

基于 GC-IMS 的白酒特征分析及鉴别

鲁祥凯¹, 杨彪^{1*}, 孙莹², 贺金娜³, 樊保民¹, 孙辉¹, 孙啸涛⁴

¹北京工商大学化学与材料工程学院 北京 100048

²江苏今世缘酒业股份有限公司 江苏淮安 223411

³山东海能科学仪器有限公司 山东德州 253000

⁴北京工商大学轻工科学技术学院 北京 100048)

摘要 探讨一种在未对挥发组分定性、定量的条件下,实现白酒准确鉴别的简便方法。选用两种香型共 10 种白酒,每种白酒取 2 个勾调批次的样品。直接原样顶空制样,通过气相色谱-离子迁移谱法(GC-IMS)分析,获得可视化的指纹图谱,并进一步提取白酒主要特征建立 Fisher 线性判别模型。从白酒样品中共检出 134 种挥发组分。指纹图谱可直观反映白酒样品中微量挥发组分的含量差异和勾调的批次稳定性。在未完全剖析和定量所有组分前提下,采用逐步回归分析分别对基于全部 134 种挥发组分、相关性分析选择的 39 种特征组分以及主成分分析提取的 6 个主成分建立的 3 个 Fisher 线性判别模型进行优化。优化后的相关性分析模型(优化为 21 种组分特征)和主成分分析模型对训练集的分类正确率和留一法交叉验证分类正确率均为 100%,对测试集样品的判别正确率均为 95%。由全部特征经逐步回归分析优化的模型,共提取 24 个组分特征,建立的 Fisher 线性判别模型效果最优。其对训练集分类正确率为 100%,留一法交叉验证分类正确率为 100%,对测试集酒样预测正确率达 100%。GC-IMS 方法可为白酒生产过程中的质量控制、市场监管及白酒原产地保护提供技术支持,结合模式识别方法可对白酒进行准确鉴别。

关键词 白酒; 气相色谱-离子迁移谱; 指纹图谱; 特征提取; 线性判别

文章编号 1009-7848(2023)01-0278-18 **DOI:** 10.16429/j.1009-7848.2023.01.027

白酒中丰富的微量组分及独特的风味,来源于酿酒原料的差异以及发酵过程中微生物群落的多样性和特异性^[1]。白酒的人工品评鉴别是基于对组分的非定性、非量化的感官判断,判别标准因人而异^[2-3]。白酒的仪器鉴别主要通过来自原料和发酵产物的差异性特征组分信息来进行判别。自 1964 年初次采用层析法定性白酒中的一些微量成分以来,已发展出光谱、色谱结合质谱以及传感器等仪器分析方法,结合模式识别可对白酒的香型、等级、年份以及产地进行鉴别^[4-10]。

根据微量无机元素或同位素含量等反映酿酒原料特征的信息可较为准确地鉴别白酒^[11-14]。姜涛等^[15]检测黔、川两地白酒中 15 个特征元素,根据其中 6 种元素(Mn、Ga、Sc、V、Na 和 Cs)含量建立了两种判别模型,对两地白酒的产地鉴别正确率分别为 94.7%和 100%。李清亮等^[11]依据 12 种无

机元素聚类分析结果,成功将 12 个品牌共 175 份白酒样品的分为 4 类。然而,生产批次、储存时间以及粮源变化会导致信息的不稳定,因此该方法仅作参考。

通过对白酒中某些组分的定性、定量分析或者采集特定范围内的光谱也可鉴别白酒^[16-21]。霍丹群等^[22]采用气相色谱分析技术,定性、定量分析 7 种不同产地白酒中的 10 种组分,通过主成分分析建立的判别模型对未知酒样鉴别正确率为 93.9%。徐睿等^[23]通过多种处理方法对 8 种白酒(金沙蓝钻酒和其它白酒)的近红外光谱进行处理,最终以 4 119.20~9 881.46 cm⁻¹ 区间的光谱特征建立金沙蓝钻酒相似度匹配模型,鉴别正确率达 100%。白酒中某些物质与传感器的特异性响应也可用于白酒的鉴别^[24-28]。Li 等^[29]使用电子鼻结合多维尺度支持向量机算法实现对 10 个品牌白酒 98.3%的正确鉴别。田婷等^[30]采用电子鼻结合主成分分析实现对酱香型白酒分类识别。然而,传感器主要是针对白酒中特定物质种类进行响应,无法做到全方位分析,有时还受到酒精的影响^[10]。上述常用的白酒仪器鉴别方法沿袭了白酒风味分析的

收稿日期: 2022-01-17

基金项目: “十三五”国家重点研发计划项目重点专项
(2018YFC1604101)

第一作者: 鲁祥凯,男,硕士生

通信作者: 杨彪 E-mail: ybiao@btbu.edu.cn

原理,需对白酒中确定的风味组分进行定性、定量或半定量检测分析,进而提取特征建立鉴别模型,存在局限性。目前,尚未建立类似人工品评的仪器方法,可以在未定性具体组分情况下,仅依据组分宏观差异来鉴别白酒。

近年来,气相色谱-离子迁移谱法(GC-IMS)凭借高分离效率、高灵敏度以及响应速度快的优势逐渐成为检测分析挥发性有机物的有效方法^[31-33]。GC-IMS 检测获取的风味特征指纹图谱,不仅能可视化分析微量组分差异,还能提取主要特征并建立鉴别模型。Chen 等^[34]依据指纹图谱区分 3 个产地共 122 个黄酒样品,建立二次判别模型,对训练集产地分类正确率为 97.47%, 预测集正确率为 95.35%。Zhou 等^[35]从不同年份的白云边陈酿酒的特征风味指纹图谱中选取 60 个特征,结合主成分分析可将 3 年和 5 年陈酿完全区分,而 15 年和 20 年陈酿因过于相似而无法完全区分。

本文将 2 种香型(含 2 种酒精度)的 10 种白酒直接顶空制样进行 GC-IMS 分析,获得指纹图谱,在未定性、定量前提下,通过指纹图谱中各微量组分相对含量提取白酒特征,建立 Fisher 线性判别模型,以实现针对不同香型、酒精度以及不同档次白酒的准确鉴别。

1 材料与方法

1.1 材料与主要仪器

白酒样品取自江苏某酒企。选取 42 度和 52 度共 10 种不同档次和香型的白酒,为确保样品的代表性,每种白酒分别取 2 个不同勾调批次的样品。白酒样品采用 A-B-C 代码表示,其中 A 代表

酒精度,B 代表酒样序号,C 代表勾调批次。如:42-1-2 表示 42 度白酒 1 号第 2 个勾调批次样品。42 度白酒的 1~5 号样品为浓香型,6 号为清雅酱香型;52 度白酒中 1~3 号样品为浓香型,4 号为清雅酱香型。

气相色谱-离子迁移谱仪(FlavourSpec 1 H1-100053),德国 GAS 公司,配套美国 Restek 公司 Rtx-Wax 色谱柱(30 m×0.53 mm,1 μm)以及瑞士 CTC 公司自动顶空进样装置(Combi PAL)。

1.2 测试方法

样品于室温下密封避光储存。测试时采用顶空制样并对制样条件进行优化。优化后的顶空制样条件为:取原始白酒样品 0.5 mL 置于 20 mL 顶空瓶中密封,70 °C 加热振荡孵化 15 min,由顶空瓶上部空间取样 200 μL 用于检测,顶空进样针温度为 85 °C。

设定 GC-IMS 测试条件:气相色谱柱温度 60 °C,漂移管温度 45 °C。气相色谱载气流速:0~2 min:2 mL/min;2~10 min:2~10 mL/min;10~30 min:10~90 mL/min;30~40 min:90 mL/min。离子迁移谱漂移气恒定流速 150 mL/min。载气和漂移气均为氮气(纯度 99.999%)。每个白酒样品重复测 3 次。

2 结果与分析

2.1 白酒挥发组分分析

从白酒样品中共检测到水和乙醇之外的 134 种挥发组分(含二聚体),见表 1。由已有数据库可定性其中 65 种物质,其余 69 种为未定物。

表 1 白酒中检测到的挥发组分
Table 1 Volatile substances in Baijiu

序号	RI	Rt/s	Dt/ms	英文名称	中文名称	CAS
1	730.6	208.84	1.06	unknow 1	未定物 1	N.A.
2	739.4	212.52	1.13	unknow 2	未定物 2	N.A.
3	740.1	212.83	1.11	unknow 3	未定物 3	N.A.
4	747.8	216.09	1.18	unknow 4	未定物 4	N.A.
5	751.4	217.67	1.23	unknow 5	未定物 5	N.A.
6	752.4	218.08	1.16	unknow 6	未定物 6	N.A.
7	752.4	218.08	1.06	unknow 7	未定物 7	N.A.
8	757.9	220.48	1.28	2,3-dimethyl-1,3-butadiene	2,3-二甲基-1,3-丁二烯	513-81-5
9	760.3	221.53	1.33	Z-2-heptene	顺-2-庚烯	6443-92-1

(续表 1)

