

基于模糊证据理论的水质环境状态估计

徐立中¹,林志贵²,黄凤辰¹

(1. 河海大学 计算机及信息工程学院,江苏 南京 210098;2. 天津工业大学 信息与通信工程学院,天津 300160)

摘要:从信息融合的角度,将湖泊水质环境多因子综合评价过程看成是一个多源数据融合处理与状态估计过程,提出基于相似度的模糊证据理论,评估湖泊水体富营养状况.通过引入模糊集合的相似度,一方面,确定模糊焦元的权值,并修改证据模型;另一方面,确定模糊焦元对类别焦元的信任函数的贡献因子,并建立状态估计与评价模型,使其能够处理湖泊水体富营养状况的突发性带来的某些焦点元素的显著变化问题.对太湖水体富营养状况的评估做了实证分析,并与营养状态指数评价方法进行比较,结果一致,说明基于相似度的模糊证据理论的估计方法可行.

关键词:信息融合;模糊证据理论;相似度;水体富营养化评价

中图分类号: TP118, X824 **文献标识码:** A **文章编号:** 1673-4785(2007)03-0079-05

Water environmental quality assessment based on fuzzy evidence theory

XU Li-zhong¹, LIN Zhi-gui², Huang Feng-chen¹

(1. College of Computer and Information Engineering, Hohai University, Nanjing 210098, China; 2. School of Information & Communication Engineering, Tianjin Polytechnic University, Tianjin 300160, China)

Abstract: The process of assessing lake water quality, as viewed from the perspective of data fusion, is a process involving multi-source data fusion and state estimation. This paper proposes using fuzzy-evidence theory to assess the eutrophication of lake water based on degrees of similarity. On one hand, by introducing degrees of similarity with fuzzy sets, the weights of fuzzy focal elements are determined and the evidence model is modified. On the other hand, by using the degree of similarity with fuzzy sets, the contribution factors of fuzzy focal elements to the belief functions of focal elements of the class are determined, and models of state estimation and assessment are established which can handle problems due to obvious changes in some focal elements induced by abrupt changes of lake water eutrophication. This proposed method was tested by water eutrophication assessment of Taihu lake in China. Results were analyzed and compared with those obtained by the trophic status index method, indicating that the proposed method based on degree of similarity is feasible for water quality assessment.

Key words: information fusion; fuzzy-evidence theory; similarity degree; assessment of water body eutrophication

湖泊水质环境的复杂性及湖泊水体富营养状况变化的突发性,可能带来监测数据的模糊性、不确定性.如何有效地处理具有不确定性的水质监测数据,对湖泊水体富营养化状况进行综合评价,显得越来越重要.近年来国内外学者提出了特征法、营养状态指数法、模糊综合评判法、人工神经网络模型等多种

评价模型^[1-2].由于影响富营养化程度的因素很多,评价因素与富营养化等级之间的关系复杂,各等级之间的关系模糊,并且这些方法均有其适用条件和局限性,至今尚未形成统一的、确定的评价模型.

D-S 证据理论作为一种不精确推理策略,在不确定性的表示、量测及组合方面的优势受到广泛的重视,文献[3]将 D-S 证据理论应用于水质监测数据融合与湖泊水体富营养化状况评估.考虑到融合对象、目标具有模糊性和不精确性,因为传统的证据

收稿日期:2006-11-23.

基金项目:国家自然科学基金资助项目(60374033);教育部科学技术研究重点项目(107057);江苏省高技术研究重大项目(BG2006003).

理论主要考虑“非此及彼”的现象. 仅依靠 D-S 证据理论难以奏效, 因此, 需要将 D-S 证据理论推广到模糊集, 这种推广主要考虑当辨识框架中的焦元具有模糊性概念(如水域水质富营养化时如何评价区分富、中、贫区^[41])时, 如何对证据进行组合问题. 目前的文献[5 - 9]在 D - S 证据理论推广到模糊集方法中, 存在着信任函数对某些焦点元素的显著变化不敏感问题. 因此, 文中通过引入模糊集合的相似度, 提出基于相似度的模糊证据理论, 使其能够处理湖泊水体富营养状况的突发性带来的某些焦点元素的显著变化问题.

