

# 引入 SPA 的 6 级模糊模式识别模型在水质综合评价中的应用

周向华<sup>1</sup>, 杨侃<sup>1</sup>, 蒋立伟<sup>2</sup>, 陈静<sup>2</sup>, 刘炜<sup>3</sup>, 朱大伟<sup>3</sup>

(1. 河海大学水文水资源学院, 江苏南京 210098; 2. 宜兴水利农机局, 江苏宜兴 214207;

3. 江苏省水利勘测设计研究院有限公司, 江苏扬州 225127)

**摘要:**为了解决普遍意义上的水质模糊识别理论不能评价指标数据超过 5 级标准值和最大隶属度原则的缺点, 本文提出 6 级水质计算方法来改进模糊模式识别模型。再者, 引入 Set Pair Analysis (SPA) 理论处理等级标准间的不确定性信息。从而构建引入 SPA 的 6 级模糊模式识别模型, 将其运用于河流实例中。评价结果表明: 引入 SPA 的 6 级模糊模式识别模型更能准确地确定水质综合等级, 为治理河流水质污染提供理论支持。

**关键词:**模糊模式识别模型; 集对分析理论; 6 级相对最优隶属度; 水质评价

中图分类号: X824; X522

文献标识码: A

文章编号: 1672-643X(2016)06-0001-06

## Application of 6 – level fuzzy pattern recognition model based on set pair analysis in river water quality evaluation

ZHOU Xianghua<sup>1</sup>, YANG Kan<sup>1</sup>, JIANG Liwei<sup>2</sup>, CHEN Jing<sup>2</sup>, LIU Wei<sup>3</sup>, ZHU Dawei<sup>3</sup>

(1. College of Hydrology and Water Resources, Hohai University, Nanjing 210098, China;

2. Yixing Agricultural Machinery Bureau of Water Conservancy, Yixing 214207, China;

3. Water Resources Survey and Design Research Institute Co., Ltd., Yangzhou 225127, China)

**Abstract:** In order to solve the general evaluation shortcomings of the fuzzy recognition theory on 5 level standard values and the maximum membership degree principle, this study puts forward the calculation method of 6 level water quality to improve the model of fuzzy pattern recognition, furthermore, the Pair Analysis Set (SPA) theory was introduced to deal with the uncertainty information between the different standard levels, then the fuzzy pattern recognition model was established to be applied in river systems. The evaluation results showed that the 6 level fuzzy pattern recognition model based on SPA was more accurate to determine the water quality comprehensive grade, which could provide theoretical support for the treatment of river water pollution.

**Key words:** fuzzy pattern recognition model; set pair analysis; 6 – level relative optimal membership degree; water quality evaluation

## 1 研究背景

近年来,随着城镇工业化程度的不断加深,在流经城镇的河段,有机污染物排放量增加,导致水生生态系统恶化,严重威胁着居民的饮用水水质安全。恢复河流生态健康已成为重中之重。而如何开展合理的河流污染评价是其中的首要工作。国内开展此类研究起于 20 世纪 80 年代<sup>[1]</sup>。截至当前,确定权重

主要包括主成分分析法、污染物超标法和最优权法等<sup>[2]</sup>。其中,模糊模式识别理论由于充分考虑到河流系统的复杂多变性,利用最优相对隶属度处理不确定性问题而获得了普遍的认可和推广。

但是,传统意义上,该理论包含两个缺点。第一,目前,水质级别上限为 5 级,要求样本数据不能超过 5 类水质的规定值。然而,在水质较差的城镇河段,当部分指标数据超出 5 类水质标准时,传统的

收稿日期:2016-06-26; 修回日期:2016-08-15

基金项目:国家重点基础研究发展计划(973 计划)(2012CB417006);“十一五”国家科技支撑计划(2009BAC56B03)

