

基于机器学习和电子舌技术的白酒掺假鉴别

程铁辕, 夏于林, 张莹

成都海关技术中心宜宾分部国家酒类检测重点实验室(宜宾 644000)

摘要 采用电子舌技术, 针对国内白酒市场上几种常见类型的掺假白酒, 对14个白酒样品的酸味、苦味、涩味、苦味回味、涩味回味、鲜味、丰富性、咸味共8个指标进行分析。测定结果显示, 掺假白酒与原酒、成品酒在苦味、涩味、苦味回味和鲜味、丰富性等指标之间存在一定程度的差异, 表明这些指标在白酒掺假鉴别方面具有应用价值。将14个白酒样品按照原酒、成品酒和掺假白酒分为3类, 结合酸味等8个指标的电子舌试验数据, 采用机器学习KNN、决策树2种分类器进行分析, 通过交叉验证比较两种分类器的准确率, Python分析结果显示, KNN、决策树等模型的准确率分别为0.900 0和0.866 7, 表明KNN的性能相对更高。综合比较2种分类器准确率, 选择KNN作为最终分类器, 该模型预测准确率为100%。

关键词 电子舌技术; 白酒; 鉴别; KNN; 决策树

Identification of Adulteration of Chinese Liquor Based on Machine Learning and Electronic Tongue Technology

CHENG Tiejuan, XIA Yulin, ZHANG Ying

The State Key Lab of Liquor Products Test of Chengdu Customs Technology Center Yibin Branch (Yibin 644000)

Abstract The electronic tongue technology was used to target several common types of adulteration of Chinese liquor in the domestic market, starting from 8 indicators such as sourness, bitterness, astringency, aftertaste-B, aftertaste-A, umami, richness and saltiness. About 14 Chinese liquor samples were analyzed, and the results showed that there was a certain degree of difference in the bitterness, astringency, aftertaste-B and umami, richness and other indicators of adulterate liquor and original liquor and bottled liquor, indicating that these indexes had certain application value in the identification of liquor adulteration. At the same time, 14 Chinese liquor samples were divided into 3 categories according to the original liquor, bottled liquor and adulterate liquor. Combined with the electronic tongue experimental data of 8 indicators such as sourness, two classifiers such as KNN and decision tree in machine learning were used for analysis, and compared the accuracy of the two classifiers through cross-validation. The results of Python analysis showed that the accuracy rates of KNN and decision tree were 0.900 0 and 0.866 7, respectively, which showed that the performance of KNN was relatively higher. After comprehensive comparison, KNN was selected as the final classifier, and the accuracy of the model prediction reached 100%.

Keywords electronic tongue technology; liquor; identification; KNN; decision tree

浓香型白酒是国内白酒产品中市场占有率最高的类别, 深受国内消费者喜爱, 同时, 也是白酒掺假的重灾区, 受检测手段和监管条件制约, 白酒掺假的监管效率相对低下, 在高额利润的驱使下, 白酒掺假成为不少中小型白酒企业的半公开化的操作模式, 严重影响白酒交易市场的公平正义, 该现象引起众多国内学者关注与重视, 并针对性采用多种研究方法, 其中电子舌技术是该领域具有较好应用前景的分析检测技术。近年来, 电子舌技术发展迅猛, 越来越广泛地被应用于各类食品的感官评价研究, 其中应用于各种酒类产品的研究主要有: 贾洪锋等^[1]采用电子舌对啤酒进行区分识别研究; 周红标等^[2]研究利用电子鼻和电子舌技术快速检测白酒品质; 李付丽等^[3]将电子舌和测色仪应用于酱香型白酒质量检测方面; 郭壮等^[4]基于电子舌技术对襄阳市售米酒滋味品质进行评价, 并对