1 基于相似度的模糊证据理论

文中采用的符号: $X(x_1, x_2, \dots, x_n)$ 为辨识框架, x_i 为基元, $R^+ = [0, +\infty)$, P_X 为 X 的模糊子集的全集, P_X 中的元素为模糊焦元, μ_A, μ_B 分别为 A, B 的隶属函数, Bel 为 P_X 中元素的信任函数, pl 为 P_X 中元素的似真函数.

基于相似度的模糊证据理论的组合规则采用 Haenni 思想^[10], 即修改信任分配模型而不改变 Dempster 组合规则的形式. 证据组合之前, 需要对模糊焦元的基本信任分配值进行修正. 文中基于模糊集合之间的相似性, 确定模糊焦元 C 与模糊焦元 A 之间的相似度, 作为权值, 修正模糊焦元 A 的基本信任分配值. 定义模糊焦元 C 与模糊焦元 A 之间的相似度 (C, A) 为

定义 1 设一辨识框架 $X(x_1, x_2, \dots, x_n)$, $x_i \in R^+$ ($i = 1, 2, \dots, n$), A, C 为其上的模糊焦元, $A, C \in P_X$, $A = (\mu_A(x_1)/x_1, \mu_A(x_2)/x_2, \dots, \mu_A(x_n)/x_n)$, $C = (\mu_C(x_1)/x_1, \mu_C(x_2)/x_2, \dots, \mu_C(x_n)/x_n)$, 则模糊焦元 A, C 的相似度 (C, A) 为

$$(C, A) = 1 - \frac{1}{|X|} \sum_i |\mu_C(x_i) - \mu_A(x_i)| \quad (1)$$

式中: $|X|$ 为辨识框架 X 的基 (X 包含的基元的个数).

假设 Bel_1 和 Bel_2 是相同辨识框架 $X(x_1, x_2, \dots, x_n)$ 上的信任函数, 具有基本可信度分配函数 m_1 和 m_2 以及模糊焦元 $\{A_1, A_2, \dots, A_p\}$ 和 $\{B_1, B_2, \dots, B_q\}$, 于是, 基本可信度分配函数 $m: 2^X \rightarrow [0, 1]$ 对于所有基本信任分配的非空集 C , 有

$$m(C) = m_1 \oplus m_2(C) = \frac{A_i \cdot B_j = C (C, A_i)^{m_1} (C, B_j)^{m_2}}{1 - \sum_{A_i \cdot B_j \neq C} (1 - (C, A_i)^{m_1} (C, B_j)^{m_2}) m_1(A_i) m_2(B_j)} \quad (2)$$

式中: (C, A_i) 为模糊焦元 A_i ($i = 1, 2, \dots, p$) 的权值; (C, B_j) 为模糊焦元 B_j ($j = 1, 2, \dots, q$) 的权值.

但是, 式 (2) 不满足结合律, 即: $(m_1 \oplus m_2) \oplus m_3 \neq m_1 \oplus (m_2 \oplus m_3)$. 为了能够进行多个证据的组合, 文中采用的策略类似文献[9]中的方法. 具体分 2 步: 第 1, 利用 Dempster 组合规则连续组合, 获得新的模糊焦元 BPA 值, 不需要正规化; 第 2, 采用相似度的权值进行正规化, 目的使组合后的模糊焦元 BPA 值之和为 1. 具体规则如下:

1)

$$m \oplus m_2 \oplus \dots \oplus m_n(C) = \frac{m_1(A_1) m_2(A_2) \dots m_n(A_n)}{A_1 \cdot A_2 \cdot \dots \cdot A_n = C} m_{1,2,\dots,n}(C)$$

2) $N[m_1 \oplus m_2 \oplus \dots \oplus m_n](C) =$

$$\frac{C=B (B, A_1) (B, A_2) \dots (B, A_n) m_{1,2,\dots,n}(C)}{1 - \sum_{A_1, A_2, \dots, A_n \neq B} (1 - (B, A_1) (B, A_2) \dots (B, A_n)) m_{1,2,\dots,n}(C)}$$

式中: (B, A_i) ($i = 1, 2, \dots, n$) 采用式 (1) 计算.

