

# 模糊模式识别法在河流分类中的应用<sup>\*</sup>

郑成德 李志斌

(大连铁道学院数学教研室, 116028)

**摘要** 探讨了应用模糊模式识别法对河流进行分类的可行性,并以辽宁省河流为例,叙述了这种方法的研究程序及计算步骤。研究表明,采用模糊模式识别方法对河流进行分类,结果合理,方法可行,有一定的参考价值。

**关键词** 相对隶属度 权重 模糊模式识别 河流分类

## The Application of Fuzzy Recognition Model for Rivers Classification

Zheng Chengde Li Zhibin

(Dept of Basic Science, Dalian Railway Institute, 116028)

**Abstract** The possibility of rivers classification with Fuzzy Recognition Model is discussed in this paper and the rivers in Liaoning province are classified so as to show the procedure of this method. Results show that the method is reasonable and practicable, so it is worthy extending.

**Keywords** relative membership; weight; fuzzy recognition model; rivers classification

### 1 引言

自1965年查德提出模糊集合概念以来,模糊集理论已得到广泛应用。河流分类一向是自然地理学和水文研究中的一个重要问题,但传统描述的分类,存在主观随意性。本文首次运用模糊模式识别方法,对辽宁省的河流进行分类,试图达到客观、科学分类的目的;以期为水文区划和水利化区划提供科学依据。

### 2 模糊模式识别理论模型

#### 2.1 指标值的规范化处理

设有 $n$ 个待识别的对象组或对象集 $X = \{x_1, x_2, \dots, x_n\}$ ;又有据以识别的 $m$ 个指标组成指标集 $x_j = (x_{1j}, x_{2j}, \dots, x_{mj})^T (j = 1, 2, \dots, m)$ ;则有实测指标矩阵 $X = (x_{ij})_{m \times n}$ 。若 $m$ 项指标按 $c$ 级标准模式进行分类,则有指标标准矩阵 $Y = (y_{ic})_{m \times c}, 2 \leq c \leq m$ 。

根据下述公式求出对象指标相对隶属度 $r_{ij} (i = 1, 2, \dots, m; j = 1, 2, \dots, n)$ 。如对数值越大级别越高的指标公式为:(设级别越高模糊概念越劣)

$$r_{ij} = \begin{cases} 0, & x_{ij} > y_{i1} \\ (x_{ij} - y_{i1}) / (y_{ic} - y_{i1}), & y_{i1} < x_{ij} < y_{ic} \\ 1, & y_{ic} < x_{ij} \end{cases} \quad (1)$$

对数值愈大级别愈低的指标,公式为:

<sup>\*</sup> 本文于1996年3月14日收到

$$r_{ij} = \begin{cases} 1, & x_{ij} = y_{ic} \\ (y_{il} - x_{ij}) / (y_{il} - y_{ic}), & y_{ic} < x_{ij} < y_{il} \\ 0, & y_{il} < x_{ij} \end{cases} \quad (2)$$

于是得到对象指标相对隶属度矩阵

$$R = (r_{ij})_{m \times n}, \quad 0 \leq r_{ij} \leq 1, \quad (i = 1, 2, \dots, m; j = 1, 2, \dots, n).$$

类似地, 可得各类标准指标相对隶属度  $s_{ih}$ , ( $i = 1, 2, \dots, m; h = 1, 2, \dots, c$ )。

$$s_{ih} = (y_{ih} - y_{il}) / (y_{ic} - y_{il}), \quad y_{il} < y_{ih} < y_{ic} \quad (3)$$

同理得到标准指标相对隶属度矩阵

$$S = (s_{ih})_{m \times c}, \quad 0 \leq s_{ih} \leq 1, \quad (i = 1, 2, \dots, m; h = 1, 2, \dots, c).$$

