



R在食品标准研制中应用与展望

钟其顶

中国食品发酵工业研究院
全国食品发酵标准化中心
2009.12



一、目录

1

食品工业与食品标准

2

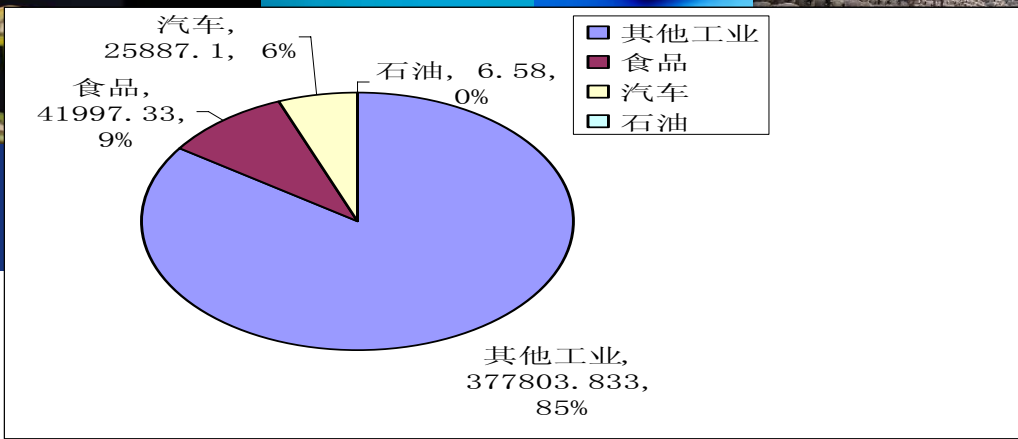
R在食品标准研究中的应用实例

3

R食品标准数据信息平台设想



食品产业概况



什么是食品产业?

- 蔬菜出口、罐头出口第一
- 肉类产量第一
- 牛奶产量第一
- 啤酒产量将近4000万吨，连续四年居世界第一
- 味精总产量近170万吨，居世界第一位
- 柠檬酸产量居世界第一

食品产业是国民经济第一大支柱产业

“民以食为天”，关注国家安全，国民身体素质



食品标准



ICS 67.160.10
X 41



中华人民共和国国家标准

GB/T 10781.1—2006
代替 GB/T 10781.1—1989, GB/T 11856.1—1989

浓香型白酒

Strong flavour Chinese spirits

2006-07-19 发布

2007 05 01 实施



中华人民共和国国家质量监督检验检疫总局 发布
中国国家标准化管理委员会

GB/T 10781.1—2006

表 2 低度酒感官要求

| 项 目 | 优 级 | 一 级 |
|-------|-------------------------|------------------|
| 色泽和外观 | 无色或微黄, 清亮透明, 无悬浮物, 无沉淀* | |
| 香 气 | 具有较浓郁的己酸乙酯为主体的复合香气 | 具有己酸乙酯为主体的复合香气 |
| 口 味 | 绵甜醇郁, 诸味协调, 余味较长 | 绵甜醇郁, 诸味协调, 诸味协调 |
| 风 味 | 具有本品典型的风味 | 具有本品典型的风味 |

* 当酒的温度低于 10℃ 时, 允许出现白色絮状物, 加热或失光, 10℃ 以上时应逐渐恢复正常。

5.2 理化要求

高度酒、低度酒的理化要求应分别符合表 3、表 4 的规定。

表 3 高度酒理化要求

| 项 目 | 比 值 | |
|-----------------|-----------|-------------------|
| | 优 级 | 一 级 |
| 酒精度(40 vol) | 43—48 | |
| 总酸(以乙酸计)(g/L) | ≥ 0.40 | 0.30 |
| 总酯(以乙酸乙酯计)(g/L) | ≥ 2.00 | 1.50 |
| 己酸乙酯(g/L) | 1.00—2.00 | ≥ 0.80—1.50 |
| 总酯量(g/L) | ≤ 0.40 | 0.40 ^a |