2 富营养化状态估计与评价模型

对于水质监测数据来说, 基于 D-S 证据理论融合处理后, 面临着评价问题, 即如何从水质监测数据融合结果进行水质状态估计与评价. 文中采用文献[4]提出的最大组合的基本信任分配 (BPA) 值的决策规则, 建立水质状态估计和评价模型.

首先从融合后的模糊焦元 BPA 计算类别焦元的信任函数值. 这个过程分 2 步: 一是确定融合后的模糊焦元 BPA 对类别焦元的信任函数 (Bel) 的贡献因子; 二是依据贡献因子计算类别焦元 Bel 的值. 信任函数贡献因子定义如下:

定义 2 设一辨识框架 $X(x_1, x_2, \dots, x_n)$, $x_i \in R^+$ ($i = 1, 2, \dots, n$), A, B 为其上的模糊焦元, $A, B \in P_X$.

$A = (\mu_A(x_1)/x_1, \mu_A(x_2)/x_2, \dots, \mu_A(x_n)/x_n)$, $B = (\mu_B(x_1)/x_1, \mu_B(x_2)/x_2, \dots, \mu_B(x_n)/x_n)$, 则模糊焦元 A 对 $Bel(B)$ 的贡献因子为

$$F(B, A) = 1 - \frac{1}{|A|} \sum_i |\mu_B(x_i) - \mu_A(x_i)| \quad (3)$$

式中: $|A_i|$ 为模糊焦元 A_i 的基 (A_i 包含的基元的个数).
由定义 2 可得模糊证据理论的信任函数:

$$Bel(B) = \sum_i F(B; A_i) \cdot m(A_i). \tag{4}$$

其次,依据式(4)计算类别焦元的信任函数值,选择其中焦元的信任函数值最大的类别作为最终估计与评价结果.

3 实验与分析

3.1 太湖水质监测数据

依据太湖实际情况及收集到的相关资料,文中选择与太湖富营养化状况直接有关的叶绿素 a (Chla)、总磷(TP)、总氮(TN)、化学需氧量(COD)、透明度(SD)作为估计与评价指标.下面以太湖 2003 年 8 月的水质监测数据为对象,取其中的 12 个监测点,具体监测数据见表 1.

表 1 太湖水质评价参数的实测数据
Table 1 Data of Taihu for water quality assessment

	Chla	TP	TN	COD	SD
	/mg · L ⁻¹	/mg · L ⁻¹	/mg · L ⁻¹	/mg · L ⁻¹	/m
五里湖心	0.068	0.15	3.85	6.5	0.30
阊江口	0.036	0.22	1.32	5.5	0.20
拖山	0.021	0.05	1.03	5.3	0.35
百渎口	0.055	0.16	1.75	7.3	0.20
沙墩港	0.014	0.03	1.35	5.7	0.70
大浦口	0.052	0.23	1.14	10.7	0.20
平台山	0.006	0.05	1.77	3.4	0.80
漫山	0.01	0.1	1.22	4.4	0.90
大雷山	0.01	0.06	1.36	6.1	0.65
小梅口	0.006	0.08	1.41	3.6	0.80
泽山	0.008	0.05	1.49	3.6	1.20
胥口	0.005	0.05	0.87	3.7	0.50

3.2 水质状态估计和评价标准的确定

目前,我国还没有完全统一的关于划分湖泊营养类型的标准.为了对太湖富营养化程度进行评价,参考相崎宇弘和郁根森 2 种标准^[11]并结合太湖具体情况,采用文献[1]提出评价太湖富营养化程度的 5 个评价指标 8 种类型的评价标准,如表 2 所示.

