

基于秩相关与聚类分析的评价员排序能力评估

Study on ranking ability evaluation for sensory assessors based on rank correlation and cluster analysis

史波林¹ 汪厚银¹ 奂 畅² 张璐璐¹

SHI Bo-lin¹ WANG Hou-yin¹ HUAN Chang² ZHANG Lu-lu¹

支瑞聪¹ 解 楠¹ 赵 镭¹ 陈修红¹

ZHI Rui-cong¹ XIE Nan¹ ZHAO Lei¹ CHEN Xiu-hong¹

(1. 中国标准化研究院食品与农业标准化研究所, 北京 100191; 2. 上海大学计算机工程与科学学院, 上海 200444)

(1. Food and Agriculture Standardization Institute, China National Institute of Standardization, Beijing 100191, China; 2. Department of Computer Engineering and Science, Shanghai University, Shanghai 200444, China)

摘要:将最常用的甜、酸基本味分别设置成 6 个浓度体系作为评价员排序能力筛选的参比样,考虑到感官疲劳的客观现象,12 位评价员对每个味道系列浓度进行 3 次重复排序,在分别采用 Spearman 检验和 Kendall 检验评估评价员排序正确性与重复性的基础上,采用聚类分析融合正确性与重复性数据特征,综合评估分析每位评价员的排序能力。发现 5 号评价员表现最差需要剔除;3 号评价员需要进一步训练;6 位评价员(编号为 2、4、9、10、11、12 号)排序能力表现优异,可直接开展检测试验;4 位评价员(编号为 1、6、7、8 号)排序能力良好。同时发现 Kendall 检验系数 W 值与 Spearman 检验系数 r_s 平均值具有整体正相关趋势,而 W 值能更精确区分评价员之间的能力差异。该研究可为感官分析实验室中评价员的筛选与维护提供理论依据和实践经验。

关键词:排序能力;感官分析;评价员;秩相关;聚类分析

Abstract: The commonly used tastes of sweet and sour were respectively set to six concentrations system as the test samples for ranking ability screening of sensory assessors. In consideration of the sensory adaption, 12 assessors were asked to rank three times to each taste. Firstly, the accuracy of every assessor was evaluated by Spearman test. And Kendall test was used to evaluate the repeatability of the assessor. Finally, the ranking ability of each assessor was evaluated and analyzed comprehensively through fusion the data characteristics of accuracy and repeatability by clustering analysis. The results were showed that No. 5 assessor need to be eliminated for his poor performance and No. 3 assessor need to be trained more, while No. 1,

6, 7, 8 assessors showed good ranking ability and No. 2, 4, 9, 10, 11, 12 could directly carry out the test because of their excellent ranking ability. It was also found that W value of Kendall test and r_s value of Spearman test were overall positively related. However, W value of Kendall test could distinguish the ability more accurately among assessors. This paper provides theoretical guidance and practical experience for assessor screening and maintenance in sensory analysis laboratory.

Keywords: ranking ability; sensory analysis; assessor; rank correlation; cluster analysis

排序法是一种要求评价员将样品按某种感官特性的强弱或对样品整体印象的好坏进行排列的分类方法^[1]。在原料评级、产品研发、市场测试等领域,均需要评价员对产品的质量、口味、总体印象等分出好坏优劣。排序法作为普遍应用的感官评价方法之一,要求评价员提供的排序结果是客观的、可靠的与稳定的,这就需要对评价员的排序能力表现进行评估分析。

