

以中位数为中心的分级相似预报法

吴绍敏 彭 沛

(数 学 系)

摘 要

本文介绍一种计算简单且效果较好的预测方法,它既可作趋势预报又可作定值预报。

(一) 引 言

分级相似预报法^[1],是一种简单易行、效果较好的统计预测方法。它既克服了在相似预报中,可能出现两个或多个差异较大的预报相似值,又可对预报对象作出趋势和定值的预报。秩数分级相似法^[2]和绝对距离分级相似法,均以各级的均值作为该级的中心点。由于观测数据的客观性,有些因子分级后出现某几级的均值很接近(即这几级的中心点很接近),这样给选择预报因子带来一定困难。为此考虑以各级的中位数作为各级的中心点,便于更合理地选择因子。同时,为了充分利用各观测值的信息,我们选择“绝对距离”为度量标准,以得到更好的预报效果。

(二) 基 本 思 想

设预报对象 y 与因子 x_1, x_2, \dots, x_s 统计相关,并设 y 与 $x_i (i=1, \overline{s})$ 有 N 个历史观测值:

$y: y_1, y_2, \dots, y_N; \quad x_i: x_{i1}, x_{i2}, \dots, x_{iN}; \quad i=1, \overline{s}$ 。各因子的当前值分别为: $x_{10}, x_{20}, \dots, x_{s0}$ 。

根据预报的需要,将 y 的历史资料分成 m 级,记为: $y^{(1)}, y^{(2)}, \dots, y^{(m)}$, 其中

$$y^{(k)} = \{ y_1^{(k)}, y_2^{(k)}, \dots, y_{N_k}^{(k)} \}, \quad k=1, \overline{m}, \quad \sum_{k=1}^m N_k = N.$$

对应于 y 的等级,将各因子的历史资料也相应地分成 m 级,记为: $x_i^{(1)}, x_i^{(2)}, \dots, x_i^{(m)}$, ($i=1, \overline{s}$), 其中

本文 1984 年 5 月 16 日收到。

$$x_i^{(k)} = \{x_{i1}^{(k)}, x_{i2}^{(k)}, \dots, x_{iN_k}^{(k)}\}, \quad i = \overline{1, s}, \quad k = \overline{1, m}, \quad \sum_{k=1}^m N_k = N.$$

并记第 i 个因子 x_i 的第 k 级的中位数为 $a_i^{(k)} = x_{i, r}^{(N_k)}$ ($i = \overline{1, s}, k = \overline{1, m}$), 若 x_i 与 y 正

相关, 则取 $r = \left[\frac{N_k}{2} \right] + 1$; 若 x_i 与 y 负相关, 则取 $r = \begin{cases} \frac{N_k}{2}, & \text{当 } N_k \text{ 为偶数时.} \\ \frac{N_k + 1}{2}, & \text{当 } N_k \text{ 为奇数时.} \end{cases}$

以 $(a_1^{(k)}, a_2^{(k)}, \dots, a_s^{(k)})$ 为第 k 级的中心点建立判别测度:

$$u(k) = \sum_{i=1}^s \frac{|x_{i0} - a_i^{(k)}|}{|\bar{x}_i|}, \quad k = \overline{1, m}.$$

其中 $\bar{x}_i = \frac{1}{N} \sum_{j=1}^N x_{ij}$.

若 $u(l) = \min_{1 \leq k \leq m} u(k)$, 则说明在“绝对距离”下, 点 $(x_{10}, x_{20}, \dots, x_{s0})$ 与点 $(a_1^{(l)}, a_2^{(l)}, \dots, a_s^{(l)})$ 最接近, 即因子的当前值与第 l 级的中心点最接近, 也说明了因子的当前值属于第 l 级. 由 y 与因子 x_1, x_2, \dots, x_s 的相关性, 可以判别与当前值 $(x_{10}, x_{20}, \dots, x_{s0})$ 相对应的 y_{N+1} (未知预报量) 应属于第 l 级. 至此我们就报出了 y 的未来趋势.

