

农业气候区划中的定性聚类方法

陶向新

(沈阳农学院)

农业气候区划工作中，经常收集到大量的定性资料，如何将这些资料充分应用到区划中，是需解决的实际问题。本文以辽宁省抚顺县气候调查资料为例，谈一下运用定性资料进行相似聚类的区划方法。

一、气候资源的定性评定

农业气候调查和考察中，定量资料是有限的，而定性资料却占有相当大的比重。并且，定量资料与定性资料同时存在的情况是普遍的，从实际出发，对定性资料进行处理，将其运用到区划中，可以节省观测时间及人力、物力，同时对解决具体问题，在某种情况下又具有独特的优点。

对气候资源的定性评定，一般说来，有以下几个方面：

1. 选择几种主要因素，对于资源条件优越的评定高分数，条件差的评低分。灾害轻微的评定高分数，灾害严重的评低分。

2. 根据各个因素的影响程度，对主要因素分级多些（也就是最高评分高一些），对次要因素分级少些。这样一来，就把不同因素的影响权重考虑进去了。

3. 定量资料和定性资料同时存在且定量资料较少时，要把定量资料变为定性资料。评定分数或等级，可以是整数也可以是小数。

抚顺县将全县热量状况，按冷、凉、温、

暖、热五级分别评为1—5分；对无霜期长短按较短、中等、较长分为三级，评分为1—3分；对水分状况按正常、半湿润、湿润分为三级，评分为0—2分。具体评分见表1。

表1 抚顺县气候资源评定表

项 目	地 点 编 号	马	海	峡	南	洋	千	李	营	黄	金	草	石	
		圈	浪	河	党	湖	斤	家	盘	旗	花	党	文	
	A	B	C	D	E	F	G	H	I	J	K	L		
热 量 评 分		2	5	4	4	4	5	2	3	5	3	4	5	
无 霜 期 评 分		1	3	3	3	2	2	1	2	3	3	3	3	
水 分 评 分		2	1	2	1	2	0	2	2	0	0	2	2	

对于相对等级或评分的确定，可以把极端差异做为临界值，将处于中间程度的状态定性分级。也可以某一对比点为基点，通过调查找出相对程度用百分比或分级的方式表示出来。还可以应用模糊数学理论，通过隶属函数、贴近度和综合评判等方法确定评分。

二、相似性的表达

在获得各地气候资源评分的基础上，分析彼此间的农业气候相似性，往往要通过计算相似系数的大小来表示两地间的相似程度。在此之前，先将气候资源这种多态变量用多个二态变量（非0则1两种状态）来表示。例如，A点和H点的气候资源可表示为表2的形式：

表2

气候资源的二态变量编码

地 点 二 态 特 征	热 量 状 况	1	2	3	4	5	无 霜 期 状 况	1	2	3	水 分 状 况	1	2	3				
		马	圈	(A)	2	1	1	0	0	0		1	0	0	2	1	1	0
营 盘(H)	3				1	1	1	0	0		2	1	1	0	2	1	1	0

对于两地的相似系数，可依夹角余弦公式来计算，即

$$\cos\theta = \frac{\sum_{i=1}^n x_i \cdot y_i}{\sqrt{\left(\sum_{i=1}^n x_i^2\right) \cdot \left(\sum_{i=1}^n y_i^2\right)}} \quad (1)$$

现将 A 和 H 两地的二态变量特征代入后，则得

$$\begin{aligned}\cos\theta_{AH} &= \frac{1 \times 1 + 1 \times 1 + 0 \times 1 + 0 \times 0 + 1}{\sqrt{1^2 + 1^2 + 1^2 + 1^2 + 1^2}} \\ &\quad \frac{+ 1 \times 0 \times 1 + 0 \times 0 + 1 \times 1 + 1 \times 1}{\sqrt{1^2 + 1^2 + 1^2 + 1^2 + 1^2 + 1^2 + 1^2}} \\ &= \frac{5}{\sqrt{5 \times 7}} = 0.845\end{aligned}$$

为了便于应用，把上面的相似系数变成另外一种形式，即公式 (2)。

$$\cos\theta = \frac{\sum_{i=1}^n \min(x_i, y_i)}{\sqrt{\left(\sum_{i=1}^n x_i\right) \cdot \left(\sum_{i=1}^n y_i\right)}} \quad (2)$$