对于评价员的排序表现,最理想的结果是评价员每次都能给出与真实品质顺序或者理论最优顺序完全一致的答案,而这样的评价员毕竟是少数,除非被排序的样品品质特征差异过大。感官评价容易受到生理、心理、环境等各种因素的影响而导致感官分析评价结果容易产生一定程度的波动。假如,对于实际秩次为(1,2,3,4,5,6)的 6 个样品,在某一次两轮重复排序试验中,3 位评价员对秩次从小到大的样品给出的结果:评价员 A 的结果分别为(1,3,2,4,5,6)和(1,2,3,5,4,6);评价员 B 的结果分别为(1,2,3,6,4,5)和(1,2,3,6,4,5);评价员 C 的结果分别为(1,2,3,5,4,6)和(1,2,3,5,4,6)。对于这 3 位评价员的表现,谁优谁劣该如何考核与评

基金项目:中国标准化研究院院长基金项目(编号:562013Y-3079, 562014Y-3348)

作者简介:史波林(1981—),男,中国标准化研究院副研究员,博士。
E-mail: shibl@cnis.gov.cn

通讯作者:赵镭

收稿日期:2015-03-03

估？这方面的研究不是很多，目前已开展的研究主要集中在欧美国家^[2]。

ISO 8587—2006、GB/T 12315—2008 采用了 Spearman 检验,通过计算出评价员实际排序结果与正确排序结果之间的秩相关系数 r_s ,来体现评价员排序结果的正确性。 r_s 值越大,评价员的排序表现越优秀。在上述例子中,可通过计算发现 A 和 C 两位评价员每次排序的 r_s 值都完全相等,然而通过观察原始排序结果,A 的稳定性显然没有 C 好,即评价员 C 比 A 表现得更加优秀。Jian Bi 等^[3]对评价小组及其成员能力评估提出组内相关系数法 (intraclass correlation coefficient, ICC),将计算的评价员排序结果 ICC 系数进行统计检验得出评价员能力等级。按此方法,B 和 C 两位评价员表现的 ICC 值均相等,但评价员 C 的答案与正确结果显然更加接近。

检测设备性能评估应包括检测结果的一致性、重复性与再现性三个方面^[4],与感官分析仪器评估的主要构成要素相似,评价员的评估也是如此。其中,再现性更多的体现评价员分析结果与评价小组内其他成员结果的差异程度^[5],主要应用于无理论顺序情况下的排序能力评估或评价小组中异常评价员的分析;而在感官分析实验室内部对评价员个人能力评估、快速筛查时,重点考察评价员表现的一致性和重复性,这两方面足以说明评价员排序能力是否优秀。为了考核感官分析实验室内评价员的表现,方便感官分析师客观筛选排序能力表现优秀的评价员,本试验提出基于秩相关检验和聚类分析的排序能力评估方法。通过秩相关检验方法分别评估评价员的一致性和重复性表现特征,采用聚类分析融合评价员的一致性和重复性并综合判断评价员排序能力。这样不仅有助于克服特殊情况下现有方法的局限性,将综合评价员的一致性和重复性评估其排序能力,也为评价小组组建时的评价员筛选提供了技术方法与理论依据。

1 材料与方法

1.1 试验材料

本研究重点在于方法探索,因此选择通用型的甜、酸基本溶液作为研究对象。甜物质为购自北京二商集团北京市糖业烟酒公司的精致绵白糖(500 g 袋装),酸物质为购自北京超益元食品添加剂经销部的食品级柠檬酸。

1.2 试验仪器

电子天平:AR5120 型,精度为 0.000 1 g,瑞士梅特勒—托利多国际股份有限公司;

评价员品评用的容器:无色无味的塑料品评杯,厦门千牛电子商务有限公司。

1.3 试验方法

考虑感官疲劳和统计意义的双重因素,对不同味道样品均配置 6 个浓度系列样品^[6],其样品浓度的难易程度根据以下原则来制备:保证 1/4 的评价员能准确排序,1/4 的评价员觉得难度很大,而剩下的 1/2 评价员只是在个别排序中把握

不准。具体浓度见表 1。

表 1 样品溶液秩次及对应浓度
Table 1 The sample solution order and corresponding concentration (g · L⁻¹)