序号	RI	Rt/s	Dt/ms	英文名称	中文名称	CAS
10	761.5	222.06	1.11	propyl ether	正丙醚	111-43-3
11	779.8	230.27	1.14	unknow 8	未定物 8	N.A.
12	793.4	236.56	1.2	unknow 9	未定物 9	N.A.
13	827	252.84	1.17	unknow 10	未定物 10	N.A.
14	827.4	253.05	1.14	unknow 11	未定物 11	N.A.
15	835.5	257.15	1.23	unknow 12	未定物 12	N.A.
16	837.7	258.3	1.28	unknow 13	未定物 13	N.A.
17	838.5	258.72	1.2	unknow 14	未定物 14	N.A.
18	844.1	261.56	1.11	propan-2-one	丙酮	67-64-1
19	852.1	265.78	1.15	1-methylethyl acetate	乙酸异丙酯	108-21-4
20	853.3	266.38	1.19	unknow 15	未定物 15	N.A.
21	872.9	276.99	1.26	unknow 16	未定物 16	N.A.
22	874.3	277.73	1.28	unknow 17	未定物 17	N.A.
23	898.9	291.63	1.34	ethyl acetate	乙酸乙酯	141-78-6
24	917.1	302.32	1.24	butan-2-one	2-丁酮	78-93-3
25	918.6	303.24	1.29	butanal	丁醛	123-72-8
26	930.1	310.22	1.37	2,5-dimethylfuran	2,5-二甲基呋喃	625-86-5
27	974	338.47	1.45	ethyl propanoate	丙酸乙酯	105-37-3
28	979.1	341.88	1.51	unknow 18	未定物 18	N.A.
29	985.5	346.29	1.57	ethyl isobutyrate	异丁酸乙酯	97-62-1
30	994.3	352.38	1.52	unknow 19	未定物 19	N.A.
31	998.4	355.92	1.42	valeraldehyde	正戊醛	110-62-3
32	999.8	357.22	1.48	propyl acetate	乙酸丙酯	109-60-4
33	999.9	357.31	1.37	2-pentanone	2-戊酮	107-87-9
34	1 006.2	363.51	1.59	unknow 20	未定物 20	N.A.
35	1 024.9	382.63	1.54	methyl 2-methylbutyrate	2-甲基丁酸甲酯	868-57-5
36	1 028.7	386.61	1.57	unknow 21	未定物 21	N.A.
37	1 029.9	387.89	1.62	isobutyl acetate	乙酸异丁酯	110-19-0
38	1 039.2	397.88	1.33	butan-2-ol	2-丁醇	78-92-2
39	1 052.4	412.56	1.56	ethyl butyrate	丁酸乙酯	105-54-4
40	1 061.3	422.73	1.4	unknow 22	未定物 22	N.A.
41	1 062.6	424.2	1.45	unknow 23	未定物 23	N.A.
42	1 066	428.19	1.6	unknow 24	未定物 24	N.A.
43	1 066.5	428.78	1.65	ethyl 2-methylbutanoate	2-甲基丁酸乙酯	7452-79-1
44	1 081.1	446.18	1.65	ethyl 3-methylbutanoate	异戊酸乙酯	108-64-5
45	1 090	457.17	1.62	unknow 25	未定物 25	N.A.
46	1 098.9	468.89	1.5	hexan-2-one	2-己酮	591-78-6
47	1 099.2	469.35	1.58	unknow 26	未定物 26	N.A.
48	1 099.7	470.19	1.53	unknow 27	未定物 27	N.A.
49	1 107.2	481.8	1.38	n-butanol	正丁醇	71-36-3
50	1 113.2	491.19	1.47	unknow 28	未定物 28	N.A.
51	1 113.2	491.19	1.35	unknow 29	未定物 29	N.A.
52	1 129.7	518.28	1.29	unknow 30	未定物 30	N.A.
53	1 130.6	519.75	1.45	unknow 31	未定物 31	N.A.
54	1 130.8	520.17	1.5	unknow 32	未定物 32	N.A.
55	1 132.2	522.46	1.56	hexanal	正己醛	66-25-1
56	1 132.3	522.69	1.59	unknow 33	未定物 33	N.A.
57	1 135.9	528.73	1.68	ethyl pentanoate	戊酸乙酯	539-82-2

(续表 1)

序号	RI	Rt/s	Dt/ms	英文名称	中文名称	CAS
58	1 136.8	530.25	1.71	unknow 34	未定物 34	N.A.
59	1 137.1	530.79	1.75	isoamyl acetate	乙酸异戊酯	129-92-2
60	1 143.5	542.01	1.52	unknow 35	未定物 35	N.A.
61	1 149.3	552.3	1.68	unknow 36	未定物 36	N.A.
62	1 156.7	565.75	1.52	3-methylbutanol	异戊醇	123-51-3
63	1 158.3	568.7	1.39	1-penten-3-ol	1-戊烯-3-醇	616-25-1
64	1 159.8	571.41	1.74	unknow 37	未定物 37	N.A.
65	1 170.7	591.99	1.74	unknow 38	未定物 38	N.A.
66	1 171	592.62	1.81	unknow 39	未定物 39	N.A.
67	1 186.3	208.84	1.06	unknow 41	未定物 41	N.A.
68	1 186.5	622.86	1.71	unknow 40	未定物 40	N.A.
69	1 186.5	623.28	1.55	unknow 42	未定物 42	N.A.
70	1 187.8	623.28	1.42	pentyl acetate	乙酸戊酯	628-63-7
71	1 194.6	625.79	1.76	heptan-2-one	2-庚酮	110-43-0
72	1 195	638.69	1.63	unknow 43	未定物 43	N.A.
73	1 196.7	639.24	1.36	unknow 44	未定物 44	N.A.
74	1 197	642.18	1.49	unknow 45	未定物 45	N.A.
75	1 198.5	642.6	1.65	unknow 47	未定物 47	N.A.
76	1 198.6	645.12	1.33	unknow 46	未定物 46	N.A.
77	1 199.4	645.33	1.21	unknow 48	未定物 48	N.A.
78	1 199.8	646.59	1.68	unknow 49	未定物 49	N.A.
79	1 200.1	647.22	1.52	2-hexanol	2-己醇	626-93-7
80	1 200.8	647.68	1.57	1,8-cineole	桉叶油醇	470-82-6
81	1 204.4	648.92	1.73	unknow 50	未定物 50	N.A.
82	1 212	654.99	1.78	unknow 52	未定物 52	N.A.
83	1 212.1	667.8	1.52	unknow 51	未定物 51	N.A.
84	1 220.7	668.01	1.66	unknow 53	未定物 53	N.A.
85	1 222.4	682.92	1.63	3-methyl-3-buten-1-ol	3-甲基-3-丁烯-1-醇	763-32-6
86	1 230.8	685.88	1.51	unknow 54	未定物 54	N.A.
87	1 232.8	700.77	1.64	butyl butanoate	丁酸丁酯	109-21-7
88	1 235.8	704.32	1.82	5-methyl-3-heptanone	5-甲基-3-庚酮	541-85-5
89	1 253.8	709.9	1.69	ethyl hexanoate	正己酸乙酯	123-66-0
90	1 266.5	743.37	1.8	4-methyl-1-pentanol	4-甲基-1-戊醇	626-89-1
91	1 273.6	767.91	1.64	unknow 55	未定物 55	N.A.
92	1 277.5	782.04	1.75	hexyl acetate	乙酸己酯	142-92-7
93	1 278.2	789.93	1.93	propyl hexanoate	己酸丙酯	626-77-7
94	1 286.5	791.29	1.86	unknow 57	未定物 57	N.A.
95	1 288.3	808.5	1.9	unknow 56	未定物 56	N.A.
96	1 289.8	812.07	1.83	unknow 58	未定物 58	N.A.
97	1 290.3	815.22	1.64	Z-3-hexenyl acetate	乙酸叶醇酯	3681-71-8
98	1 297.6	816.26	1.3	unknow 59	未定物 59	N.A.
99	1 297.9	831.81	1.57	unknow 60	未定物 60	N.A.
100	1 297.9	832.44	1.85	unknow 61	未定物 61	N.A.
101	1 304.1	832.44	1.21	unknow 63	未定物 63	N.A.
102	1 304.8	845.88	1.61	unknow 62	未定物 62	N.A.
103	1 304.8	847.35	1.65	octanal	正辛醛	124-13-0
104	1 305.2	847.45	1.4	pentanol	正戊醇	71-41-0

(续表 1)

序号	RI	Rt/s	Dt/ms	英文名称	中文名称	CAS
105	1 309.8	848.22	1.51	1-octen-3-one	1-辛烯-3-酮	4312-99-6
106	1 329.6	858.3	1.68	ethyl heptanoate	庚酸乙酯	106-30-9
107	1 330.9	903.22	1.91	pentyl butanoate	丁酸戊酯	540-18-1
108	1 335.8	906.28	1.96	unknow 64	未定物 64	N.A.
109	1 348.6	917.7	1.74	heptyl acetate	乙酸庚酯	112-06-1
110	1 357.5	948.64	1.91	2-heptanol	2-庚醇	543-49-7
111	1 361.4	970.52	1.72	ethyl lactate	乳酸乙酯	97-64-3
112	1 362.8	980.28	1.54	unknow 65	未定物 65	N.A.
113	1 363.1	983.85	2.02	unknow 66	未定物 66	N.A.
114	1 369.2	984.69	1.97	2,4,5-trimethylthiazole	2,4,5-三甲基噻唑	13623-11-5
115	1 370.1	1 000.19	1.57	3-octanol	辛醇	589-98-0
116	1 371.8	1 002.58	1.77	1-hexanol	正己醇	111-27-3
117	1 371.9	1 006.93	1.33	2-nonanone	2-壬酮	821-55-6
118	1 427.1	1 007.29	1.4	ethyl octanoate D	辛酸乙酯(D)	106-32-1
119	1 428.9	1 161.24	2.03	ethyl octanoate M	辛酸乙酯(M)	106-32-1
120	1 453.6	1 166.43	1.47	octyl acetate M	乙酸辛酯(M)	112-14-1
121	1 454.9	1 243.17	1.48	octyl acetate D	乙酸辛酯(D)	112-14-1
122	1 473.7	1 247.32	2.02	decanal	癸醛	112-31-2
123	1 474.1	1 309.07	1.53	unknow 67	未定物 67	N.A.
124	1 474.8	1 310.4	2.08	isopentyl hexanoate	己酸异戊酯	2198-61-0
125	1 485.1	1 312.87	2.14	unknow 69	未定物 69	N.A.
126	1 485.3	1 348.2	1.77	unknow 68	未定物 68	N.A.
127	1 485.5	1 348.83	1.71	furfural	糠醛	98-01-1
128	1 492.3	1 349.44	1.33	acetic acid	乙酸	64-19-7
129	1 533	1 373.56	1.15	benzaldehyde	苯甲醛	100-52-7
130	1 533.2	1 525.37	1.47	propanoic acid	丙酸	79-09-4
131	1 584.7	1 526.14	1.27	4-butanolide	4-羟基丁酸内酯	96-48-0
132	1 594.9	1 742.4	1.31	2-acetylpyrazine	2-乙酰基吡嗪	22047-25-2
133	1 633.3	1 788.84	1.21	phenylacetaldehyde	苯乙醛	122-78-1
134	1 666.7	1 975.11	1.25	ethyl decanoate	癸酸乙酯	110-38-3

