利用式(3)求取加权回归模型。

$$g(i) = \left[1 - \left| \frac{x-x(i)}{d(x)} \right|^3 \right]^3 \quad (2)$$

$$\hat{\mathbf{Y}} = \mathbf{X}(\mathbf{X}^T \mathbf{G} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{G} \mathbf{Y} \quad (3)$$

式中: x 为当前传感器响应值, $x(i)$ 为位于平滑窗口内的相邻点, $d(x)$ 为 x 到窗口内紧邻点距离最大值, $g(i)$ 为窗口内每点的权值, $\mathbf{G}$ 为权值矩阵, $\mathbf{G} = \text{diag}[g(-n) \dots g(n)]$,其中 $g(-n) \sim g(n)$ 窗口内每个点对应的权值, $\hat{\mathbf{Y}}$ 为求取的加权回归模型。

图3为TGS8100传感器滤波去噪后曲线响应图,经过中值滤波和平滑滤波后,获得了更加平滑的曲线。

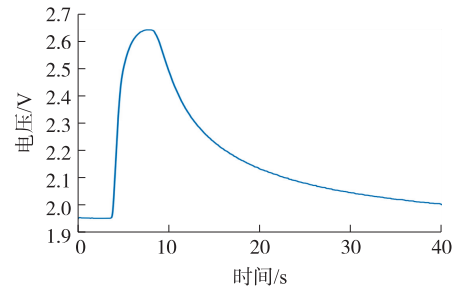


图3 TGS8100 传感器滤波后响应曲线

2 特征提取

本文提出的用于白酒检测的手持式电子鼻多域特征融合方法包括:提取基于统计学方法传感器响应曲线的时域特征,提取基于小波包分解和重构的时频域特征,提取基于OVR-CSP的空域特征,对所有特征进行归一化处理以及利用特征加权方法进行特征融合。

2.1 基于统计学分析的时域特征提取

基于统计学分析对响应曲线进行时域特征提取,计算时域波形数据的统计指标,提取其静态特征和动态特征。通过对手持式电子鼻采样方式和响应曲线特点的分析,选取了6个时域特征,分别为:一阶微分最大值 M_1 ,一阶微分最大值时刻对应的响应值 M_2 ,采样时间内的响应曲线的积分值 M_3 ,采样时间内微分平均值 M_4 ,最大响应值与响应初始时刻对应响应值的差值 M_5 ,二阶微分最大值 M_6 ,提取的时域特征表达如下:

$$M_1 = \max \left(\frac{S_{t+1} - S_{t-1}}{2\Delta t} \right) \quad (4)$$

$$M_2 = \sum_{t=1}^T S_{t_d} \quad (5)$$

$$M_3 = \sum_{t=1}^T S_t \quad (6)$$

$$M_4 = \frac{1}{T-1} \sum_{t=1}^{T-1} \frac{S_{t+1} - S_{t-1}}{2\Delta t} \quad (7)$$

$$M_5 = \max(S_t) - S_{t_0} \quad (8)$$

$$M_6 = \max \left(\frac{S_{t+1} - 2S_t + S_{t-1}}{\Delta t^2} \right) \quad (9)$$

式中: S_{t-1} 、 S_t 、 S_{t+1} 分为 $t-1$ 、 t 、 $t+1$ 时刻响应值, Δt 为采样间隔时间, t_d 代表一阶微分最大值对应的时刻, T 为采样周期, t_0 为响应初始时刻, t_s 为曲线稳态响应时刻。

对手持式电子鼻的6个传感器输出的7条响应曲线分别提取上述的6个时域特征,将其整合成单个样本时域特征向量 $f_1 \in R^{m_1 \times 1}$,其中 m_1 表示时域中提取的特征总个数,本文中 $m_1=42$ 。