溶液	各秩次浓度					
	1	2	3	4	5	6
酸	0.42	0.52	0.65	0.80	0.97	1.16
甜	12.1	14.0	16.0	18.1	20.3	22.6

采用中国标准化研究院开发的感官分析试验计算机管理系统《轻松感官分析系统》在线设计本次甜、酸两组并每组 3 轮重复的排序试验,由系统自动生成样品制备表、样品呈送表和评价员回答表,保证完全随机设计、每轮每个样品间不重复的 3 位随机编码,整个过程严格按照中国 GB/T 12315—2008 的要求进行设计。12 位评价员为实验室经过基础筛选并用于开展差别检验试验的初级评价员。按照排序试验和滋味评价技术要求^[1],由评价员对样品进行感官分析,样品评价过程中保证漱口和休息时间。

1.4 数据处理

排序试验所获得的为定序数据,在做相关性的统计分析时,都需要转化成由小到大排成的秩次。Spearman 检验用来分析两组定序数据之间的相关性,特别应用于评价员排序结果与真实顺序之间的吻合度,其最终值以检验系数 r_s 表示,具体公式见 GB/T 12315—2008。 r_s 在-1~1,负数代表负相关,在感官分析中几乎不可能出现负相关现象,一般在 0~1,越靠近 1,说明相关性越强,若为 1 则代表与真实顺序完全一致。

对于 3 组以上(含 3 组)定序数据的相关性分析,主要采用 Kendall 和谐系数 $W^{[6]}$,其值位于 0~1,越接近 1 说明评价员自身多次排序结果之间越一致,重复性越好。而对于 W 值的显著分析需要借助于自由度为 $n-1$ 的卡方检验。即将 W 转化为卡方值 $\chi^2(W)=k \times (n-1) \times W^{[7]}$,其中 k 为评价次数, n 为样品数量。

本研究采用系统聚类法综合分析评价员的排序能力,聚类分析是将数据分类到不同的类或者簇的统计方法,同一个簇中的对象具有很大相似性,不同簇间具有很大的相异性^[8]。

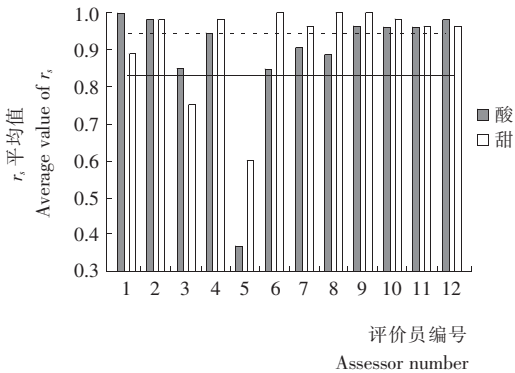
2 结果分析

2.1 基于 Spearman 检验的评价员排序正确性分析

评价员筛选的排序测试中,一次试验结果具有一定的偶然性,很难完全体现评价员个体的排序能力。但因为感官评价是通过评价员进行测试,若重复次数过多,也非常容易引起感官疲劳,评价员产生负面情绪。本研究对甜和酸两个系列溶液分别进行了 3 次重复的排序试验。图 1 是评价小组的 12 位评价员在甜味和酸味排序中秩相关系数 r_s 的平

均值。

由图1可知,评价员自身普遍对甜味的排序正确性要好



实线代表 5%显著水平线,虚线代表 1%显著水平线

图 1 评价员酸味、甜味排序结果一致性平均值
Figure 1 The average value for consistency of sour and sweet ranking results