注:N.A.表示未鉴定该物质,无CAS号。

将所有挥发性组分信号峰提取后形成指纹图谱(图 1)。其中,每个点代表一种物质,颜色越红表示物质浓度越高。横轴为样品中检测到的所有挥发组分,纵轴为同一挥发组分在不同白酒样品中的浓度差异。由图 1 可以发现,虽然酒精度和香型不同,但是指纹图谱中部蓝色区域内的组分为白酒的共有特征组分。清雅酱香型白酒 42-6 和 52-4 的图谱类似,且与其它香型白酒图谱差异明显。指纹图谱左侧黄色区域内的大部分组分,在清雅酱香型白酒中的浓度要明显低于浓香型白酒,而谱图中绿色及右侧黄色区域中 2-甲基丁酸乙酯、乙酸叶醇酯、乙酸异戊酯、2-己醇、糠醛等组分

在清雅酱香型白酒中的浓度则明显高于浓香型白酒。这反映出白酒香型不同导致的组分差异。作为低端产品的 42-2 和 52-1 浓香型白酒,在图谱两侧黄色区域中的组分浓度明显低于其它浓香型白酒。另外,每种白酒平行测试 3 次的结果无明显差异,说明 GC-IMS 法的稳定性很好。除 42-3 和 52-1 两种白酒的批次之间有较明显差异外,其它白酒的批次差异不明显,说明其勾调批次稳定性较好。以上简单分析可知,指纹图谱能够清晰直观地反映白酒样品中微量挥发组分的含量差异以及白酒勾调批次的稳定性。

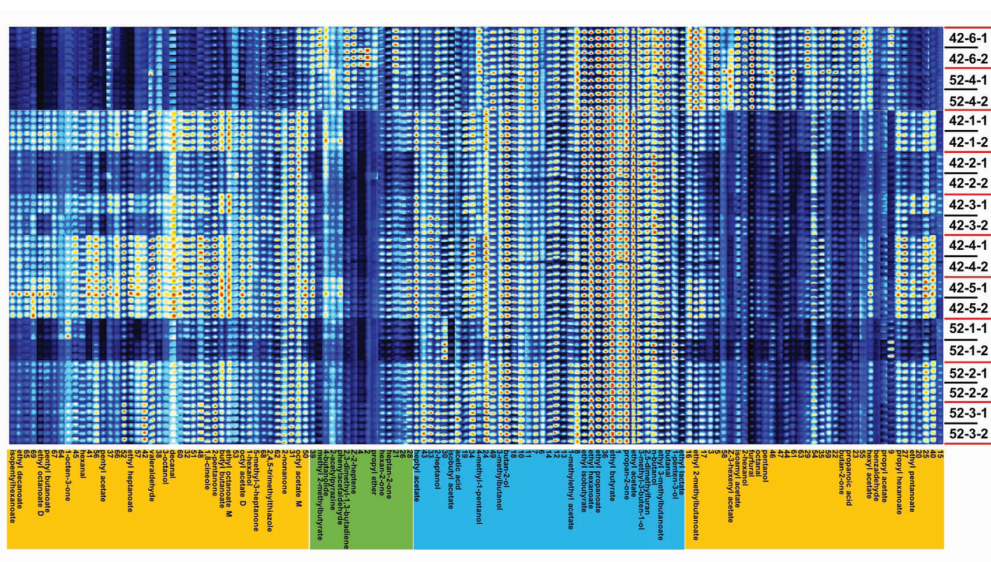


图 1 白酒样品挥发物质指纹图谱

Fig.1 Fingerprints of volatile substances in Baijiu samples

根据 GC-IMS 测试数据计算得到 10 种白酒样品的平均欧氏距离(表 2)。欧氏距离越大,两样品间差异越大,反之则相似度越高。结合指纹图谱分析可知,每个样品平行检测 3 次的结果十分稳定,欧氏距离均小于 0.65(表 2 中粗体数据)。同一样品批次之间的差异(表 2 中灰色阴影数据),除样品 42-3 和样品 52-1 的批次差异较明显(欧式距离分别为 1.680 和 1.554)外,其它样品批次差异均小于 1。因此,欧氏距离也可直观反映白酒生产过程中的勾调稳定性。由表 2 数据可知,样品 42-1、42-3、42-4 和 42-5 较为相似,42-2 与 52-2 较为相似,52-2 与 52-3 较为相似,42-6 与 52-4 较为相似,与可视化的指纹图谱结论一致。综上,采用 GC-IMS 检测白酒中挥发组分,不仅结果稳定性好,还能反映不同批次间的质量稳定性,可直接应用于白酒生产过程中的质量监控。

2.2 白酒鉴别模型

建立白酒的鉴别模型前,首先需要获取白酒的主要特征,常用的方法有特征选择和特征提取两种。两种方法均是对原有特征的降维处理,区别在于是否改变特征的属性。特征选择从原始特征中筛选子集,未改变其属性。主要是从准确性和特征维数考虑,根据相关程度或其它角度从原始数据中筛选,形成具有较高正确率的特征子集^[36-37]。特征提取则是对原始特征进行数学变换,通过线

性组合(如主成分分析)形成相互独立的新特征,有效地解决“维数灾难”^[38]。

本文基于顶空制样所测得的 60 组数据,在兼顾样品品质和香型的前提下随机取样,消除类别属性,从中划分出 20 个独立样本,作为测试集,剩余样本作为训练集。分别依据相关性分析和主成分分析对训练集数据进行主要特征提取,然后建立 Fisher 线性判别模型并进行交叉验证,最后对测试集样品进行判别验证。

2.2.1 特征选择(相关性分析) 显而易见,浓度差异显著的组分对白酒的鉴别意义更大。需要通过相关性分析,对原始数据进行特征筛选,获得白酒的鉴别特征。本文对所有组分特征进行皮尔逊双尾相关性检测,保留与组变量相关性系数 $\alpha \geq 0.4$ 且显著相关($P < 0.05$)的变量,筛除相关性弱以及相对冗余的变量,最终提取出 39 个特征组分变量。相关性分析部分见表 3。由表 3 可知,白酒类别相关性较大的 39 个组分变量在 0.01 水平相关性显著,具有统计学意义。其中有 20 个组分为正相关变量,依据相关性排序为未定物 49(0.752)、未定物 59(0.629)、未定物 3(0.609)、2-乙酰基吡嗪(0.608)、未定物 10(0.553)、未定物 8(0.535)、乙酸叶醇酯(0.533)、未定物 7(0.523)、未定物 9(0.482)、乳酸乙酯(0.469)、丙酸(0.464)、未定物 58(0.464)、未定物 12(0.447)、2-己醇(0.446)、未

表 2 白酒样品间平均欧氏距离

Table 2 The average Euclidean distance between Baijiu samples

	42-1-1	42-1-2	42-2-1	42-2-2	42-3-1	42-3-2	42-4-1	42-4-2	42-5-1	42-5-2	42-6-1	42-6-2	52-1-1	52-1-2	52-2-1	52-2-2	52-3-1	52-3-2	52-4-1	52-4-2
42-1-1	0.153	<u>0.819</u>	2.625	2.547	0.614	1.521	0.819	0.685	1.570	1.111	6.304	6.975	3.501	4.388	1.269	1.732	2.016	2.071	7.225	6.571
42-1-2	0.345	<u>0.345</u>	3.213	3.204	0.880	1.984	1.013	0.898	0.929	0.806	6.674	7.361	4.174	5.061	1.969	2.398	2.745	2.796	7.637	7.000
42-2-1	0.292	<u>0.602</u>	0.292	2.683	1.888	1.888	3.100	2.898	3.895	3.438	4.810	5.370	1.949	2.390	1.726	1.271	1.820	1.726	5.543	4.924
42-2-2	0.226	<u>0.226</u>	0.226	2.626	1.859	2.985	2.985	2.826	3.909	3.388	4.918	5.475	1.701	2.250	1.556	1.127	1.506	1.419	5.634	4.996
42-3-1	0.161	<u>0.161</u>	0.161	0.848	0.786	1.510	1.160	6.370	7.037	3.667	4.503	1.420	1.913	2.239	2.270	7.324	6.669			
42-3-2	0.375	<u>0.375</u>	0.375	1.948	1.715	2.638	2.109	5.571	6.279	3.000	3.906	1.540	1.511	2.001	1.995	6.534	5.892			
42-4-1	0.357	<u>0.558</u>	0.357	1.574	1.154	6.759	7.436	3.923	4.735	1.699	2.237	2.302	2.357	7.708	7.043					
42-4-2	0.410	<u>0.410</u>	0.410	1.527	1.121	6.589	7.274	3.819	4.638	1.595	2.074	2.247	2.297	7.541	6.882					
42-5-1	0.499	<u>0.928</u>	0.499	6.947	7.642	4.816	5.746	2.648	3.067	3.437	3.491	7.948	7.326							
42-5-2	0.226	<u>0.226</u>	0.226	6.471	7.178	4.210	5.211	2.143	2.536	2.852	2.911	7.466	6.823							
42-6-1	0.186	<u>0.903</u>	0.186	4.548	5.056	5.564	5.083	5.523	5.471	1.566	1.171									
42-6-2	0.161	<u>0.161</u>	0.161	5.012	5.384	6.170	5.683	6.095	6.040	1.107	1.054									
52-1-1	0.645	<u>1.554</u>	0.645	2.475	2.022	2.031	1.994	4.991	4.362											
52-1-2	0.149	<u>0.149</u>	0.149	3.285	2.901	2.700	2.632	5.289	4.730											
52-2-1	0.295	<u>0.295</u>	0.295	1.076	1.073	6.376	5.714													
52-2-2	0.183	<u>0.183</u>	0.183	1.067	1.018	5.849	5.195													
52-3-1	0.434	<u>0.434</u>	0.434	6.214	5.550															
52-3-2	0.361	<u>0.361</u>	0.361	6.164	5.500															
52-4-1	0.369	<u>0.369</u>	0.369	0.818																
52-4-2	0.261	<u>0.261</u>	0.261																	