2.2 基于小波包分析的时频域特征提取

小波分析是时频分析的一种,是对短时傅里叶变换局部化思想的继承和发展,不仅克服了傅里叶分析中窗口大小不能跟随频率改变的缺点,而且提供了一个随频率变化的时间频率。窗口应用多分辨率分析,通过尺度函数进行伸缩和平移,对信号进行多频带分解。小波分析将时间序列分解为低频和高频两部分,并对低频信息部分进一步分解,高频部分则不再处理。小波分析可以很好地表征信号低频信息,但会忽略高频的大量细节信息。小波包分析在小波分析的基础上,同时进一步分解低频和高频两部分,以获取更全面的信号特征。

由于 Daubechies 小波的正交性、高阶连续可导、对称性、小时频窗口和高消失矩的特性,本文选取 3 阶的 Daubechies 进行分解,经过多次验证,当分解尺度为 3 时既有良好的去噪效果,且信号不失真,同时计算量较小。本文中选取 db3 小波函数对各个传感器响应信号进行 3 层小波包分解与重构,将传感器时间响应序列 S 分解为低频信息 a 和高频信息 d 两部分,下一层分解中,继续将低频信息 a 和高频信息 d 部分分解为 aa, ad, da, dd 四部分,以此类推,经过三层分解后可得到 $aaa, aad, ada, add, daa, dad, dda, ddd$ 共 8 个频带,图 4 为在各个尺度上分解与重构后得到的小波树图,分别提取第三层小波重构后能量的最大值及小波包能量熵值作为白酒类别的时频域特征。其中小波包能量熵值是基于小波包分解系数计算得到熵值,该值是对信号中小波分解重构后能量分布的平均复杂度表征。

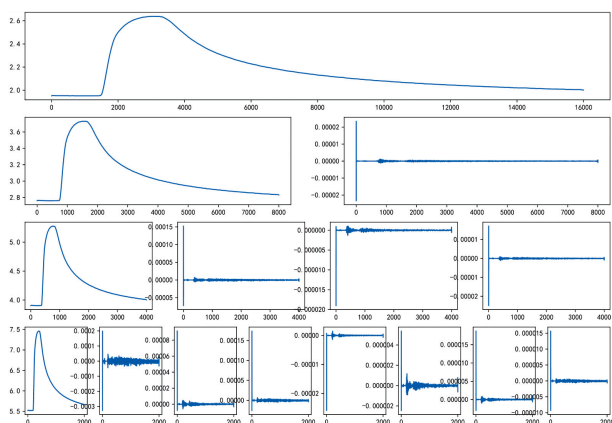


图 4 传感器响应信号 3 层小波包分解重构结果

基于小波包分析的小波包能量特征和小波包能量熵特征的提取步骤如下:

①选择 db3 小波函数对预处理后传感器响应曲线进行 3 层小波包分解,获取第三层各个节点对应的小波包系数,分别为 d_{3k} ,其中 $k=1,2,\dots,8$ 。

②利用所求的第 3 层各节点小波包系数重构各节点信号 $S_{3k}, k=1,2,\dots,8$ 。

③利用小波包重构信号计算第三层小波包各节点能量 E_{3k} 和总能量 E ,其中:

$$E_{3k} = \sum_{t=1}^n |S_{3k}(t)|^2, \quad k=1,2,\dots,8 \quad (10)$$

式中: $S_{3k}(t)$ 为第 3 层第 k 个节点重构信号的第 t 个数据, $t=1,2,\dots,n$ 为重构节点的数据个数。

$$E = \sum_{k=1}^8 E_{3k} \quad (11)$$

④计算小波包总能量熵值 W_{EE} :

$$P_{3k} = \frac{E_{3k}}{E}, \quad k=1,2,\dots,8 \quad (12)$$

$$W_{EE} = - \sum_{k=1}^8 P_{3k} \log P_{3k} \quad (13)$$

式中: P_{3k} 为各节点小波归一化能量。

⑤获取时频域特征小波包能量最大值 $E_{\max}$:

$$E_{\max} = \max[E_{3k}], \quad k=1,2,\dots,8 \quad (14)$$

对手持式电子鼻的 6 个传感器输出的 7 条响应曲线分别提取上述的 2 个时频域特征,将其整合成单个样本时频域特征向量 $f_2 \in R^{m_2 \times 1}$,其中 m_2 表示时频域中提取的特征总个数,本文中 $m_2 = 14$ 。

