

水文要素模糊因果聚类预报及其应用

周念来, 吴泽宁

(郑州工业大学水利与环境工程学院, 河南 郑州 450002)

摘 要: 根据水文预报中不确定性因素显著, 但预测对象与预测因子之间因果关系明显的特点, 运用模糊聚类分析对各次历史洪水资料进行分类, 以解决不确定性因素的影响, 然后通过提取同类洪水资料的模糊特征建立预测因子与预测对象之间的因果关系, 从而建立模型, 并用实例验证了模型和方法的可行性.

关键词: 水文预报; 模糊聚类分析; 隶属函数; 拉格朗日插值

中图分类号: TV 124 文献标识码: A

0 引言

水文预报一直是水资源工作者致力于研究的问题. 根据水量守恒原理, 地理、地质和气象因素与洪水水情存在的因果关系, 提出由因及果的预报模式. 但由于地理、地质和气象因素的不确定性——随机性和模糊性, 使得直接由预测因子与预测对象之间建立对应关系并求其结构变得较为困难, 即使勉强为之, 预测精度也得不到保证.

怎样既能利用水文系统中的因果关系, 又能克服预测因子的随机性和模糊性带来的困难? 本文尝试利用模糊因果聚类分析的方法来解决此问题. 此方法运用模糊数学、数理统计及数值计算的知识对各次历史洪水记录进行分类, 认为同一类中历史洪水记录的不确定性因素的影响程度相同, 这样就可以在同一类中建立预测因子和预测对象的因果关系, 并进行预测.

1 水文模糊因果聚类预报

1.1 原因因子筛选

设共有 T 次历史洪水资料, $z_i = (x_{i1}, x_{i2}, \dots, x_{in}, y_i)$, $i = 1, 2, \dots, T$; 其中, $x_{i1}, x_{i2}, \dots, x_{in}$ 和 y_i 表示通过相关分析得到的 n 个与预测对象相关程度大的预测因子和预测对象的第 i 次历史洪水的对应值, 且第 j 个因子与预测对象的相关系数为 ρ_j .

1.2 数据预处理

采用中心—标准化法对历史数据进行预处理. x_{ij}, y_i 经过预处理后变为 x'_{ij}, y'_i . 为简便, 用 $x_{(n+1)}$ 和 $x'_{(n+1)}$ 分别表示 y_i, y'_i .

$$\bar{x}_j = \frac{1}{T} \sum_{i=1}^T x_{ij} \quad (j = 1, 2, \dots, n+1); \quad (1)$$

$$s_j^2 = \frac{1}{T-1} \sum_{i=1}^T (x_{ij} - \bar{x}_j)^2 \quad (j = 1, 2, \dots, n+1); \quad (2)$$

$$x'_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j} \quad (i = 1, 2, \dots, T; j = 1, 2, \dots, n+1). \quad (3)$$

1.3 模糊因果聚类

求第 h, g 次洪水资料 z_h 与 z_g 的相关系数的绝对值 r_{hg} , 并作为其相关程度^[1], 且近似用权重 w_k 考虑各因子对相关程度的影响.

$$r_{hg} = \frac{\sum_{k=1}^{n+1} [w_k (x'_{hk} - \bar{x}'_h) (x'_{gk} - \bar{x}'_g)]}{\sqrt{\sum_{k=1}^{n+1} [w_k (x'_{hk} - \bar{x}'_h)^2]} \sqrt{\sum_{k=1}^{n+1} [w_k (x'_{gk} - \bar{x}'_g)^2]}} \quad (h = 1, 2, \dots, T; g = 1, 2, \dots, T), \quad (4)$$

式中, $w_k = \frac{\rho_k}{\sum_{k=1}^n \rho_k + 1}$; $w_{n+1} = \frac{1}{\sum_{k=1}^n \rho_k + 1}$;

$$\bar{x}'_h = \sum_{k=1}^n x'_{hk} \quad (k = 1, 2, \dots, n).$$

构成模糊相似矩阵 $R = (r_{ij})_{T \times T}$, 给定阈值并采用直接聚类法中的编网法^[2]进行聚类分析, 得出最

佳聚类为： $U_1, U_2 \dots, U_m$ (m 表示类别个数)。

1.4 提取模糊特征

对应同一聚类的原因因子集

$$V_t = \begin{Bmatrix} x_{t_1 1} & x_{t_1 2} & \cdots & x_{t_1 n} \\ x_{t_2 1} & x_{t_2 2} & \cdots & x_{t_2 n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{t_f 1} & x_{t_f 2} & \cdots & x_{t_f n} \end{Bmatrix}.$$