其中

$$s_1 = (s_{11}, s_{21}, \dots, s_{m1})^T = (1, 1, \dots, 1)^T;$$

$$s_c = (s_{1c}, s_{2c}, \dots, s_{mc})^T = (0, 0, \dots, 0)^T.$$

### 2.2 权重的确定

各识别对象指标权重的确定是重要、困难的一环, 可以用层次分析法或目前较为流行的指数加权法确定。在模糊集合论中隶属度也可定义为权重。矩阵  $R_{m \times n}$  给出待识别对象的  $m$  个指标超过 1 级标准模式的相对隶属度值, 因此该矩阵不仅描述了待识别对象各指标的超标情况, 而且表示了各指标的权重, 其按列归一化矩阵  $W = (w_{ij})_{m \times n}$ , 称为超标权重矩阵。满足

$$\sum_{i=1}^m w_{ij} = 1, \quad j = 1, 2, \dots, n.$$

其中

$$w_{ij} = r_{ij} / \sum_{i=1}^m r_{ij} \quad i = 1, 2, \dots, m; j = 1, 2, \dots, n \quad (4)$$

上式表示了不同对象各指标的权重是不同的。它根据不同对象各指标超标情况确定相应对象指标的权重, 就避免了其它方法只用一个权重划分对象指标的不合理之处, 据此我们认为用上述公式计算权重更合理、更可靠。

### 2.3 建立模糊模式识别理论模型

将  $n$  个对象根据其  $m$  个指标特征按  $c$  类标准模式加以识别, 则模糊识别矩阵为:

$$U = (u_{hj})_{c \times n}, \quad \text{式中 } u_{hj} \text{ 为对象 } j \text{ 隶属于标准模式 } h \text{ 的相对隶属度, 满足条件: } 0 \leq u_{hj} \leq 1,$$

$$\sum_{h=1}^c u_{hj} = 1 \quad (5)$$

于是得到对象  $j$  的  $m$  个指标的权向量  $w_j = (w_{1j}, w_{2j}, \dots, w_{mj})$ , 指标特征向量  $r_j = (r_{1j}, r_{2j}, \dots, r_{mj})^T$ ,  $h$  类标准模式  $m$  个指标特征向量  $s_h = (s_{1h}, s_{2h}, \dots, s_{mh})^T$

设第  $j$  个对象与  $h$  类标准模式的差异用广义距离

$$d_{hj} = |w_j |r_j - s_h| = \left[ \sum_{j=1}^m (w_{ij} |r_{ij} - s_{ih}|)^p \right]^{\frac{1}{p}} \quad \text{表示.}$$

式中  $p$  为距离参数,  $p = 1$  表示海明距离,  $p = 2$  表示欧氏距离。

根据模糊识别矩阵  $U_{c \times n}$ , 第  $j$  个对象同时以相对隶属度  $u_{hj}$  隶属于第  $h$  类,  $h = 1, 2, \dots, c$ ;  $\sum_{h=1}^c u_{hj} = 1$ , 故可以将  $u_{hj}$  作为对象  $j$  关于  $h$  类标准的权重, 引入广义权距离,  $D_{hj} = u_{hj} \cdot d_{hj}$ 。广义权距离更合理地描述了识别对象  $j$  与标准模式  $h$  之间的差异。

为求解最优模糊识别矩阵, 我们提出目标函数为: 全体对象对于各标准模式间的权广义距离平方和最小。即:

$$\min \{F(u_{hj})\} = \min \left\{ \sum_{j=1}^n \sum_{h=1}^c \left[ u_{hj} \left( \sum_{i=1}^m (w_{ij} |r_{ij} - s_{ih}|)^p \right)^{\frac{1}{p}} \right]^2 \right\}$$

$$= \sum_{j=1}^n \min \left\{ \sum_{h=1}^c u_{hj}^2 \left[ \sum_{i=1}^m (w_{ij} |r_{ij} - s_{ih}|)^p \right]^{\frac{2}{p}} \right\} \tag{6}$$

目标函数式(6)应满足等式约束条件式(5)。

为此,可以构造L agrange 函数,将等式约束求极值化为无条件极值问题。相应的L agrange 函数为:

$$L(u_{hj}, \lambda) = \sum_{h=0}^c u_{hj}^2 \left[ \sum_{i=1}^m (w_{ij} |r_{ij} - s_{ih}|)^p \right]^{\frac{2}{p}} - \lambda \left( \sum_{h=1}^c u_{hj} - 1 \right)$$

