

模糊聚类分析 在大白菜品质分类上应用

张虹

(黑龙江省农科院园艺研究所·哈尔滨)

提 要

本文以8个大白菜品种(品系)为材料,根据大白菜15个主要品质数量性状进行聚类分析,其分类结果与实际情况相符。初步证明了此种分类方法在实践中是可以应用的。

前 言

六十年代,一些生物学家曾将聚类分析方法应用于生物分类,七十年代以来,已发展为数量分类学,根据一些主要性状资料,将不同的植物分成不同类群。随着人民生活水平的日益提高,蔬菜的品质愈来愈受到人们的重视。育种目标也逐渐从丰产,转到抗病、质优为主要目标。由于蔬菜品种繁多,在目前情况下,对每一个品种均进行评判还有一定的困难,所以有必要将其聚类,进行品质评判较为简单。

本试验试图对8个大白菜品种、品系进行品质分类,其目的在于了解它们之间的数量差异及亲缘关系,作为选配亲本的参考依据。并为进一步大量品种资源的品质分类和模糊数学在蔬菜育种应用作些探索性的试验。

材料和方法

为了便于与实际情况相比较,我们选用了参加区域试验的品种、品系。见表1。

表1 试验材料

序号	原代号	亲本名称	来 源	主 要 特 性
1	CK	牡丹江一号	牡丹江	品质较优良、较抗病
2	78	"	哈尔滨	品质差
3	10	玉 杯	日 本	品质一般、较抗病
4	参照	二 牛 心	哈尔滨	品质优良
5	94	牡丹江一号、5号	佳木斯	品质较优良、较抗病
6	35	4.8自交系、牡丹江一号	哈尔滨	" "
7	18	二 牛 心	"	品质优良
8	04	二 牛 心	"	品质优良

• 本文承蒙李光池、康永春、徐华老师的帮助表示感谢

1985~1986年, 8个品系于我所试验田, 点播、行株距 0.7×0.4m, 行长 10 m, 4 行区、随机区组、三次重复。收获时, 每小区中间两行、拔取 5 株, 于我所化验室进行化验分析, 其项目有: 水分、可溶性固形物、可滴定酸度, 还原型 vc、折光糖、还原性糖、可溶性糖、蔗糖、粗纤维。将结果列于表 2。田间调查叶色、叶柄色、帮叶比、整齐度、净菜重、味道。

表 2 大白菜 8 个品种 15 个性状的化学分析结果和观察值

品 种	X	性 状	原 代 号	水 分 %	可溶性固形物 %	可滴定酸度 %	还原型 vc mg/100g	折 光 糖 %	还原性糖 %	可溶性糖 %	蔗 糖 %	粗 纤维 %	叶 色	叶 柄 色	帮 叶 比	整 齐 度	净 菜 重	味 道
序 号																		
1	CK		93.01	564	0.21	15.92	5.75	2.81	3.05	0.23	0.48	73.3	62.2	0.10	91.4	3.41	92.0	
2	78		93.59	4.65	0.22	16.32	5.00	2.68	2.84	0.15	0.48	80.0	40.0	0.1	02	90.7	2.85	88.2
3	10		93.56	5.46	0.21	19.37	5.80	3.11	3.29	0.17	0.49	71.1	62.2	0.98	92.8	3.86	90.1	
4	参照		93.37	4.96	0.22	19.50	5.25	2.81	3.02	0.20	0.46	48.9	53.3	0.10	90.8	3.00	95.5	
5	94		93.67	4.54	0.22	18.04	4.60	2.50	2.57	0.07	0.49	68.9	62.2	0.95	95.2	3.69	89.7	
6	35		93.90	4.54	0.22	16.65	4.60	2.48	2.64	0.15	0.54	68.9	62.2	0.84	92.7	3.27	91.5	
7	18		92.62	5.60	0.22	18.57	6.00	3.20	3.44	0.23	0.52	33.3	35.6	1.06	81.5	3.39	92.1	
8	04		93.09	4.78	0.22	20.82	5.25	2.93	3.15	0.21	0.55	71.1	62.2	0.99	89.0	3.11	93.0	