($t = 1 \dots, m$; $t_1, t_2 \dots, t_f$ 为 t 类所包含的历史洪水记录),运用数理统计和模糊数学建立模糊集 A_t 以表示其特征,称为原因特征. 可以用它来判别待预测的洪水记录是否属于此类,并用它作为中间量来建立预测因子和预测对象的对应关系.

计算 V_t 的均值和方差

$$\bar{x}_{ij} = \frac{1}{f} \sum_{q=1}^f x_{t_q j} \quad (j = 1, 2, \dots, n), \tag{5}$$

$$\sigma_{ij}^2 = \frac{1}{f} \sum_{q=1}^f (x_{t_q j} - \bar{x}_{ij})^2 \quad (j = 1, 2, \dots, n) \tag{6}$$

然后,根据 \bar{x}_i 和 σ_{ij}^2 建立 V_t 上的模糊集 $A_i(x)$. 这里建立正态模糊集

$$A_i(x) = \sum_{j=1}^n w_j e^{-\frac{1}{2\sigma_{ij}^2}(x_j - \bar{x}_{ij})^2}. \tag{7}$$

式中,以权重 $w_j = \frac{\rho_j}{\sum_{j=1}^n \rho_j}$ 表示各预测因子所作的贡献不同,模糊集 A_i 就是类 V_t 的特征. 这样,可由 $\{U_1, U_2 \dots, U_m\}$ 中各类的模糊特征值构成模糊特征表

$$\begin{bmatrix} U_1 & U_2 & \cdots & U_m \\ A_1 & A_2 & \cdots & A_m \end{bmatrix}.$$

1.5 预测方法

(1) 确定归属类:首先确定待预测洪水记录的归属类. 将待预报记录的预测因子 $\{x_{s_1}, x_{s_2}, \dots, x_{s_n}\}$ 带入模糊特征表,得到 $A = (A_1^s, A_2^s, \dots, A_m^s)$. 根据最大隶属原则选出 A_c^s , 即可得到其归属类为第 c 类.

(2) 求预测值:将 V_c 中的各次历史洪水资料分别带入 A_c 得其隶属于该类的隶属度集 $A_c^0 = (A_{c_1}, A_{c_2}, \dots, A_{c_L})$ (L 为 c 类历史洪水记录数). 由于水文系统中预测因子与预测对象之间的因果关

系, A_c^0 与 U_c 类的预测对象集 $\{y_{c_1}, y_{c_2}, \dots, y_{c_L}\}$ 存在着因果关系. 因此,将点集 $\{(A_{c_1}, y_{c_1}), \dots, (A_{c_L}, y_{c_L})\}$ 模拟成曲线,以 A_c^s 插值求出预测值. 现对插值进行计算简化.

根据 $y_{c_q} > \bar{y}_c$ 或 $y_{c_q} \leq \bar{y}_c$ ($\bar{y}_c = \sum_{q=1}^{L_1} y_{c_q}$) 将隶属度集 A_c^0 分成两类,根据待预测记录的预测因子 $\{x_{s_1}, x_{s_2}, \dots, x_{s_n}\}$ 与 V_c 平均值 $\{\bar{x}_{c_1}, \dots, \bar{x}_{c_n}\}$ 的关系 (如根据 $\sum_{j=1}^n w_j (x_{s_j} - \bar{x}_{c_j})$ 的正负) 近似判别 $\hat{y}_s > \bar{y}_c$ 或 $\hat{y}_s \leq \bar{y}_c$. 在上述隶属度集分类中选择在同一类别隶属度与 A_c^s 最相近的 A_{c_h}, A_{c_g} ($1 \leq h, g \leq L$) 所对应的两组值 $(A_{c_h}, y_{c_h}), (A_{c_g}, y_{c_g})$, 加上隶属度函数的顶点 $(1, \bar{y}_c)$ 共 3 点构成拉格朗日插值多项式^[3], 然后将 A_c^s 带入求出 \hat{y}_s :