解之,得:  $(h= 1, 2, \dots, c; j= 1, 2, \dots, n)$

$$u_{hj} = 1 \left/ \sum_{k=1}^c \left[ \frac{\sum_{i=1}^m (w_{ij} |r_{ij} - s_{ih}|)^p}{\sum_{i=1}^m (w_{ij} |r_{ij} - s_{ik}|)^p} \right]^{\frac{2}{p}} \right. \tag{7}$$

公式(7)即为我们建立的模糊模式识别理论模型。

### 2.4 运用模糊模式识别模型的原则

应该指出,用式(7)进行识别,  $k$  的取值范围是动态的,要根据矩阵  $R_{m \times n}$  中对象  $j$  的  $m$  个指标与矩阵  $S_{m \times c}$  作比较后决定。

首先将对象  $j$  指标  $i$  的相对隶属度  $r_{ij}$  逐个地与矩阵  $S_{m \times c}$  中指标  $i$  的各类标准相对隶属度值  $s_{i1}, s_{i2}, \dots, s_{ic}$  进行比较,若落入  $s_i$  各类标准相对隶属度值相邻区间内的最小类别  $m_j$ , 最大类别  $M_j$ , 则  $m_j, M_j$  就是  $k$  的取值范围,即  $k = m_j, m_j + 1, m_j + 2, \dots, M_j$ 。由于在类别  $m_j$  与  $M_j$  之间可能出现  $m$  个指标无一归属的类别,但  $k$  的取值仍然保持以 1 为等差的正整数序列,  $h$  取值也以 1 为间隔的正整数列:  $m_j, m_j + 1, m_j + 2, \dots, M_j$ , 且  $1 < m_j < \dots < M_j < c$ , 而不考虑出现间断值。

式(7)中  $h$  的取值范围与  $k$  相同,显然有:

$$u_{hj} = 0, \text{ 当 } h < m_j \text{ 或 } h > M_j, \text{ 且 } \sum_{h=m_j}^{M_j} u_{hj} = 1$$

则模糊模式识别理论模型(7)的完整形式为

$$u_{hj} = \begin{cases} 1, & h = m_j = M_j \\ 1 \left/ \sum_{k=m_j}^{M_j} \left[ \frac{\sum_{i=1}^m (w_{ij} |r_{ij} - s_{ih}|)^p}{\sum_{i=1}^m (w_{ij} |r_{ij} - s_{ik}|)^p} \right]^{\frac{2}{p}} \right., & m_j < h < M_j \\ 0, & h < m_j \text{ 或 } h > M_j \end{cases} \tag{8}$$

$h = 1, 2, \dots, c; j = 1, 2, \dots, m$

若  $u_{lj} = \max_h \{u_{hj}\}$  (最大隶属度原则) (9)

则待识别对象  $j$  划归  $l$  类  $(1 \leq l \leq c)$ 。

## 3. 辽宁省河流概况和水文要素的选择

### 3.1 辽宁省河流概况

辽宁省境内大小河流约 200 条,其中流域面积在 1000km<sup>2</sup> 以上的干支流河流有 45 条。本文选出 15 条有代表性的河流进行模糊模式识别。

辽河发源于河北省七老图山脉光头山,经内蒙、吉林,由东辽河和西辽河相汇于我省福德店附近称辽河。在我省境内主要支流有:招苏台河、清河、柳河、秀水河、养息牧河等。招苏台河位于辽北,是辽河东侧支流;清河是铁岭站以北辽河东侧支流;柳河与养息牧河均为辽河西侧支流,分别于新民、巨流河附近注入辽河干流。在我省境内辽河总流域面积 41836 km<sup>2</sup>,河长约 480km。浑河发源于清原滚马岭,总流域面积 11480km<sup>2</sup>,河长 415km;太子河发源于新宾红石砬子山,总流域面积 13880km<sup>2</sup>,河长 413km;两河相汇于三岔河称大辽河。绕阳河发源于阜新骆驼山,下游与双台子河汇流后注入渤海,河长 283km,流域面积

