

22个马铃薯品种(系)的模糊聚类分析

严凤喜 王凤义 陈伊里

(东北农学院)

摘 要

本文应用模糊聚类, 对马铃薯品种(系)进行了分类。聚类过程为: 计算相似矩阵 R ; R 取幂, 至 $R^{2k} = R^k$, 得到模糊等价关系 R^* ; 选取适当的 λ 进行聚类。

$$\text{本文的相似系数 } r_{ij} = \frac{\sum (x_{ik} - \bar{x}_i)(x_{jk} - \bar{x}_j)}{\sqrt{\sum (x_{ik} - \bar{x}_i)^2} \sqrt{\sum (x_{jk} - \bar{x}_j)^2}}$$

聚类结果表明, 普通栽培种与新型栽培种存在明显差异, 说明模糊聚类在马铃薯品种(系)分类中是一种有效的方法

近二十年来, 一些生物学家将模糊数学应用于生物学的许多领域, 取得了令人高兴的研究成果。例如, 利用模糊聚类方法, 对种质资源进行分门别类, 这样能为育种取材提供较为客观的标准。但迄今为止, 马铃薯种质资源的数量分类进展不大, 马铃薯品种

(系)的聚类分析尚未见报道。本文运用模糊聚类方法, 用11个农艺性状对22个马铃薯品种进行了聚类分析, 试图探索马铃薯品种(系)的数量分类, 以便为杂交育种时合理选配亲本提供一定依据, 同时对马铃薯种质资源的数量化研究做些尝试。

to 1984 in order to investigate the effect of whole potato seeds on growth, development and yield of potatoes. The results indicated that the yield of planting whole potato seeds was almost equal to or higher than that of using cut seed pieces, especially in 1980, the dry year, planting whole potato seeds showed high resistance to dry conditions. Moreover, the emergence of whole potato seeds was early and potato plants tended to senesce early. The photosynthetic momentum of plants from whole potato seeds was 109.6% and 182.8% up on that of plants from cut seed pieces in 1981 and 1980 respectively. As to marketability of potato tubers, the average rate of marketable potato tubers in 6 years was 74.1% for using whole potato seeds and 80.0% for planting cut seed pieces if tubers more than 112.5g/tuber were considered as ware potatoes, while if the tubers more than 50g/tuber were regarded as ware potatoes the former was 97.3% and the latter was 98.2%.

材料和方法

供试材料共22个: 普通栽培种12个, 包括从国外引进的Muncel、Aron、Temsea、Dstara、Semenic、Superior以及国内育成的品种(系): 克新1号、东农303、克新4号、东农3042、四川801-5、T8024; 新型栽培种品系9个: NEA10005、NEA10015-13、NEA10015-15、NS12-156-6、AR1-88、AR1-257、AR83-8-7、AR1-172、NEA10037; 一个T8024×m6的无性系(无性3代, 代号是85-26)。

试验于1937年在东北农学院试验站(哈尔滨)实施。试验采用随机区组设计, 小区行长5米, 双行区, 每小区定植40株, 3次重复。调查、测定性状11个: (1) 苗期生长势: 本试验中, 试材的播种期是4月27日(哈尔滨地区的正常播期为4月底左右), 在6月5日目测调查苗期生长势(哈尔滨地区一般在5月底左右出苗), 没有出苗的定为1级, 生长最旺盛、植株最大的定为5级, 居中者分别为2、3、4级。(2) 中期生长势: 按植株大小、繁茂程度, 也分为5级, 植株生长最弱者为1级, 生长最旺盛者为5级。种植84天(7月20日)后调查。(3) 叶形指数(宽/长): 种植85天(7月21日)后调查。(4) 主茎数: 种植100天(8月5日)后调查。(5) 成熟度: 种植115天(8月20日)后调查, 按1~5级, 熟期愈晚, 根据植株地上部分的枯黄程度目测调查。收获后的考种项目: (6) 单株块茎重(g), (7) 单株块茎数, (8) 商品薯率(不小于5.5 cm的块茎重量/总块茎重量), (9) 芽眼深浅(1~4级): 芽眼愈深, 分级指标愈大, (10) 块茎型记(1~5级): 按1~5级, 薯形依次变劣, (11) 块茎比重: 用水中比重法测定。