$$\begin{aligned} \hat{y}_s = & \bar{y}_c \frac{(A_c^s - A_{c_h})(y_{c_g} - A_{c_g})}{(1 - A_{c_h})(y_{c_g} - A_{c_g})} + y_{c_h} \cdot \\ & \frac{(A_c^s - A_{c_g})(y_{c_h} - A_{c_h})}{(A_{c_h} - 1)(y_{c_h} - A_{c_h})} + y_{c_g} \frac{(A_{c_h} - 1)(y_{c_h} - A_{c_h})}{(A_{c_h} - 1)(y_{c_h} - A_{c_h})}. \end{aligned} \tag{8}$$

2 实例验证

2.1 基本情况

间河是淮河一条支流,流经正阳、息县和淮滨 3 县,流域面积 898 平方公里. 间河上包信水文站,具有水位、雨量和流量的观测资料,其上游有王勿桥雨量站. 本例收集了数场洪水记录数据并建立模型,通过相关因素来预报某场降雨产生的包信洪峰流量.

2.2 洪峰流量预报

(1) 因素选取及基本资料:通过相关分析得出王勿桥和包信最大日降雨量、次降雨量与包信洪峰流量相关程度大,因而选取资料见表 1.

各预测因子与预测对象的相关系数

$$\rho_1 = 0.926, \rho_2 = 0.855, \rho_3 = 0.824, \rho_4 = 0.771.$$

(2) 采用数据中心—标准化法得到数据,见表 2 所示.

(3) 求出模糊相似矩阵,见表 3 所示.

表 1 间河降雨量原始数据

记录号	时间	王勿桥最大 日降雨量/mm	包信最大 日降雨量/mm	王勿桥次降 雨量/mm	包信次降 雨量/mm	包信洪峰 流量/(m³/s)
Z ₁	1965 年 3 月 23 日	28.2	23.7	66.8	63.6	39.1
Z ₂	1965 年 4 月 28 日	41.5	36.0	41.5	40.7	85.0
Z ₃	1965 年 7 月 10 日	44.0	45.9	56.0	60.9	122.0

续表						
记录号	时间	王勿桥最大日 降雨量/mm	包信最大日 降雨量/mm	王勿桥次降 雨量/mm	包信次降 雨量/mm	包信洪峰 流量/(m ³ /s)
Z ₄	1963 年 5 月 8 日	32.0	11.5	47.0	11.5	84.6
Z ₅	1963 年 5 月 27 日	20.0	13.4	44.5	27.4	70.4
Z ₆	1963 年 6 月 4 日	52.0	35.4	53.0	46.6	163.0
Z ₇	1963 年 7 月 9 日	113.5	64.8	184.0	122.0	376.0
Z ₈	1963 年 8 月 7 日	25.9	38.1	32.6	82.4	109.0
Z ₉	1963 年 8 月 17 日	38.4	29.8	84.1	72.0	195.0
Z ₁₀	1968 年 7 月 1 日	46.7	61.8	75.1	128.3	128.0
Z ₁₁	1968 年 10 月 9 日	35.1	28.9	76.0	61.6	120.0
Z ₁₂	1968 年 10 月 13 日	35.1	28.9	120.3	115.0	155.0
Z ₁₃	1975 年 7 月 5 日	130.8	107.7	148.7	160.3	584.0
Z ₁₄	1975 年 7 月 4 日	12.3	45.5	17.9	52.6	61.5
Z ₁₅	1975 年 8 月 15 日	49.0	25.1	49.1	30.1	66.5
Z ₁₆	1975 年 9 月 20 日	39.8	28.5	39.8	28.5	49.9
Z ₁₇	1975 年 10 月 26 日	16.3	18.4	46.3	47.2	51.7