9950km<sup>2</sup>。鸭绿江发源于吉林长白山主峰白头山,为中朝两国界河,经丹东注入黄海,在我省境内流域面积 16620km<sup>2</sup>,河长约 200km。大凌河发源于凌源打鹿沟,于锦县注入渤海,河长 397km,流域面积 23550km<sup>2</sup>。小凌河发源于朝阳助安喀喇山,于锦州注入渤海,河长 206km,流域面积 5480km<sup>2</sup>。大洋河发源于岫岩新开岭,河长 202km,流域面积 6200km<sup>2</sup>。碧流河发源于盖州七盘岭南麓,河长 159km,流域面积 2820km<sup>2</sup>。六股河位于渤海西岸,河长 153km,流域面积 3080km<sup>2</sup>。

### 3.2 水文要素的选择

影响河流分类的因素很多,如流量的变化,泥沙的大小,冰冻期的长短等。其中各种形式的流量变化是特别重要的,它能反映出不同自然地理环境的河流特性。而影响辽宁省河川径流形成和变化的因素归纳起来有三方面,即气候因素,流域的下垫面因素,人类改造自然因素。其中气候因素是主要的,通常说河流是气候的产物。因此,本文选取以下水文要素作为辽宁省河流分类的依据:

A. 多年平均降水量(mm), B.  $P=75\%$ 时多年平均降水量(mm), C. 连续最大四个月多年平均降水量(mm); D. 多年平均水面蒸发量(mm); E. 多年平均径流深(mm), F.  $P=75\%$ 时每平方公里多年平均径流量( $\times 10^4\text{t}/\text{km}^2$ ), G. 汛期(6~9月)每平方公里多年平均径流量( $\times 10^4\text{t}/\text{km}^2$ ), H. 多年平均悬移质含沙量( $\text{kg}/\text{m}^3$ ), I. 水质有机污染综合等级, J. 水质五毒污染综合等级。

## 4. 辽宁省河流分类的模糊模式识别

### 4.1 水文资料

将待识别的 15 条河流代表测站水文要素特征值列于表 1:

表 1 各河水文要素值(单位同前)

序号	河名	站名	要素A	要素B	要素C	要素D	要素E	要素F	要素G	要素H	要素I	要素J
1	辽河	铁岭	712 0	625 3	521 2	1450 7	36 2	2 020	2 683	3 87	2	4
2	浑河	抚顺	777. 5	671. 3	559. 4	1088 1	298 9	20 335	21 083	0 19	5	3
3	太子河	本溪	781. 4	684 2	554 7	1391 5	382 7	27 058	27 521	0 27	5	3
4	招苏台河	王宝庆	522 3	454 8	392 7	1476 5	85 6	4 626	4 264	2 44	4	3
5	清河	开原	699 1	597. 1	575 0	1489 5	232 0	14 396	18 081	1 02	1	2
6	秀水河	彭家堡	589 6	531. 3	450 0	1790 6	45 1	1 803	3 711	2 61	4	3
7	养息牧河	小荒地	582 9	526 2	453 2	1659 1	51. 7	2 471	4 716	5 24	5	3
8	柳河	新民	628 3	526 2	479 0	1659 1	59. 8	1 042	3 655	22 6	3	3
9	绕阳河	王回窝堡	619 2	497. 4	473 0	1698 8	54 1	1 517	4 578	0 41	4	2
10	大凌河	锦县	584 5	507. 2	456 3	1763 7	88 7	5 337	6 421	13 3	5	4
11	小凌河	锦州	576. 4	456 0	448 5	1763 7	130 7	6 549	10 562	6 40	4	3
12	鸭绿江	荒沟	1193 0	977. 9	917. 3	1065 4	457. 8	36 629	32 660	0 041	1	1
13	大洋河	沙里寨	1053 2	977. 9	817. 8	1005 1	445 5	28 482	33 817	0 30	1	2
14	碧流河	小宋家屯	721 1	575 3	550 0	1387 3	354 6	21 994	30 075	0 53	1	1
15	六股河	绥中	619 2	482 8	488 3	1648 5	201 0	9 840	18 185	1 72	3	2