于酸味,这不但与被测试样品浓度有关,重要的是酸味的后味相对难以祛除,若漱口不充分或休息时间不到位,其评价结果容易产生更大误差。1 号评价员的 3 次甜味排序均值虽然未达到 1%的显著水平,但也通过了 5%显著水平检验,而其它 9 位评价员的甜味排序表现出了较为优秀的正确性,其 r_s 平均值的检验水平都超过了 1%显著水平。在酸味排序中,7 位评价员(1、2、4、9、10、11、12 号)表现出了较为优秀的排序正确性(r_s 平均值超过 1%显著水平);4 位评价员(3、6、7、8 号)排序结果的正确性表现尚可(r_s 平均值超过 5%显著水平);而 5 号评价员表现最差,甚至比自身的甜味排序一致性表现还差。甜味排序表现欠佳的评价员人数(3 号与 5 号)要多于酸味的(5 号),评价员 3 与 5 的排序结果见表 2。由表 2 可知,其中 5 号评价员对甜味和酸味的排序表现都是最差的,其每次排序结果的正确率非常低,从而导致 r_s 平均值也偏低。其中 3 号评价员的甜味排序正确率也偏低,其 r_s 平均值也未达到 5%的显著水平线。

表 2 评价员 3 与 5 的排序结果
Table 2 The ranking results of No. 3 and 5 assessors

样品系列	评价员	样品秩次						r_s	r_s 平均值	W
		1	2	3	4	5	6			
甜	3 号	1	2	5	6	3	4	0.54	0.752	0.771
		1	2	4	3	5	6	0.94		
		2	1	4	5	3	6	0.77		
	5 号	1	2	3	4	5	6	1	0.600	0.429
		1	3	2	4	5	6	0.94		
		3	2	6	5	4	1	-0.14		
酸	5 号	3	2	4	5	1	6	0.37	0.371	0.492
		4	2	3	1	5	6	0.49		
		3	4	2	5	1	6	0.26		

在酸味排序中的 1 号评价员和在甜味排序中的 6、8、9 号评价员,每次评价结果都与真实排序高度一致,充分体现了他们的高灵敏度,显然他们在这些试验中的重复性也是最高的。这样的评价员是最理想的,但并不是每位评价员都能做到。

如 2 号评价员的甜和酸溶液排序、4 与 10 号评价员的甜溶液排序、12 号评价员的酸溶液排序,在各自 3 次重复试验中,都只有其中一次的相邻样品排序颠倒的现象,这种情况在实际的排序试验中是经常出现的,主要是感官疲劳或注意力不集中所致,其评价员的真实排序能力也非常不错。

2.2 基于 Kendall 检验的评价员排序稳定性分析

在评价员评估中,若只是采用每次排序后 Spearman 检验 r_s 平均值的处理方式,会掩盖很多信息而降低评估精度。如在甜溶液排序中,评价员 7、11、12 的 r_s 平均值均相同,然而 12 号评价员显得相对更加稳定,因为 2 次秩次颠倒都出

现在样品实际秩次为“5”和“6”上(见表 3)。因此,需要在考察评价员正确性的基础上,继续考察其稳定性的表现,以充分说明评价员结果的可信度到底有多高。若每位评价员每次评价结果均与真实排序高度一致,则足以说明其稳定性,这也是感官分析师最理想的目标,但这种现象是极少数的,不完全符合真实状况。

Kendall 系数 W 值能综合考察评价员多次重复排序所体现的重复性。而对于评价员排序稳定性的显著分析是基于 W 值的卡方检验。由于排序重复次数不超过 7 次时,卡方检验不存在 1%显著水平下的临界值^[7,9]。基于 6 个样品 3 次重复排序试验的卡方检验在 5%临界水平下的临界值,推导出对应的 W 值临界值为 0.738,也就是 W 值在不低于 0.738 的时候,其评价员的稳定性表现才算合格,评价小组的 12 位评价员在甜味和酸味排序中 3 次重复排序后的 W 值见图 2。