模糊聚类^[1]的过程分三步进行:

1. 计算相似矩阵 $\tilde{R}$: 相似系数的计算, 原则上可借鉴普通聚类分析计算相似系数的方法, 可有多种选择。本文依据

$$r = \frac{\sum (x_{ik} - \bar{x}_i)(x_{jk} - \bar{x}_j)}{\sqrt{\sum (x_{ik} - \bar{x}_i)^2} \sqrt{\sum (x_{jk} - \bar{x}_j)^2}}$$

计算相似系数, 进而建立模糊相似矩阵 $\tilde{R}$ 。

2. 确立模糊等价关系 $\tilde{R}^*$: 模糊相似矩阵 $\tilde{R}$ 仅具有自反性和对称性, 不满足传递性, 不是模糊等价关系, 一般不能依此聚类, 通过求 $\tilde{R}$ 的幂, $\tilde{R}^2, \tilde{R}^4, \tilde{R}^8, \dots$, 直至 $\tilde{R}^k \stackrel{\Delta}{=} \tilde{R}^*$, 便能得到模糊等价关系 $\tilde{R}^*$ 。

3. 聚类: 选取任一置信水平 λ , 可得到动态聚类图。至此, 聚类过程可视为结束。

由于本试验涉及的品种较多, 数据庞大, 计算复杂, 为了能简单明了起见, 用一虚拟数据说明计算过程。假设用5个品种的6个性状进行聚类分析, 首先是计算相似矩阵, 这步与其他聚类方法类似, 故此处略去原始数据及相似系数的计算, 直接列出模糊相似矩阵。

$$\tilde{R} = \begin{pmatrix} 1 & 0.8 & 0.6 & 0.1 & 0.2 \\ 0.8 & 1 & 0.8 & 0.2 & 0.85 \\ 0.6 & 0.8 & 1 & 0 & 0.9 \\ 0.1 & 0.2 & 0 & 1 & 0.1 \\ 0.2 & 0.85 & 0.9 & 0.1 & 1 \end{pmatrix}$$

显然, $\tilde{R}$ 满足自反性和对称性, 不满足传递性, 故应对其改造。取 $\tilde{R}$ 的幂 $\tilde{R}^2, \tilde{R}^4, \dots$,

计算得 $\tilde{R}^2 = \tilde{R}^4 \stackrel{\Delta}{=} \tilde{R}^*$, 于是模糊等价矩阵为

$$\tilde{R}^* = \begin{pmatrix} 1 & 0.8 & 0.8 & 0.2 & 0.8 \\ 0.8 & 1 & 0.85 & 0.2 & 0.85 \\ 0.8 & 0.85 & 1 & 0.2 & 0.9 \\ 0.2 & 0.2 & 0.2 & 1 & 0.2 \\ 0.8 & 0.85 & 0.9 & 0.2 & 1 \end{pmatrix}$$

取 λ 聚类, 当 $0.85 < \lambda \leq 0.9$ 时, R_{λ}^* 为一普通矩阵

$$R_{\lambda}^* = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \end{pmatrix}$$

根据 R_{λ}^* 5 个品种可分为四类 $\{x_1\}$ 、 $\{x_2\}$ 、 $\{x_3, x_5\}$ 、 $\{x_4\}$ 。依照上述的计算程序, 本文的计算过程在APPLE II 微机上完成。

结果及分析

通过计算多元的相似系数, 得到模糊相似矩阵 $\tilde{R}$, $\tilde{R}$ 是一个 22×22 的矩阵, 由于 $\tilde{R}$ 不满足传递性, 由 $\tilde{R}$ 取幂, 有 $\tilde{R}^3 = \tilde{R}^4 = \tilde{R}^*$, 于是有模糊等价关系 $\tilde{R}^*$ (限于篇幅, 文中略去了 $\tilde{R}$)。