^a 酒精度 41 vol—49 vol 的酒, 总酯量可小于或等于 0.30 g/L。

表 4 低度酒理化要求

| 项 目 | 比 值 | |
|-----------------|-----------|-----------|
| | 优 级 | 一 级 |
| 酒精度(20 vol) | 25—40 | |
| 总酸(以乙酸计)(g/L) | ≥ 0.30 | 0.25 |
| 总酯(以乙酸乙酯计)(g/L) | ≥ 1.50 | 1.00 |
| 己酸乙酯(g/L) | 0.70—2.00 | 0.40—2.00 |
| 总酯量(g/L) | ≤ 0.70 | 0.70 |

5.3 卫生要求

应符合 GB 2757 的规定。

5.4 净含量

按国家质量监督检验检疫总局[2005]第 75 号令执行。

6 检验方法

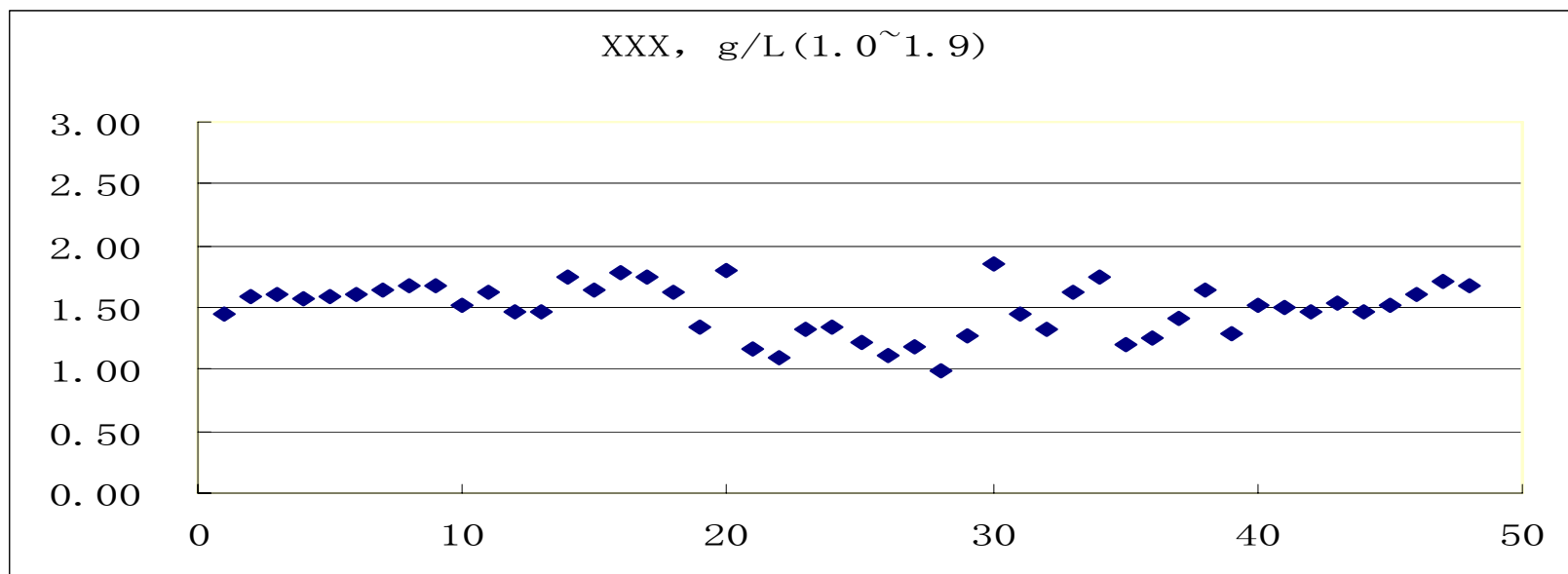
感官要求、理化要求的检验按 GB/T 10345 执行。

净含量的检验按 JJF 1070 执行。



标准制定

❖ 收集样品，测定数据，统计分析，制定标准值。





R在标准研究中实际应用案例

- ❖ 标准数据-----仪器检测
- ❖ 评价标准-----R对付数据-无所不能
- ❖ R在小曲白酒特征指标研究应用
- ❖ R在葡萄酒产地识别技术研究应用



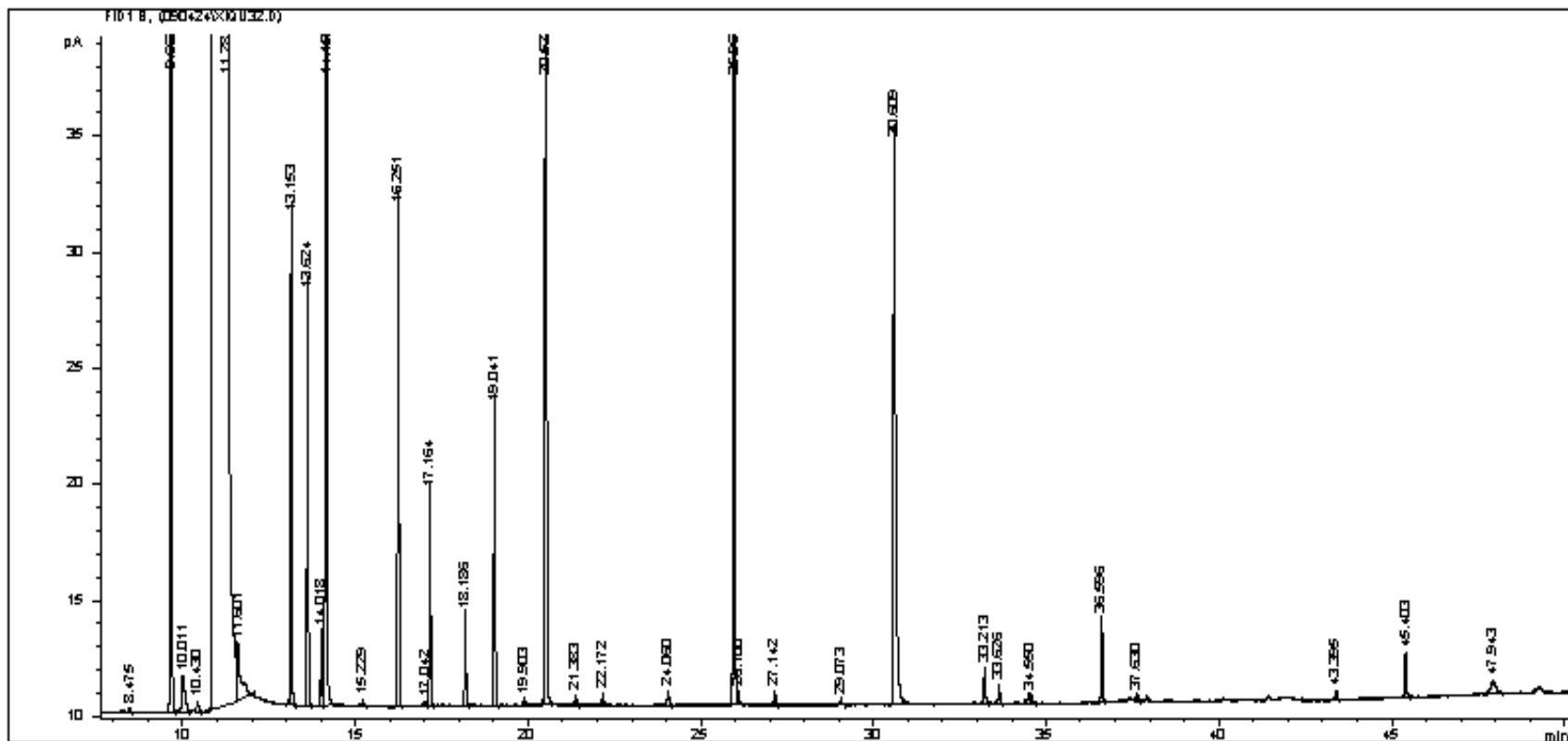
R在小曲白酒特征指标研究应用

❖ 目标:

- 1、找出小曲白酒与几大香型白酒具明显差异质量指标;
- 2、找出小曲白酒与大曲清香白酒风味差异指标;
- 3、乙酸乙酯或乳酸乙酯是否能反映出小曲白酒的质量特征。