注：表中数据均×10⁴。粗体数据为同一样品平行测试的欧氏距离，下划线数据为同一种酒不同批次样品之间的欧氏距离。

定物 5(0.437)、2,3-二甲基-1,3-丁二烯(0.430)、3-甲基-3-丁烯-1-醇(0.427)、乙酸异戊酯(0.411)、未定物 4(0.409)、正戊醇(0.407)；另外 19 个组分为负相关变量，依据相关性排序为乙酸辛酯(M)(0.754)、未定物 53(0.690)、异丁酸乙酯(0.606)、辛酸乙酯(M)(0.583)、癸酸乙酯(0.583)、未定物 19 (0.569)、2,4,5-三甲基噻唑 (0.553)、未定物 60 (0.545)、50(0.490)、26(0.477)、36(0.472)、31(0.462)、65(0.445)、5-甲基-3-庚酮(0.439)、未定物 69(0.428)、68(0.425)、33(0.420)、丁酸戊酯(0.417)、未定物 64(0.414)。通过相关性分析可知，白酒样品间的差异主要来源于更微量的组分(非主体风味物质)，而不是相对含量更高的物质(主体风味物质)。

表 3 白酒特征相关性分析(部分)

Table 3 Correlation analysis of Baijiu characteristics (part)

	未定物 3	异丁酸乙酯	未定物 49	未定物 53	未定物 59	乙酸辛酯(M)	...	G
未定物 3	1							
异丁酸乙酯	-0.260	1						
未定物 49	0.0649**	-0.277	1					
未定物 53	-0.233	0.620**	-0.409**	1				
未定物 59	0.859**	-0.229	0.755**	-0.434**	1			
乙酸辛酯(M)	-0.586**	0.434**	-0.746**	0.709**	-0.742**	1		
...	
G	0.609**	-0.606**	0.752**	-0.690**	0.629**	-0.754**	...	1

注:**. 在 0.01 级别(双尾),相关性显著;*. 在 0.05 级别(双尾),相关性显著。

2.2.2 特征提取(主成分分析) 主成分分析可将原始特征变换成新的特征向量,且能够保证相互之间正交。为分析各特征对白酒组分的贡献,本文采用最大正交旋转变换,使每个变量中解释的因子量最少。以特征值大于 1 为阈值,共选取 6 个主成分,相关矩阵特征值见表 4。6 个主成分可以反映原特征信息的 92.648%,基本保留了原始特征,具有合理性。

以表 1 中各物质顺序依次分别记作变量 X_1 、 X_2 、 X_3 、...、 X_{134} ,通过主成分分析获得原始变量与各主成分之间的相关程度,见表 5。通过主成分分析获得 6 个相互独立的综合指标模型:

$PC1 = 0.970 X_{11} + 0.948 X_5 + 0.944 X_3 + \dots - 0.376 X_{114}$,可反映原数据标准变异的 53.530%,代表变量为未定物 8、丙酸、正戊醇等;

$PC2 = 0.046 X_{11} + \dots + 0.961 X_{107} + 0.951 X_{112} + \dots + 0.372 X_{114}$,可反映原数据标准变异的 19.987%,代表变量为丁酸戊酯、未定物 65、癸酸乙酯等;

$PC3 = -0.040 X_{11} + \dots + 0.910 X_{88} + 0.846$

表 4 相关矩阵特征值

Table 4 Eigen values of the correlation matrix

成分	特征值	方差 贡献率/%	累积方差 贡献率/%
PC1	20.877	53.530	53.530
PC2	7.795	19.987	73.517
PC3	2.783	7.136	80.653
PC4	2.340	6.000	86.653
PC5	1.325	3.397	90.050
PC6	1.013	2.598	92.648

$X_{108} + \dots + 0.448 X_{114}$,可反映原数据标准变异的 7.136%,代表变量为 5-甲基-3-庚酮、未定物 64、未定物 33 等;

$PC4 = -0.136 X_{11} + \dots + 0.920 X_{29} + 0.654 X_{47} + \dots - 0.376 X_{114}$,可反映原数据标准变异的 6.000%,代表变量为异丁酸乙酯、未定物 26 等;

$PC5 = 0.137 X_{11} - 0.126 X_5 + 0.131 X_3 + \dots + 0.058 X_{114}$,可反映原数据标准变异的 3.397%;

$PC6 = -0.067 X_{11} - 0.141 X_5 - 0.012 X_3 + \dots + 0.518 X_{114}$,可反映原数据标准变异的 2.598%。

表 5 原始变量与各主成分之间相关程度

Table 5 Correlation between original variables and principal components

变量	物质	主成分					
		PC1	PC2	PC3	PC4	PC5	PC6
X_{11}	未定物 8	0.970	0.046	-0.040	-0.136	0.137	-0.067
X_5	未定物 5	0.948	0.005	-0.197	0.052	-0.126	-0.141
X_3	未定物 3	0.944	0.053	-0.086	-0.200	0.131	-0.012
X_7	未定物 7	0.940	0.121	-0.039	-0.077	0.159	-0.138
X_{130}	丙酸	0.912	-0.149	-0.194	0.149	-0.151	0.131
X_{98}	未定物 59	0.869	-0.277	-0.179	-0.141	-0.060	0.161

(续表 5)

变量	物质	主成分					
		PC1	PC2	PC3	PC4	PC5	PC6
X_8	2,3-二甲基-1,3-丁二烯	0.854	-0.249	-0.277	0.083	-0.274	-0.189
X_{104}	正戊醇	0.817	-0.365	-0.216	0.198	-0.296	-0.120
X_{132}	2-乙酰基吡嗪	0.816	0.025	-0.212	-0.185	0.122	0.296
X_{79}	2-己醇	0.760	-0.500	-0.176	0.216	-0.291	-0.005
X_4	未定物 4	0.749	-0.317	-0.271	0.086	-0.157	-0.287
X_{59}	乙酸异戊酯	0.728	-0.535	0.000	0.034	-0.267	-0.237
X_{13}	未定物 10	0.714	0.081	0.171	-0.258	0.578	0.029
X_{53}	未定物 31	-0.709	0.367	0.412	-0.005	0.370	0.058
X_{97}	乙酸叶醇酯	0.706	-0.518	-0.253	0.098	-0.132	0.188
X_{96}	未定物 58	0.684	-0.557	-0.186	0.198	-0.062	0.269
X_{78}	未定物 49	0.673	-0.531	0.063	-0.135	0.232	0.329
X_{120}	乙酸辛酯(M)	-0.585	0.484	0.312	0.341	0.079	-0.310
X_{107}	丁酸戊酯	-0.073	0.961	0.065	0.075	0.152	0.006
X_{112}	未定物 65	-0.070	0.951	0.116	0.090	0.019	0.004
X_{99}	未定物 60	-0.162	0.887	0.222	0.282	0.035	0.022
X_{125}	未定物 69	0.031	0.861	0.207	0.279	-0.193	0.171
X_{134}	癸酸乙酯	-0.200	0.852	0.340	0.149	-0.081	0.050
X_{81}	未定物 50	-0.071	0.838	0.287	0.326	0.269	-0.042
X_{61}	未定物 36	-0.163	0.826	0.314	0.255	0.274	0.001
X_{119}	辛酸乙酯(M)	-0.357	0.823	0.234	0.164	0.254	-0.042
X_{30}	未定物 19	-0.191	0.811	0.009	0.374	0.369	-0.034
X_{85}	3-甲基-3-丁烯-1-醇	0.506	-0.756	0.042	0.162	-0.274	0.102
X_{15}	未定物 12	0.062	-0.691	-0.122	0.194	0.200	0.015
X_{126}	未定物 68	0.156	0.660	0.264	0.559	-0.217	0.159
X_{111}	乳酸乙酯	0.503	-0.652	-0.308	0.157	-0.249	0.330
X_{88}	5-甲基-3-庚酮	-0.124	0.303	0.910	0.106	-0.048	-0.074
X_{108}	未定物 64	-0.385	0.182	0.846	0.068	0.045	0.117
X_{56}	未定物 33	-0.406	0.409	0.765	-0.040	0.175	-0.045
X_{12}	未定物 9	0.272	-0.455	-0.531	-0.285	-0.481	-0.106
X_{29}	异丁酸乙酯	-0.079	0.217	0.020	0.920	-0.007	0.067
X_{47}	未定物 26	0.316	0.512	0.033	0.654	-0.183	-0.359
X_{84}	未定物 53	-0.178	0.539	0.328	0.572	0.306	0.032
X_{114}	2,4,5-三甲基噻唑	-0.376	0.372	0.448	0.412	0.058	0.518

注:粗体数据表示该物质对应变量与对应主成分关系紧密。

计算各主成分得分,以 PC1、PC2 和 PC3 为坐标轴,绘制主成分三维得分图(图 2),可以对白酒进行分类。42-6 和 52-4 为清雅酱香型白酒,聚类区域与其它浓香型白酒相距较远,且这两种酒之间区分较为明显。浓香型白酒相距较近,除了 42-3 和 52-1 批次差异较大,导致同类酒样本分散外,其它酒聚类范围均较小,然而,也发现 42-1 和 42-5 相距十分近。由此可知,基于主成分分析的

分类方法能够反映白酒风味的相似性与典型性。

2.2.3 Fisher 线性判别模型 Fisher 线性判别是典则判别中常用的一种方法,主要是将高维空间内样本坐标投影至低维空间内进行分类,投影原则是同类差异最小化,异类差异最大化,分类完成后获得每类的组质心坐标,对未知样本进行鉴别时同样采用就近原则进行判别^[9]。本文分别将全部 134 个组分特征数据,根据相关性筛选出的 39