作者简介:周向华(1992-),女,湖北潜江人,硕士研究生,研究方向为水文学及水资源。

通讯作者:杨侃(1965-),男,江苏南京人,博士,教授,博士生导师,研究方向为水文学及水资源。

水质模糊识别理论应用受到限制,评价结果与实际情况不相符。第二,水质评价模型多采用最大隶属度原则。该原则不仅不能解决隶属度大小相差不大或者相等的问题,而且难以处理评价标准中存在的确定性信息,计算水质综合指数的大小。集对分析(SPA)理论<sup>[3]</sup>能够有效解决这些问题。高苏蒂等<sup>[4]</sup>结合模糊逻辑理论与集对分析理论,较好地将其应用到了构建水质综合指数评价的集对分析模型中;杨红卫等<sup>[5]</sup>将该方法与最差因子法作比较,分析得出集对分析能更清楚地解决水质污染程度的问题;孟宪萌等<sup>[6]</sup>利用基于熵权的集对分析模型得出了该方法有利于解决各评价指标的不确定性;周宇哲等<sup>[7]</sup>运用改进的集对分析模型解决了水环境的不确定性和模糊性。

因此,为解决上述两个问题,一方面,本文提出改进的基于模糊模式识别理论的6级水质综合评价方法。另一方面,引入SPA理论处理评价标准中存在的确定性信息,对水质评价结果进行数据分析。

## 2 引入 SPA 的模糊模式识别模型

模糊模式识别理论的中心思想是根据模糊数学中的相对隶属度及隶属度函数理论,参考连续统中定义指标的相对隶属函数,计算最优值,然后运用判定法则来确定相关对象的综合等级。

### 2.1 无量纲化处理

假设  $X$  为样本集矩阵,  $m$  为样本总数,  $n$  为指标总数,则样本  $i$  在指标  $j$  上的数据表示成  $x_{ij}$  ( $i = 1, 2, \dots, m; j = 1, 2, \dots, n$ )。同样地,标准集矩阵  $Y$ , 在指标  $j$  以及等级  $h$  上的标准值表示成  $y_{hj}$  ( $h = 1, 2, \dots, c; j = 1, 2, \dots, n$ )。其中,  $c$  表示评价等级数。根据模糊模式识别理论对矩阵  $X, Y$  进行无量纲化处理,得到无量纲化矩阵  $R, S$ 。对于效益型指标,计算公式为:

$$r_{ij} = \begin{cases} 1 & (x_{ij} \geq y_{1j}) \\ \frac{x_{ij} - y_{cj}}{y_{1j} - y_{cj}} & (y_{1j} > x_{ij} > y_{cj}) \\ 0 & (x_{ij} \leq y_{cj}) \end{cases} \quad (1)$$

对于成本型指标,计算公式为

$$r_{ij} = \begin{cases} 1 & (x_{ij} \leq y_{1j}) \\ \frac{x_{ij} - y_{cj}}{y_{1j} - y_{cj}} & (y_{1j} < x_{ij} < y_{cj}) \\ 0 & (x_{ij} \geq y_{cj}) \end{cases} \quad (2)$$

对于评价标准集  $Y$ ,无量纲化过程为

$$S_{hj} = \begin{cases} 1 & (h = 1) \\ \frac{y_{hj} - y_{cj}}{y_{1j} - y_{cj}} & (h = 2, 3, 4) \\ 0 & (h = 5) \end{cases} \quad (3)$$

### 2.2 指标权重确定

目前,从主观角度来说,确定指标权重的代表性方法为AHP法,其理论简单,计算简捷,应用全面。然而这种方法决策者的主观意识占有决定性的地位,严重忽略了客观数据的重要性<sup>[8]</sup>。而客观性的代表方法之一是污染物超标法,由于其完全依据样本数据获取指标对河流水质的污染程度,更为符合实际情况。因此,成为目前推广最为普遍的一种权重计算方法<sup>[9]</sup>。

其基本思想是依据污染物浓度超标度<sup>[10-11]</sup>的大小,确定不同样本情况下的各指标权重。若污染物超标程度  $b_{ij}$  越大,则相应的权重  $w_{ij}$  越大。反之,亦然。 $b_{ij}$  的计算公式如下:

$$b_{ij} = \begin{cases} \frac{x_{ij}}{a_j} & j \text{ 为非 DO 指标} \\ \frac{DO_f - x_{ij}}{DO_f - a_j} & j \text{ 为 DO 指标} \end{cases} \quad (4)$$