根据 (2) 式计算 A 与 H 间的相似系数，与 (1) 式结果相同，同时更便于代入有小数评分的特征值。

$$\cos\theta_{AH} = \frac{\min(2, 3) + \min(1, 2) + \min(2, 2)}{\sqrt{2+1+2} \cdot \sqrt{3+2+2}}$$

表 3

初始相似聚类

地 中 心 点	A	B	C	D	E	F	G	H	I	J	K	L	初步归类
H	0.845*	0.756	0.882	0.802	0.935	0.714*	0.845*	1.000	0.668	0.772	0.882	0.837	A.G.H
K	0.745	0.889*	1.000*	0.943	0.943*	0.756	0.745	0.882	0.825*	0.816*	1.000	0.949	C.D.E.J.K
L	0.707*	0.949	0.949	0.894	0.894*	0.837	0.707	0.837*	0.894	0.775	0.949	1.000*	B.E.I.L

3. 初始相似聚类，按“相似系数最大”原则，择近归类。一个地点应归入与其相似程度最大的中心点那一类里，也就是说从该地点对几个中心点的相似系数中，挑取最大者做为归类目标。例如，A 点同 H、K、L 三个中心点的相似系数分别为：0.845、0.745、0.707。其中，A 点同 H 点的相似系数 (0.845) 最大，则 A 点应归入以 H 为中心的这一类，标上 * 号。则得到 A、G、H 为一类；C、D、E、J、K 为一类；B、F、I、L 为一类，这

$$= \frac{2 + 1 + 2}{\sqrt{5 \times 7}} = 0.845$$

依此类推，任何两个地点的相似系数均可求得。只要确定它们的定性评分，无论包括多少种要素，都使相似性得到定量化表示。

三、定性相似聚类

既然农业气候相似性可以得到数量化表示，所以以相似系数为统计量，对限定范围内的不同地点，就可以进行相似聚类了。其主要步骤如下：

1. 选择代表性地点做为中心点。这些中心点的选择，是在农业气候调查的基础上，综合气候代表性、农业生产情况、地理环境等方面特征，凭实际经验选定的。

一般说来，中心点的气候特点，反映了不同区域的气候差异，因此，一方面具有区域的代表性，同时，彼此间气候差异也较大。中心点的数目，就是最终分区的数目。可通过若干种结果对比来选择最佳点。

本文引例中，是以营盘、章党、石文（即 H、K、L）为三个区域的代表点的。

2. 分别计算各个地点同中心点之间的相似系数〔依公式 (2) 进行计算〕。详见表 3 所示。

便是初始分类。

4. 计算初始分类后各类中心点的特征值。用各类所包括的地点各个特征值的平均值做为类中心点的特征值。例如，以 A、G、H 的特征值求得它们的类中心 H' 的特征值，如表 4。

5. 逐步归类，以初始分类的类中心为基点，分别计算各个地点对于这些中心点的相似系数。重复择近归类步骤，得到逐步归类结果，如表 5 所示。

表 4

初始聚类类中心特征值

要 素 地 点	A	G	H	H'	C	D	E	J	K	K'	B	F	I	L	L'
热 量	2	2	3	2.3	4	4	4	3	4	3.8	5	5	5	5	5.0
无 霜 期	1	1	2	1.3	3	3	2	3	3	2.8	3	2	3	3	2.75
水 分	2	2	2	2.0	2	1	2	0	2	1.4	1	0	0	2	0.75

表 5

逐步归类结果表

中心点 地 点	A	B	C	D	E	F	G	H	I	J	K	L	逐步归类
H'	0.945*	0.648	0.789	0.687	0.837	0.575	0.945*	0.894*	0.538	0.621	0.789	0.748	A.G.H
K'	0.696	0.896*	0.943	0.950*	0.900*	0.775	0.696	0.855	0.825*	0.837	0.943*	0.894	C.D.E.J.K
L'	0.575*	0.972	0.857	0.910	0.819	0.907*	0.575	0.745	0.940*	0.805	0.857	0.922*	B.F.I.L

6. 若逐步归类同初始归类结果相同，则将其做为最终归类结果，划出分区。本文引例中，两个聚类结果相同，故将抚顺县划分为三个农业气候区域。如果逐步聚类同初始

聚类结果不同，应将中心点进行调整或重选，重复上述步骤，直到达到两个聚类结果相同为止。