2.3 基于 OVR-CSP 的空域特征提取

CSP 是一种用于两分类任务下的空间滤波算法,目前在脑电领域尤其是运动想象特征提取方面已广泛应用,但在电子鼻的分类识别领域应用较少。CSP 的原理是构造一个空间滤波器,两类信号通过这个滤波器后,一类信号的方差最大,另一类信号的方差最小,从而实现对两种不同类别信号的区分。然而对于白酒种类多样的问题, CSP 无法完成分类任务,因此本文采用 OVR-CSP 方法实现多种白酒种类的识别。对于文中的 6 种白酒分类而言, OVR-CSP 是将其分成 6 个一对一的两分类问题,依次计算每个滤波器,共需要构建 6 个空间滤波器。基于 OVR-CSP 的空域特征提取步骤如下:

①将其中某一类白酒视为一类信号,其余白酒种类视为另一类信号,求取两种信号的空间滤波器。计算两类白酒信号矩阵的空间协方差均值矩阵:

$$\bar{R}_s = \frac{1}{n} \sum_{i=1}^n \frac{X_{s,i} X_{s,i}^T}{\text{trace}(X_{s,i} X_{s,i}^T)}, \quad s=1,2 \quad (15)$$

式中: $\bar{R}_s$ 为第 s 类白酒信号的空间协方差均值矩阵, $X_{s,i}$ 表示第 s 类酒的第 i 次实验, $X_{s,i}^T$ 为 $X_{s,i}$ 的转置, n 表示同一种酒的样本数量, $\text{trace}()$ 表示矩阵的迹。

计算两类白酒混合空间的协方差矩阵 R , 并进

行特征值分解:

$$\mathbf{R} = \bar{\mathbf{R}}_1 + \bar{\mathbf{R}}_2 = \mathbf{U} \boldsymbol{\Lambda} \mathbf{U}^T \quad (16)$$

式中: $\bar{\mathbf{R}}_1$ 和 $\bar{\mathbf{R}}_2$ 是两类白酒信号的均值协方差矩阵, $\boldsymbol{\Lambda}$ 是 $\mathbf{R}$ 的特征值, $\mathbf{U}$ 是 $\mathbf{R}$ 的特征向量。

求取混合协方差矩阵的白化矩阵 $\mathbf{W}$:

$$\mathbf{W} = \sqrt{\boldsymbol{\Lambda}^{-1}} \mathbf{U}^T \quad (17)$$

利用求取的白化矩阵对两类白酒的协方差均值矩阵进行白化处理:

$$\begin{cases} \mathbf{S}_1 = \mathbf{W} \bar{\mathbf{R}}_1 \mathbf{W}^T \\ \mathbf{S}_2 = \mathbf{W} \bar{\mathbf{R}}_2 \mathbf{W}^T \end{cases} \quad (18)$$

对白化后的矩阵进行特征值分解:

$$\begin{cases} \mathbf{S}_1 = \mathbf{U}_s \boldsymbol{\Lambda}_1 \mathbf{U}_s^T \\ \mathbf{S}_2 = \mathbf{U}_s \boldsymbol{\Lambda}_2 \mathbf{U}_s^T \end{cases} \quad (19)$$

式中: $\mathbf{U}_s$ 是 $\mathbf{S}_1$ 和 $\mathbf{S}_2$ 的共同特征向量, $\boldsymbol{\Lambda}_1$ 和 $\boldsymbol{\Lambda}_2$ 对应 $\mathbf{S}_1$ 和 $\mathbf{S}_2$ 的特征值, 且 $\boldsymbol{\Lambda}_1$ 和 $\boldsymbol{\Lambda}_2$ 的和为单位矩阵。因此, 当 $\mathbf{S}_1$ 的特征值最大时, $\mathbf{S}_2$ 特征值最小, 两类信号的协方差区别最大, 因此可构造出空间滤波器:

$$\mathbf{Q} = \mathbf{U}_s^T \mathbf{W} \quad (20)$$

按照此种方法构造出所有白酒类别的空间滤波器族 $\mathbf{Q}_c, c=1, 2, \dots, 6$ 。

②空域特征信号提取。

将白酒样本矩阵经过空间滤波投影得到特征矩阵:

$$\mathbf{Z}_j = \mathbf{Q}_c \mathbf{X}_j, c=1, 2, \dots, 6, j=1, 2, \dots, N \quad (21)$$