但实际工作中, 往往还需要对预报对象作定值估计. 为此, 我们将属于第 l 级的各历史资料列出, 以“绝对距离” (广义的) 为相似度, 找出与因子当前值最相似的时段, 从而作出 y 的定值预报.

(三) 具体实施步骤

1. 确定预报对象 y 利用相关系数检验法, 选取 s 个与 y 统计相关而彼此之间相关关系尽可能小的因素, 作为预报因子, 记为 x_1, x_2, \dots, x_s . 设它们均有 N 个历史观测值.

为了更充分地发挥各因子在预报中的作用, 在作预报之前, 将各因子的观测数据 (包括当前值) 作适当的处理. 如各因子相应地减去一常数 c_i ($i = \overline{1, s}$), 适当控制各因子 x_i ($i = \overline{1, s}$) 的均值 \bar{x}_i (数据已经处理), 使判别测度的被加项: $\frac{|x_{i0} - a_i^{(k)}|}{|\bar{x}_i|}$, ($i = \overline{1, s}$) 的数量级相当. 记各因子经处理后的数据为:

$$x'_{ij} = x_{ij} - c_i,$$

$$x'_{i0} = x_{i0} - c_i,$$

$$i = \overline{1, s},$$

$$j = \overline{1, N}$$

并列表 1.

• $x_{ir}^{(N_k)}$ 为因子 x_i 的第 k 级按小到大顺序排列的第 r 个顺序统计量.

2. 根据预报的需要, 将预报对象 y 的历史资料按由小到大的顺序适当地分成 m 级, 对应于 y 的等级, 将各因子的历史资料也自然地分成 m 级 (表 2)。

表 1

编 号	y	x_1	x_2	x_s
1	y_1	x'_{11}	x'_{21}	x'_{s1}
2	y_2	x_{12}	x'_{22}	x'_{s2}
\vdots	\vdots	\vdots	\vdots		\vdots
\vdots	\vdots	\vdots	\vdots		\vdots
N	y_N	x'_{1N}	x'_{2N}	x'_{sN}
当前值		x'_{10}	x'_{20}	x'_{s0}

表 2

级 别	编 号	y	x_1	x_2	x_s
1	$b_1^{(1)}$	$y_1^{(1)}$	$x_{11}^{(1)}$	$x_{21}^{(1)}$	$x_{s1}^{(1)}$
	$b_2^{(1)}$	$y_2^{(1)}$	$x_{12}^{(1)}$	$x_{22}^{(1)}$	$x_{s2}^{(1)}$
	\vdots	\vdots	\vdots	\vdots		\vdots
	$b_{N_1}^{(1)}$	$y_{N_1}^{(1)}$	$x_{1N_1}^{(1)}$	$x_{2N_1}^{(1)}$	$x_{sN_1}^{(1)}$
2	$b_1^{(2)}$	$y_1^{(2)}$	$x_{11}^{(2)}$	$x_{21}^{(2)}$	$x_{s1}^{(2)}$
	$b_2^{(2)}$	$y_2^{(2)}$	$x_{12}^{(2)}$	$x_{22}^{(2)}$	$x_{s2}^{(2)}$
	\vdots	\vdots	\vdots	\vdots		\vdots
	$b_{N_2}^{(2)}$	$y_{N_2}^{(2)}$	$x_{1N_2}^{(2)}$	$x_{2N_2}^{(2)}$	$x_{sN_2}^{(2)}$
\vdots	\vdots	\vdots	\vdots	\vdots		\vdots
	\vdots	\vdots	\vdots	\vdots		\vdots
m	$b_1^{(m)}$	$y_1^{(m)}$	$x_{11}^{(m)}$	$x_{21}^{(m)}$	$x_{s1}^{(m)}$
	$b_2^{(m)}$	$y_2^{(m)}$	$x_{12}^{(m)}$	$x_{22}^{(m)}$	$x_{s2}^{(m)}$
	\vdots	\vdots	\vdots	\vdots		\vdots
	$b_{N_m}^{(m)}$	$y_{N_m}^{(m)}$	$x_{1N_m}^{(m)}$	$x_{2N_m}^{(m)}$	$x_{sN_m}^{(m)}$
当前值			x'_{10}	x'_{20}	x'_{s0}