表 3 评价员 7、11、12 的甜溶液排序结果

Table 3 The ranking results of No. 7, 11 and 12 assessor for sweet solutions

评价员	样品秩次						r_s	r_s 平均值	W
	1	2	3	4	5	6			
7 号	1	2	3	4	6	5	0.94	0.962	0.949
	2	1	3	4	5	6	0.94		
	1	2	3	4	5	6	1		
11 号	1	2	3	5	4	6	0.94	0.962	0.949
	1	2	3	4	5	6	1		
	2	1	3	4	5	6	0.94		
12 号	1	2	3	4	5	6	1	0.962	0.975
	1	2	3	4	6	5	0.94		
	1	2	3	4	6	5	0.94		

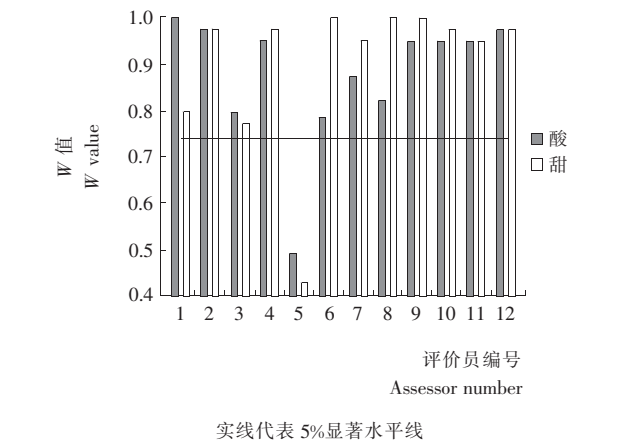


图 2 评价员酸味、甜味排序结果重复性 W 值

Figure 2 The W value of repeatable ranking results for sour and sweet solutions

由于甜度排序的正确性普遍好于酸味样品,由图 2 可知,甜度重复排序试验的重复性也基本优于酸味样品。而 5 号评价员在甜和酸的两个排序试验中的稳定性表现均最差,而他的一致性和重复性并未体现一定的正相关性,并且在

甜样品的 3 次重复排序时,表现一次不如一次,第一次排序与正确结果完全吻合,第二次排序颠倒了正确秩次为 2 和 3 的相邻两个样品,而最后一次排序更乱,导致 5 号评价员的排序正确性甜味好于酸味,而稳定性反之(见表 2)。评价员 7、11、12 的甜度正确性表现比较好,并且 r_s 平均值均相同,但通过 W 值能更精确的发现评价员 12 号的重复性要好于其他两位评价员。同理评价员 3 和 6 号的酸度正确性一致,但 3 号评价员的稳定性更优于 6 号,然而 6 号评价员的酸度排序却一次比一次好,这是很值得进一步关注的现象(见表 4)。同时,W 值分析能将评价员之间的水平差异体现更加明显,如酸试验中评价员 7 的正确性比 8 提高 2.2%,而相应的稳定性却提高了 6.2%,这在图 1 与 2 中也能很直观地观察到。由此可见,虽然重复性分析结果与正确结果之间具有一定的正相关性,但重复性能更加精细的区分评价员之间的细致变化,可提高评价员能力的区分精度。

2.3 基于聚类分析的评价员排序能力评估

当每次重复排序试验结果都与真实排序一致时,自然也体现了高度的可重复性,这样只考察 r_s 是否为“1”就可以了。由于感官分析试验是将评价员作为仪器,因心理、生理

表 4 评价员 3 与 6 的酸溶液排序结果

Table 4 The ranking results of No. 3 and 6 assessors for sour solutions

评价员	样品秩次						r_s	r_s 平均值	W
	1	2	3	4	5	6			
3 号	1	2	4	5	3	6	0.83	0.848	0.797
	2	1	3	4	6	5	0.89		
	1	2	5	3	4	6	0.83		
4 号	4	2	1	3	5	6	0.60	0.848	0.784
	1	2	4	3	5	6	0.94		
	1	2	3	4	5	6	1.00		