式中:  $a_j$  为第  $j$  个指标的标准数值;  $DO_f$  为饱和溶解氧。

计算各指标权重  $w_{ij}$  的公式为

$$w_{ij} = \frac{b_{ij}}{\sum_{j=1}^n b_{ij}} \quad (5)$$

### 2.3 改进的6级指标隶属度

根据模糊模式识别理论,水质综合评价中关键的步骤是计算各样本数据的指标隶属度。但是,由于水质数据中存在超出5类水质标准值的情况,因此,通过式(3)对标准集  $Y$  进行的无量纲化已经不适用。本文提出,取评价标准中相邻等级的平均值作为6级标准值  $s_{hj}^*$ ,其计算公式为

$$s_{hj}^* = \begin{cases} 1 & (h = 1) \\ \frac{1/2(y_{(h-1)j} + y_{hj}) - y_{cj}}{y_{1j} - y_{cj}} & (h = 2, 3, 4, 5) \\ 0 & (h = 6) \end{cases} \quad (6)$$

接下来,利用第  $i$  个样本,第  $j$  个指标的无量纲化值  $r_{ij}$  与相应标准集的无量纲化值  $s_{hj}^*$  进行比较,确定所属级别  $h_{ij}$ ,公式为:

$$h_{ij} = \begin{cases} 1 & (r_{ij} = 1) \\ t & (s_{(t+1)j} > r_{ij} \geq s_{tj} \text{ 且 } r_{ij} \neq 0) \\ 6 & (r_{ij} = 0) \end{cases} \quad (7)$$

紧接着,找出各样本  $h_{ij}$  的最大值  $\max_i = \max(h_{ij})(j = 1, 2, \dots, n)$  和最小值  $\min_i = \min(h_{ij})(j = 1, 2, \dots, n)$ , 则每个样本所处的水质范

围为  $[\min_i, \max_i]$ 。第  $i$  个样本对不同水质等级  $h$  的最优相对隶属度  $u_{ih}$  的计算公式如下:

$$u_{ih} = \begin{cases} \left[ \frac{\sum_{j=1}^n [w_{ij}(r_{ij} - s_{hj}^*)]^p}{\sum_{j=1}^n [w_{ij}(r_{ij} - s_{kj}^*)]^p} \right]^{\frac{2}{p}} & (\min_i \leq h \leq \max_i) \\ 0 & (h < \min_i \text{ 或 } h > \max_i) \end{cases} \quad (8)$$

式中:  $u_{ih}$  取值越大,表示所属等级  $h$  的程度越高;  $p$  为距离参数,  $p = 1$  反映一维距离,  $p = 2$  反映二维距离。

## 2.4 引入SPA的水质综合评价

传统的水质模糊识别理论在判定样本水质级别时,多采用最大隶属度原则。该原则不仅不能解决隶属度大小相差不大或者相等的问题,而且不同河段的水质评价结果难以进行数值比较。而SPA主要用于解决不确定性问题<sup>[3]</sup>。其核心思想是把确定、不确定视为一个系统,将确定性分为“同一”与“对立”两个方面,将不确定性称为“差异”,从同、异、反三方面分析事物及系统<sup>[12]</sup>。

SPA将样本和分级标准看成两个相互关联的集合,通过联系数计算公式,对样本的综合水质等级做定性、定量的描述。联系数计算公式如下:

$$D_i = u_{i1} \cdot i_1 + u_{i2} \cdot i_2 + u_{i3} \cdot i_3 + u_{i4} \cdot i_4 + u_{i5} \cdot i_5 + u_{i6} \cdot i_6 \quad (9)$$