3.3 证据获取

评价指标具有不同的量纲,且类型不同,故指标间具有不可公度性.因此,评价时要先消除量纲不同的影响,同时,结合模糊证据理论的特点,需将每一个指标的监测数据转化为相应的证据,依据监测人

员或专家系统的经验,确定每种证据对应的基本可信度分配值.

表 2 太湖富营养化程度评价标准
Table 2 Eutrophication assessment criteria of Taihu

营养类型	Chla	TP	TN	COD	SD
	/mg · L ⁻¹	/mg · L ⁻¹	/mg · L ⁻¹	/mg · L ⁻¹	/m
贫营养	0.0016	0.0046	0.079	0.48	8.00
贫 - 中营养	0.0041	0.0100	0.160	0.96	4.40
中营养	0.0100	0.0230	0.310	1.80	2.40
中 - 富营养	0.0260	0.0500	0.650	3.60	1.30
富营养	0.0640	0.1100	1.200	7.10	0.73
重富营养	0.1600	0.2500	2.300	14.00	0.40
严重富营养	0.4000	0.5550	4.600	27.00	0.22
异常富营养	1.0000	1.2300	9.100	54.00	0.12

选择辨识框架为 $\Theta = \{1, 2, 3\}$, 1 表示贫营养; 2 表示中营养; 3 表示富营养. 相应的模糊子集为 $\{A_1, A_2, A_3, A_4, A_5, A_6, A_7, A_8\}$, 其具体数值及代表的水质类别如下:

- $A_1 = \{1/1, 0.50/2, 0.25/3\}$, 贫营养;
- $A_2 = \{0.65/1, 0.55/2, 0.25/3\}$, 贫 - 中营养;
- $A_3 = \{0.5/1, 1/2, 0.5/3\}$, 中营养
- $A_4 = \{0.25/1, 0.65/2, 0.55/3\}$, 中 - 富营养;
- $A_5 = \{0.25/1, 0.5/2, 1/3\}$, 富营养;
- $A_6 = \{0.1/1, 0.2/2, 1/3\}$, 重富营养;
- $A_7 = \{0.1/2, 1/3\}$, 严重富营养;
- $A_8 = \{1/3\}$, 异常富营养.

以太湖的五里湖心为例,依据监测人员或专家系统的经验,将水质参数 Chla、TP、TN、COD、SD 的数据,转化为证据的基本信任分配(BPA),其值见表 3.

表 3 证据的基本信任分配(BPA)
Table 3 BPA of evidences

证据	A ₁	A ₂	A ₃	A ₄	A ₅	A ₆	A ₇	A ₈
Chla	0	0	0	0.1	0.8	0.1	0	0
TP	0	0	0	0.2	0.6	0.2	0	0
TN	0	0	0	0	0.1	0.3	0.5	0.1
COD	0	0	0	0.2	0.7	0.1	0	0
SD	0	0	0	0	0	0.3	0.6	0.1

3. 4 分析和讨论

从表 3 可以看出,证据数为 5,模糊焦元个数为 8,其组合后得到不同焦元个数为 218. 考虑到模糊焦元 A_1 、 A_2 、 A_3 相对于这 5 个证据的 BPA 值都为 0,如果组合式(2)分子中含有 A_i ($i = 1, 2, 3$)中任何一个,则组合结果为 0,即如果有模糊焦元 A_1 、 A_2 、 A_3

之一参加组合,其结果不变且都为 0,因此,文中不考虑模糊焦元 A_1 、 A_2 、 A_3 参加组合情况. 对于剩下的 5 个模糊焦元 A_4 、 A_5 、 A_6 、 A_7 、 A_8 ,其相应的 5 个证据依据式(1)、(2)进行组合,得到不同的组合焦元个数为 31. 由表 3 证据 BPA 分布的特点,最后得到 9 种不同的组合焦元 C_1 、 C_2 、...、 C_9 ,其 BPA 见表 4.

