

利用决策树模型探讨小曲白酒质量特征成分

钟其顶,田亚琼,熊正河

(中国食品发酵工业研究院,北京 100027)

摘要: 主要应用分类树、Bagging 两种基于决策树的方法对小曲白酒与其他香型白酒的特征风味物质进行验证对比分析,得出小曲白酒明显区别于其他香型白酒。并在此基础上,探讨小曲白酒特征成分指标,发现小曲白酒特征指标主要体现在乙酸乙酯、正戊醇和异戊醇等成分上,在制订小曲白酒国家标准时可参考。同时,结合实际,证实了利用数据处理模型开展特征分析的可行性和科学性。

关键词: 小曲白酒; 风格; 分类; 数据处理

中图分类号:TS262.3;TS261.4;TP27

文献标识码:A

文章编号:1001-9286(2010)08-0028-04

Analysis of Characteristic Flavoring Compositions of Xiaoqu Liquor by Use of Decision Tree Modeling

ZHONG Qi-ding, TIAN Ya-qiong and XIONG Zheng-he

(China Food Fermentation Industry Research Institute, Beijing 100027, China)

Abstract: Comparative analysis of characteristic flavoring compositions between Xiaoqu liquor and liquor of other flavor types was carried out by use of decision tree modeling-based methods including classification and bagging. The results showed that there was significant difference between Xiaoqu liquor and liquor of other flavor types. On that basis, the characteristic compositions indexes of Xiaoqu liquor were investigated and it was found that the characteristic indexes of Xiaoqu liquor were mainly embodied in ethyl acetate, n-pentanol and isopentanol etc., which could be used as reference for the formulation of national standards of Xiaoqu liquor. Moreover, the feasibility of using data processing models to analyse characteristic indexes was proved in this paper. (Tran. by YUE Yang)

Key words: Xiaoqu liquor; style; classification; data processing

小曲白酒在我国发展历史悠久,是具有传统特色的一个酒种,主要分布在重庆、四川、贵州、云南及两湖、两广地区。小曲白酒主要以粮谷为原料、小曲为糖化发酵剂,经蒸煮、培菌糖化、续糟发酵、蒸馏、陈酿、勾兑而成。与传统大曲香型白酒相比,小曲白酒独具特色:用曲量少,出酒率高,发酵周期短,其醇香清雅、糟香突出、回甜、协调,余味清爽。经过多年发展,小曲白酒在产品结构、生产工艺、技术装备等各个方面都取得了长足进步,并呈现规模化生产。随着国际交流扩大,人们饮酒习惯正在向淡雅风味转变,小曲酒口感风格正好符合这一发展趋势,饮用人群逐渐扩大,全国小曲白酒年产量已达到 300 万 t,成为中国白酒不可忽视的后起之秀。

在我国目前的白酒国家标准体系中,产品质量标准以大曲为发酵剂各类香型白酒标准为主,尚缺乏小曲白酒国家标准或行业标准,小曲白酒生产和销售主要执行地方标准和企业标准。经中国酿酒工业协会、全国食品工业标准化技术委员会酿酒分技术委员会推荐,国家标

准化管理委员会批准了《小曲白酒》国家标准计划(项目编号为:20081939—T-607),由此,开展小曲白酒国家标准研究与制订工作。本文围绕确定小曲白酒风味特征物质,利用数据处理手段,开展系列探索性的分析与研究。

1 研究目标

小曲白酒以清爽淡雅口味著称,与清香型白酒和米香型白酒较相似,南方地区常称其为小曲清香型白酒。本文测定不同产地、不同企业、不同质量等级的小曲白酒的风味成分,对比其他香型白酒,其中重点研究小曲白酒与清香型白酒、米香型白酒之间的差异性指标,通过数据处理手段,找出小曲白酒特征风味指标,为制订《小曲白酒》国家标准提供技术支撑。具体目标:

- ①小曲白酒与其他香型白酒是否具有明显特征差异;
- ②小曲白酒与清香型白酒、米香型白酒是否实现有效区分;
- ③清香型白酒的特征成分乙酸乙酯或乳酸乙酯是否

收稿日期:2010-05-07

能反映小曲白酒的质量特征。

2 研究数据与方法

2.1 研究数据

①全国食品发酵标准化中心征集不同生产企业 23 个小曲白酒样品,采用 GC-FID 测得风味数据。

②全国食品发酵标准化中心积累的浓香型、清香型、凤香型、米香型和芝麻香型白酒数据,挑选近 5 年的 250 个制标数据,其中清香型白酒数据包括了部分小曲白酒。

③考虑到标准的可操作性,根据白酒微量香气成分含量多少,主要考虑选择醇类和酯类物质作为特征风味物质研究,风味物质包括甲酸乙酯、乙酸乙酯、丁酸乙酯、己酸乙酯、乳酸乙酯、正丙醇、异丁醇、正丁醇、异戊醇、正戊醇等 10 种。