$$u_{i1} + u_{i2} + u_{i3} + u_{i4} + u_{i5} + u_{i6} = 1 \quad (10)$$

根据联系数隶属范围,依据“均分原则”<sup>[12-13]</sup>,首先把阈值空间  $[-1.0, 1.0]$  平均分割为5份,从右到左,将6个节点值赋给  $i_1, i_2, i_3, i_4, i_5$  和  $i_6$ 。同理,采用“均分原则”<sup>[13-14]</sup>,把取值空间平均分割为6份,形成6个子空间,分别为  $[-1, -0.667)$ ,  $[-0.667, -0.333)$ ,  $[-0.333, 0)$ ,  $[0, 0.333)$ ,  $[0.333, 0.667)$ ,  $[0.667, 1]$ 。具体等级判定标准为:当  $D_i < -0.667$  时,样本  $i$  的水质综合等级为VI;当  $-0.667 \leq D_i < -0.333$  时,样本  $i$  的水质综合等级为V;当  $-0.333 \leq D_i < 0$  时,样本  $i$  的水质综合等级为IV;当  $0 \leq D_i < 0.333$  时,样本  $i$  的水质综合等级为III;当  $0.333 \leq D_i < 0.667$  时,样本  $i$  的水质综合等级为II;当  $0.667 \leq D_i \leq 1$  时,样本  $i$  的水质综合等级为I。由于综合评价值  $D_i$  的取值范围为  $[-1.0, 1.0]$ ,6级水质综合指数的取值范围为  $[0.0, 6.0]$ ,因此,在计算水质综合指数方面,需要对综合评价值  $D_i$  进行数值转化。

## 3 水质实例应用与分析

### 3.1 河流概况

某河位于某市东南部,全长约11.5 km,河底高程1 m,河底宽20~40 m,最小生态水位2.17 m,属于市级河道。该河流经主要城区,但随着城镇工业化程度的不断加深,有机污染物排放量的加大,河段水质趋于恶化,生态系统遭受破坏。鉴于此,市政府计划在未来一段时间内,采取相应的治理措施,保护城镇居民饮用水的水质安全,以期恢复当地良好的水生态环境。因此,进行实时的水质监测和合理的综合评价成为首要工作。

本次水质等级采用《地表水环境质量标准》,具体详见表1。2013-2015年的水质检测数据,详见表2。数据的采样时间与频率是严格按照河流的具体情况进行的采样。中小河流的采样每年不得少于6次,分别在丰水期、枯水期和平水期进行采样,每期采样两次。数据本身没有规律,对结果也没有影响。

表1 地表水环境质量标准

评价等级	DO	COD <sub>Mn</sub>	COD	BOD <sub>5</sub>	NH <sub>3</sub> -N	TP	氟化物	挥发酚
$y_1$	7.5	2	15	3	0.015	0.02	1	0.002
$y_2$	6	4	15	3	0.5	0.1	1	0.002
$y_3$	5	6	20	4	1.0	0.2	1	0.005
$y_4$	3	10	30	6	1.5	0.3	1.5	0.01
$y_5$	2	15	40	10	2.0	0.4	1.5	0.1

由表2可见,在断面2013-2015年的水质监测数据中,指标BOD<sub>5</sub>和NH<sub>3</sub>-N出现超出5类水质标准值的情况。因此,原有的5级评价模型不再适用,需开展相关的改进工作。

### 3.2 无量纲化处理

在获取监测断面的水质数据后,按照公式(1)对数据进行无量纲化处理,处理结果  $r_{ij}$  见表3。而根据公式(3)、公式(6)分别得到评价标准集  $Y$  的5级、6级无量纲化结果  $s_{hj}$  和  $s_{hj}^*$ , 见表4和表5。



表5 水质等级标准的6级无量纲化

$s^*$	DO	COD <sub>Mn</sub>	COD	BOD <sub>5</sub>	NH <sub>3</sub> -N	TP	氟化物	挥发酚
I	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
II	0.864	0.923	1.000	1.000	0.878	0.895	1.000	1.000
III	0.636	0.769	0.900	0.929	0.630	0.658	1.000	0.985
IV	0.364	0.538	0.600	0.714	0.378	0.395	0.500	0.944
V	0.091	0.192	0.200	0.286	0.126	0.132	0.000	0.459
VI	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

由表5可以看出,改进后的6级地表水环境质量标准的无量纲化数值能够很好地解决指标数据中出现的超出5类水质标准值的情况。

因此,在之后的求解相对最优隶属度方面,更加准确、合理。

### 3.3 确定指标权重系数

本次研究河流所属的水功能区水质目标为III类。因此,首先在表1《地表水环境质量标准》的基础上,根据公式(4)计算样本各指标的污染物浓度超标度 $b_{ij}$ 。再者,通过公式(5)求得不同样本情况下的各指标权重 $w_{ij}$ ,计算结果见表6。