GC-FID 色谱图



8.475 甲酸乙酯 · 9.662 乙酸乙酯 · 10.011 甲醇 · 10.430 异戊醛 · 11.284 乙醇 · 13.153 叔戊醇 (内标 1) · 13.624 仲丁醇 · 14.018 丁酸乙酯 · 14.159 丙醇 · 16.251 异丁醇 · 17.164 戊酸乙酯 · 18.186 丁醇 · 19.041 乙酸正戊酯 (内标 2) · 20.523 异戊醇 · 21.383 己酸乙酯 · 22.172 正戊醇 · 24.060 酯噻 · 25.967 乳酸乙酯 · 26.100 己醇 · 30.609 乙酸 · 33.213 丙酸 · 34.550 异丁酸 · 36.596 丁酸 · 43.395 己酸 · 45.403 苯乙醇 · 47.943



数据分析

❖ 数据来源:

其中浓香型、清香型、凤香型、特香型和芝麻香型的白酒数据**250**个，组代号为**1**；

小曲白酒数据**23**个，组代号为**2**；

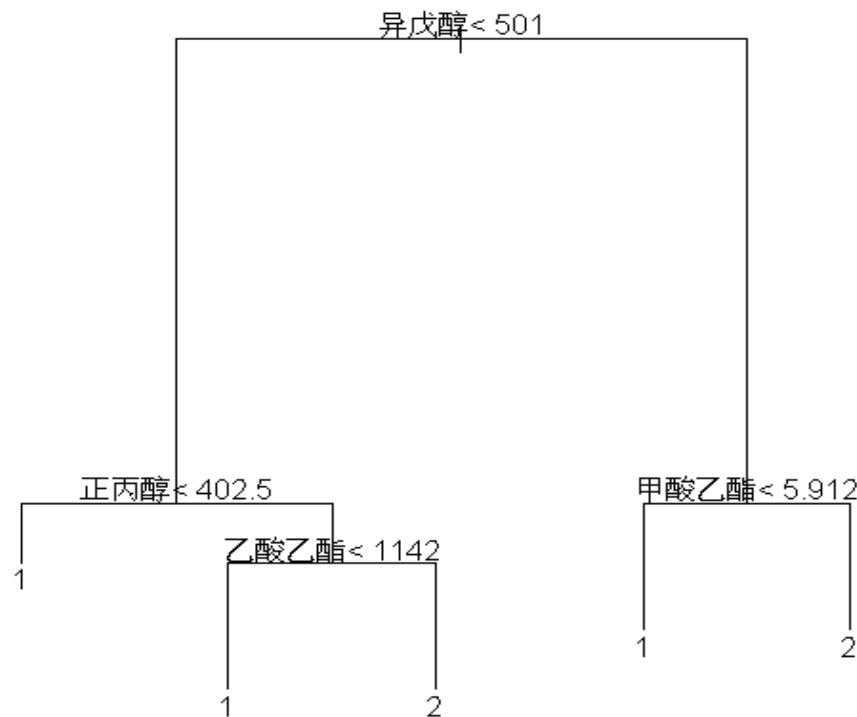
共**273**个样品



数据处理初步分析

- ❖ true
- ❖ pred 1 2
- ❖ 1 247 3
- ❖ 2 3 20

- ❖ 错分率: 2.1%;
- ❖ 准确率: 97.9%





数据分析

❖ 数据来源:

其中浓香型、清香型、凤香型、特香型和芝麻香型的白酒数据**250**个，组代号为**1**；

小曲白酒数据**23**个，组代号为**2**；

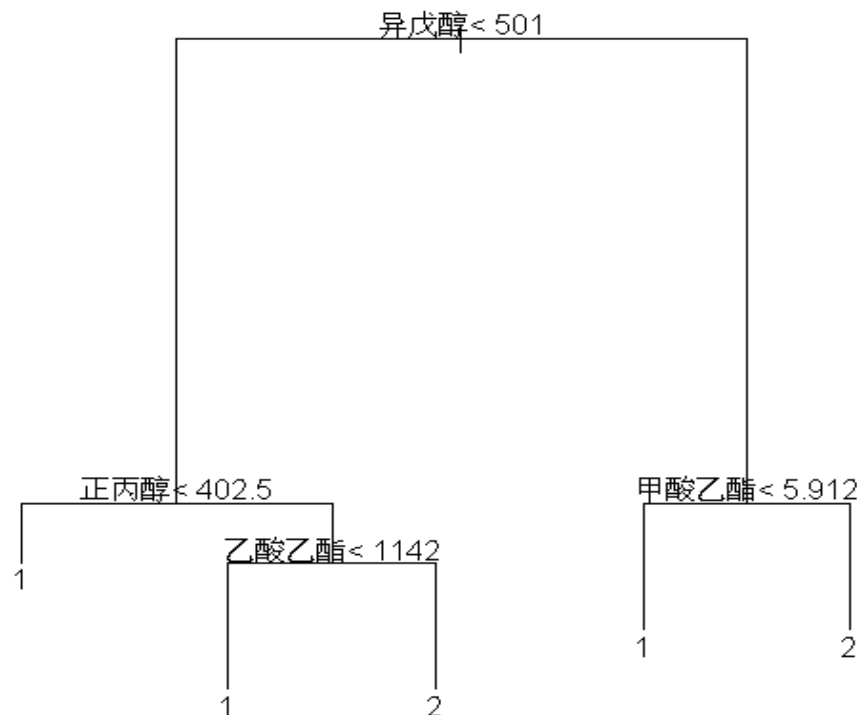
共**273**个样品



数据处理初步分析

- ❖ true
- ❖ pred 1 2
- ❖ 1 247 3
- ❖ 2 3 20

- ❖ 错分率: 2.1%;
- ❖ 准确率: 97.9%





数据处理手段

❖ 决策树

决策树是数据挖掘的一种重要方法。所谓的决策树就像一棵倒长的树，树中有分根节点、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。

优点：

- ❖ (1) 决策树方法不需要假设先验概率分布，使其具有更好的灵活性和鲁棒性。
- ❖ (2) 决策树方法不仅可以利用连续实数或离散的数值样本。
- ❖ (3) 决策树方法产生的决策树或产生式规则集具有结构简单直观，容易理解，以及计算效率高的特点。
- ❖ (4) 决策树方法能够有效的抑制训练样本噪音和解决属性缺失问题。

缺点：

- ❖ (1) 决策树算法在产生规则的时候采用了局部的贪婪方法，每次只选取一个属性进行分析构造决策树，所以它们在产生的分类规则往往相当复杂。
- ❖ (2) 在决策树的学习中，由于分类器过于复杂，则它可能会过于适应噪声，从而导致过度拟合 (**overfit**) 的问题。



Bagging 自助整合法或袋装法

- ❖ **1996年**，由**Breiman**提出的一种集成学习方法，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，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。

在集合多个决策树基础上，提高了预测的准确性，并能找出在多个数据指标属性的重要程度。



特定香型数据分析

❖ 数据来源:

其中米香白酒数据**13**个，组代号为**1**;

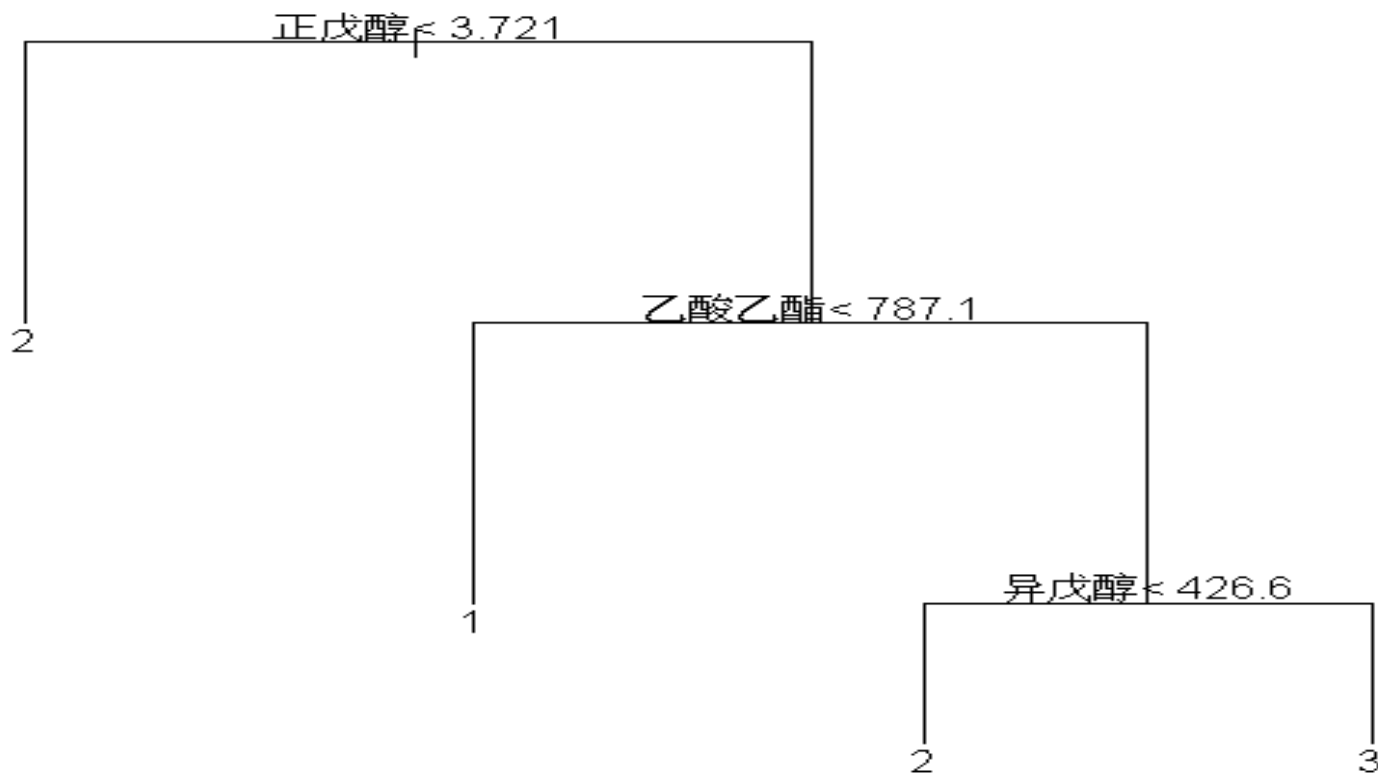
清香白酒数据**48**个，组代号为**2**;