根据表 2 的测试数据, 用模糊聚类分析进行分类。模糊聚类分析一般有以下三个步骤

- 1、把各代表点的统计指标的数据标准化, 以便于分析和比较, 我们可以采用如下公式来进行:
- $$x = \frac{x' - \bar{x}'}{C}$$
- 式中: x' ——原始数据
 \bar{x}' ——原始数据的均值
 C ——原始数据的标准差

这里所说的原始数据就是指表 2 的数据。我们把这一工作与下面两步的计算都放在一起进行。

- 2、计算品种之间的相似程度。即根据标准化后的数据, 我们可以选择适当的公式计算出品种之间的相似程度。相似程度的计算方法很多, 在这里我们选取夹角余弦法:

即:

$$r_{ij} = \frac{\sum_{k=1}^m x_{ik} \cdot x_{jk}}{\sqrt{\left[\sum_{k=1}^m x_{ik}^2\right]\left[\sum_{k=1}^m x_{jk}^2\right]}}$$

式中: r_{ij} ——第*i*个品种和第*j*个品种之间的相似程度
 x_{ik} ——第*i*个品种的第*k*个指标标准化后的值

3、聚类
计算出*r_{ij}*之后, 可以形成一个相似矩阵*R*

即*R* =

$$\begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{n1} & r_{n2} & \cdots & r_{nn} \end{bmatrix}$$

此矩阵满足于自反性和对称性,但不一定满足传递性,所以需要对R进行改造,使其同时满足以上三个条件,才可进行聚类。具体方法是对满足以上三个条件R,通过选取一个 λ 可以得到一个 λ 截矩阵 R_λ , R_λ 就决定一种分类。当 λ 取不同值时,可以得到一个动态的聚类过程,当 λ 由大变小时,分类由细变粗。

结果与分析

本文根据以上三个步骤的计算方法,编制了计算机程序,将表2中的数据输入后,其计算结果如下:

当 $0.98 < \lambda \leq 1$ 时,分成8类,即 $\{1\}$ 、 $\{2\}$ 、 $\{3\}$ 、 $\{4\}$ 、 $\{5\}$ 、 $\{6\}$ 、 $\{7\}$ 、 $\{8\}$;

当 $0.972 < \lambda \leq 0.98$ 和 $0.965 < \lambda \leq 0.972$ 时,均分成7类,即: $\{1\}$ 、 $\{2\}$ 、 $\{3\}$ 、 $\{4\}$ 、 $\{5, 6\}$ 、 $\{7\}$ 、 $\{8\}$ 。

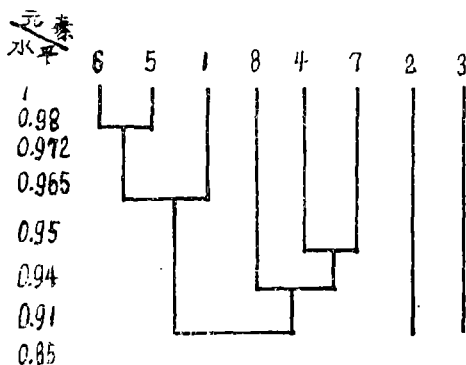
当 $0.95 < \lambda \leq 0.965$ 时,分成6类,即: $\{1, 5, 6\}$ 、 $\{2\}$ 、 $\{3\}$ 、 $\{4\}$ 、 $\{7\}$ 、 $\{8\}$ 。

当 $0.94 < \lambda \leq 0.95$ 时,分成5类,即: $\{1, 5, 6\}$ 、 $\{2\}$ 、 $\{3\}$ 、 $\{4, 7\}$ 、 $\{8\}$ 。

当 $0.91 < \lambda \leq 0.94$ 时,分成4类,即 $\{1, 5, 6\}$ 、 $\{2\}$ 、 $\{4, 7, 8\}$ 、 $\{3\}$ 。

当 $0.85 < \lambda \leq 0.91$ 时,分成3类,即 $\{1, 4, 5, 6, 7, 8\}$ 、 $\{2\}$ 、 $\{3\}$ 。

从整个聚类(分类)过程,使我们可以看到,所研究的大白菜品种在品质上的距离。试将聚类过程划成一个动态聚类图,可以更为直观地观察到品种之间的品质距离求品质相似程度。如图所示