当 $\lambda > 0.3116$ 时, 可得到 $R_{\lambda}^* > 0.3116$ 为一普通矩阵, 同样, 可得 $R_{\lambda}^* > 0.7697$, 文中将 $R_{\lambda}^* > 0.3116$ 和 $R_{\lambda}^* > 0.7697$ 合并一起, 得到表 1。

表 1. 22 个材料的矩阵分类表

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1	1	1	1	1	1	1	1	1	0	*	0	0	0	0	0	0	0	0	0	0	0
1								1	0	*	0										0
1								1	0	*	0										0
1								1	0	*	0										0
1								1	0	*	0										0
1								1	0	*	0										0
1								1	0	*	0										0
1								1	0	*	0										0
1								1	0	*	0										0
1	1	1	1	1	1	1	1	1	0	*	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	1	*	1	1	1	1	1	1	1	1	1	1
*	*	*	*	*	*	*	*	*	*	1	*	*	*	*	*	*	*	*	*	*	*
0	0	0	0	0	0	0	0	0	0	1	*	1	1	1	1	1	1	1	1	1	1
0								0	1	*	1										1
0								0	1	*	1										1
0								0	1	*	1										1
0								0	1	*	1										1
0								0	1	*	1										1
0								0	1	*	1										1
0								0	1	*	1										1
0								0	1	*	1										1
0								0	1	*	1										1
0	0	0	0	0	0	0	0	0	0	1	*	1	1	1	1	1	1	1	1	1	1

$\sim R^* =$

1	.8875	.8888	.8514	.7865	.7865	.8888	.8721	.8743	.7697
.8875	1	.8875	.8514	.7865	.7865	.8875	.8721	.8743	.7697
.8888	.8875	1	.8514	.7865	.7865	.8937	.8721	.8743	.7697
.8514	.8514	.8514	1	.7865	.7865	.8514	.8514	.8514	.7697
.7865	.7865	.7865	.7865	1	.8537	.7865	.7865	.7865	.7697
.7865	.7865	.7865	.7865	.8537	1	.7865	.7865	.7865	.7524
.8888	.8875	.8937	.8514	.7865	.7865	1	.8721	.8743	.7697
.8721	.8721	.8721	.8514	.7865	.7865	.8721	1	.8721	.7697
.8743	.8743	.8743	.8514	.7865	.7865	.8743	.8721	1	.7697
.7697	.7697	.7697	.7697	.7697	.7524	.7697	.7697	.7697	1
.3116	.3116	.3116	.3116	.3116	.3116	.3116	.3116	.3116	.3116
.7697	.7697	.7697	.7697	.7697	.7616	.7697	.7697	.7697	.8227
.7697	.7697	.7697	.7697	.7697	.7621	.7697	.7697	.7697	.8255
.7697	.7697	.7697	.7697	.7697	.7524	.7697	.7697	.7697	.8095
.7697	.7697	.7697	.7697	.7697	.7524	.7697	.7697	.7697	.8095
.7697	.7697	.7697	.7697	.7697	.7697	.7697	.7697	.7697	.8255
.7697	.7697	.7697	.7697	.7697	.7524	.7697	.7697	.7697	.8341
.7697	.7697	.7697	.7697	.7697	.7697	.7697	.7697	.7697	.8255
.7697	.7697	.7697	.7697	.7697	.7697	.7697	.7697	.7697	.8255
.7697	.7697	.7697	.7697	.7697	.7524	.7697	.7697	.7697	.8459
.7697	.7697	.7697	.7697	.7697	.7697	.7697	.7697	.7697	.8255
.7697	.7697	.7697	.7697	.7697	.7697	.7697	.7697	.7697	.8255