小曲白酒数据**23**个，组代号为**3**;

共**84**个样品



分类树分析（一）





分类树分析

> predclass

```

1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26
1  1  1  1  1  1  1  1  1  1  1  1  2  2  2  2  2  2  2  2  2  2  1  2  2  2
27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52
2  2  2  2  2  2  2  2  2  2  2  2  2  2  2  2  2  2  2  2  2  2  3  2  2
53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78
2  2  2  2  2  2  2  2  2  2  2  3  3  3  3  3  3  3  3  3  3  3  3  3  3
79 80 81 82 83 84
3  3  3  3  3  3
  
```

预测准确率: **93%**;

其中出错**13, 23, 50, 62, 63, 64**

| | | true | | |
|------|---|------|----|----|
| | | 1 | 2 | 3 |
| pred | 1 | 12 | 1 | 0 |
| | 2 | 1 | 46 | 3 |
| | 3 | 0 | 1 | 20 |



Bagging 自助整合法或袋装法

```
[1] "1" "1" "1" "1" "1" "1" "1" "1" "1" "1" "1" "1" "1" "2" "2" "2" "2" "2" "2" "2"
[20] "2" "2" "2" "2" "2" "2" "2" "2" "2" "2" "2" "2" "2" "2" "2" "2" "2" "2" "2"
[39] "2" "2" "2" "2" "2" "2" "2" "2" "2" "2" "2" "2" "3" "2" "2" "2" "2" "2" "2"
[58] "2" "2" "2" "2" "2" "2" "3" "3" "3" "3" "3" "3" "3" "3" "3" "3" "3" "3" "3"
[77] "3" "3" "3" "3" "3" "3" "3" "3"
```