其中, $b_j^{(k)} (j = \overline{1, N_k}, k = \overline{1, m})$ 为 $y_j^{(k)}$ 在表 1 中所对应的序号, $\{y_j^{(k)}: j = \overline{1, N_k}, k = \overline{1, m}\} = \{y_j: j = \overline{1, N}\}; \{x_{ij}^{(k)}: j = \overline{1, N_k}, k = \overline{1, m}\} = \{x_{ij}: j = \overline{1, N}\}, i = \overline{1, s}.$

3. 计算各因子的均值

$$\bar{x}_i = \frac{1}{N} \sum_{j=1}^N x'_{ij}, \quad i = \overline{1, s}.$$

确定各级的中位数 ($a_1^{(k)}, a_2^{(k)}, \dots, a_s^{(k)}$), $k = \overline{1, m}$ (表 3)。

表 3

级 别	x_1	x_2	...	x_s
1	$a_1^{(1)}$	$a_2^{(1)}$...	$a_s^{(1)}$
2	$a_1^{(2)}$	$a_2^{(2)}$...	$a_s^{(2)}$
\vdots	\vdots	\vdots		\vdots
m	$a_1^{(m)}$	$a_2^{(m)}$...	$a_s^{(m)}$
均 值	\bar{x}_1	\bar{x}_2	...	\bar{x}_s

4. 建立判别测度

$$u(k) = \sum_{i=1}^s \frac{|x'_{i0} - a_i^{(k)}|}{|\bar{x}_i|}, \quad k = \overline{1, m} \tag{1}$$

5. 预报 y 的趋势

若 $u(l) = \min_{1 \leq k \leq m} u(k)$, 则预报 y 属于第 l 级。

6. 预报趋势的效果检验

将因子的各个历史资料, 作为当前值, 代入式 (1), 回报 y 所属的级别, 看是否与实况相符。若回报准确率较高, 则进行下一步的相似定值预报, 否则另选因子或采用其他预报方法。

7. 作 y 的相似定值预报

若已预报出 y 属于第 l 级; 则将各因子属于第 l 级的观测数据列成表 4。计算各历史资料与当前值的“绝对距离”:

$$d_j = \sum_{i=1}^s \frac{|x_i^{(l)} - x'_{i0}|}{|\bar{x}_i|}, \quad j = \overline{1, N_l}$$

列入表 4。

表 4

编 号	$x_1^{(l)}$	$x_2^{(l)}$...	$x_s^{(l)}$	d_j
1	$x_{11}^{(l)}$	$x_{21}^{(l)}$...	$x_{s1}^{(l)}$	d_1
2	$x_{12}^{(l)}$	$x_{22}^{(l)}$...	$x_{s2}^{(l)}$	d_2
\vdots	\vdots	\vdots		\vdots	\vdots
N_l	$x_{1N_l}^{(l)}$	$x_{2N_l}^{(l)}$...	$x_{sN_l}^{(l)}$	d_{N_l}

注: 在作 y 的相似预报时, 亦可只挑出对趋势预报起主导作用的因子 (即各级中位数分的较开的因子)。

若 $d_k = \min_{1 \leq j \leq l} d_j$, 则预报 $\hat{y}_{N+1} \approx y_k^{(l)}$.