### 3.4 求解相对最优隶属度

第一步,根据公式(7),比较水质样本的无量纲化值 $r_{ij}$ 与标准集无量纲化值 $s_{hj}$ ,确定各样本数据的级别 $h_{ij}$ ,并求出所属级别的最大值 $\max_i$ 与最小值 $\min_i$ ,结果见表7。

第二步,根据模糊模式识别理论,利用公式(8),计算不同样本在等级 $h$ 上的最优相对隶属度 $u_{ih}$ ,结果见表8。

### 3.5 SPA水质综合评价

将上述的最优相对隶属度 $u_{ih}$ 代入公式(9),计算得到各样本的水质综合评价值 $D_i$ ,并依据判定原则进行水质等级确定,两种方法的评价结果见表9。

由表9所示,将SPA理论引入模糊模式识别理论后,可以很方便地计算水质综合评价值 $D_i$ 和水质综合指数值。同时,其计算结果可以很系统地进行统计,方便样本之间的对比分析。而最大隶属度原则下的最优相对隶属度,只能反映局部情况,忽略了评价等级标准中的不确定性信息。由此可见,引入SPA的模糊模式识别模型能够很好地解决最大隶属度原则造成的隶属度大小相差不大或者相同的问题,更能够准确地处理等级评价标准中存在的确定性信息。同时,就等级判定而言,引入SPA的水质评价模型比最大隶属度原则的评价结果更好,水质污染程度更低。但是,鉴于我国实行最为严格的水质评价标准-单因子评价法,本次构建的引入SPA的模糊模式识别模型在水质级别判定方面,只作为理论参考提出。

表6 各样本的指标权重计算结果

序号	日期	DO	COD <sub>Mn</sub>	COD	BOD <sub>5</sub>	NH <sub>3</sub> -N	TP	氟化物	挥发酚
1	2013-01-04	0.07	0.10	0.11	0.27	0.20	0.14	0.05	0.06
2	2013-03-01	0.08	0.10	0.11	0.12	0.26	0.13	0.06	0.15
3	2013-05-02	0.10	0.14	0.19	0.18	0.10	0.18	0.10	0.02
4	2013-07-01	0.19	0.13	0.11	0.18	0.13	0.14	0.10	0.03
5	2013-09-02	0.23	0.19	0.15	0.21	0.04	0.11	0.04	0.03
6	2013-11-01	0.20	0.15	0.17	0.14	0.02	0.13	0.15	0.04
7	2014-01-02	0.04	0.13	0.12	0.24	0.19	0.13	0.10	0.05
8	2014-03-03	0.04	0.08	0.13	0.20	0.34	0.12	0.08	0.02
9	2014-05-05	0.12	0.15	0.15	0.27	0.10	0.10	0.09	0.03
10	2014-07-01	0.20	0.13	0.14	0.13	0.12	0.12	0.13	0.03
11	2014-09-01	0.23	0.16	0.16	0.13	0.07	0.11	0.12	0.03
12	2014-11-03	0.17	0.15	0.15	0.10	0.11	0.17	0.12	0.03
13	2015-01-04	0.08	0.15	0.12	0.16	0.16	0.18	0.11	0.03
14	2015-03-03	0.00	0.08	0.09	0.14	0.41	0.20	0.05	0.02
15	2015-05-04	0.16	0.14	0.16	0.18	0.01	0.20	0.12	0.03

表7 各样本所属级别范围的上、下限

上下限	样本序号														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
min	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
max	6	6	5	4	5	4	5	6	5	4	4	4	4	6	4