和环境所带来的影响,使得不同评价员之间的表现有一定差异;且同一评价员多次结果中也会有偶然性,因此只单方面考察正确性(一致性)或稳定性(重复性)都比较片面。在评价员筛选和评估中,感官分析师都希望评价员的表现既有好的正确性,也有高的稳定性。所以,需要融合该两大指标综合判断和分析评价员的排序能力。本研究对每个味道排序试验结果分别进行聚类分析,对同一味道都以评价员的一致性和重复性为变量进行评价员排序能力分类,其聚类树结果见图 3。

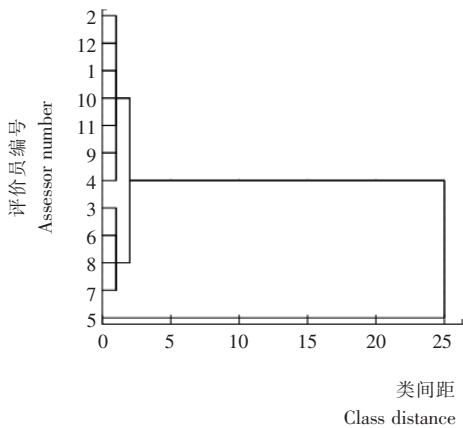
由图 3 可知,不同类的评价员排序能力的水平。在酸味排序能力中,评价员的能力可分为 3 类,第一类为 1、2、4、9、10、11、12 号评价员,第二类为 3、6、7、8 号评价员,第三类为 5 号评价员。在甜味排序能力中,评价员的能力可分为 4 类,第一类为 2、4、6、7、8、9、10、11、12 号评价员,第二类为 1 号评价员,第三类为 3 号评价员,第四类为 5 号评价员。聚类分析有机结合了一致性和重复性的数值特征,共同反映评价员能力的优劣,并将其聚成不同的类别,类别之间的距离可数值化,即类别数值可视化,这为感官分析师对评价员的管

理提供了可靠的依据。

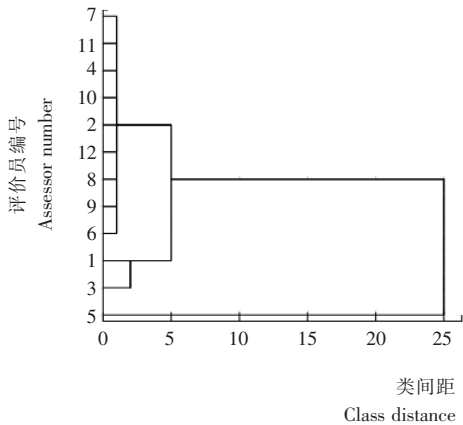
众所周知, r_s 平均值越大评价员一致性越好; W 值越大评价员重复性越好。图 4 是以一致性为横坐标、重复性为纵坐标的评价员排序能力散点图。图 4 中代表评价员的散点越靠近坐标系右上角,说明评价员排序能力越好。根据 r_s 平均值和 W 值的显著临界值线,将评价员性能分成了几类。其中甜味排序能力图 4(b)中评价员能力分层较为明显,将 12 位评价员分成了四块;酸味排序能力图 4(a)将 12 位评价员的能力分为三块。此结论与聚类分析结果完全吻合,即第一类评价员表现最优秀,第二类良好,第三类需进一步训练,第四类最差,可直接淘汰。只是此处理方式可以在同类中再进一步细分评价员的性能高低。由于本结果是同一天试验所得,而在评价员性能的不同天长期监测中,每一类评价员的能力是允许在一定范围内波动的,因此聚类分析结果更具有实用性。