从图中可以看出,选取 $0.91 < \lambda \leq 0.94$ 时,所研究的品种从品质上可分成四类,即 $\{1, 5, 6\}$ 、 $\{4, 7, 8\}$ 、 $\{2\}$ 、 $\{3\}$ 。这一结果与实际比较相符。实际中,1、5、6三个品种(品系)都是牡丹江一号系统的,其中,6是4、8,牡丹江一号三个亲本的杂交后代系选得到的,5是从牡丹江一号和5号品种中选出的品系,1是牡丹江一号品种。2品系虽来源牡丹江一号品种,但经过长期的培育和选择,在品质上已有明显的差异,分属于不同的两类。8、4、7三个品种(品系)都是二牛心系统的,其中8、7来源于二牛心。3是从日本品种玉杯中选出的,故品种间的品质距离较远。

通过以生产上公认的优质二牛心品种,较优质的牡丹江一号品种为代表,进行品质分

类,可将不同品种分为不同的类别。8、4、7是品质最优良的一类,6、5、1是较优良、较抗病的一类。因此,在蔬菜品质育种中,应该选择那些分离距离比较远,品质又较优良的品种,或自交系作为杂交亲本。

结束语

1、大白菜品质方面的性状有许多是模糊的,所以应用模糊聚类进行品质分类比较适宜。由于每个实际分类问题都具有自身的特点,所以合理选择相似程度的计算公式是非常重要的,在具体分类时应结合专家们的实际经验选择夹角余弦法来计算相似程度,其分类结果与我们的实际体验比较相符。

2、在蔬菜品质育种中,优良亲本选配是非常重要的。模糊聚类分析方法不仅可以进行数量分类,而且能为选配优良亲本提供理论依据和新的信息。在原始材料较多的情况下,有必要进行聚类分析,以便更好地了解 and 掌握原始材料,合理的选配优良亲本。(收稿时间为1987年10月18日)参考文献略。

温室表层加热保温效果

本文表层加热指在温室表面上加一薄层流动温水来提高保温性的技术。其效果是既可以把温水的热量传入室内,又能捕获温室向外界大气放出的热量。

试验所用温室宽8.7米,长12.6米,高4.2米,屋顶坡度1:2,南北延长,铝制骨架,屋顶和侧面复以玻璃,山墙为硬质板。温水由水泵输送,四根带孔的管子象图1那样将水分开,外侧两个供水的水流量约为内侧的二倍,管道的水压以1.0米为宜。温室表面积206平方米,表层加热面积167平方米占总面积的81%。

温水的流量必须保持 $0.4\text{L}/\text{S}\cdot\text{m}^2$,为使供水量充足,水泵的动力每平方米温室的表面积为 6.56W ,全温室共用 1100W 。

使用表面加热的结果表明,加热比不加热外界气温要低 1.9°C ,为保证室内恒定温度,就必须使表面加热达到 $6.03\text{W}/\text{m}^2\cdot^\circ\text{C}$,电气取暖设备达到 $9.04\text{W}/\text{m}^2\cdot^\circ\text{C}$ 以提高温度,也就是说用表面加热方式能节省能源33%,如果考虑水泵输出功率的话,也能节省30%。

白天使用表面加热方式就不如夜间表现明显,温室内所有取暖设备的功率在表面加热时要消耗 $2.78\text{W}/\text{m}^2\cdot^\circ\text{C}$ 停下来时消耗 $3.54\text{W}/\text{m}^2\cdot^\circ\text{C}$,因此表面加热的节能率包括水泵输出功率在内为21%,昼夜综合起来的节能率为30%,包括水泵仅为26%。(中国农科院黑龙江省现代化研究所杨丽华译薛彦斌校)

欢迎订阅《农业科学实验》

《农业科学实验》是面向农村、半农半牧区全国发行的科普月刊。

《农业科学实验》是农林牧副渔综合性的科普读物。为满足两户和广大农民发展商品生产的需要,宣传和推广勤劳致富的典型经验,先进可行的种植业、养殖业及经营管理方面的技术和经验,及时提供科技和经济信息,并根据读者喜爱开设栏目。《农业科学实验》所刊稿件,注重实用、新颖。凭本刊总期数(剪下贴在来信右上角)可免费为农牧民订户刊登购销信息(要加盖村委会公章)。《农业科学实验》每月一日出版,每期定价0.26元。本刊代号16—38,全国各地邮电局(所)均可订阅