个组分特征数据以及特征提取的 6 个主成分作为 Fisher 线性判别分析的数据集, 分别建立白酒的 Fisher 线性判别模型。这 3 个模型对训练样本的分类正确率均为 100%, 而交叉验证分类正确率依次为 17.5%, 67.5% 和 90%。说明对原始数据进行特征提取, 可有效提高模型鉴别正确率, 然而同时也发现模型缺乏优化, 效果较差。本文结合逐步回归分析对判别模型引入的特征进一步优化。其基本思路是将变量依次选入, 当其对因变量影响显著时保留, 不显著则移除, 继续引入新的变量, 若被移除的变量在新变量引入后变得重要时, 则重新被引入^[40]。通过不断拟合, 直至达到鉴别效果最优的模型。本文中条件设置为: F 概率 P 值选入和移除阈值分别为 0.05 和 0.1。

将全部 134 个特征数据采用逐步回归分析拟合 26 次, 获得最优模型, 记作 M1。M1 采用 24 个主要组分特征, 分别为未定物 8、9、17、20、22、29、31、47、51、53、56、59、65, 乙酸乙酯、2,5-二甲基咪唑、丙酸乙酯、2-甲基丁酸甲酯、2-己酮、异戊醇、乙酸戊酯、正辛醛、乳酸乙酯、正己醇和 2-壬酮。Wilk's Lambda 检验结果 $P=0.000$, 说明该模型具有统计学意义。取 9 个判别函数, 其特征值分别为 207474.591, 7530.331, 1512.763, 374.611, 337.366, 49.578, 45.563, 34.382 和 7.277, 分别能够解释模型方差变化的 95.449%, 3.464%, 0.696%, 0.172%, 0.155%, 0.023%, 0.021%, 0.016% 和 0.003%。前 2 个 Fisher 线性判别函数总共可解释 98.913% 的方差变化, 包含了主要的信息, 可描述各特征的差异与联系。以前 2 个 Fisher 线性判别函数作散点图 (图 3)。图 3 中 10 种白酒彼此分隔, 同种白酒聚敛程度十分好, 基本以组质心为中心聚拢, 且测试

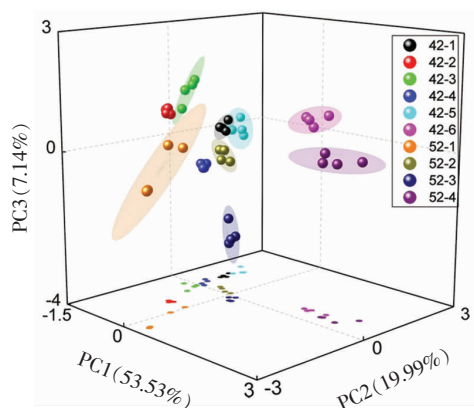


图 2 白酒样品主成分三维得分图

Fig.2 Three-dimensional score map of principal components in Baijiu samples

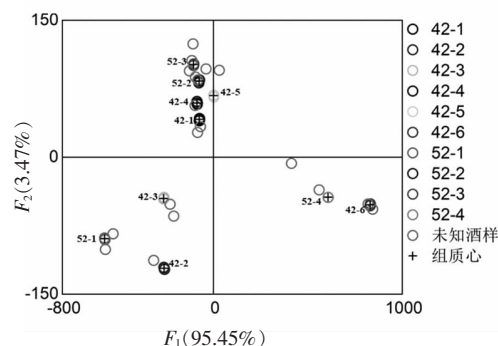


图 3 模型 M1 的 Fisher 线性判别函数散点图

Fig.3 The scatter plot of Fisher's linear discriminant function for model M1

集的白酒样品基本分布于各组质心附近。该模型对训练集鉴别正确率达 100%, 采用留一法交叉验证正确率也为 100%。所建立模型的 Fisher 线性判别函数的系数见表 6, 各组质心见表 7。

表 6 模型 M1 的 Fisher 线性判别函数系数

Table 6 Fisher's linear discriminant function coefficient for model M1

变量	物质	F_1	F_2	F_3	F_4	F_5	F_6	F_7	F_8	F_9
X_{11}	未定物 8	0.285	-0.106	0.015	-0.034	0.056	0.028	0.030	-0.064	0.000
X_{12}	未定物 9	0.368	0.285	-0.033	0.242	0.037	0.078	0.092	0.021	0.002
X_{22}	未定物 17	-0.007	0.064	0.025	0.031	-0.012	-0.006	0.005	0.013	0.004
X_{23}	乙酸乙酯	-0.003	0.000	0.005	0.010	0.004	0.000	0.002	-0.002	-0.001
X_{26}	2,5-二甲基咪唑	0.207	-0.070	0.057	-0.004	0.036	0.000	0.033	0.018	0.014
X_{27}	丙酸乙酯	0.041	0.045	-0.017	-0.002	-0.010	-0.001	-0.005	0.005	-0.004
X_{34}	未定物 20	0.262	0.148	0.096	0.047	-0.025	-0.036	-0.057	0.004	-0.017

(续表 6)

变量	物质	F_1	F_2	F_3	F_4	F_5	F_6	F_7	F_8	F_9
X_{35}	2-甲基丁酸甲酯	0.020	-0.162	0.027	-0.040	0.070	-0.005	-0.039	0.000	0.021
X_{40}	未定物 22	-0.419	-0.013	-0.105	-0.204	-0.170	-0.095	-0.059	0.014	-0.012
X_{46}	2-己酮	-0.198	0.167	-0.132	-0.086	0.081	-0.027	0.012	0.012	-0.034
X_{51}	未定物 29	0.628	0.118	-0.005	0.125	0.033	0.049	0.042	-0.011	-0.003
X_{53}	未定物 31	0.007	-0.034	-0.028	0.009	0.018	0.013	0.009	-0.002	0.011
X_{62}	异戊醇	-0.125	0.073	-0.014	0.025	-0.021	-0.019	-0.022	0.032	-0.009
X_{70}	乙酸戊酯	-0.238	-0.136	-0.038	-0.002	0.107	-0.007	0.066	0.029	0.021
X_{75}	未定物 47	0.999	-0.440	-0.036	-0.019	0.168	0.112	0.098	-0.007	0.008
X_{83}	未定物 51	-0.095	0.071	-0.001	-0.021	-0.079	0.002	-0.027	0.015	0.002
X_{84}	未定物 53	0.330	-0.088	0.054	-0.022	0.023	0.040	0.034	-0.028	0.000
X_{95}	未定物 56	0.220	0.058	-0.014	0.000	0.014	0.016	-0.006	-0.005	0.003
X_{98}	未定物 59	0.058	-0.054	0.024	0.007	-0.019	-0.005	0.011	-0.025	-0.002
X_{103}	正辛醛	0.016	0.069	0.016	0.019	-0.034	-0.015	-0.015	0.001	-0.007
X_{111}	乳酸乙酯	-0.004	0.007	-0.002	0.000	0.000	0.000	-0.001	0.001	0.000
X_{112}	未定物 65	0.047	-0.086	0.083	-0.029	-0.078	0.005	-0.001	-0.051	0.004
X_{116}	正己醇	0.011	-0.042	0.090	-0.015	0.001	-0.008	-0.002	-0.002	-0.006
X_{117}	2-壬酮	0.035	0.015	-0.002	0.008	0.008	0.007	0.004	0.004	0.001
C	常量	-1 651.784	-261.285	-257.640	-310.737	-178.647	-84.080	-123.979	4.373	24.801

表 7 模型 M1 中各组质心坐标

Table 7 The centroid coordinates of each group in model M1

组类别	F_1	F_2	F_3	F_4	F_5	F_6	F_7	F_8	F_9
1	-73.975	41.398	44.938	4.867	-24.440	1.514	0.794	-7.886	-3.258
2	-260.726	-121.698	15.772	-7.062	14.792	11.423	2.418	-3.693	1.602
3	-263.233	-45.270	29.684	-14.106	-11.610	-1.340	-10.089	8.983	-0.318
4	-86.215	59.228	5.198	-21.790	31.456	-5.300	2.630	-0.150	-3.089
5	1.386	67.304	45.690	5.488	0.493	-5.742	5.498	0.999	4.923
6	830.274	-52.297	3.801	19.135	12.819	-2.831	-7.512	-2.552	-0.073
7	-575.244	-89.370	-42.462	18.143	-4.100	-9.368	2.328	-1.514	-0.365
8	-75.315	83.284	-16.494	27.997	4.785	8.503	2.552	7.391	-1.405
9	-104.440	101.372	-54.967	-13.239	-5.283	2.834	-7.279	-5.011	2.304
10	607.488	-43.951	-31.160	-19.434	-18.912	0.307	8.659	3.433	-0.320

根据表 6 中系数建立模型判别函数:

$$\begin{aligned}
 F_1 &= 0.285 X_{11} + 0.368 X_{12} - 0.007 X_{22} + \dots + \\
 &0.035 X_{117} - 1651.784 \\
 F_2 &= -0.106 X_{11} + 0.285 X_{12} + 0.064 X_{22} + \dots + \\
 &0.015 X_{117} - 261.285 \\
 F_3 &= 0.015 X_{11} - 0.033 X_{12} + 0.025 X_{22} + \dots - \\
 &0.002 X_{117} - 257.640 \\
 F_4 &= -0.034 X_{11} + 0.242 X_{12} + 0.031 X_{22} + \dots + \\
 &0.008 X_{117} - 310.737
 \end{aligned}$$

$$\begin{aligned}
 F_5 &= 0.056 X_{11} + 0.037 X_{12} - 0.012 X_{22} + \dots + \\
 &0.008 X_{117} - 178.647
 \end{aligned}$$

$$\begin{aligned}
 F_6 &= 0.028 X_{11} + 0.078 X_{12} - 0.006 X_{22} + \dots + \\
 &0.007 X_{117} - 84.080
 \end{aligned}$$

$$\begin{aligned}
 F_7 &= 0.030 X_{11} + 0.092 X_{12} + 0.005 X_{22} + \dots + \\
 &0.004 X_{117} - 123.979
 \end{aligned}$$

$$\begin{aligned}
 F_8 &= -0.064 X_{11} + 0.021 X_{12} + 0.013 X_{22} + \dots + \\
 &0.004 X_{117} + 4.373
 \end{aligned}$$

$$\begin{aligned}
 F_9 &= 0.002 X_{12} + 0.004 X_{22} - 0.001 X_{23} + \dots +
 \end{aligned}$$