式中: $\mathbf{X}_j$ 是第 j 个白酒样本, $\mathbf{Z}_j$ 是第 j 个白酒样本经过空间滤波器 $\mathbf{Q}_c$ 滤波投影得到的特征矩阵, N 为白酒样本个数。

取特征矩阵 $\mathbf{Z}_j$ 的所有行的方差作为信号矩阵的空域特征值:

$$v_i = \log \left\{ \text{VAR}[\mathbf{Z}_j(i)] / \sum_{i=1}^k \text{VAR}[\mathbf{Z}_j(i)] \right\} \quad (22)$$

$$i=1, 2, \dots, k$$

式中: $\mathbf{Z}_j(i)$ 表示特征矩阵 $\mathbf{Z}_j$ 的第 i 行, k 为特征矩阵 $\mathbf{Z}_j$ 的总行数, 即电子鼻传感器阵列输出响应曲线条数, 本文中 $k=7$ 。

按照上式分别求取白酒样本在不同滤波器下的特征值, 将其整合成单个样本的空域特征向量 $\mathbf{f}_3 \in \mathbf{R}^{m_3 \times 1}$, 其中 m_3 表示空域中提取的特征总个数, 本文中 $m_3=42$ 。

2.4 多域特征加权融合

提取全部白酒样本的时域特征、时频域特征和空域特征, 生成时域特征、时频域特征和空域特征矩阵 $\mathbf{F}_1, \mathbf{F}_2, \mathbf{F}_3$:

$$\mathbf{F}_1 = [f_{1,1} \quad \dots \quad f_{1,j} \quad \dots \quad f_{1,N}]^T, \mathbf{F}_1 \in \mathbf{R}^{N \times m_1} \quad (23)$$

$$\mathbf{F}_2 = [f_{2,1} \quad \dots \quad f_{2,j} \quad \dots \quad f_{2,N}]^T, \mathbf{F}_2 \in \mathbf{R}^{N \times m_2} \quad (24)$$

$$\mathbf{F}_3 = [f_{3,1} \quad \dots \quad f_{3,j} \quad \dots \quad f_{3,N}]^T, \mathbf{F}_3 \in \mathbf{R}^{N \times m_3} \quad (25)$$

式中: $f_{1,j}, f_{2,j}$ 和 $f_{3,j}$ 分别为时域、时频域和空域下提取的第 j 个白酒样本的特征向量, 其中 $j=1, 2, \dots, N, N$ 为白酒样本个数。

将时域特征、时频域特征和空域特征矩阵组合生成多域特征矩阵 $\mathbf{F}$:

$$\mathbf{F} = [\mathbf{F}_1 \quad \mathbf{F}_2 \quad \mathbf{F}_3], \mathbf{F} \in \mathbf{R}^{N \times m} \quad (26)$$

式中: m 为多域特征矩阵所含特征个数, 本文中 $m=m_1+m_2+m_3=98, N$ 为白酒样本个数。

在进行特征加权融合前, 为减少各个特征值量纲不统一对后续数据处理的影响, 对特征矩阵进行归一化处理, 归一化公式如下所示:

$$a_i = \frac{a_i - a_{\min}}{a_{\max} - a_{\min}} \quad (27)$$

式中: a_i 表示特征矩阵中的某个特征的一个特征值, $a_{\min}$ 为这个特征值中最小的值, $a_{\max}$ 为特征值中最大的值。

特征加权融合方法的原理是根据一定的准则去评价每个特征值的重要程度, 并根据他们的重要程度去赋予不同的权值, 求取其权值矩阵。本文中采用样本特征的可分度函数去计算特征权值, 对特征以及不同类间的相关性进行量化。某个特征的可分度越大, 则赋予其更高的权重值, 否则则赋予其更低的权重值。定义特征的可分度函数为:

$$D_i = \frac{\sum_{j=1, k=1}^c (\mu_{ij} - \mu_{ik})^2}{\sum_{p=1}^c \sigma_{ip}^2} \quad (28)$$