(四) 实 例

根据龙溪地区气象台 1955—1982 年的实测资料, 利用该方法预报龙溪地区 1983 年稳定通过 20.0℃ 的开始期。

1. 设预报对象——龙溪地区稳定通过 20.0℃ 开始期为 y 。为便于计算, 将 y 的历史资料进行数量化(令 5 月 1 日为 0) 其数据见表 5。

表 5

年 份	y	x_1	x_2	x_3	x_4	x_5
1956	0	11.3	0.4	1.0	233.8	19.3
57	-14	14.8	5.1	3.1	0.0	20.2
58	-14	16.4	14.2	4.0	80.8	22.0
59	-9	19.1	21.2	3.8	49.5	18.6
60	23	10.1	12.2	4.4	180.7	16.7
61	-2	15.4	9.5	5.3	93.0	17.0
62	-10	11.9	4.1	6.4	154.0	19.2
63	-19	19.6	0.0	5.6	46.2	21.3
64	-21	17.2	0.0	9.2	13.5	25.2
65	4	11.3	68.6	3.2	57.6	16.3
66	-11	14.6	0.3	4.6	1.2	21.5
67	-4	15.8	0.4	5.9	3.4	19.4
68	0	8.7	24.0	5.0	87.2	18.7
69	-11	10.6	7.1	5.8	59.4	17.4
70	0	7.2	83.8	9.2	0.0	18.0
71	12	9.1	39.3	0.4	157.6	14.2
72	-6	15.1	31.6	4.1	21.9	17.9
73	12	7.0	24.0	4.1	41.0	14.7
74	-18	7.2	12.5	5.2	91.0	16.5
75	23	10.9	3.4	5.1	33.4	13.6
76	-14	15.8	3.2	6.1	36.4	14.2
77	-9	15.1	1.3	5.1	96.8	17.0
78	-22	8.3	22.5	5.6	32.2	18.9
79	4	10.3	13.5	1.4	183.5	15.1
80	11	6.2	41.3	0.2	127.2	17.9
81	5	1.9	32.4	2.3	119.6	17.5
82	-5	13.1	17.3	0.8	53.5	17.3
当前值		11.6	0.7	5.9	47.5	17.1

利用相关系数检验法, 选取 5 个预报因子如下:

x_1 —龙溪地区一月份下旬极端最低气压 P_D ; $r_{1y} = -0.4536$

x_2 —龙溪地区一月份下旬雨量 R ; $r_{2y} = 0.3890$

x_3 —龙溪地区五月份中旬平均绝对湿度 e ; $r_{3y} = -0.4030$

x_4 —龙溪地区八月份下旬雨量 R ; $r_{4y} = 0.4042$

x_5 —龙溪地区十月份中旬极端最高气压 P_G ; $r_{5y} = -0.6146$

根据各因子的波动情况, 取 $C_1 = C_5 = 1000$, $C_2 = C_3 = 0$, $C_4 = 20$. 计算 $x'_{ij} = x_{ij} - C_i$, $x'_{i0} = x_{i0} - C_i$, $i = \overline{1, 5}$, $j = \overline{1, 27}$, 列入表 5.

2. 分级 (三级)

- $y < -10$, 称为第一级,
- $-10 \leq y < 1$, 称为第二级,
- $1 \leq y$, 称为第三级,

对应于 y 的级别, 将因子 x_1, x_2, \dots, x_5 也相应地分成三级, 并列入表 6.

表 6

级 别	年 份	y	x_1	x_2	x_3	x_4	x_5	回 报
1	1978	-22	8.3	22.5	5.6	32.2	18.9	1
	64	-21	17.2	0.0	9.2	13.5	25.2	1
	63	-19	19.6	0.0	5.6	46.2	21.3	1
	74	-18	7.2	12.5	5.2	91.0	16.5	2
	57	-14	14.8	5.1	3.1	0.0	20.2	1
	58	-14	16.4	14.2	4.0	80.8	22.0	1
	76	-14	15.8	32.1	6.1	36.4	14.2	1
	66	-11	14.6	0.3	4.6	1.2	21.5	1
	69	-11	10.6	7.1	5.8	59.4	17.4	1
2	62	-10	11.9	4.1	6.4	154.0	19.2	2
	59	-9	19.1	21.2	3.8	49.5	18.6	2
	77	-9	15.1	11.3	5.1	96.8	17.0	2
	72	-6	15.1	31.6	4.1	21.9	17.9	2
	82	-5	13.1	17.3	0.8	53.5	17.3	2
	67	-4	15.8	0.4	5.9	3.4	19.4	1
	61	-2	15.4	9.5	5.3	93.0	17.0	2
	56	0	11.3	0.4	1.0	233.8	19.3	2
	68	0	8.7	24.0	5.0	87.2	18.7	2
3	70	0	7.2	83.8	9.2	0.0	18.0	2
	65	4	11.3	68.6	3.2	57.6	16.3	3
	79	4	10.3	13.5	1.4	183.5	15.1	3
	81	5	1.9	32.4	2.3	119.6	17.5	3
	80	11	6.2	41.3	0.2	127.2	17.9	3
	71	12	9.1	39.3	0.4	157.6	14.2	3
	73	12	7.0	24.0	4.1	41.0	14.7	2
	60	23	10.1	12.2	4.4	180.7	16.7	2
	75	23	10.9	3.4	5.1	33.4	13.6	2
	当前值		11.6	0.7	5.9	47.5	17.1	