注:最后两项要素 I, J 特征值为 1980 年监测数据。

### 4.2 河流分类标准

为了对辽宁省河流进行模糊模式识别,根据辽宁省河流水文特性,参考全国河流水文要素特征值特



$$S = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.250 & 0.250 & 0.25 & 0.25 & 0.25 & 0.25 & 0.25 & 0.25 & 0.25 & 0.25 \\ 0.50 & 0.50 & 0.50 & 0.50 & 0.50 & 0.50 & 0.50 & 0.50 & 0.50 & 0.50 \\ 0.75 & 0.75 & 0.75 & 0.75 & 0.75 & 0.75 & 0.75 & 0.75 & 0.75 & 0.75 \\ 1.00 & 1.0 & 1.0 & 1.0 & 1.0 & 1.0 & 1.0 & 1.0 & 1.0 & 1.0 \end{pmatrix}^T$$

将矩阵  $R_{10 \times 15}$  按列归一化, 得各河流诸水文要素超标权重矩阵  $W_{10 \times 15}$  (其中鸭绿江各项指标均未超标, 故赋予等权重)。

$$W = \begin{pmatrix} 0.095 & 0.127 & 0.134 & 0.128 & 0.160 & 0.116 & 0.111 & 0.102 \\ 0.082 & 0.112 & 0.112 & 0.126 & 0.148 & 0.099 & 0.095 & 0.098 \\ 0.101 & 0.151 & 0.165 & 0.128 & 0.131 & 0.109 & 0.102 & 0.097 \\ 0.085 & 0 & 0.131 & 0.080 & 0.152 & 0.124 & 0.109 & 0.112 \\ 0.145 & 0.078 & 0.020 & 0.115 & 0.115 & 0.124 & 0.117 & 0.117 \\ 0.145 & 0.074 & 0.010 & 0.115 & 0.122 & 0.124 & 0.115 & 0.120 \\ 0.141 & 0.080 & 0.024 & 0.117 & 0.099 & 0.116 & 0.106 & 0.113 \\ 0.061 & 0 & 0 & 0.031 & 0.015 & 0.033 & 0.069 & 0.121 \\ 0.036 & 0.252 & 0.269 & 0.096 & 0 & 0.093 & 0.117 & 0.060 \\ 0.109 & 0.126 & 0.135 & 0.064 & 0.058 & 0.062 & 0.059 & 0.060 \\ 0.116 & 0.105 & 0.119 & 0.10 & 0 & 0.225 & 0.142 \\ 0.118 & 0.095 & 0.121 & 0.10 & 0 & 0.244 & 0.151 \\ 0.110 & 0.095 & 0.108 & 0.10 & 0 & 0.222 & 0.128 \\ 0.134 & 0.111 & 0.123 & 0.10 & 0 & 0.170 & 0.150 \\ 0.133 & 0.099 & 0.096 & 0.10 & 0 & 0.055 & 0.096 \\ 0.134 & 0.097 & 0.102 & 0.10 & 0 & 0.082 & 0.115 \\ 0.122 & 0.093 & 0.086 & 0.10 & 0 & 0 & 0.070 \\ 0 & 0.111 & 0.091 & 0.10 & 0 & 0.002 & 0.025 \\ 0.100 & 0.111 & 0.092 & 0.10 & 0 & 0 & 0.082 \\ 0.033 & 0.083 & 0.062 & 0.10 & 1.00 & 0 & 0.041 \end{pmatrix}$$

对不同的样本  $j, k$  的上限值  $M_j$ , 下限值  $m_j$  不一定相同。现以  $j=1$  的辽河为例简述  $M_{j, m_j}$  的确定。当  $j=1$  时由矩阵  $R_{10 \times 15}$  可知:

$$r_{j=1} = (r_{11}, r_{21}, \dots, r_{10,1})^T = (0.655, 0.562, 0.697, 0.585, 1.00, 0.999, 0.976, 0.421, 0.25, 0.75)^T$$

将向量  $r_{j=1}$  中的  $r_{11}=0.655$  与矩阵  $S_{10 \times 5}$  的第一行元素值作一比较, 可知  $r_{11}$  介于 0.500 与 0.750 之间, 即落入 3, 4 级之间。将  $r_{21}=0.562$  与矩阵  $S_{10 \times 5}$  中第二行元素值比较,  $r_{21}$  处于 0.500 与 0.750 之间, 即落入 3, 4 级之间。类似地, 可知  $r_{31}$  落入 3, 4 级之间, ...,  $r_{10,1}$  落入 4 级。综合起来, 可知  $k$  的上限值  $M_{1=5}$ , 下限值  $m_1=2$ 。