3 结论与展望

在本研究的评价员排序能力筛选中,综合考虑评价员感官疲劳和体现性能的统计意义,设置了甜酸两个系列、分别 6



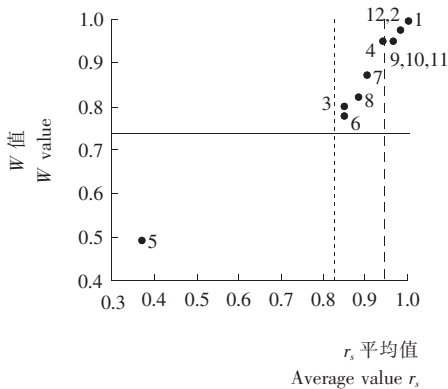
(a) 酸溶液



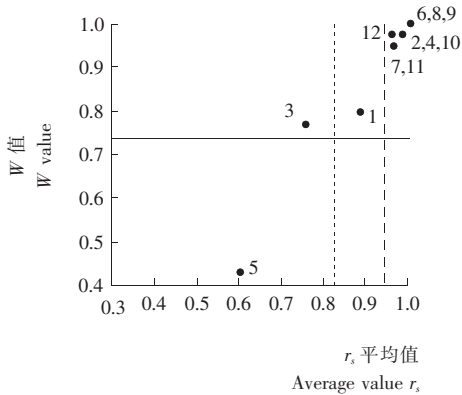
(b) 甜溶液

图 3 评价员排序能力聚类树

Figure 3 The clustering tree for ranking ability of assessors



(a) 酸溶液



(b) 甜溶液

实线代表 W 值的 5% 显著水平线, 短虚线代表 r_s 平均值的 5% 显著水平线, 长虚线代表 1% 显著水平线

图 4 评价员排序能力散点图

Figure 4 The scatter diagram for ranking ability of assessors

(下转第 221 页)

固定化脂肪酶添加量为 8.5%、米糠油与油酸甘油酯质量比为 14,此时 DAG 产率为 58.31%,验证实际条件预测的理论值为 60.36%与实际值的相对误差为 3.4%,因此响应分析的模型是可靠的^[14]。

3 结论

反应温度、反应时间和反应物质量比对 DAG 产率影响显著,因素的主次顺序依次为反应温度、反应时间和米糠油与油酸甘油酯质量比、固定化脂肪酶添加量。试验通过优化制备富含甘二酯的米糠油的工艺提高米糠油中的甘二酯,反应结束后,使用固体催化剂易于分离且可以重新利用,比较适合制备中低含量的甘油二酯产品,反应原料容易获得,适合大规模生产应用。在今后的研究中应考虑固定化后的脂肪酶重复利用次数对 DAG 产率影响及 pH 对固定酶活性的影响。

参考文献

1 Zhong Nan-jing, Gui Zhong-yu, Li Xu, et al. Solvent-free enzymatic synthesis of 1, 3-Diacylglycerols by direct esterification of glycerol with saturated fatty acids[J]. Lipids in Health and Disease, 2013, 12(1): 65~68.

2 李磊. 高酸价米糠油酶法催化制备富含甘油二酯米糠油的研究[D]. 无锡:江南大学,2010.

3 黄莉莉,苏宜香. 甘油二酯的功能及安全性研究现状[J]. 国外医学(卫生学分册),2007,34(2):94~98.

4 魏金霞,毕艳兰. 甘二酯的制取与应用研究进展[J]. 粮油加工与食品机械,2005(2):45~48.

5 Beis I, Juffer A, Weckström M, et al. Physical properties of model phosphatidylcholine bilayers containing diacylglycerol (DAG): a molecular dynamics simulations study[J]. Biophysical Journal, 2013, 104(2): 663~664.

6 Kiełczyński P, Szalewski M, Balcerzak A, et al. Investigation of high-pressure phase transitions in DAG (diacylglycerol) oil using the Bleustein - Gulyaev ultrasonic wave method[J]. Food Research International, 2012, 49(1): 60~64.

7 胡永涛,刘钟栋,杨菁. 单甘酯、甘二酯高纯品的生产、理化性质及特殊用途[J]. 中国食品添加剂,2009(1):57~64.

8 于济洋,李新华,赵前程,等. 甘二酯制备的研究进展[J]. 中国油脂,2007,32(11):12~15.