0.001 $X_{117} + 24.801$

根据 Fisher 线性判别函数 $F_1 \sim F_9$, 将未知酒样 (测试集) 的 24 个特征数据代入函数中, 计算函数值, 即为样本坐标, 然后分别计算与 10 个组质心

的距离, 将其鉴别为最近距离的组 (G1~G10)。对测试集酒样鉴别正确率达 100%, 鉴别结果见表 8。

表 8 测试集酒样与模型 M1 中各组质心距离以及鉴别结果

Table 8 The distance between samples in test set and centroids of each group and identification results in model M1

	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	预测	实际
unknow1	12.47	252.07	213.06	80.29	82.52	905.25	529.79	91.97	132.64	686.27	1	1
unknow2	20.37	238.41	197.60	85.89	99.06	919.96	515.17	100.94	137.07	700.70	1	1
unknow3	292.49	57.80	93.26	288.89	367.84	1148.74	266.98	313.98	309.32	928.26	2	2
unknow4	274.78	261.36	304.29	291.13	293.39	880.09	578.00	322.06	352.89	670.24	2	2
unknow5	175.47	84.69	58.75	183.51	251.03	1042.14	377.31	213.74	219.14	821.44	3	3
unknow6	184.99	84.00	35.47	187.34	261.90	1060.88	357.77	215.92	215.41	839.53	3	3
unknow7	72.41	274.55	224.92	39.61	99.40	924.14	526.69	61.10	75.21	704.44	4	4
unknow8	84.84	245.16	203.34	14.40	118.33	935.23	505.60	77.20	86.09	714.51	4	4
unknow9	123.24	366.05	328.24	136.28	41.89	814.27	640.52	128.03	173.73	599.75	5	5
unknow10	71.36	314.04	267.39	87.06	53.63	883.25	572.51	69.12	116.38	666.48	5	5
unknow11	898.69	1082.06	1082.61	912.69	826.80	16.64	1395.42	904.68	938.25	219.42	6	6
unknow12	923.63	1106.02	1107.14	937.54	851.79	16.62	1419.71	929.66	963.25	243.92	6	6
unknow13	478.53	277.56	276.95	471.39	557.16	1362.25	53.78	486.12	468.83	1140.57	7	7
unknow14	526.33	319.71	324.29	516.56	604.57	1404.34	13.03	531.40	510.86	1181.87	7	7
unknow15	92.50	271.78	224.35	68.82	122.83	936.96	512.18	22.39	54.77	717.12	8	8
unknow16	108.89	259.85	206.78	83.60	150.77	969.38	485.03	54.05	58.13	749.55	8	8
unknow17	138.84	297.66	245.06	92.78	160.09	957.46	515.16	80.17	28.89	736.22	9	9
unknow18	130.28	280.68	229.58	89.64	160.52	960.23	501.82	73.69	12.42	738.12	9	9
unknow19	498.25	688.76	682.10	511.05	429.88	426.98	993.34	501.59	530.96	199.47	10	10
unknow20	643.29	827.17	825.18	655.94	573.81	279.39	1136.99	648.74	678.91	49.44	10	10
正确率 / %											100	

注: 粗体数据为未知酒样对应不同组别的质心距离的最小值, 用来判定未知酒样的组别归属。

将经相关性分析筛选出的 39 个特征采用逐步回归分析拟合 21 次, 获得最优模型 M2。M2 中采用 21 个主要特征, 分别为未定物 3、5、8、9、10、12、19、33、36、50、53、58、65、69、2,3-二甲基-1,3-丁二烯、乙酸异戊酯、2-己醇、5-甲基-3-庚酮、2,4,5-三甲基噻唑、丙酸和癸酸乙酯。Wilk's Lambda 检验结果 $P=0.000$, 说明该模型具有统计学意义。取 9 个判别函数, 其特征值分别为 8 927.756, 1 063.998, 269.675, 75.069, 54.977, 19.170, 15.420, 3.956 和 2.795, 分别能够解释模型方差变化的 85.574%, 10.199%, 2.585%, 0.720%, 0.527%, 0.184%, 0.148%, 0.038% 和 0.027%。前 2 个 Fisher 线性判别函数共可解释 95.772% 的方差

变化, 包含了主要的信息, 并可描述各特征的差异与联系。以前两个 Fisher 线性判别函数作散点图 (图 4)。图 4 中, 10 种白酒彼此分隔, 同种白酒聚敛十分好, 基本以组质心为中心聚拢。与图 3 不同的是, 图 4 中测试集中未知酒样的分布相对混乱, 可以预见其对未知酒样鉴别正确率会降低。该模型对训练集鉴别正确率达 100%, 采用留一法交叉验证正确率为 100%。所建立的模型 Fisher 线性判别函数系数见表 9, 各组质心见表 10。根据 9 个判别函数计算出的训练集未知酒样坐标并鉴别, 其结果见表 11。采用 21 个主要特征建立的判别模型对未知酒样鉴别正确率为 95%, 只有 52-4 被错误预测为 42-6。

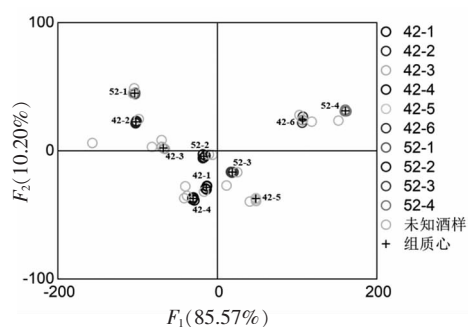


图4 模型 M2 的 Fisher 线性判别函数散点图

Fig.4 The scatter plot of Fisher's linear discriminant function for model M2

表9 模型 M2 的 Fisher 线性判别函数系数

Table 9 Fisher's linear discriminant function coefficient for model M2

变量	物质	F_1	F_2	F_3	F_4	F_5	F_6	F_7	F_8	F_9
X_3	未定物 3	0.369	0.011	-0.015	0.047	-0.042	0.052	0.005	0.034	-0.004
X_5	未定物 5	-0.055	-0.038	0.008	-0.011	0.010	-0.024	-0.011	-0.016	0.007
X_8	2,3-二甲基-1,3-丁二烯	0.553	0.040	-0.127	-0.095	-0.047	0.101	0.083	0.063	-0.021
X_{11}	未定物 8	0.080	-0.061	-0.048	-0.004	0.003	-0.005	0.004	-0.008	-0.015
X_{12}	未定物 9	0.176	0.086	-0.015	0.016	0.019	-0.023	0.052	0.013	0.005
X_{13}	未定物 10	-0.078	0.033	0.016	0.008	-0.001	-0.022	0.002	-0.013	0.010
X_{15}	未定物 12	0.051	0.039	0.026	0.043	0.016	0.042	0.064	0.007	0.024
X_{30}	未定物 19	0.045	0.018	0.029	0.023	0.016	0.013	0.021	0.023	0.005
X_{56}	未定物 33	0.059	-0.020	-0.019	0.024	-0.014	0.011	0.008	0.001	0.014
X_{59}	乙酸异戊酯	-0.006	-0.002	0.006	-0.001	0.004	0.001	0.001	0.002	-0.001
X_{61}	未定物 36	0.008	0.010	-0.009	0.000	0.002	0.000	0.006	0.003	0.000
X_{79}	2-己醇	0.098	0.095	0.058	0.030	0.008	-0.004	-0.034	-0.013	0.019
X_{81}	未定物 50	0.005	-0.012	0.007	-0.001	-0.001	0.001	-0.004	-0.004	-0.001
X_{84}	未定物 53	-0.072	-0.005	0.014	0.006	0.000	-0.020	-0.009	-0.014	-0.010
X_{88}	5-甲基-3-庚酮	-0.074	0.055	0.003	0.010	0.039	0.005	0.007	0.033	-0.012
X_{96}	未定物 58	0.440	0.088	0.036	0.046	-0.097	-0.006	-0.011	0.037	-0.009
X_{112}	未定物 65	0.011	-0.015	-0.032	-0.041	-0.010	-0.019	-0.013	0.024	0.005
X_{114}	2,4,5-三甲基噻唑	-0.002	-0.005	0.001	-0.005	0.000	-0.002	0.001	0.000	0.000
X_{125}	未定物 69	0.013	0.013	0.023	-0.010	-0.015	0.006	0.005	-0.005	-0.005
X_{130}	丙酸	-0.073	0.026	-0.004	0.027	0.012	0.011	0.017	0.009	0.006
X_{134}	癸酸乙酯	0.007	-0.007	0.000	0.008	0.004	0.002	-0.003	-0.004	0.001
C	常量	-331.721	-191.563	113.399	-76.098	-62.262	9.224	-171.638	-72.314	-14.944

表10 模型 M2 中各组质心坐标

Table 10 The centroid coordinates of each group in model M2

组类别	F_1	F_2	F_3	F_4	F_5	F_6	F_7	F_8	F_9
1	-13.751	-28.692	10.168	-1.623	-0.661	-7.925	-0.926	-0.222	-2.471
2	-101.985	22.480	13.473	5.099	1.611	2.467	0.662	3.741	-0.667
3	-67.672	2.048	20.092	2.430	-0.415	4.253	-4.061	-3.224	0.198
4	-30.381	-37.413	-15.554	-0.964	9.005	-0.676	-4.677	1.058	1.848

(续表 10)

组类别	F_1	F_2	F_3	F_4	F_5	F_6	F_7	F_8	F_9
5	48.283	-37.374	7.902	-12.616	-8.363	3.748	2.620	0.914	0.675
6	106.824	24.164	1.930	-6.535	13.380	1.752	2.674	-0.695	-0.972
7	-103.505	44.738	-19.213	-10.440	-4.762	-2.247	0.116	-0.765	0.256
8	-17.034	-4.338	4.116	9.369	0.454	-3.517	6.538	-1.064	2.218
9	18.692	-16.713	-25.192	10.108	-3.774	4.194	1.255	-0.593	-1.954
10	160.530	31.099	2.278	5.171	-6.475	-2.050	-4.201	0.849	0.869