表 4 证据组合后模糊焦元的 BPA 值
Table 4 BPA values of fuzzy focal elements after evidences combination

	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9
BPA	0. 000 0	0. 000 0	0. 007 45	0. 019 7	0. 072 85	0. 000 0	0. 010 04	0. 203 5	0. 086 6

表 4 中, $C_1 = \{0. 25/1, 0. 65/2, 0. 55/3\}$, $C_2 = \{0. 25/1, 0. 5/2, 1/3\}$, $C_3 = \{0. 1/1, 0. 2/2, 1/3\}$, $C_4 = \{0. 1/2, 1/3\}$, $C_5 = \{1/3\}$, $C_6 = \{0. 25/1, 0. 5/2, 0. 55/3\}$, $C_7 = \{0. 1/1, 0. 2/2, 0. 55/3\}$, $C_8 = \{0. 1/2,$

$0. 55/3\}$, $C_9 = \{0. 55/3\}$.
得到焦元 C_1 、 C_2 、...、 C_9 的 BPA 之后,采用式(3)、(4)计算类别焦元 $A_1 \sim A_8$ 的 Bel 值,见表 5.

表 5 证据组合后模糊焦元的信任函数值
Table 5 Believe function values of fuzzy focal elements after evidences combination

	A_1	A_2	A_3	A_4	A_5	A_6	A_7
Bel	0. 227 11	0. 223 28	0. 240 52	0. 296 93	0. 266 61	0. 302 72	0. 312 71

从表 5 看出,焦元 A_7 的 Bel 值最大,因此,太湖区域五里湖心的富营养状况为严重富营养. 与太湖的五里湖心的富营养状况评价过程类似,得到其他太湖的 11 个位置的富营养状况,见表 6. 从表 6 看出,2003 年 8 月太湖区域富营养化状况分布为整个区域基本上都属于中富营养化状态,北部区域比南部区域富营养状况严重,东部情况较好,这一结果符合实际情况.

表 6 太湖 12 个位置的富营养化评价结果比较
Table 6 Eutrophication assessment results of 12 points in Taihu

位置	TSIM	文中方法	位置	TSIM	文中方法
五里湖心	富营养	严重富营养	平台山	中营养	中营养
闻江口	富营养	富营养	漫山	中营养	中 - 富营养
拖山	富营养	富营养	大雷山	中营养	中 - 富营养
百渎口	富营养	重富营养	小梅口	中营养	中营养
位置	TSIM	文中方法	位置	TSIM	文中方法
五里湖心	富营养	严重富营养	平台山	中营养	中营养

环境监测部门采用的 TSIM 方法对这 12 个位置的评价结果,见表 6. 从表 6 评价结果看出,文中的评价方法与 TSIM 方法基本一致. 环境监测部门将区域富营养状况分为 3 个等级:富营养、中营养、贫营养,而文中将其分为 8 个等级($A_1 \sim A_8$),相应地比环境监测部门分的等级要细,因此出现细微的差异. 如果采用环境监测部门对区域富营养状况分为富营养、中营养、贫营养 3 个等级的情况下,依据文中的方法,由表 5 看出,五里湖心的富营养状况为富营养 ($Bel(A_5) = \max\{Bel(A_1), Bel(A_3), Bel(A_5)\}$),与 TSIM 方法一致,其他 11 个位置的富营养状况评价结果与 TSIM 方法一致. 通过分析比较表明,基于相似度的模糊证据理论进行区域富营养化状况评价与 TSIM 方法一致,这说明应用文中方法得到的估计与评价结果是可靠的.

