



模糊集分类判别的预报方法

马 万 庆

(宁夏石嘴山市气象局)

最近，我们在做长期天气预报中，应用模糊数学对资料进行了处理，并结合“判别分析的非参数方法”得出了模糊集分类判别的预报方法。该方法主要是计算出各交点的距离，更切合实际地描述了空间点的分布情况，预报效果较好。

一、基本思路

在长期天气预报中，常常要做出降水量、温度等项目的分级预报。模糊集分类判别的预报方法是为解决这类问题而提出的。这种方法以模糊集理论为基础，从各元素对模糊集的隶属度出发，计算各样本点的距离。

若有几个模糊集的一组隶属度，记为： $M = \{\mu(x_{11}), \mu(x_{21}), \dots, \mu(x_{n1})\}$ ，({ } 表示列向量)，对该组各隶属度取交运算(取小)的结果，可以看作是几维空间的一个点，记作： $M_{min} = \wedge[\mu(x_{11}), \dots, \mu(x_{n1})]$ 。同理，对几个模糊集的另一组隶属度取交运算得出另一个点，记作： $N_{min} = \wedge[\mu(x_{12}), \mu(x_{22}), \dots, \mu(x_{n2})]$ 。两个空间点的距离为：

$$d_{ij} = \sqrt{(M_{min} - N_{min})^2} \quad (1)$$

以上， $i, j = 1, 2, \dots, T$ 。假如在 T 个历史样品中所考虑的预报对象分为 A、B、C 三类，即： $T = T(A) + T(B) + T(C)$ 。这样，我们就可以根据(1)式，在 T 个样品中计算出 n 个模糊集的 T 组隶属度的交集点与待报年的一组隶属度的交集点的 T 个距离。如果因子选择得好，隶属函数确定得好，几种类别的点集可以是几个不同的点集。同类的点集彼此都应该比较靠近，即彼此之间的距离也比较小；而不同类的点集距离比较大。所以由公式(1)计算出来的距离的大小，就反映了待报年空间点与各类中每一点的靠近程度。距离最小的点落在哪一类，就可以认为该点属于哪一类。

二、操作步骤

- 选取预报因子。
- 确定隶属函数，计算各元素的隶属度。
- 按所需的级别分类，并将隶属度分类列表。
- 判断待报年预报量所属类别(有五种判断方法)。

(1) 最近空间点距离法：按公式(1)计算出各点的距离后，选取最小的距离进行判断。最小距离的点落在哪一类别中，就预报待报年为该类。如最近空间点落在 A 类中，就报待报年为 A 类，余类推。

(2) K 点距离法：这种方法是在方法(1)选出的最近点附近选出与此点最靠近的 K 个点进行判断分析。若 K 个点中以 A 类占多数，则报待报年为 A 类，否则报待报年为其它类别。 K 的数目常采用以下近似公式来计算：

$$K = (\frac{T}{2})^{\frac{4}{4+n}} \quad (2)$$

式中 T 为样本容量， n 为预报因子个数。

(3) 平均距离法：在方法(1)计算出的 T 个距离中，求出各类距离的平均值，选取最小平均距离所属的一类为待报年的类别。

(4) 标准交集距离法：先计算出各类中 n 个模糊集隶属度的平均值，分别对各类的 n 个平均隶属度取交运算，再分别计算这些交运算结果与待报年 n 个隶属度的交运算结果的距离，这些距离称为标准交集距离。以最小距离所属类别作为待报年的类别。

(5) 并集距离法：首先对各类中 n 个模糊集的隶属度分别进行并运算，再计算这些并运算结果与待报年的 n 个隶属度的并的距离(简称为并集距离)，挑取最小距离所属类别作为待报年的类别。

三、预报实例

应用模糊集分类判别的预报方法做陶乐县1981年9月份的月降水量预报(用该县1959—1981年共22年资料，即 $T = 22$)。

1. 选取 3 个预报因子： x_1 为 7 月降水量； x_2 为 8 月上旬降水量； x_3 为 5 月上旬日平均气温，即 $n = 3$ 。 x_1 和 x_2 与预报量为正相关， x_3 为反相关。

2. 建立隶属函数关系式如下：

$$\mu_A(x_1) = \begin{cases} 1 & x_{11} \geq 63.8 \\ \frac{x_{11} - 6.0}{63.8 - 6.0} & 6.0 < x_{11} < 63.8 \\ 0 & x_{11} \leq 6.0 \end{cases}$$

$$\mu_B(x_2) = \begin{cases} 1 & x_{2i} \geq 86.2 \\ \frac{x_{2i} - 0.2}{86.2 - 0.2} & 0.2 < x_{2i} < 86.2 \\ 0 & x_{2i} \leq 0.2 \end{cases}$$

$$\mu_C(x_3) = \begin{cases} 1 & x_{3i} \leq 12.3 \\ 1 - \frac{x_{3i} - 12.3}{19.8 - 12.3} & 12.3 < x_{3i} < 19.8 \\ 0 & x_{3i} \geq 19.8 \end{cases}$$

3. 计算各元素的隶属度。

4. 把 22 年(1959—1980年)的 9 月份雨量按偏少(A类)、正常(B类)、偏多(C类)分类。按 A、B、C 三类将历年隶属度依类列表(附表)。1981 年为待报年，各因子的隶属度为 0.66, 0.02, 0.64。