根据表 9 中系数建立模型判别函数:

$$F_1 = 0.369 X_3 - 0.055 X_5 + 0.553 X_8 + \dots + 0.007 X_{134} - 331.721$$

$$F_2 = 0.011 X_3 - 0.038 X_5 + 0.040 X_8 + \dots - 0.007 X_{134} - 191.563$$

$$F_3 = 0.015 X_3 + 0.008 X_5 - 0.127 X_8 + \dots - 0.004 X_{130} + 113.399$$

$$F_4 = 0.047 X_3 - 0.011 X_5 - 0.095 X_8 + \dots + 0.008 X_{134} - 76.098$$

$$F_5 = 0.042 X_3 + 0.010 X_5 - 0.047 X_8 + \dots +$$

$$0.004 X_{134} - 62.262$$

$$F_6 = 0.052 X_3 - 0.024 X_5 - 0.101 X_8 + \dots + 0.002 X_{134} + 9.224$$

$$F_7 = 0.005 X_3 - 0.011 X_5 - 0.083 X_8 + \dots - 0.003 X_{134} - 171.638$$

$$F_8 = 0.034 X_3 - 0.016 X_5 + 0.063 X_8 + \dots - 0.004 X_{134} - 72.314$$

$$F_9 = -0.004 X_3 + 0.007 X_5 - 0.021 X_8 + \dots + 0.001 X_{134} - 14.944$$

表 11 未知酒样与模型 M2 中各组质心距离以及鉴别结果

Table 11 The distance between unknown samples and centroids of each group and identification results in model M2

	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	预测	实际
unknow1	5.01	102.20	63.85	31.41	68.03	137.02	119.71	32.15	54.79	188.79	1	1
unknow2	26.24	87.94	50.74	32.25	88.51	157.89	108.62	43.48	74.23	210.28	1	1
unknow3	101.09	5.26	40.41	97.13	160.58	206.33	40.90	87.60	130.23	259.63	2	2
unknow4	150.05	67.48	94.20	143.34	213.71	267.92	85.59	145.79	187.16	320.33	2	2
unknow5	77.12	31.50	15.84	77.21	138.55	192.12	65.02	69.98	114.56	245.32	3	3
unknow6	68.70	36.13	8.59	70.19	128.23	179.01	63.24	57.88	102.61	232.47	3	3
unknow7	41.07	86.12	57.15	15.56	95.86	157.25	98.66	43.46	64.56	211.13	4	4
unknow8	39.25	89.88	58.79	12.22	94.44	161.50	103.78	48.44	67.89	214.53	4	4
unknow9	57.13	156.50	117.01	78.71	12.18	94.73	170.14	71.26	53.01	140.45	5	5
unknow10	65.49	163.66	124.82	84.99	4.48	88.64	176.11	77.77	54.77	133.73	5	5
unknow11	132.71	209.56	177.42	151.62	87.58	4.50	212.92	128.80	103.56	58.57	6	6
unknow12	142.83	221.24	188.81	161.62	94.56	13.82	224.43	140.03	113.16	48.49	6	6
unknow13	122.48	42.80	71.17	115.14	177.05	214.17	8.33	105.59	141.07	266.52	7	7
unknow14	122.92	46.08	73.11	113.47	177.22	215.88	5.39	107.40	141.04	268.58	7	7
unknow15	32.91	100.26	65.67	50.74	70.73	119.17	113.87	11.82	42.52	171.07	8	8
unknow16	29.81	95.34	58.92	52.28	72.33	123.14	111.21	12.38	50.80	175.08	8	8
unknow17	42.61	128.42	93.57	47.09	52.15	112.83	137.14	44.83	14.11	161.87	9	9
unknow18	55.44	138.19	104.54	63.26	49.68	98.56	144.09	52.87	7.43	146.37	9	9
unknow19	130.61	206.08	174.38	149.96	88.13	20.56	208.53	124.88	99.82	59.24	6	10
unknow20	174.22	254.52	221.73	193.07	121.63	49.62	257.01	171.97	141.76	14.01	10	10
正确率/%												95

注:粗体数据为未知酒样对应不同组别的质心距离的最小值,用来判定未知酒样的组别归属。

此外,采用主成分分析得到的6个主分量作为建模数据,逐步回归分析拟合结果与直接输入结果一致,对训练集和验证集分类正确率分别为100%和90%,对测试集中未知酒样的鉴别正确率为95%。

本文选取的白酒样本既有香型导致的组分差异,也有勾调批次的差异,具有体系的复杂性。因此相关性特征选择和主成分分析提取的特征建立的Fisher线性判别模型,对测试集样品的留一法交叉验证正确率未达到100%。以逐步回归分析从全部特征中提取未定物8,2,5-二甲基呋喃和丙酸乙酯等24个主要特征建立的Fisher线性判别模型,其训练集、验证集和测试集分类正确率均为100%;根据相关性筛选出的39个组分特征数据,经逐步回归分析优化为未定物3,2,3-二甲基-1,3-丁二烯和乙酸异戊酯等21个主要特征建立模型,其训练集、验证集和测试集分类正确率分别为100%,100%和95%;以6个主分量建立的鉴别模型,对训练集、验证集和测试集分类正确率分别为100%,90%和95%。以上结果表明,采用逐步回归分析对判别模型中的主要特征进行优化,全部组分特征中经逐步回归筛选出24个特征建立的Fisher判别模型对白酒样品的鉴别效果最佳,其性能上优于预先进行特征提取建立的模型。

3 结论

白酒不经预处理,直接顶空制样进行GC-IMS分析,所得指纹图谱能够直观反映白酒样品间的相似性与组分差异,以及勾调批次的稳定性。通过逐步回归分析提取主要特征可以明显改善Fisher判别模型的性能。以全部134个挥发组分结合逐步回归分析拟合,提取出24个主要组分建立的Fisher线性判别模型性能最佳,对训练集分类正确率为100%。采用留一法交叉验证分类正确率也为100%,相比于直接Fisher线性判别建模正确率提升了82.5%,对测试集中未知酒样的鉴别正确率也达到100%。

GC-IMS方法无需对微量组分进行定性及定量,依据各物质相对含量即可提取白酒的主要特征,并建立模型实现白酒的准确鉴别。GC-IMS检测方法简便易操作,可为白酒生产过程中质量控

制、市场监管及白酒原产地保护提供技术支持,为不同地域、香型以及不同年份的白酒分类鉴别提供参考。

参 考 文 献

- [1] LIU H L, SUN B G. Effect of fermentation processing on the flavor of Baijiu[J]. Journal of Agricultural and Food Chemistry, 2018, 66(22): 5425-5432.
- [2] 吴广黔,曹文涛.白酒品评技巧的心得体会[J].酿酒科技,2019(6):78-81.
WU G Q, CAO W T. Experience of Baijiu sensory evaluation[J]. Liquor-Making Science & Technology, 2019(6): 78-81.
- [3] 李玉勤,葛向阳,孙庆海.白酒品评浅析[J].酿酒,2017,44(2):33-35.
LI Y Q, GE X Y, SUN Q H. Simple analysis of Baijiu sensory evaluation[J]. Liquor Making, 2017, 44(2): 33-35.
- [4] ZHANG Y, KUN H, LU Z M, et al. Colorimetric artificial nose for Baijiu identification[J]. Flavour and Fragrance Journal, 2012, 27(2): 165-170.
- [5] VERONIKA U, JANA S. Determination of geographical origin of alcoholic beverages using ultraviolet, visible and infrared spectroscopy: A review[J]. Spectrochimica Acta Part a-Molecular and Biomolecular Spectroscopy, 2015, 148: 131-137.
- [6] SONG X B, HOU M, LI Z, et al. Multi-element analysis of Baijiu (Chinese liquors) by ICP-MS and their classification according to geographical Origin [J]. Food Quality and Safety, 2018, 2(1): 43-49.
- [7] CHENG P Y, FAN W L, YAN X. Determination of Chinese liquors from different geographic origins by combination of mass spectrometry and chemometric technique[J]. Food Control, 2014, 35(1): 153-158.
- [8] ZHU Z W, QUE L Z, CHEN G Q, et al. Year discrimination of mild aroma Chinese liquors using three-dimensional fluorescence spectroscopy combined with parallel factor and neural network [J]. Spectroscopy and Spectral Analysis, 2015, 35(9): 2573-2577.
- [9] 李娜,程伟,张杰,等.白酒原产地分析鉴别技术研究进展[J].酿酒科技,2018(6):116-121.
LI N, CHENG W, ZHANG J. Research progress in