市售与农家自酿孝感米酒滋味品质进行比较研究^[5]; 门洪等^[6]基于电子鼻/舌融合技术对白酒类别进行辨识; 巴特尔达赖等^[7]通过电子舌技术对真假两种伊力老陈酒10年进行分析; 马泽亮等^[8]采用电子舌法检测橙汁及白酒的品牌及纯度; Men等^[9]利用E-Tongue and E-Nose挖掘啤酒风味信息分类中的数据融合特征; 沈馨等^[10]应用电子舌对柑橘酒滋味品质进行评价。从文献来看, 电子鼻和电子舌技术在白酒品质辨识方面应用较为广泛, 采用的数学分析方法主要有主成分分析、多元方差分析、K均值法、支持向量机法, 按照学习方法的分类, 大多数属于无监督学习(如聚类问题等), 极少数为监督学习(如分类问题等), 总体来看, 在分类及预测上应用的数学分析方法较为单一。鉴于电子鼻数据在白酒风味辨识上的数据刻画效果略差, 主要基于电子舌数据, 选用并比较机器学习

KNN、决策树2种分类器的效果,以期从数学分析角度丰富电子舌数据处理方法,为白酒品质鉴别等研究领域提供更多的可供选择的数据处理方法。

1 材料与方 法

1.1 样品制备与前处理

样品制备:浓香型白酒样品共14个,编号为1~14,其中,1号为2009年一级原酒;2号为2018年一级原酒;3号为2018年优级原酒;4号为2019年一级原酒;5号为2019年优级原酒;6号为掺假白酒(高酯酒);7号为掺假白酒(己酸乙酯掺假);8号为30%己酸乙酯掺假酒+70% 2019年一级原酒(体积比);9号为50%己酸乙酯掺假酒+50% 2019年一级原酒(体积比);10号为70%己酸乙酯掺假酒+30% 2019年一级原酒(体积比);11~14号为同类别不同批次52°五粮浓香型成品酒。白酒原酒及成品酒购自川南地区白酒企业;掺假白酒中高酯酒及己酸乙酯为网上购买。

样品前处理:分别量取30 mL各白酒样品倒入250 mL烧杯中,加入60 mL纯净水稀释混匀后待测。

1.2 主要仪器设备与试剂

味觉分析系统(简称电子舌,日本INSENT公司,

型号TS-5000Z),该设备具有广域选择特异性的人工脂膜传感器,可以模拟生物活体的味觉感受机理,通过检测各种味物质和人工脂膜之间的静电作用或疏水性相互作用产生的膜电势的变化,实现对5种基本味(酸、涩、苦、咸、鲜)和甜味的评价。

测试用液:reference溶液(人工唾液)为KCl+酒石酸;负极清洗液为水+乙醇+HCl;正极清洗液为KCl+水+乙醇+KOH。

1.3 数据处理

采用Python软件进行数据处理。

2 结果与讨论

2.1 试验结果

白酒样品的电子舌试验数据如表1所示,所有数据均是以人工唾液(参比溶液)为标准的绝对输出值,电子舌测试人工唾液的状态模拟人口腔中只有唾液时的状态。其中,tasteless为无味点,即参比溶液的输出,参比溶液(reference)由KCl和酒石酸组成味觉值,故酸味的无味点为-13,咸味的无味点为-6,以此为基准,样品的味觉值低于tasteless时,即认为样品无该味道或者味道很弱,反之则有。

表1 白酒样品的电子舌试验数据

序号	酸味	苦味	涩味	苦味回味	涩味回味	鲜味	丰富性	咸味
tasteless(无味点)	-13	0	0	0	0	0	0	-6
1	19.71	-3.1	-4.21	0.98	1.34	0.86	2.79	-10.37
2	20.02	-2.02	-2.9	1.2	1.63	0.53	2.58	-11.77
3	18.65	-2.64	-3.6	1.09	1.64	1.33	2.83	-11.46
4	17.39	-4.55	-4.99	0.73	1.34	1.76	3	-12.46
5	18.28	-3.9	-4.22	0.91	1.51	1.42	2.82	-12.38
6	13.32	1.38	-1.56	0.87	1.69	3.39	4.27	-10.44
7	-16.2	3.03	-5.77	-0.91	0.28	5.73	3.64	-15.5
8	20.66	-4.4	-4.73	0.59	1.11	0.44	1.8	-15.06
9	19.5	-4.3	-4.91	0.44	0.94	0.57	1.83	-15.94
10	16.84	-4.46	-5.79	0.24	0.73	1.15	2.14	-17.01
11	22.5	-0.36	-2.96	0.94	1.27	-0.99	1.44	-16.75
12	22.61	-0.3	-2.78	0.96	1.37	-1.1	1.43	-16.89
13	22.67	-0.24	-2.65	0.93	1.34	-1.17	1.39	-16.99
14	22.88	-0.13	-2.47	0.91	1.34	-1.4	1.28	-17.09