附表

类 别	编 号	年 代	$\mu_A(x_1)$	$\mu_B(x_2)$	$\mu_C(x_3)$
A类 $R_9 \leq 18.0$ 毫米	1	1963	1	0.05	0.52
	2	1965	0.13	0.14	0.64
	3	1966	0.56	0	0.52
	4	1970	0	0.39	0.43
	5	1972	0.09	0.03	0
	6	1974	0.15	0.20	0.56
	7	1975	0.44	0.07	0.92
	8	1979	1	0.13	0.91
	9	1980	0.07	0.28	0.69
平均			0.38	0.14	0.58
B 类 18.1 < $R_9 < 33.4$ 毫米	10	1961	0.98	0.32	0.60
	11	1964	0.33	0.17	0.49
	12	1967	1	0.26	0.67
	13	1968	0.30	0.95	0.77
	14	1969	0.31	0.24	0.12
	15	1971	0.03	0	0.75
平均			0.49	0.32	0.57
C类 $\geq R_9 33.5$ 毫米	16	1959	0.73	0.76	0.72
	17	1960	0.55	0.16	0.73
	18	1962	0.39	0.06	1
	19	1973	0.63	0.59	0.75
	20	1976	0.62	0.66	0.92
	21	1977	1	0.28	0.64
	22	1978	1	1	0.73
平均			0.70	0.50	0.78

5. 预报

(1) 最近空间点距离法：计算历年三个因子隶属度的交运算的结果与待报年的三个因子隶属度的交运算结果的距离。在 22 个交集距离中，最小的距离是 0.02，有四个点，由于四个点中有三个点落在 A 类，而 A 类为偏少类，故报 1981 年的 9 月份雨量偏少。

(2) K 点距离法：此例中， $T = 22$, $n = 3$ ，依据(2)式：

$$K = \left(\frac{T}{2}\right)^{\frac{4}{4+n}} = \left(\frac{22}{2}\right)^{\frac{4}{4+3}} = (11)^{\frac{4}{7}} \approx 4$$

最靠近 0.02 的 4 个点中有 3 个点落在 A 类别中，故预报 1981 年 9 月份雨量偏少。

(3) 平均距离法：由计算得出各类的平均距离 A 类是 0.06, B 类为 0.14, C 类为 0.43。因为 $A < B < C$ ，故预报 1981 年 9 月份雨量属 A 类(偏少)。

(4) 标准交集距离法：由历年隶属度分类列表(附表)得 A 类平均隶属度： $\mu_A(x_1) = 0.38$; $\mu_B(x_2) = 0.14$; $\mu_C(x_3) = 0.58$ 。

三个平均隶属度的交运算结果为：

$$\mu_{DA}(x) = \wedge [\mu_A(x_1), \mu_B(x_2), \mu_C(x_3)] \\ = \wedge [0.38, 0.14, 0.58] = 0.14$$

$$\text{于是 } d_{A\text{标}} = \sqrt{[\mu_{DA}(x) - \mu_{D_0}(x)]^2} \\ = \sqrt{(0.14 - 0.02)^2} = 0.12$$

同理可得：

$$d_{B\text{标}} = 0.30$$

$$d_{C\text{标}} = 0.48$$

因为 $d_{A\text{标}} < d_{B\text{标}} < d_{C\text{标}}$ ，所以预报 1981 年 9 月份雨量为 A 类(偏少)。

(5) 并集距离法：由附表得知：A 类的 $\mu_A(x_1)$ 取并运算结果为 1; $\mu_B(x_2)$ 的并为 0.39; $\mu_C(x_3)$ 的并为 0.92。

所以 $d_{A\text{并}} =$

$$\sqrt{(1 - 0.66)^2 + (0.39 - 0.02)^2 + (0.92 - 0.64)^2} \\ = \sqrt{0.3309} = 0.58$$

$$d_{B\text{并}} = 1.01$$

$$d_{C\text{并}} = 1.1$$

由于 $d_{A\text{并}} < d_{B\text{并}} < d_{C\text{并}}$ ， $d_{A\text{并}}$ 最小，所以报 1981 年 9 月份雨量为偏少类。

以上五种方法都报 1981 年 9 月份降水量为 A 类，所以预报陶乐县该月降水量偏少。实况为 7.1 毫米，属于偏少类，预报正确。

四、小结

我们运用模糊集分类判别的预报方法做了 12 次预报检验(两次预报，十次试报)，预报准确率均在 10/12 以上。五种判断方法中，以最近空间点距离法、并集距离法和 K 点距离法效果较好，标准交集距离法和平均距离法稍差些。

模糊集分类判别的预报方法运用了模糊集原理，吸取了“判别分析的非参数方法”的判断原则和优点。又由于在计算各种距离前，进行了隶属度的变换，使所计算的各种距离更好地描述了各空间点的分布情况，因此，预报效果比判别分析的非参数法要好一些。

这种方法还有以下几个优点：①工作量少，方法简便易于操作；②因子与样本容量的关系不大；③有五种判断方法，有利于分析预报。

使用该方法的关键是选取好的预报因子，确定恰当的隶属函数。在预报中，遇到五种判断方法结果不一时，可以首先考虑 K 点距离法、最近空间点距离法和并集距离法的预报结果，并结合其它预报方法来综合分析。