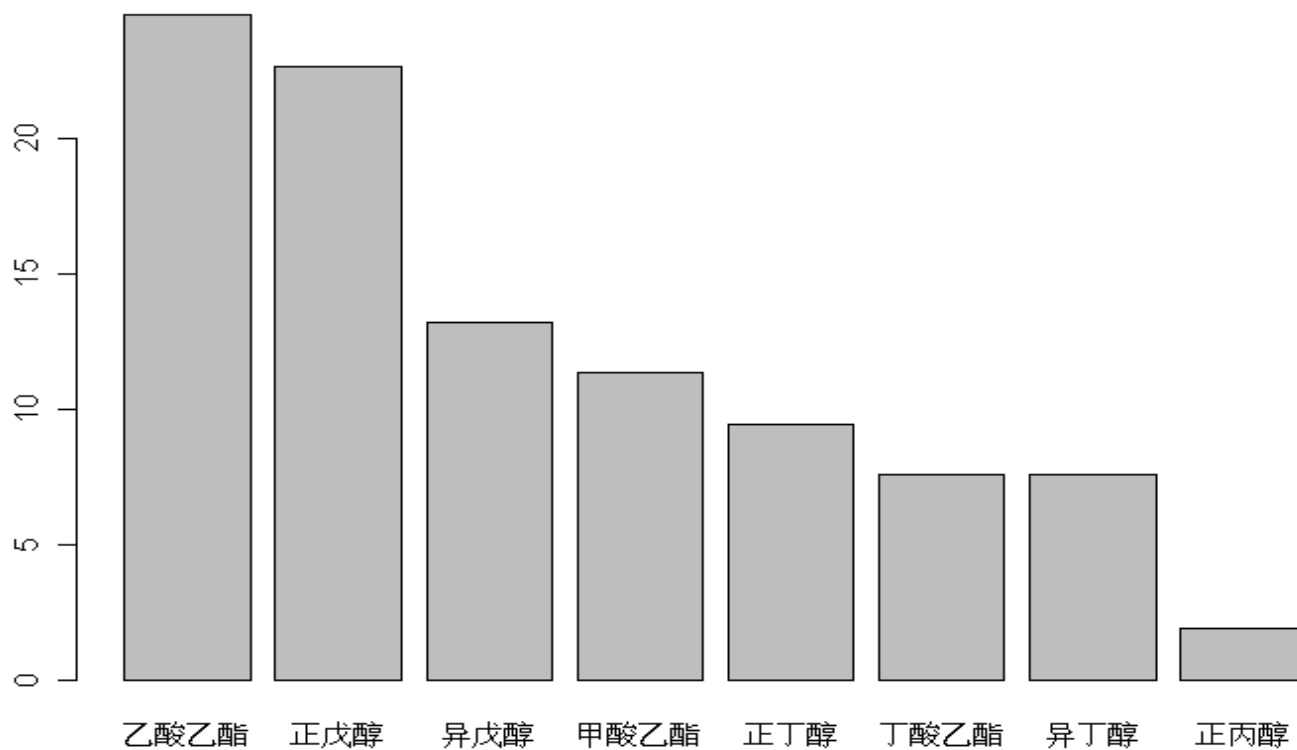
| | true | | |
|------|------|----|----|
| pred | 1 | 2 | 3 |
| 1 | 12 | 0 | 0 |
| 2 | 1 | 47 | 2 |
| 3 | 0 | 1 | 21 |

分类准确率达到: **95.3%**

预测出错样品: **13, 50, 62, 63**



Bagging 自助整合法或袋装法





结论

- ❖ 1、数据分析结果表明，小曲白酒有别于其他香型白酒，具有独特风格，小曲白酒确实具有明显的共性特征；
- ❖ 2、小曲白酒特征性指标排序，乙酸乙酯、正戊醇、异戊醇、甲酸乙酯、正丁醇
- ❖ 3、小曲白酒和米香型白酒，乙酸乙酯差异明显，乳酸乙酯未能作为重要指标。
- ❖ 4、小曲白酒与大曲清香白酒，主要以正戊醇、异戊醇等醇类物质差异明显。



酒类等级、品种和产地技术标准研究与制定

❖ 2006年，新版GB15037-2006葡萄酒颁布实施。首次将品种葡萄酒、产地葡萄酒定义进行了规定，逐步与OIV国际葡萄酒法规

3.3

年份葡萄酒 vintage wines

所标注的年份是指葡萄采摘的年份，其中年份葡萄酒所占比例不低于酒含量的80%(体积分数)。

3.4

品种葡萄酒 varietal wines

用所标注的葡萄品种酿制的酒所占比例不低于酒含量的75%(体积分数)。

3.5

产地葡萄酒 original wines

用所标注的产地葡萄酿制的酒所占比例不低于酒含量的80%(体积分数)。

注：所有产品中均不得添加合成着色剂、甜味剂、香精、增稠剂。



酒类产地识别技术模型构建

- ❖ 采用ICP-AES和ICP-MS测定葡萄酒51种无机元素数据。根据检测限和测量准确度要求，对所检测的无机元素及其同位素数据进行筛选，去除20组含量小于0.1ppb无机元素数据。