3. 计算各因子的均值

$$\bar{x}_i = \frac{1}{27} \sum_{j=1}^{27} x_{ij}, \quad i = \overline{1, 5}$$

并确定各级的中位数, 列表 7.

表 7

级 别	x_1	x_2	x_3	x_4	x_5
1	14.8	7.1	5.6	36.4	20.2
2	13.1	17.2	5.0	87.2	18.0
3	9.1	32.3	2.3	127.2	15.1
均 值	12.0000	19.7074	4.3296	76.0889	17.9852

4. 建立判别测度并作 y 的趋势预报

$$u(1) = \sum_{i=1}^5 \frac{|x'_{i0} - a_i^{(1)}|}{|\bar{x}_i|} = 0.9790$$

$$u(2) = \sum_{i=1}^5 \frac{|x'_{i0} - a_i^{(2)}|}{|\bar{x}_i|} = 1.7470$$

$$u(3) = \sum_{i=1}^5 \frac{|x'_{i0} - a_i^{(3)}|}{|\bar{x}_i|} = 3.8070$$

$$\therefore u(1) = \min_{1 \leq k \leq 3} u(k) = 0.9790$$

\therefore 预报 \hat{y}_{83} 属于第一级, 即 $\hat{y}_{83} < -10$.

5. 预报趋势效果检验

经回报, 可得回报准确率 $P = \frac{22}{27} = 81\%$

6. 作“绝对距离”相似定值预报

将各因子属于第 1 级的历史资料列出, 并计算“绝对距离”:

$$d_j = \sum_{i=1}^5 \frac{|x_i^{(1)} - x'_{i0}|}{|\bar{x}_i|}, \quad j = \overline{1, 9}$$

其结果列表 8.

表 8

年 份	j	$x_1^{(1)}$	$x_2^{(1)}$	$x_3^{(1)}$	$x_4^{(1)}$	$x_5^{(1)}$	d_j
1978	1	8.3	22.5	5.6	32.2	18.9	1.7516
64	2	17.2	0.0	9.2	13.5	25.2	2.1616
63	3	19.6	0.0	5.6	46.2	21.3	1.0221
74	4	7.2	12.5	5.2	91.0	16.5	1.7322
57	5	14.8	5.1	3.1	0.0	20.2	1.9333
58	6	16.4	14.2	4.0	80.8	22.0	2.2340
76	7	15.8	32.1	6.1	36.4	14.2	2.2966
65	8	14.6	0.3	4.6	1.2	21.5	1.4237
69	9	10.6	7.1	5.8	59.4	17.4	0.6043

$$\therefore d_0 = \min_{1 \leq j \leq 9} d_j = 0.6043$$

而对应于 y 的年号为 69 年, 因此预报 $y_{83} \approx y_{69}$, 即龙溪地区 1983 年稳定通过 20.0°C 开始期约在 4 月 20 日左右, 与 83 年实况 4 月 19 日基本相符。

参 考 文 献

- [1] 吴绍敏, 几种相似预报方法, 福建师范大学学报(自然科学版), 1(1979)。
 [2] 彭沛、黄丽影, 秩数分级相似预报法, 华侨大学学报, 2(1983)。

A Method of Classified-Resembled Prediction With the Median As the center

Wu Shaumin Pen Pei

Abstract

A method of prediction is presented in this paper. By using this method, a better effect of prediction may be reached and the computation will be simplified. It may be used both as tendency prediction and numerical prediction.