2.2 研究方法

随着现代数据处理技术的发展,数据统计分类方法在很多领域都有着广泛的应用。在以往白酒研究中,多次提到 LDA、PCR、Cluster 等统计分类算法,但是这些研究往往基于正态分布假设,存在假定过强的问题。因此,引入非参数统计将是有助于确定小曲白酒特征物质。

2.2.1 决策树

决策树的构造方法最先源自 Hunt's Algorithm^[1]。所谓的决策树就像一棵倒长的树,分根节点、叶节点、终节点。它把特征空间划分为一系列的矩形区域,然后在每一个区域中拟合一个简单的模型^[2]。目前,主要的算法有 ID3、C5.0 和 CART。

决策树有很多优点^[3]。首先,它的图形输出能够直观地展示各种可能的结果及其误判的频率,这种展示帮助人们有效地做出决策。其次,它的效率较高,能够很快清楚地表示各种可能的决策。再次,而且当获得新的信息时,可以较容易修正决策树。虽然决策树很容易计算,但有必要设定不使其过分生长的停止规则或者修剪方法。缺点是每次分叉只和前一次分叉有关,而且不考虑对以后的影响。

构造树的方法如下:寻找一个变量使得依照该变量的某个条件把数据分成纯度较大的 2 个或几个数据子集,而用其他变量所划分的子集不如该变量划分得那样纯。也就是说,根据某种局部最优性来选择变量,而非纯度的度量主要采用以下 3 种方法^[2]:

$$\textcircled{1} \text{熵(entropy): } I = - \sum_{i=1}^K p(i|t) \log_2 p(i|t)$$

$$\textcircled{2} \text{Gini 指数(Gini index): } \sum_{i=1}^K p(i|t) \{1 - p(i|t)\}$$

$$= 1 - \sum_{i=1}^K \{p(i|t)\}^2$$

③误分率 (misclassification rate 或 classification error): $1 - \max_i \{p(i|t)\}$

其中, $y_1, \dots, y_k$, 表示 k 种类别的标记,用 $p(i|t)$ 表示在第 t 个节点处的样本点中属于第 i 类的比例。当有了非纯度的度量,我们就可选择使得非纯度下降最快的树。

2.2.2 Bagging 法

Bagging(bootstrap aggregating, 一般中译为自助整合法或袋装法)是由 Breiman(1996)提出的一种集成学习方法,通过采用有放回随机取样技术(Bootstrap 取样)获取训练集,然后使用该数据集生成集成学习中的个体。在这种方法中,集成个体间的差异性是通过 Bootstrap 重取样技术获得的,即它是通过训练数据的随机性及独立性来提供集成中个体的差异性,该方法主要用于不稳定的学习算法,这里的不稳定是指当训练集中数据有微小的变化时,会导致模型有很大的变化,例如决策树。为了减少个体所产生的方差,通过对这些个体的预测值投票,从而减少了泛化误差。

它主要是对决策树中的训练样本进行多次放回抽样,然后,对每个样本生成一个决策树,这样,每棵树都对一个新的观测值产生一个预测。然后由这些树的分类结果的多数据(“投票”)产生 Bagging 的分类^[3]。

3 结果与讨论

为了能够准确地判别小曲白酒特征风味指标,本文分别采用决策树和 Bagging 两类非参数算法构建酒类分类模型,再代入模型进行仿真,根据仿真结果进行结果分析与讨论。

3.1 小曲白酒与典型香型白酒比较

将浓香型、清香型、凤香型、米香型和芝麻香型白酒 250 个普查数据标记为 1,其中 23 个小曲白酒普查数据标记为 2,采用决策树进行分析,构建树见图 1,仿真结果见表 1。

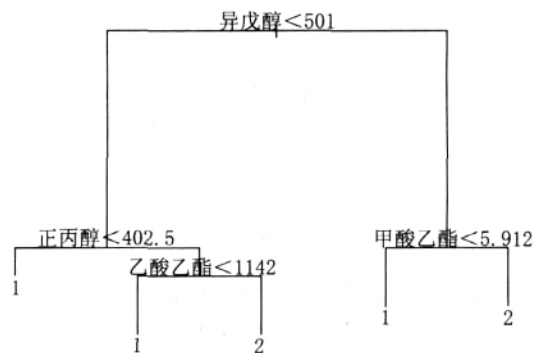


图 1 小曲白酒与典型香型白酒构建决策树

由表 1 可知,典型香型白酒(浓香型、清香型、凤香

表1 决策树模型仿真小曲白酒与典型香型白酒
获得预测结果

预测	标记	
	1	2
1	247	3
2	3	20

型、米香型等)和小曲白酒在决策树模型下,预测准确率可达到 97.9 %,误分率为 2.1 %,说明小曲白酒与我国典型香型白酒具有明显差异,从数据上阐明了小曲白酒独具一帜,因此提出并制订《小曲白酒》国家标准,具有科学意义。此外,从图 1 可知,异戊醇、正丙醇、甲酸乙酯和乙酸乙酯是区分小曲白酒的主要风味指标,小曲白酒中醇类(异戊醇,正丙醇)和酯类(甲酸乙酯,乙酸乙酯)明显偏高,与其小曲白酒发酵温度高、周期短工艺引起的醇高酯高特点相符,如实反映了小曲白酒的工艺风格与特征。