表 2 降雨量数据标准化					mm
王勿桥最大日降雨量	包信最大日降雨量	王勿桥次降雨量	包信次降雨量	流量/(m ³ /s)	
- 0.527	- 0.608	- 0.063	- 0.098	- 0.763	
- 0.103	- 0.079	- 0.643	- 0.653	- 0.431	
- 0.023	0.3461	- 0.310	- 0.164	- 0.164	
- 0.406	- 1.132	- 0.517	- 1.359	- 0.434	
- 0.786	- 1.050	- 0.574	- 0.974	- 0.537	
0.2314	- 1.132	- 0.379	- 0.510	0.1318	
2.1926	1.1587	2.6214	1.3142	1.6704	
- 0.600	0.0108	- 0.846	0.3560	- 0.258	
- 0.202	- 0.345	0.3328	0.1043	0.3629	
0.0624	1.0297	0.1266	1.4667	- 0.120	
- 0.307	- 0.384	0.1472	- 0.147	- 0.178	
- 0.307	- 0.384	1.1621	1.1448	0.0740	
2.7442	3.0029	1.8127	2.2410	3.1728	
- 1.034	0.3289	- 1.183	- 0.365	- 0.601	
0.1358	- 0.547	- 0.468	- 0.909	- 0.565	
- 0.157	- 0.401	- 0.681	- 0.948	- 0.685	
- 0.906	- 0.836	- 0.533	- 0.495	- 0.672	

表 3 间河降雨量模糊相似矩阵																	
	Z ₁	Z ₂	Z ₃	Z ₄	Z ₅	Z ₆	Z ₇	Z ₈	Z ₉	Z ₁₀	Z ₁₁	Z ₁₂	Z ₁₃	Z ₁₄	Z ₁₅	Z ₁₆	Z ₁₇
Z ₁	1.00	0.91	0.79	0.11	0.38	0.81	0.36	0.08	0.78	0.13	0.85	0.95	0.96	0.43	0.46	0.74	0.88
Z ₂	0.91	1.00	0.81	0.17	0.43	0.84	0.23	0.04	0.92	0.16	0.85	0.9	0.85	0.33	0.63	0.89	0.98
Z ₃	0.79	0.81	1.00	0.36	0.77	0.41	0.67	0.40	0.90	0.35	0.86	0.81	0.85	0.80	0.08	0.50	0.69
Z ₄	0.11	0.17	0.36	1.00	0.78	0.59	0.87	0.93	0.17	0.99	0.25	0.17	0.13	0.80	0.81	0.56	0.33
Z ₅	0.38	0.43	0.77	0.78	1.00	0.01	0.90	0.82	0.73	0.81	0.76	0.41	0.59	0.86	0.29	0.04	0.30
Z ₆	0.81	0.84	0.41	0.59	0.01	1.00	0.15	0.36	0.61	0.57	0.62	0.86	0.68	0.13	0.87	0.93	0.92
Z ₇	0.36	0.23	0.67	0.87	0.90	0.15	1.00	0.93	0.49	0.86	0.64	0.27	0.59	0.93	0.55	0.21	0.08
Z ₈	0.08	0.04	0.40	0.93	0.82	0.36	0.93	1.00	0.26	0.94	0.45	0.01	0.34	0.78	0.68	0.45	0.17
Z ₉	0.78	0.92	0.90	0.17	0.73	0.61	0.49	0.26	1.00	0.19	0.93	0.89	0.82	0.55	0.39	0.71	0.85
Z ₁₀	0.13	0.16	0.35	0.99	0.81	0.57	0.86	0.94	0.19	1.00	0.28	0.17	0.12	0.77	0.77	0.54	0.31
Z ₁₁	0.85	0.85	0.86	0.25	0.76	0.62	0.64	0.45	0.93	0.28	1.00	0.87	0.93	0.61	0.28	0.58	0.79
Z ₁₂	0.95	0.99	0.81	0.17	0.41	0.86	0.27	0.01	0.89	0.17	0.87	1.00	0.90	0.35	0.60	0.87	0.97
Z ₁₃	0.96	0.85	0.85	0.13	0.59	0.68	0.59	0.34	0.82	0.12	0.93	0.90	1.00	0.60	0.27	0.59	0.79
Z ₁₄	0.43	0.33	0.80	0.80	0.86	0.13	0.93	0.78	0.55	0.77	0.61	0.35	0.60	1.00	0.51	0.10	0.16
Z ₁₅	0.46	0.63	0.08	0.81	0.29	0.87	0.55	0.68	0.39	0.77	0.28	0.60	0.27	0.51	1.00	0.90	0.76
Z ₁₆	0.74	0.89	0.50	0.56	0.04	0.93	0.21	0.45	0.71	0.54	0.58	0.87	0.59	0.10	0.90	1.00	0.95
Z ₁₇	0.88	0.98	0.69	0.33	0.30	0.92	0.08	0.17	0.85	0.31	0.79	0.97	0.79	0.16	0.76	0.95	1.00