式中: μ_{ij} 表示样本中所有属于 j 类白酒的第 i 个特征值的均值, μ_{ik} 表示样本中所有属于 k 类白酒的第 i 个特征值的均值, σ_{ip} 表示样本中所有属于第 p 类白酒的第 i 个特征值的方差, c 表示白酒种类数量, 文中 c 等于 6。

D_i 值越大, 表示第 i 个特征值对区分各类白酒样本集的作用越大, 通过可分度函数, 对特征矩阵中的所有特征分量进行计算, 定义特征的权重系数:

$$\begin{cases} w_i = \frac{D_i}{\alpha} \\ \sum_{i=1}^m w_i = 1 \end{cases} \quad (29)$$

式中: α 为常量, m 为特征矩阵维数, w_i 为第 i 个特征值的权重系数。

则特征权值对角矩阵 $\mathbf{P}$ 和特征加权矩阵 $\mathbf{F}_w$ 为:

$$P = \text{diag}[w_1 \cdots w_i \cdots w_m] \quad (30)$$

$$F_w = FP \quad (31)$$

式中: $w_1 \sim w_m$ 为根据式(29)求得的各个特征值的权重系数。

基于上述可分度函数,计算每个特征值的归一化权值,最终得到所有特征值可分度权值,并按照权值大小降序排列。表 3 给出按照降序排列后的前 12 个特征及对应的权值,符号 $f_{a,b}$ 及其下面的数字分别表示特征及对应权值。其中,下角标 a 取 1、2、3,分别对应时域、时频域和空域下提取的特征;下角标 b 对应 a 值为 1、2 和 3 时的取值范围分别是 1~42、1~14 和 1~42,表示为时域、时频域及空域下的第 b 个特征。通过相关性分析,可以直观地看出不同特征值对白酒类别识别的重要程度,为每个特征值赋予不同权值。

表 3 按可分度函数权值大小降序排列的前 12 个特征

$f_{3,14}$	$f_{3,10}$	$f_{3,39}$	$f_{1,9}$	$f_{3,24}$	$f_{1,32}$
0.048 7	0.042 1	0.036 8	0.036 8	0.036 7	0.034 7
$f_{2,13}$	$f_{1,3}$	$f_{2,4}$	$f_{3,20}$	$f_{1,19}$	$f_{3,2}$
0.034 6	0.032 3	0.027 0	0.024 5	0.024 0	0.021 3

3 结果与分析

3.1 单一特征和多域特征对比

为验证本文所述方法的有效性,采集 6 种白酒共 395 个样本,其中每种白酒样本 60~70 个。由于样本数量有限,所以采用交叉验证的方式进行分类器的性能评估。每次随机选择 35 个样本作为测试样本,其余为训练样本,迭代次数 20 次,取 20 次模型下识别结果的平均值作为识别准确率。选取 SVM 分类器作为分类器,单一特征和多域特征下的平均分类准确率如表 4 所示。

表 4 不同特征的平均准确率

识别准确率	F_1	F_2	F_3	F_w
国窖 1573	82.65%	79.31%	96.15%	98.15%
西凤	98.17%	98.46%	100%	100%
汾酒	98.10%	98.75%	100%	100%
剑南春	71.96%	75.00%	89.22%	95.00%
五粮液	94.20%	90.59%	98.80%	100%
飞天茅台	82.14%	91.07%	91.82%	93.13%
总正确率	87.50%	89.33%	95.83%	97.33%