.3116	.7697	.7697	.7697	.7697	.7697	.7697	.7697	.7697	.7697	.7697	.7697
.3116	.7697	.7697	.7697	.7697	.7697	.7697	.7697	.7697	.7697	.7697	.7697
.3116	.7697	.7697	.7697	.7697	.7697	.7697	.7697	.7697	.7697	.7697	.7697
.3116	.7697	.7697	.7697	.7697	.7697	.7697	.7697	.7697	.7697	.7697	.7697
.3116	.7697	.7697	.7697	.7697	.7697	.7697	.7697	.7697	.7697	.7697	.7697
.3116	.7616	.7621	.7524	.7524	.7697	.7524	.7697	.7697	.7524	.7697	.7697
.3116	.7697	.7697	.7697	.7697	.7697	.7697	.7697	.7697	.7697	.7697	.7697
.3116	.7697	.7697	.7697	.7697	.7697	.7697	.7697	.7697	.7697	.7697	.7697
.3116	.7697	.7697	.7697	.7697	.7697	.7697	.7697	.7697	.7697	.7697	.7697
.3116	.8227	.8255	.8095	.8095	.8255	.8341	.8255	.8255	.8459	.8255	.8255
1	.3116	.3116	.3116	.3116	.3116	.3116	.3116	.3116	.3116	.3116	.3116
.3116	1	.8787	.8228	.8922	.8922	.8255	.8541	.8541	.8227	.8612	.8541
.3116	.8787	1	.8228	.8787	.8787	.8255	.8541	.8541	.8255	.8612	.8541
.3116	.8228	.8228	1	.8228	.8228	.8077	.8228	.8228	.79	.8228	.8228
.3116	.8922	.8787	.8228	1	.8961	.8255	.833	.8541	.8105	.8612	.8363
.3116	.8922	.8787	.8228	.8361	1	.8255	.8375	.8541	.8255	.8612	.8375
.3116	.8255	.8255	.8077	.8255	.8255	1	.8255	.8255	.8459	.8255	.8255
.3116	.8541	.8541	.8228	.833	.8375	.8255	1	.8997	.8255	.8541	.8607
.3116	.8541	.8541	.8228	.8541	.8541	.8255	.8997	1	.8255	.8541	.8607
.3116	.8227	.8255	.79	.8105	.8255	.8459	.8255	.8255	1	.8255	.8255
.3116	.8612	.8612	.8228	.8612	.8612	.8255	.8541	.8541	.8255	1	.8541
.3116	.8541	.8541	.8228	.8363	.8375	.8255	.8607	.8607	.8255	.8541	1

由 R^* 对22个供试材料进行动态聚类分析, 其结果见图1。

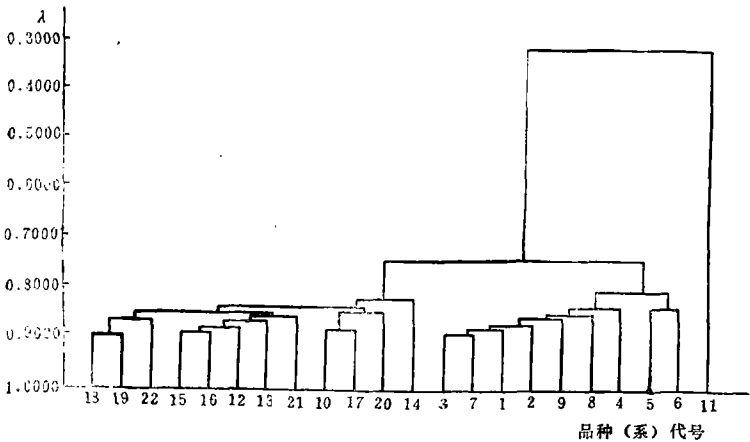


图1. 22个品种(系)的动态聚类图

1—NEA10015-13; 2—NEA10015-13; 3—NS12-156-6; 4—NEA10005;
5—AR1-88; 6—AR1-257; 7—AR83-8-7; 8—AR1-172; 9—
NEA10037-15; 10—Muncel; 11—85-26; 12—Dstara; 13—Aron;
14—Semenic; 15—Superior; 16—Temsea; 17—克新4号; 18—克新1号;
19—东农3042; 20—东农303; 21—川801-5; 22—T8024