3.2 小曲白酒与清香型、米香型白酒比较

为了进一步考察小曲白酒与清香型和米香型白酒之间的差异,选择清香型、米香型白酒和小曲白酒 84 个普查数据进行分析,其中米香型白酒数据 13 个,标记为 1;清香型白酒数据 48 个,标记为 2;小曲白酒数据 23 个,标记为 3。

3.2.1 决策树

采用决策树进行分析,构建树见图 2,仿真结果见表2。

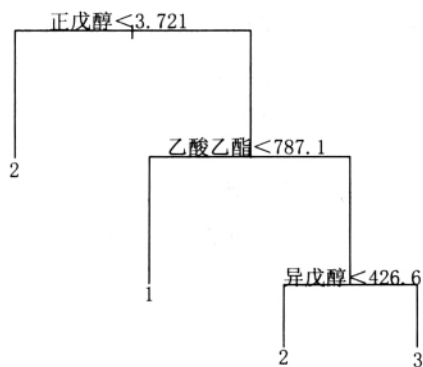


图2 米香型白酒、清香型白酒和小曲白酒构建决策树

表2 采用决策树模型仿真预测结果

预测	1 米香型	2 清香型	3 小曲白酒
1 米香型	12	1	0
2 清香型	1	46	3
3 小曲白酒	0	1	20

由表 2 可知,米香型、清香型白酒和小曲白酒在决策树模型下,预测准确率可达到 93 %,其中 1 个米香型白酒错分为清香型白酒,1 个清香型白酒错分为米香型白酒,1 个清香型白酒错分为小曲白酒,3 个小曲白酒错分为清香型白酒。经溯源发现,错分白酒质量信息见

表3。

表3 错分白酒样品质量信息

编号	原分类	预测分类	酒样信息
13	1	2	三花酒(企标)
23	2	3	高粱酒
50	2	3	特制江津酒
62	3	2	劲牌毛铺纯谷酒
63	3	2	劲牌毛铺纯谷酒
64	3	2	劲牌毛铺老酒

3.2.2 Bagging

为了进一步提高米香型白酒、清香型白酒和小曲白酒的预测分类的准确性,考虑采用精度更高的 Bagging 分类技术构建模型,预测结果见表 4。

表4 采用 Bagging 模型仿真预测结果

预测	1 米香型	2 清香型	3 小曲白酒
1 米香型	12	0	0
2 清香型	1	47	2
3 小曲白酒	0	1	21

由表 4 可知,米香型白酒、清香型白酒和小曲白酒在 Bagging 模型下,预测准确率可达到 95.3 %,其中 1 个米香型白酒错分为清香型白酒,1 个清香型白酒错分为小曲白酒,2 个小曲白酒错分为清香型白酒。经溯源发现,错分白酒质量信息见表 5。

表5 错分白酒样品质量信息

编号	原分类	预测分类	酒样信息
13	1	2	三花酒(企标)
50	2	3	特制江津酒
62	3	2	劲牌毛铺纯谷酒
63	3	2	劲牌毛铺纯谷酒

从错分样品信息看出,桂林三花酒厂执行企标生产的米香型白酒被错分为清香型,重庆江津酒厂生产的特制清香型白酒错分为小曲白酒,劲牌酒厂生产的毛铺纯谷酒错分为清香型白酒。分析发现,江津酒厂产品本属于小曲白酒工艺,劲牌毛铺亦采用清香工艺发酵,这说明白酒生产工艺与香型之间存在着密切的关系,也从侧面证实了数据模型在白酒分类使用的可行性和科学性。

为了进一步考察小曲白酒的特征指标,采用 Bagging 模型将其特征指标按分类贡献度大小进行排列,具体见图 3。

从图 3 可知,小曲白酒与米香型、清香型白酒在乙酸乙酯、高级醇(正戊醇和异戊醇)方面具有明显特征,建议在小曲白酒国家标准特征指标中可作为参考依据。

4 结论与展望

通过采用分类树、Bagging 两种基于决策树的方法对小曲白酒与其他香型白酒构建分类模型,得到如下几

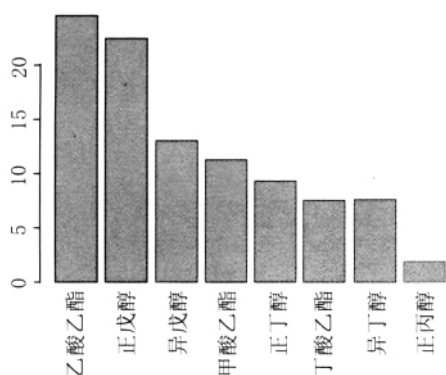


图3 Bagging 分类模型中指标贡献度排列

点结论:

- 4.1 小曲白酒明显区别于其他香型白酒, 风格独具一帜;
- 4.2 小曲白酒不属于清香型白酒和米香型白酒, 制订小曲白酒国家标准不仅有现实意义, 也具有科学依据;
- 4.3 小曲白酒特征指标集中体现在乙酸乙酯、正戊醇和异戊醇等重要指标, 在制订小曲白酒国家标准时可参考。

(上接第27页)

- 3.2 经液相色谱分析, 米浆水中含有大量有机酸, 其中乳酸占的比例比较高, 约占酸类物质浓度的 50% 以上, 并且添加了米浆水后, 发酵时各类有机酸浓度提高, 相应促进了各类酯类物质浓度的提高。
- 3.3 添加米浆水后, 不仅总酸含量提高, 氨基酸态氮含量也有了较大的增加, 对于提高黄酒质量有一定的帮助。此外, 酒精度等指标也有略微的提高。
- 3.4 通过对氨基酸含量变化的分析, 发现添加米浆水, 在一定程度上增加了发酵醪中的氨基酸含量, 不仅使最终的氨基酸含量增加, 还间接促进了 β -苯乙醇含量的提高。

本次研究普查的数据尚有一定的局限性, 在制订小曲白酒国家标准过程中, 还应采集更多数据, 进行更为周密的科学研究, 在此基础上, 完成国家标准的制订。

致谢 本研究得到科技部国际科技合作(2008DFA32880)项目“我国重要食品真实性判定前沿技术标准研究”和中国食品发酵工业研究院青年资金“数据驱动酿造质量管理标准体系”项目资助, 在此表示感谢。

参考文献:

- [1] Hunt E.B., Marin J. and Stone P.J., Experiments in Induction [M]. New York, Academic Press, 1966.
- [2] Trevor Hastie, Robert Tibshirani, Jerome Friedman. The Elements of Statistical Learning: Data Mining, Inference, and Prediction [M]. Springer, 2001.
- [3] 吴喜之. 统计学: 从概念到数据分析 [M]. 北京: 高等教育出版社, 2008.
- [4] Breiman, Leo., J.H. Friedman. Classification and Regression Trees [M]. Monterey, 1984.

参考文献:

- [1] 杨百荣. 绍兴黄酒生产中米浆酸的控制技术 [J]. 中国酿造, 2002, (5): 32.
- [2] 毛青钟. 黄酒浸米浆水及其微生物变化和作用 [J]. 酿酒科技, 2004, (3): 73-76.
- [3] 鲁玉龙, 祁华宝. 黄酒酿制米浆废水的处理 [J]. 工业用水与废水, 2002, (2): 50-51.
- [4] 俞卫华. 米浆水循环利用处理方法的研究 [D]. 杭州: 浙江工业大学, 2002.
- [5] 傅金泉. 黄酒生产技术 [M]. 北京: 中国轻工业出版社, 2005. 317-373.
- [6] 薛洁, 王昇静, 刘岩, 孟世靖. 黄酒酿造中米浆水回用技术的研究 [J]. 酿酒科技, 2009, (9): 17-19.

河南开展白酒品酒师国家资格鉴定

本刊讯 为配合 2010 届国家白酒评酒委员换届, 7 月 9~12 日, 河南省酒业协会、中国酿酒工业协会职业技能鉴定站, 在郑州进行河南省白酒《品酒师》国家职业资格培训、鉴定。来自河南各地的 130 多名学员参加了白酒品酒师考试和技能鉴定。

中国酿酒工业协会酒精分会副秘书长张国红介绍, 品酒师系列职业技能鉴定今年首次举办, 以后将每 3 年举办 1 次。

参加此次鉴定的学员要通过理论考试和实践操作两部分测试, 其中理论考试占总成绩的 20%, 实践操作共分 8 轮进行, 占总成绩的 80%。

据悉, 本次品酒师职业技能培训和鉴定受到河南省各白酒企业的高度重视, 宋河、宝丰、仰韶、张弓、赊店、林河、君利、杜康、富平春、皇沟等重点白酒企业均派出技术骨干参加考试, 河南君利酒业有限公司总经理杨朝阳等企业负责人也参加了培训和考试。此外, 参加本次考试的 130 多名学员中, 有 40 多名女学员。(小小)