(4) 根据给定的阈值 $\lambda_0 = 0.91$,采用直接聚类法中的编网法进行聚类分析 ,得出最佳聚类为 :
 $U_1 = \{Z_1 , Z_2 , Z_3 , Z_9 , Z_{11} , Z_{12} , Z_{13} , Z_{17}\}; U_2 = \{Z_6 , Z_{15} , Z_{16}\}; U_3 = \{Z_4 , Z_8 , Z_{10}\}; U_4 = \{Z_5 , Z_7 , Z_{14}\};$

表 4 降雨量分类后各类均值

类别	\bar{x}_1	\bar{x}_2	\bar{x}_3	\bar{x}_4	\bar{y}_i
U_1	46.2	39.9	78.0	77.7	169.0
U_2	46.9	29.7	47.3	35.1	93.1
U_3	34.9	37.1	51.6	74.1	107.2
U_4	48.6	41.2	82.1	67.3	169.3

(5) 求模糊特征值 :得到的各类的均值和方差值如表 4、表 5 所示.

(6) 选择 3 场洪水进行预测 ,其资料及预测结果如表 6 所示.

表 5 降雨量分类后各类方差

类别	σ_{i1}^2	σ_{i2}^2	σ_{i3}^2	σ_{i4}^2
U_1	1088.7	714.6	1218.1	1410.8
U_2	26.7	18.4	30.7	66.9
U_3	76.2	422.1	311.5	2308.4
U_4	2115.9	449.4	5306.3	1600.1

表 6 阎河包信站洪峰流量预测结果

时间	王勿桥最大日 降雨量/mm	包信最大日 降雨量/mm	王勿桥次降 雨量/mm	包信次降 雨量/mm	包信洪峰流量 预测值/(m ³ /s)	包信洪峰流量 真实值/(m ³ /s)	预测精 度/%
1965 年 7 月 22 日	54.8	69.4	54.9	69.7	346	367	94.3
1963 年 8 月 17 日	43.0	31.9	65.1	55.3	180	231	77.8
1975 年 8 月 8 日	127.8	107.0	155.4	179.1	637	638	99.8

3 结 语

本文对如何克服水文预报中模糊性和随机性因素影响显著的困难进行了探讨 ,提出了方法和模型 ,并在资料较少的情况下建模取得了较好效果 ,此方法同样适应其它流域的水文要素短期预报 .本例选取的相关预测因子对降雨历时不到一天的情况预报困难 ,笔者认为可以根据气象情况

来对降雨历时人为扩展至一天后进行预报.

参 考 文 献 :

[1] 韩立岩 ,汪培庄 .应用模糊数学[M].北京 :首都经济贸易大学出版社 ,1998.
[2] 闫家杰 ,赵万忠 ,迟凤起 .模糊数学基础及应用初阶[M].郑州 :河南教育出版社 ,1993.
[3] 段银田 ,孟树锁 .数值算法和程序过程[M].北京 :海洋出版社 ,1993.

Hydrological Forecasting Based on Fuzzy Causal Classification and it 's Application

ZHOU Nian - lai , WU Ze - ning

(College of Hydraulilc & Environmental Engineering Zhengzhou University of Technology Zhengzhou 450002 ,China)

Abstract : Based on remarkable indefinite factors and causality in hydrology ,the paper applies the method of fuzzy classification and classifies various historical floods to overcome indefinite factors effect ,then sets up causality of forecasting factor and forecasting result by working out fuzzy character ,and at last makes a forecast .Feasibility of the model and method is validated by a practical application.

Key words hydrological forecasting ; fuzzy cluster analysis ; membership functions ; Lagrange 's inverting value