表8 各样本在不同等级上的最优相对隶属度

等级	样本序号														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
I	0.06	0.06	0.08	0.07	0.05	0.07	0.09	0.04	0.15	0.08	0.09	0.06	0.10	0.04	0.08
II	0.07	0.08	0.09	0.11	0.07	0.13	0.11	0.05	0.17	0.14	0.14	0.10	0.15	0.05	0.12
III	0.09	0.17	0.14	0.29	0.14	0.69	0.17	0.09	0.24	0.64	0.47	0.72	0.55	0.09	0.57
IV	0.16	0.38	0.37	0.53	0.60	0.10	0.34	0.22	0.33	0.15	0.30	0.12	0.20	0.24	0.23
V	0.34	0.20	0.32	0.00	0.14	0.00	0.28	0.36	0.12	0.00	0.00	0.00	0.00	0.36	0.00
VI	0.28	0.10	0.00	0.00	0.00	0.00	0.00	0.23	0.00	0.00	0.00	0.00	0.00	0.22	0.00

表9 两种方法的评价结果比较分析

方法	水质指数	样本序号														
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
	$D_i$	-0.40	-0.15	-0.10	0.09	-0.09	0.27	-0.05	-0.41	0.16	0.26	0.21	0.24	0.26	-0.41	0.22
SPA	综合指数	4.19	3.44	3.31	2.74	3.26	2.19	3.14	4.22	2.52	2.23	2.37	2.29	2.22	4.22	2.33
	等级	V	IV	IV	III	IV	III	IV	V	III	III	III	III	III	V	III
最大隶属度	等级	V	IV	IV	IV	IV	III	IV	V	IV	III	III	III	III	V	III

## 4 结论与展望

(1) 计算结果显示,引入 SPA 的 6 级模糊模式识别模型能够很好地解决水质评价标准中存在的 uncertainty 问题,合理地确定各样本指标的相对隶属度。

(2) 集对分析模型可以计算水质综合指数的大小,实现样本间的比较与分析,为恢复河流良好的生态系统提供理论支持。

(3) 该模型仍然存在不足之处。比如采用污染物超标法确定指标权重,此方法不仅忽略了各样本之间的相互关联性,而且当评价样本较多时,计算指标权重的工作量较大。有待于在今后的研究中得到解决。

### 参考文献:

- [1] 董哲仁. 河流健康的内涵[J]. 中国水利,2005(4):15-18.
- [2] 崔嘉宇,张宁红,郁建桥,等. 改进的模糊综合评价法在太湖水质评价中的应用[J]. 环境工程学报,2015,9(11):5357-5363.
- [3] 赵克勤. 集对分析及其初步应用[M]. 杭州:浙江科学技术出版社,2000.
- [4] 高苏蒂,方达宪,潘争伟. 基于集对分析的河流水质综合指数评价模型[J]. 合肥工业大学学报(自然科学版),2015(5):654-658.
- [5] 杨红卫,王晓燕,李萍,等. 基于集对分析理论的南京市饮用水水源地水质评价[J]. 水电能源科学,2013(6):66-68.
- [6] 孟宪萌,胡和平. 基于熵权的集对分析模型在水质综合评价中的应用[J]. 水利学报,2009,40(3):257-262.
- [7] 周宇哲,高茂庭. 水质评价中模糊-集对分析法的改进[J]. 环境工程学报,2015,9(2):749-755.
- [8] 李浩,杨侃,陈静,等. 灰色三角白化权集对分析模型在河流健康评价中的应用[J]. 水电能源科学,2015(8):33-36.
- [9] 韩晓刚,黄廷林,陈秀珍. 改进的模糊综合评价法及在给水厂原水水质评价中的应用[J]. 环境科学学报,2013,33(5):1513-1518.
- [10] 马玉杰,郑西来,李永霞,等. 地下水质量模糊综合评判法的改进与应用[J]. 中国矿业大学学报,2009,38(5):745-750.
- [11] 徐健,吴玮,黄天寅,等. 改进的模糊综合评价法在同里古镇水质评价中的应用[J]. 河海大学学报(自然科学版),2014,42(2):143-149.
- [12] 邱林,唐红强,陈海涛,等. 集对分析法在地下水水质评价中的应用[J]. 节水灌溉,2007(1):13-15.
- [13] 汪新凡. 投资风险综合评价的联系数学模型[J]. 统计与决策,2006(19):19-20.
- [14] 汪新凡. 基于属性数学和联系数学的煤灰结渣特性综合评判模型及应用[J]. 锅炉技术,2006,37(6):47-50.