从图1可以看出, 当 $\lambda < 0.8105$ 时, 供试材料分为三类。而这三类正好将新型栽培种与普通栽培种以及 $4x \times 2x$ 的后代区别开来, 这说明数量分类的方法是可行的。表2给出了三种类型诸性状的平均值和标准差。

由表2即知, 第一类, 新型栽培种, 其特点有: 叶形指数小 (0.47 ± 0.046), 单株块茎数多 (11.43 ± 2.50), 单株产量较低 (267.21 ± 43.23), 商品薯率低 (0.24 ± 0.052); 第二类, 普通栽培种, 主要特点是

表2. 各种类型的11个性状的平均值和标准差

性 状	新 型 栽 培 种	普 通 栽 培 种	T8024 $\times$ m6的 无性系
	1、2、3、4、5、6、7、8、9、	10、12、13、14、15、16、17、18、 19、20、21、22、	11
苗期生长势	2.2 $\pm$ 0.707	2.3 $\pm$ 0.951	1
中期生长势	3.33 $\pm$ 1.145	3.75 $\pm$ 0.865	3
叶形指数	0.47 $\pm$ 0.046	0.71 $\pm$ 0.071	0.76
主茎数	2.86 $\pm$ 0.357	3.01 $\pm$ 0.651	3
成熟度	3.55 $\pm$ 0.72	2.25 $\pm$ 0.97	3
单株块茎重(g)	267.21 $\pm$ 43.23	448.33 $\pm$ 43.52	121.1
单株块茎数	11.43 $\pm$ 2.51	6.17 $\pm$ 1.998	4.1
商品薯率	0.24 $\pm$ 0.052	0.73 $\pm$ 0.103	0.21
芽眼深浅	2.6 $\pm$ 0.14	1.5 $\pm$ 0.17	3
块茎型记	3 $\pm$ 0.352	2.25 $\pm$ 0.813	1.5
块茎比重	1.0667 $\pm$ 0.013	1.067 $\pm$ 0.008	1.0682

丰产性好, 商品薯率高, 其单株产量为 (448.33 ± 43.52) , 单株块茎数为 6.17 ± 1.908 , 叶形指数为 0.71 ± 0.0703 ; 第三类, 呈 $T3321 \times m6$ 的一个无性系 (无性3代)。

这个无性系块茎数少, 薯形差, 商品薯率很低, 与其它品种 (系) 相比, 在生产上的利用价值很低。这不难解释, 除了向杂交后代转移某些控制特殊优良性状的遗传物质外, $F \times D$ 类型的杂交组合直接被用于生产还有许多有待克服的困难。如何提高二倍体的生产价值, 是国内外马铃薯育种家感兴趣的课题^[5]。

讨 论

就所取11个性状来看, 新型栽培种除薯较多外, 并没表现其它优势。这至少有两方面的原因, 一是本研究尚欠深入, 所选性状具有局限性, 如新型栽培种具有较强抗逆性等, 而本试验由于没有详尽考察这些性状, 文中用来分类的11个性状就没有包含这些种间差异明显的性状, 这为以后深入研究提出了课题; 同时也应看到问题的另一面, 抛开这些差别明显的性状不论, 其分类结果都令人满意, 这种分类方法的可靠性是无庸置疑的; 二是新型栽培种直接用于生产确有一段距离, 国内除怀薯6号 (利用新型栽培种作杂交亲本之一) 具有较好的综合性状和对青枯病的田间抗性用于生产外, 新型栽培种尚未直接在生产中利用。但这并不是说新型栽培种不能用于生产, 恰恰相反, 许多实验结果表明, 新型栽培种与普通栽培种之间的杂交组合类型在生产中可望有广阔的利用前景^[2-4]。