2.2 确定有效的味觉指标

以参比溶液的输出为“0”,除酸味和咸味外,其他指标的无味点均为0,将大于无味点的味觉项目作为评价对象。因基准液由氯化钾与酒石酸配制而成,故基准溶液中含有少量的酸和盐,酸味和咸味的无味点分别为-13和-6。无味点以下的项目可认为该样品没有此味道或者味道很弱,尽管白酒的咸味和涩味在无味点以下,但两者在分类时仍可作为有效指标。图1为白酒样品有效味觉指标雷达图。

2.3 白酒各项味觉指标分析

2.3.1 苦味、涩味回味和苦味回味

通过味觉值可知,大部分白酒样品的苦味值在无

味点以下,苦味回味值略高于无味点,大部分数值位于0~1之间。其中,6号和7号白酒具有明显的苦味,6号苦味值为1.38,7号苦味值为3.03,这种明显的苦味可能与6号、7号白酒为完全掺假白酒有关。此外,对于8号、9号和10号部分掺假白酒而言,相较于4号样品,掺假后的白酒中苦味回味有所降低,也与实际相符。由图2可见,11~14号4个白酒在图中聚集在一起,表明四者在苦味方面基本无差异。

将涩味回味与苦味、苦味回味共同分析可知,如图3所示,图中气泡的大小反映涩味回味值的大小,结合味觉值可知,7号白酒的涩味回味在14种白酒样品中最小,其他13种酒在涩味回味方面非常接近。大

部分白酒的涩味回味值位于1.3~1.7, 7号、9号和10号的涩味回味值低于1。

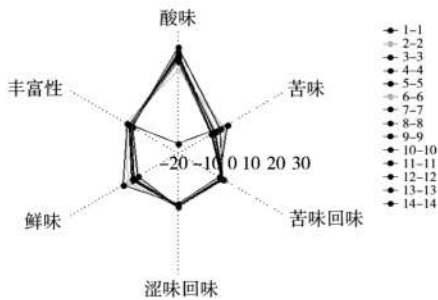


图1 1~14号白酒样品有效味觉指标雷达图

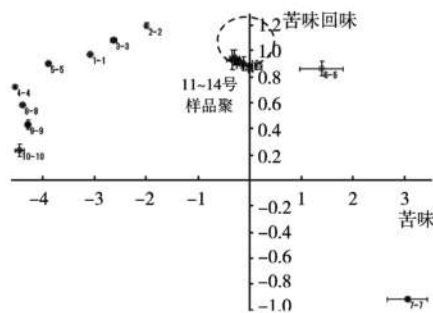
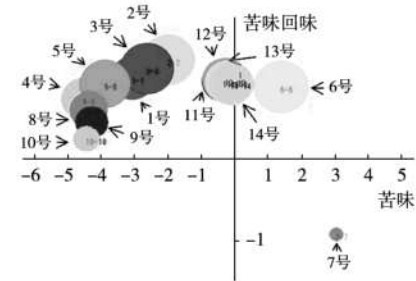


图2 1~14号白酒样品苦味和苦味回味散点图



注：气泡大小反映涩味回味值大小。

图3 白酒样品苦味、苦味回味和涩味回味气泡图

2.3.2 酸味

测定结果显示, 白酒样品会引起酸味传感器较强烈的响应, 如图4所示, 7号样品的酸味最小, 酸味值为-16, 低于酸味无味点。其余13个白酒样品的酸味均很大, 酸味值均在10以上, 其中, 6号白酒的酸味相对较小, 酸味值在15以下, 其他12个白酒样品酸味值均在15以上。

将白酒的酸味制成折线图(图5), 其中未包含7号样品。11~14号4个白酒样品在酸味方面基本一致且酸味值较强, 其余样品在酸味方面的差异也比较明显, 酸味值相差1个刻度以上, 样品中酸味由小到大的排列顺序是7<6<10<4<5<3<9<1<2<8<11~14。