距离参数  $P$  采用为 2。

于是对于  $j=1$ , 模型(8)变为

$$u_{h1} = \begin{cases} 0, & h < 2 \\ 1 / \left( \sum_{k=2}^5 \frac{\sum_{i=1}^{10} (w_{i1} |r_{i1} - s_{ih}|)^2}{\sum_{i=1}^{10} (w_{i1} |r_{i1} - s_{ik}|)^2} \right), & 2 \leq h \leq 5 \end{cases}$$

将有关数据代入上式, 得到:

$$u_{11} = 0, u_{21} = 0.098, u_{31} = 0.167, u_{41} = 0.461, u_{51} = 0.274$$

对  $j = 2, 3, \dots, 15$  进行计算, 得到辽宁省 15 条河流对各个级别的相对隶属度矩阵  $U_{5 \times 15}$ :

$$U_{5 \times 15} = \begin{pmatrix} 0 & 0.097 & 0.161 & 0.039 & 0.053 & 0 & 0 & 0 \\ 0.098 & 0.133 & 0.389 & 0.048 & 0.102 & 0.059 & 0 & 0 \\ 0.167 & 0.354 & 0.246 & 0.098 & 0.609 & 0.097 & 0.119 & 0.134 \\ 0.461 & 0.266 & 0.118 & 0.448 & 0.236 & 0.496 & 0.368 & 0.406 \\ 0.274 & 0.150 & 0.086 & 0.367 & 0 & 0.348 & 0.513 & 0.460 \\ 0.028 & 0 & 0 & 1.00 & 0.995 & 0.086 & 0.051 \\ 0.035 & 0 & 0 & 0 & 0.005 & 0.757 & 0.069 \\ 0.071 & 0 & 0.151 & 0 & 0 & 0.157 & 0.532 \\ 0.557 & 0.370 & 0.517 & 0 & 0 & 0 & 0.164 \\ 0.309 & 0.630 & 0.322 & 0 & 0 & 0 & 0.184 \end{pmatrix}$$

由最大隶属度原则, 按矩阵  $U_{5 \times 15}$  计算辽宁省河流分类级别为: 鸭绿江、大洋河为 1 级, 水资源极丰富; 太子河、碧流河为 2 级, 水资源较丰富; 浑河、清河、六股河为 3 级, 水资源尚丰富; 辽河、招苏台河、秀水河、绕阳河、小凌河为 4 级, 水资源中丰富; 养息牧河、柳河、大凌河为 5 级, 水资源较干枯。

## 5 结语

1. 本文的模糊模式识别法给出了各河流隶属于各级别的隶属度, 这对于我们正确、全面地掌握河流水资源丰富状况, 实施科学管理大有裨益。

2. 模糊模式识别法对河流这一复杂的模糊体系具有独到之处, 它强调了河流的模糊值, 较为深刻地刻画了河流的客观属性。它杜绝了主观臆断性, 对河流分类与实际情况相符。

3. 模糊模式识别法充分考虑河流各等级之间的模糊性, 包括人们对其认识的模糊性, 同时刻画了各水文要素的不同权重, 较好地描述了河流“丰与枯”的相对性和渐变性, 并且能客观、全面、正确地表达多因素共同作用的综合结果, 分辨率较高。同时能反映出处于同一级别不同河流之间的“丰枯”差异, 可为管理、决策者提供具体的科学依据。

4. 模糊模式识别计算方法简便, 结论符合客观实际, 因而是河流分类的一种理想、有效的决策方法, 值得在生产、实践中推广使用。

## 参考文献

- 1 贺仲雄. 模糊数学及其应用. 天津: 天津科学技术出版社, 1983.
- 2 陈守煜. 系统模糊决策理论与应用. 大连: 大连理工大学出版社, 1994.
- 3 田锡堂. 辽宁省河流的模糊聚类分析. 沈阳: 辽宁师范大学学报(自然科学版), 1989, (2)
- 4 辽宁省水文总站. 辽宁省地表水资源, 1984(内部资料)