在普通栽培种中, 从国外引入的品种较之国内育成的品种 (系) 表现欠佳, 这种差异在聚类图中可反应出来。造成这种差异的原因可能是: 这些品种因不太适应哈尔滨地

区的环境条件而表现较差; 另外, 这些品种育成推广多年, 也许因病毒积累, 种薯退化, 而不如国内育成品种。

本文仅限于一些形态性状和经济性状, 没有涉及生化特性和遗传基因, 研究有待深入。此外, 试验所取种薯都是东北农学院保存的原始材料和品种, 这些材料因病毒积累程度不用, 退化程度是不一样的。因此, 试材的产量性状及其有关的农艺性状的表现, 在一定程度上为病毒影响所波动, 从而影响试验的准确性, 在条件允许的情况下, 可用脱毒薯作试材 (一般情况下, 这样做并非必要)。尽管如此, 本文的结果与专家评判颇相一致, 这说明了模糊聚类分析方法是可靠的。这种方法在育种上的利用也是明显的。例如, 从动态聚类图中可以看出, 每个类型内的各个品种 (系) 的杂交组合可能不理想, 因为它们的性状接近, 异质程度差。但不同类型之间的交配, 会有明显的优势, 如20 (东农303) 与新型栽培种 (3、7、1、2、9、8、4、5、6) 之间的杂交组合, 从其后代中获得有利用价值材料的可能性远大于东农303所在类型内品种间的杂交后代。即使在同一类型中配制杂交组合, 也应该选用差别大的。这种方法在杂交育种中可以在一定程度上帮助克服选配亲本的盲目性, 因此, 模糊聚类分析方法在育种上是实用的。此外, 在马铃薯品种资源的研究中, 对于数量众多的原始材料以及繁杂的性状数据, 这种方法就有助于从这些数据中找出规律, 对材料客观地分门别类。随着电子计算机应用领域的扩大, 这种方法在马铃薯种质资源的计算机管理中也能起到一定作用。总之, 有理由相信, 模糊数学对于马铃薯种质资源的研究和利用、马铃薯育种的选材将是一种有用工具。

参考文献

- [1] 汪培庄: 模糊数学简介(I), 《数学的实践和认识》, 1980, 2: 45~59。
 [2] 李德纯、李景华: 普通栽培与新型栽培种配合力的研究, 《东北农学院学报》, 1982, 3。
 [3] 肖志敏、李景华: 近缘栽培种间杂种优势主要经

济性状的配合力分析, 《东北农学院研究生学位论文》, 1985。

- [4] Glendinning, D. R.: *Neo-Tuberosum* and its potential value in breeding Potato Res. 1975, 18: 351—362.
 [5] Pavak, J. J.: Recent and future developments of the potato in the world—A Workshop Amer. Potato Journal 1987, 64(11): 599—602。

A STUDY ON FUZZY CLUSTER ANALYSIS OF TWENTY-TWO POTATO CULTIVARS (CLONES)

Yan Fengxi, Wang Fengyi and Chen Yili
(Northeast Agricultural Collge, Harbin, China)

ABSTRACT

Fuzzy cluster analysis was used to study 22 potato cultivars(clones). The procedure of this method was as follows:

First, calculating similarity matrix $\tilde{R}$: Various similarity coefficients may be used here.

Second, transforming $\tilde{R}$ to fuzzy equivalent relation R^* : Make $\tilde{R}^2, \tilde{R}^4, \tilde{R}^8, \dots$, if $\tilde{R}^k = \tilde{R}^{2k} = R^*$, R^* was a fuzzy equivalent matrix.

Third, clustering: select suitable value of λ and cluster. If $\lambda < 0.7697$, 22 cultivars (clones) can be clustered three groups.

The results of fuzzy clustering indicated that there were significant differences between *S. tuberosum* and *Neo-Tuberosum*. These showed that fuzzy cluster method was useful for numerical classification of potato cultivars (clones).