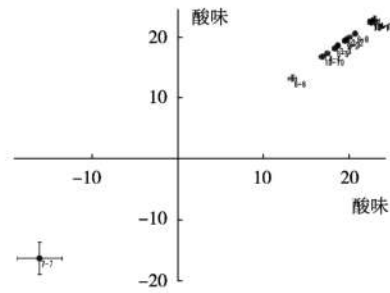


图4 1~14号白酒样品酸味散点图

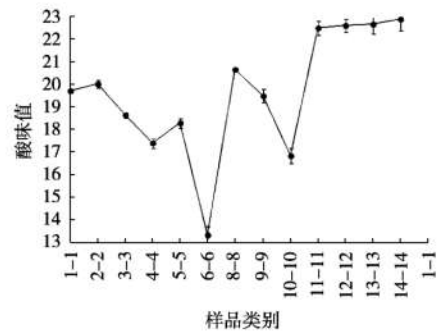


图5 1~14号白酒样品酸味折线图

2.3.3 鲜味和丰富性

通常情况下, 鲜味是由氨基酸、核酸等产生, 丰富性是鲜味的回味, 用于反映鲜味的持久性, 又称为鲜味持久度, 常用于反映鱼汤、海鲜类产品鲜味的持久度。测定结果显示(图6), 多数白酒样品具有相对较大的鲜味和丰富性, 其中, 11~14号4个白酒样品在鲜味和丰富性方面非常接近, 鲜味和丰富性均很小, 且鲜味在无味点以下; 7号和6号白酒样品的鲜味和丰富性在14个白酒样品中最大, 表明完全掺假白酒的鲜味和丰富性均可能更高; 其余8个白酒样品的鲜味集中在0~2, 丰富性集中在1~3。相较于4号样品, 白酒掺假后, 其鲜味和丰富性均有所下降。

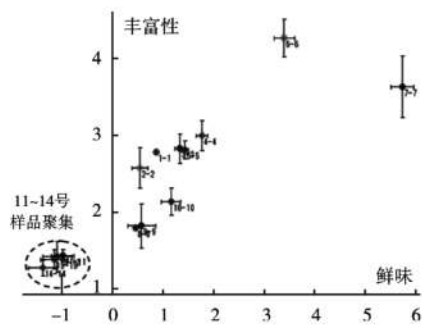


图6 白酒样品鲜味和丰富性散点图

2.4 数据重现性分析

所有白酒样品检测4次, 即4次循环, 将第1次循环去掉后, 取后3次的平均值, 如图7所示, 传感器响应稳定, 有较好的重现性, 因此, 可以认为试验数据

有效。

3 分类和预测

测试分类器主要选用KNN、决策树2种，经测试采用8个指标效果较好。未选用SVM（支持向量机法）分类器的原因是SVM主要是解决二分类问题；未选用朴素贝叶斯分类器的原因是每种指标都必须相互独立，而试验中部分指标具有相关性。具体过程：选择最佳参数构建KNN模型，决策树模型参数全部保持默认，不用传入参数，如图8所示，n_neighbors（最近邻）=3时，模型效果最佳，所以选择最优的n_neighbors（最近邻）=3传入模型；分别对2种分类器进行交叉验证，交叉验证是将原始数据集划分为若干份（试验中划分为4份），进行4次训练，每次取1份作为测试数据，剩余3份作为训练数据，并将4个模型的预测值做平均。交叉验证主要用于评估模型的预测性能，尤其是训练好的模型在新数据上的表现，可在一定程度上减小过拟合，且可从有限的数据中获取尽可能多的有效信息。KNN和决策树2种分类器的准确率分别为0.900 0和0.866 7，通过对分类器准确率的比较，发现KNN的性能较高，综合比较2种分类器的准确率后，选择KNN作为最终分类器。使用KNN进行模型分类与预测，模型预测的准确率达到100%（12条作为训练数据，2条作为测试数据，试验结果即为2条测试数据都分类正确）。