从表 4 中可以看出,使用本文的多域融合特征 F_w 时,能达到 97.33% 的平均正确率,与单一的时域、时频域和空域特征相比,平均识别正确率分别提高了 9.83、8 和 1.5 个百分点。因为与单一域特征相

比,多域特征加权融合实现了信号不同域下特征的互补,同时根据可分度函数计算每个特征值对白酒分类的不同重要程度设置特征权值,从而获得了更好的分类识别效果。

3.2 不同分类算法对比

选取三种常用的分类识别算法 SVM、KNN 和 BP-ANN 进行多域特征下的白酒种类识别,由于多域特征加权融合后的特征维数仍然较大,因此选取核主成分分析(Kernel Principal Component Analysis, KPCA)对特征空间进行降维,减少特征冗余的不利影响,提高分类器的效率。表 5 所示为不同分类器下各类白酒的分类结果。

表 5 不同分类算法的平均准确率

识别准确率	SVM	KNN	BP-ANN
国窖 1573	98.15%	96.36	97.89%
西凤	100%	100%	100%
汾酒	100%	100%	100%
剑南春	95.00%	96.72%	98.39%
五粮液	100%	99.03%	99.15%
飞天茅台	93.13%	100%	100%
总正确率	97.33%	98.71%	99.23%

从表 5 中可以看出,多域融合特征经 KPCA 降维后,BP-ANN 的准确率最高,比 SVM 和 KNN 分别多 1.9 和 0.52 个百分点。但在训练样本、测试样本数量相同、迭代次数相同的情况下,BP-ANN 的识别时间为 77.48 s,SVM 的识别时间为 4.19 s,KNN 的识别时间为 2.57 s,考虑到白酒用手持式电子鼻的在线快速识别要求,虽然 BP-ANN 识别准确率最高,但运行时间比其他两种分类算法长很多,而 KNN 的识别准确率也很高,运行时间最短,所以建议选择 KNN 作为分类识别算法。

4 结论

本文综合考虑了白酒识别用电子鼻传感器响应曲线的时域、时频域和空域特征,基于多域特征提取方法和自制的用于白酒检测的手持式电子鼻完成了 6 种白酒的采样和识别。在特征提取方面,分别基于统计学方法、小波包分析方法和 OVR-CSP 法提取了每个传感器响应曲线的时域、时频域和空域特征,并运用特征加权融合方法对三类特征进行了融合。实验结果表明,在相同分类算法的前提下,本文提出的多域特征加权融合方法相较于单一域方法的平均准确率更高。最后,分析比较了 SVM、KNN、BP-ANN 三种分类算法对多域融合特征的分类性能,实验表明,三种分类算法的识别率均能达到 95% 以

上,其中 BP-ANN 准确率最高,但运行时间最长。KNN 算法运行时间最短,且识别准确率很高。针对本文的实验结果,建议选择 KNN 作为分类器用于白酒的在线识别。

参考文献:

- [1] 安明哲,廖勤俭,李杨华,等. 高效液相色谱法在白酒分析中的应用[J]. 酿酒科技,2019(11):39-42.
- [2] Shen F, Yang D, Ying Y, et al. Discrimination Between Shaoxing Wines and Other Chinese Rice Wines by Near-Infrared Spectroscopy and Chemometrics[J]. Food & Bioprocess Technology, 2012, 5(2): 786-795.
- [3] Gardner J W, Bartlett P N. A Brief History of Electronic Noses[J]. Sensors and Actuators; B. Chemical, 1994, 18(1-3): 210-211.
- [4] Capelli L, Sironi S, Rosso R Del. Electronic Noses for Environmental Monitoring Applications[J]. Sensors (Switzerland), 2014, 14(11): 19979-20007.
- [5] 周亦斌,王俊. 电子鼻在食品感官检测中的应用进展[J]. 食品与发酵工业, 2004, 30(002): 129-132.
- [6] 朱先约,宗永立,李炎强,等. 电子鼻技术及其在烟草行业中的应用[J]. 中国烟草学报, 2008, 14(4): 66-70.
- [7] Gebicki J, Szulczynski B, Kaminski M. Determination of Authenticity of Brand Perfume Using Electronic Nose Prototypes[J]. Measurement Science and Technology, 2015, 26(12): 125103.
- [8] Cebi N, Yilmaz M T, Sagdic O. A Rapid ATR-FTIR Spectroscopic Method for Detection of Sibutramine Adulteration in Tea and Coffee Based on Hierarchical Cluster and Principal Component Analyses [J]. Food Chemistry, 2017, 229: 517-526.
- [9] Pornpanomchai C, Suthamsmai N. Beer Classification by Electronic Nose[C]//Proceedings of the 2008 International Conference on Wavelet Analysis and Pattern Recognition (ICWAPR), Hong Kong, China; IEEE, 2008: 333-338.
- [10] Qi P F, Meng Q H, Zhou Y, et al. A Portable E-Nose System for Classification of Chinese Liquor[C]//2015 IEEE Sensors, Busan, South Korea; IEEE, 2015: 1-4.
- [11] Duuta R, Hines E L, Gardner J W, et al. Bacteria Classification Using Cyranose 320 Electronic Nose[J]. Biomedical Engineering Online, 2002, 1(1): 4.
- [12] Li H, Chen Q, Zhao J, et al. Non-Destructive Evaluation of Pork Freshness Using a Portable Electronic Nose (E-Nose) Based on a Colorimetric Sensor Array[J]. Analytical Methods, 2014, 6(16): 6271-6277.
- [13] Li Z H, Meng Q H, Qi P, et al. Boosting-Based One-Class SVM for Recognizing True-Fake Chinese Liquors Using Electronic Noses [C]//Proceedings of the World Congress on Intelligent Control and Automation (WCICA), Guilin, China; IEEE, 2016: 1408-1413.
- [14] 闫明,段发阶. 曲线的特征识别方法研究[J]. 传感技术学报, 2006, 19(03): 724-727.
- [15] 崔瑶. 基于电子鼻的特征提取及模式分类方法研究[D]. 北京:北京化工大学, 2011.
- [16] 亓培锋,孟庆浩,井雅琪,等. 用于白酒识别的电子鼻数据分析与参数优化[J]. 天津大学学报(自然科学与工程技术版), 2015, 48(7): 643-651.
- [17] 邹小波,赵杰文,吴守一. 气体传感器阵列中特征参数的提取与优化[J]. 传感技术学报, 2002, 15(4): 282-286.
- [18] Lozano J, García A, García C J, et al. Wine Classification with Gas Sensors Combined with Independent Component Analysis and Neural Networks[C]//Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Salamanca, Spain; Springer, 2009: 1280-1287.
- [19] Zeng H, Li Q, Gu Y. New Pattern Recognition System in the E-Nose for Chinese Spirit Identification [J]. Chinese Physics B, 2016, 25(2): 024201.
- [20] 李金金,孙哲华,孟庆浩. 基于手持式电子鼻的白酒识别[J]. 食品与发酵工业, 2019, 45(24): 218-222.
- [21] 殷勇,申晓鹏,于慧春. 基于 KECA+FDA 的白酒电子鼻多特征鉴别方法[J]. 农业机械学报, 2018, 49(4): 374-380.
- [22] Djemal R, Bazyed A G, Belwafi K, et al. Three-Class EEG Based Motor Imagery Classification Using Phase-Space Reconstruction Technique[J]. Brain Sciences, 2016, 6(3): 36.



王 茜(1995—),女,现为天津大学控制工程专业在读硕士研究生,主要研究方向为模式识别与机器学习, 15732115228@163.com;



孟庆浩(1968—),男,工学博士,天津大学电气自动化与信息工程学院教授,博士生导师,主要从事机器人感知与导航、仿生嗅觉和情感计算等方面的研究, qh_meng@tju.edu.cn;



靳荔成(1989—),男,工学硕士,工程师,主要研究方向为机器人与自主系统、电子鼻技术等, jinlc@tju.edu.cn。