- analysis & identification technology of the origin of Baijiu [J]. *Liquor-Making Science & Technology*, 2018(6): 116-121.
- [10] 鲁祥凯, 杨彪, 樊保民, 等. 白酒地域特征鉴别技术研究进展[J]. *食品科学*, 2021, 42(15): 278-284. LU X K, YANG B, FAN B M, et al. Advances in technologies for identification of geographical characteristics of Baijiu[J]. *Food Science*, 2021, 42(15): 278-284.
- [11] 李清亮, 赵振宇, 曾稳稳, 等. 基于无机元素的白酒聚类分析[J]. *酿酒科技*, 2020(8): 46-50. LI Q L, ZHAO Z Y, ZENG W W, et al. Cluster analysis of Baijiu based on inorganic elements [J]. *Liquor-Making Science & Technology*, 2020 (8): 46-50.
- [12] 杨雯懿, 郝婧, 田佳于, 等. 利用 ICP-MS/MS 测定白酒中无机元素的主成分分析及聚类分析[J]. *食品与发酵工业*, 2020, 46(3): 257-263. YANG W Y, HAO J, TIAN J Y, et al. Principal component analysis and cluster analysis of inorganic elements in Baijiu using inductively coupled plasma tandem mass spectrometry[J]. *Food and Fermentation Industries*, 2020, 46(3): 257-263.
- [13] 李贺贺, 胡萧梅, 孙宝国, 等. 碳稳定同位素在白酒真实性中的应用[J]. *中国食品学报*, 2019, 19(1): 183-189. LI H H, HU X M, SUN B G, et al. Application of carbon stable isotope in the authenticity of Baijiu [J]. *Journal of Chinese Institute of Food Science and Technology*, 2019, 19(1): 183-189.
- [14] ZHANG J, TIAN Z Q, MA Y Q, et al. Origin identification of the sauce-flavor Chinese Baijiu by organic acids, trace elements, and the stable carbon isotope ratio [J]. *Journal of Food Quality*, 2019, 2019: 1-7.
- [15] 姜涛, 肖唐付, 宁增平, 等. 黔川典型白酒多元素组成及其原产地判定研究[J]. *地球与环境*, 2013, 41(5): 529-535. JIANG T, XIAO T F, NING Z P, et al. Multi-element composition and geographical origin discrimination of the selected liquors[J]. *Earth and Environment*, 2013, 41(5): 529-535.
- [16] LI Z, WANG P P, HUANG C C, et al. Application of vis/NIR spectroscopy for Chinese liquor discrimination [J]. *Food Analytical Methods*, 2014, 7(6): 1337-1344.
- [17] 李娜, 程伟, 张杰, 等. 基于紫外可见分析光谱法对金种子白酒的研究[J]. *酿酒*, 2018, 45(6): 63-65. LI N, CHENG W, ZHANG J, et al. Research of golden seed liquor based on UV-Vis spectroscopy[J]. *Liquor Making*, 2018, 45(6): 63-65.
- [18] ZHU Z W, CHEN G Q, WU Y M, et al. Discrimination of brands of strong aroma type liquors using synchronous fluorescence spectroscopy and chemometrics methods[J]. *Journal of Applied Spectroscopy*, 2019, 85(6): 1101-1106.
- [19] 朱焯炜, 阙立志, 吴亚敏, 等. 三维荧光光谱结合 PARAFAC 和 GA 对中国白酒品牌的鉴别[J]. *中国激光*, 2015, 42(6): 315-320. ZHU Z W, QUE L Z, WU Y M, et al. Identification of Chinese liquors by three-dimensional fluorescence spectra combined with PARAFAC and genetic algorithm[J]. *Chinese Journal of Lasers*, 2015, 42(6): 315-320.
- [20] 杨建磊, 朱拓, 武浩. 基于三维荧光光谱特性的白酒聚类分析研究[J]. *光电子·激光*, 2009, 20(4): 495-498. YANG J L, ZHU T, WU H. Cluster analysis of distilled spirit based on three-dimensional fluorescent spectra[J]. *Journal of Optoelectronics·Laser*, 2009, 20(4): 495-498.
- [21] XU R Y, ZHU Z W, HU Y J, et al. The discrimination of Chinese strong aroma type liquors with three-dimensional fluorescence spectroscopy combined with principal component analysis and support vector machine[J]. *Spectroscopy and Spectral Analysis*, 2016, 36(4): 1021-1026.
- [22] 霍丹群, 张苗苗, 侯长军, 等. 基于主成分分析和判别分析的白酒品牌鉴别方法[J]. *农业工程学报*, 2011, 27(S2): 297-301. HUO D Q, ZHANG M M, HOU C J, et al. Recognition method of liquor brands based on principal component and linear discriminant analysis[J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2011, 27(S2): 297-301.
- [23] 徐睿, 谭红, 何锦林, 等. 金沙蓝钻酒近红外相似度匹配分析模型的建立[J]. *食品科技*, 2014, 39(11): 315-318. XU R, TAN H, HE J L, et al. Establishment of NIR similarity matching analysis model of Sands blue diamond wine[J]. *Food Science and Technolo-*

- gy, 2014, 39(11): 315–318.
- [24] 霍丹群, 宋春霞, 李俊杰, 等. 一种可视化快速鉴别白酒的新方法研究[J]. 现代食品科技, 2014, 30(5): 267–272.
- HUO D Q, SONG C X, LI J J, et al. Development of a colorimetric sensor array for rapid discrimination of Chinese liquors[J]. *Modern Food Science and Technology*, 2014, 30(5): 267–272.
- [25] 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): 168–173.
- [26] WANG Y F, YIN Y, GE F, et al. A drift correction method of E-nose data based on wavelet packet decomposition and no-load data: Case study on the robust identification of Chinese spirits[J]. *Sensors and Actuators B-Chemical*, 2019, 292: 217–224.
- [27] LI J W, YANG P, ZHANG J, et al. A metal ion-regulated colorimetric sensor array: Discriminating Chinese Baijiu from other beverages [J]. *Analytical Methods*, 2019, 11(41): 5274–5281.
- [28] WU X H, ZHU J, WU B, et al. Discrimination of Chinese liquors based on electronic nose and fuzzy discriminant principal component analysis[J]. *Foods*, 2019, 8(1): 38–51.
- [29] LI Q, GU Y, JIA J. Classification of multiple Chinese liquors by means of a QCM-based E-nose and MDS-SVM classifier[J]. *Sensors*, 2017, 17(2): 272–286.
- [30] 田婷, 邱树毅, 聆吉, 等. 电子鼻技术对不同轮次酱香型白酒的区分与识别[J]. 中国酿造, 2017, 36(10): 71–75.
- TIAN T, QIU S Y, LING J, et al. Discrimination and identification of different rounds Moutai-flavor Baijiu by electronic nose[J]. *China Brewing*, 2017, 36(10): 71–75.
- [31] HE W, REN F, WANG Y Q, et al. Application of GC-IMS in detection of food flavor substances [J]. *IOP Conference Series: Earth and Environmental Science*, 2020, 545(1): 012030.
- [32] WITKIEWICZ Z, PERYCZ U, MAZIEJUK M, et al. Coupling gas chromatography with ion mobility spectrometry[J]. *LC GC Europe*, 2016, 29(6): 294–303.
- [33] WANG S Q, CHEN H T, SUN B G. Recent progress in food flavor analysis using gas chromatography-ion mobility spectrometry (GC-IMS)[J]. *Food Chemistry*, 2020, 315: 126158.
- [34] CHEN T, QI X P, CHEN M J, et al. Discrimination of Chinese yellow wine from different origins based on flavor fingerprint[J]. *Acta Chromatographica*, 2020, 32(2): 139–144.
- [35] ZHOU R, CHEN X, XIA Y, et al. Research on the application of liquid-liquid extraction-gas chromatography-mass spectrometry (LLE-GC-MS) and headspace-gas chromatography-ion mobility spectrometry (HS-GC-IMS) in distinguishing the Baiyunbian aged liquors[J]. *International Journal of Food Engineering*, 2021, 17(2): 83–96.
- [36] JOHN G H, RON K, KARL P. Irrelevant features and the subset selection problem[J]. *Machine Learning Proceedings*, 1994, 1994: 121–129.
- [37] DASH M, LIU H. Feature selection for classification [J]. *Journal of Machine Learning Research*, 1997, 1(1/2/3/4): 131–156.
- [38] 李郅琴, 杜建强, 聂斌, 等. 特征选择方法综述[J]. *计算机工程与应用*, 2019, 55(24): 10–19.
- Li Z Q, DU J Q, NIE B, et al. Summary of feature selection methods[J]. *Computer Engineering and Applications*, 2019, 55(24): 10–19.
- [39] TEBBENS J D, SCHLESINGER P. Improving implementation of linear discriminant analysis for the high dimension/small sample size problem[J]. *Computational Statistics & Data Analysis*, 2007, 52(1): 423–437.
- [40] 游士兵, 严研. 逐步回归分析法及其应用[J]. *统计与决策*, 2017(14): 31–35.
- YOU S B, YAN Y. Stepwise regression analysis and its application[J]. *Statistics & Decision*, 2017(14): 31–35.

Feature Analysis and Identification of Baijiu Based on Gas Chromatography-Ion Migration Spectrometry

Lu Xiangkai¹, Yang Biao^{1*}, Sun Ying², He Jinna³, Fan Baomin¹, Sun Hui¹, Sun Xiaotao⁴

(¹College of Chemistry and Materials Engineering, Beijing Technology and Business University, Beijing 100048

²Jiangsu King's Luck Brewery Joint-Stock Co., Ltd., Huaian 223411, Jiangsu

³Shandong Hanon Instrument Co., Ltd., Dezhou 253000, Shandong

⁴School of Light Industry, Beijing Technology and Business University, Beijing 100048)

Abstract A simple method for accurate identification of Baijiu was developed without qualitative and quantitative determination of volatile components. A total of 10 kinds of Baijiu with two flavor types were selected, and each of them has 2 samples from different blending batches. After directly headspace sample preparation, gas chromatography-ion mobility spectrometry (GC-IMS) was used to obtain the visualized fingerprints, and further extract the main features of Baijiu to establish Fisher linear discriminant models. A total of 134 volatile components were detected in the samples. The fingerprints can directly reflect the content difference of trace volatile components and the batch stability of blending. Under the premise of not fully analyzing and quantifying all components, stepwise regression analysis is used to optimize the three Fisher linear discriminant models which are respectively based on all 134 volatile components, 39 characteristic components selected by correlation analysis, and the 6 principal components obtained from principal component analysis. For the optimized correlation analysis model (the component features are reduced to 21) and the principal component analysis model, the classification accuracy rates and leave-one-out cross validation classification accuracy rates for the training set are 100%, and the discrimination accuracy rates for the test set are both 95%. The optimized Fisher linear discriminant model based on all the features has the best result, and 24 component features are extracted. The model's classification accuracy rate of the training set is 100%, the leave-one-out method cross-validation classification accuracy rate is 100%, and the prediction accuracy rate of the Baijiu samples in the test set reaches 100%. The GC-IMS is significant for the quality control, market supervision and the origin protection of Baijiu, and it can accurately identify Baijiu combining with pattern recognition.

Keywords Baijiu; GC-IMS; fingerprint; feature extraction; linear discriminant