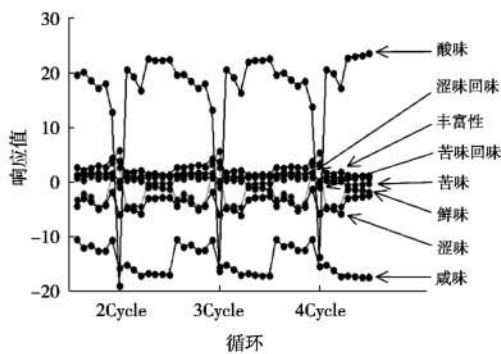


图7 白酒样品的后3次循环折线图

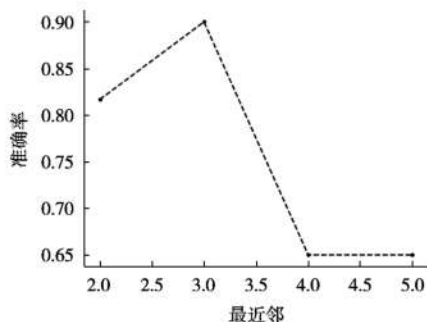
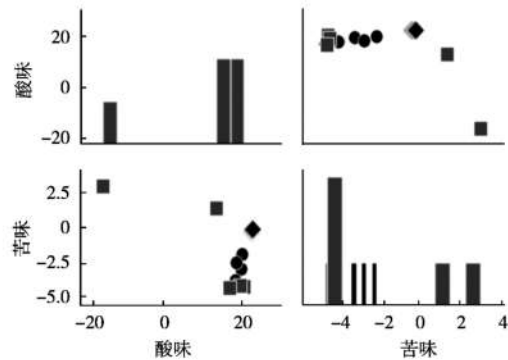


图8 KNN模型准确率折线图

图9为散点图，其中矩阵的对角线是每个特征的

直方图，可以直观看出3个类别能够明显区分开。



注：限于篇幅仅选取了部分散点图展示。

图9 散点图

4 结论

基于电子舌技术，采用味觉分析系统对14个白酒样品的苦味等8个指标进行检测分析，并根据原酒、成品酒和掺假白酒3类样品，采用并比较机器学习KNN、决策树2种分类器，综合比较后，发现KNN的性能相对更高，以KNN作为最终分类器，针对试验数据，模型预测的准确率达到100%，表明机器学习和电子舌技术相结合在白酒掺假鉴别中具有一定的应用前景。由于选取的数据量较小，模型预测结果与现有样本量关系很大，后续研究将选取更多样本进行分析。

参考文献：

- [1] 贾洪锋, 梁爱华, 何江红, 等. 电子舌对啤酒的区分识别研究[J]. 食品科学, 2011, 32(24): 252-255.
- [2] 周红标, 李珊. 基于电子鼻和电子舌的白酒检测[J]. 粮油食品科技, 2014, 22(6): 78-82.
- [3] 李付丽, 吴鑫颖, 陈仁远, 等. 电子舌和测色仪在酱香型白酒质量检测方面的应用[J]. 酿酒科技, 2015, 249(3): 77-80.
- [4] 郭壮, 汤尚文, 王玉荣, 等. 基于电子舌技术的襄阳市售米酒滋味品质评价[J]. 食品工业科技, 2015, 36(15): 289-293.
- [5] 郭壮, 汤尚文, 蔡宏宇, 等. 市售与农家自酿孝感米酒滋味品质的比较研究[J]. 食品工业, 2015, 36(11): 185-188.
- [6] 门洪, 张晓婷, 丁力超, 等. 基于电子鼻/舌融合技术的白酒类别辨识[J]. 现代食品科技, 2016, 32(5): 283-288.
- [7] 巴特尔多赖, 王锡昌, 吴娜, 等. 电子舌技术在真假伊力老陈酒鉴别中的应用[J]. 食品工业科技, 2017, 38(9): 290-293.
- [8] 马泽亮, 殷廷家, 国婷婷, 等. 采用电子舌法检测橙汁及白酒的品牌及纯度[J]. 食品工业科技, 2018, 39(8): 190-194.
- [9] MEN H, SHI Y, FU S L, et al. Mining feature of data fusion in the classification of beer flavor information using e-tongue and e-nose[J]. Sensors, 2017, 17(7): 1656-1672.
- [10] 沈馨, 马佳佳, 刘文汇, 等. 浓香型白酒窖泥中乳酸菌的分离鉴定及其在柑橘酒中的应用[J]. 中国酿造, 2018, 37(7): 42-46.