9 魏明,赵世光,钱森和. 超声辅助水酶法提取米糠油的研究[J]. 中国油脂,2014,39(10):6~9.

10 欧阳建勋. 米糠油资源开发应用探讨[J]. 粮食科技与经济, 2011,36(3):24~26.

11 钟南京,李琳,李冰,等. 甘油酯的液相色谱分析[J]. 现代食品科技,2012,28(1):123~126.

12 胡士恒. 玉米油酶法合成甘油二酯研究[D]. 长春:吉林大学, 2013.

13 郭希娟,马萍,张桂芳. 响应面法在南瓜多糖超声提取中的应用 [J]. 粮油加工,2010(8):143~146.

14 魏金霞. 响应面法优化溶剂体系下大豆油甘油解制备甘二酯的研究[J]. 农业机械,2011(8):46~48.

(上接第 75 页)

个浓度体系的参比样。采用 Spearman 检验后的 r_s 均值法,能够客观评估不同评价员在多次排序中所体现的正确性。由于酸味的后味相对难以祛除,评价员自身对甜味排序的正确性普遍要好于酸味,不过甜味排序表现欠佳的评价员人数要多于酸味。同时通过 Kendall 检验能够评估评价员自身多次排序试验之间的重复性,整体上 W 值与正确性结果之间体现了正相关性的趋势,但 W 值更加精细的区分了评价员之间的细致变化,特别是对 r_s 平均值相同的评价员能够区分相互间的重复表现,提高了评价员能力的区分精度。针对试验管理者感官分析师所要求的既正确又稳定的评价员综合能力,采用聚类分析法可直接将评价员分成不同等级类别,甜味排序能力可分成 4 档,酸味排序能力可分成 3 档,体现了很强的实用性和科学性,为感官分析实验室的评价员筛选和管理提供了可靠的方法。

一种科学的评价员性能评估方法,需要依托于适宜的参比样体系。因为样品评价难度过高,远远超过评价小组团队的差别阈值,使得绝大多数评价员不能正确评价,也就失去了评估和筛选的意义;而样品评价难度过低,所有评价员的每次重复结果都最佳,也起不到筛选更优秀评价员的目的。同时,感官分析结果是评价小组最终的结果,而非受某个评价员的结果所左右,在统计意义的范围内,评价小组内允许个别评价员某几次表现欠佳,因此,可将筛选出的优秀评价员组建成优秀的评价小组队伍。而单独对于评价小组性能

评估还需要进一步深入探讨,包括评价员性能与评价小组之间的关系、不同评价小组之间的性能对比等。

参考文献

1 赵镭,刘文. 感官分析技术应用指南[M]. 北京:中国轻工业出版社,2011.

2 史波林,赵镭,汪厚银,等. 感官分析评价小组及成员表现评估技术动态分析[J]. 食品科学,2014, 35(8): 29~35.

3 Jian Bi, Kuesten C. Intraclass Correlation Coefficient (ICC): A Framework for monitoring and assessing performance of trained sensory panels and panelists[J]. Journal of Sensory Studies, 2012(27): 352~364.

4 郑日昌,孙大强. 心理测量与测验[M]. 北京:中国人民大学出版社,2008.

5 中国合格评定国家认可中心, 国家认证认可监督管理委员会, 中国计量科学研究院,等. GB/T 16291.1—2012 感官分析选拔、培训与管理评价员一般导则第 1 部分:优选评价员[S]. 北京:中国标准出版社,2012.

6 甘怡群,张轶文,邹玲. 心理与行为科学统计[M]. 北京:北京大学出版社,2009.

7 鲍贵. 评估者之间信度分析:从理论到实践[J]. 外语电化教学, 2010(132): 21~25.

8 李春林,陈旭红. 应用多元统计分析[M]. 北京:清华大学出版社,2013.

9 王静龙. 非参数统计分析[M]. 北京:高等教育出版社,2012.