4 结束语

综合评价湖泊水体富营养化状况是一个多源数据融合处理与状态估计、识别过程. 文中提出一种基于相似度模糊证据理论的湖泊水体富营养状况评估

方法,使其能够处理湖泊水体富营养变化状况的突发性带来的某些焦点元素显著变化的问题,并能够准确评估水质富营养化的富、中、贫状态.

通过对太湖区域 12 个监测点进行富营养化状况评估,得出的评估结果符合实际情况.与环境监测部门常常采用的营养状态指数法(TSIM)进行评价的结果进行比较,其结果基本一致,这说明本方法是可行的,评估结果是可信、可靠的.但是从实验分析看出,采用文中新方法,如果在证据比较多的情况下,组合后得到不同焦元的个数急剧增加,将带来证据组合的计算量呈指数级增加,这是本方法需要改进的问题.

参考文献:

[1]任 黎,董增川,李少华. 人工神经网络模型在太湖富营养化评价中的应用[J]. 河海大学学报(自然科学版), 2004,32(2):147 - 150.
REN Li, DONG Zengchuan, LI Shaohua. Application of artificial neural network model to assessment of Taihu Lake eutrophication[J]. Journal of Hohai University (Natural Sciences), 2004,32(2):147 - 150.

[2]徐立中,王建颖,黄凤辰. 基于神经网络的水环境监测数据融合处理与评价模型[J]. 测试技术学报, 2004,18(增刊):195 - 198.
XU Lizhong, WANG Jianying, HUANG Fengchen. Water environment monitoring data fusion based on neural network[J]. Journal of Test and Measurement Technology, 2004,18(Supp):195 - 198.

[3]林志贵,徐立中,黄凤辰. 基于 D-S 理论的多源水质监测数据融合处理[J]. 计算机工程与应用, 2004,40(10):3 - 5.
LIN Zhigui, XU Lizhong, HUANG Fengchen. Multi-source water quality monitoring data fusion based on D-S theory[J]. Computer Engineering and Applications, 2004,40(10):3 - 5.

[4]LIN Z G, XU L Z, HUANG F C, et al. multi-source monitoring data fusion and assessment model on water environment[A]. In: Proceedings of the Third International Conference on Machine Learning and Cybernetics [C]. Shanghai, 2004.

[5]ISHIZU KA M, FU K S, YAO J. Inference procedures and uncertainty for the problem-reduction method[J]. Information Sciences, 1982, 28(2):179 - 206.

[6]OGAWA H, FU K S. An inexact inference for damage assessment of existing structures[J]. International Journal of Man-Machine Studies, 1985, 22(3):295 - 306.

[7]YAGER R R. Generalized probabilities of fuzzy events from fuzzy belief structures[J]. Information Sciences, 1982, 28(1):45 - 62.

[8]YEN J. Generalizing the Dempster-Shafer theory to fuzzy sets[J]. IEEE Transactions on System, Man and Cybernetics, 1990, 20(3):559 - 570.

[9]YANG M S, CHEN T C, WU K L. Generalized belief function, plausibility function, and Dempster's combinational rule to fuzzy sets[J]. International Journal of Intelligent Systems, 2003, 18(5):925 - 937.

[10]HAENNI R. Comments on "About the belief function combination and the conflict management problem"[J]. Information Fusion, 2002, 3(4):237 - 239.

[11]金相灿. 中国湖泊富营养化[M]. 北京:中国环境科学出版社,1990.

作者简介:



徐立中,男,1958 年生,博士,教授,博士生导师,主要研究方向为信息获取与遥测遥控、信息处理与智能系统.曾获省部级科技进步二等奖 2 项,三等奖 5 项,申请国家发明专利 2 项,发表论文 90 余篇,出版专著 4 部.

E-mail:lzhxu@hhu.edu.cn.



林志贵,男,1974 年生,博士,讲师,主要研究方向为智能信息处理、嵌入式系统及应用.



黄凤辰,男,1964 年生,硕士,副教授,主要研究方向为遥测遥控、信号与信息处理及应用.