样本数据

产地A:10个样本

产地B:6个样本

产地C:6个样本



非参数统计方法

- ❖ 分类树;
- ❖ Bagging;
- ❖ Adaboost
- ❖ 随机森林



各方法判别效果比较—交叉验证

拟合:

| tree | bagging | adaboost | randomforest |
|--------|---------|----------|--------------|
| 0.2667 | 0 | 0 | 0.1333 |

预测:

| | | | | | | | |
|----------------|---|---|---|---|---|---|---|
| 实际产地 | A | A | A | B | B | C | C |
| tree预测 | A | A | A | C | C | B | B |
| bagging预测 | A | A | A | B | B | A | A |
| adaboost预测 | A | A | A | B | B | C | C |
| randomforest预测 | A | A | A | B | B | C | C |



R食品标准研究数据平台设想

- ❖ 生物信息学平台
- ❖ 比如：美国生物技术信息中心（NCBI）
genbank
- ❖ 开发新的数据库搜索和多重序列对齐算法，建立非冗余序列数据库，序列相似性的统计显著性评估的数学模型，和文本检索的矢量模型；采用**BLAST**是一个NCBI开发的序列相似搜索程序，还可作为鉴别基因和遗传特点的手段。**BLAST**能够在小于**15**秒的时间内对整个**DNA**数据库执行序列搜索。
- ❖ **Web-R** 在线分析



R食品标准研究数据平台设想

欧盟联合研究中心 下属酒类办公室



欧洲葡萄酒数据银行

European Office for Wine Alcohol and Spirit Drinks (BEVABS)

Communication (COM(93) 360) from the Commission
European Parliament established the European Office

Spirit Drinks (BEVABS) at the JRC to combat major frauds in the wine area.
This is now extended to a horizontal Public Security and Antifraud activity
and to the development of advanced analytical methods to verify the origin
and authenticity of various commodities.

BEVABS and its partners are working to enhance and promote both quality and labelling. This includes ensuring consistency with EU regulatory systems such as the Protected Designation of Origin (PDO), the Protected Geographical Indication (PGI), the Traditional Speciality Guaranteed (TSG) and organic farming



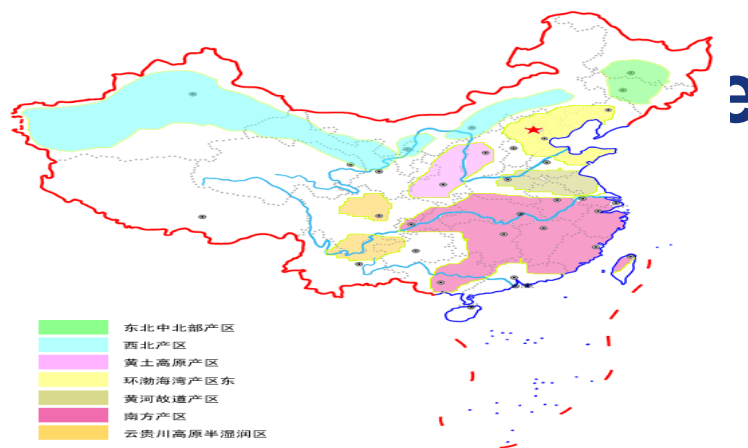
欧盟葡萄酒数据银行(wine c

Authenticating EU wine

The JRC manages a wine databank containing details of thousands of wines produced in the EU since 1991. This includes results of isotopic analyses, as well as more than 80 other parameters relating to geographical origin, year of production, grape type, winemaking process, chemical analysis, soil composition, weather conditions, etc.



中国酒类数据银行 Databank)





食品标准数据信息平台

- ❖ 检测数据社会共享，节约大量科研资源；
- ❖ R数学处理模型网络后台，极大推动数据模型在食品领域运用。
- ❖ 食品技术标准实施与监督可行性，解决市场难以规范技术难题



- ❖ 食品工业对国民经济贡献度很大
- ❖ R在食品领域应用价值很高
- ❖ R在食品领域应用前景很好

衷心希望对食品领域感兴趣同仁，请多多交流与指导



钟其顶 Email:
Zhong.qiding@gmail.com

Thank You !