

文章编号: 1000-5862(2020)04-0394-09

基于多源数据融合的突发水污染事故可靠预警方法

傅 韬¹, 谭德坤^{2*}, 付雪峰², 涂振宇², 王 晖²

(1. 江西省防汛信息中心, 江西 南昌 330009; 2. 南昌工程学院江西省水信息协同感知与智能处理重点实验室, 江西 南昌 330099)

摘要: 在突发水污染事故自动监测领域中, 传感器节点监测数据的异常是影响自动监测系统预警可靠性的重要原因。考虑到多传感器信息之间的互补性和相关性, 该文提出了一种基于多源数据融合的突发水污染事故可靠预警方法。基于改进的 D-S 证据理论, 利用综合权重对节点证据进行加权修正, 并用 D-S 融合规则对多源数据进行两两融合, 最终根据融合结果对突发水污染事故进行预警决策。案例分析及实验结果表明: 与传统方法相比, 该方法能得到可靠度更高、聚焦性更好的预警结论。

关键词: 突发水污染事故; 异常数据; D-S 理论; 多源数据融合; 预测预警

中图分类号: TP 18; TP 391 **文献标志码:** A **DOI:** 10.16357/j.cnki.issn1000-5862.2020.04.11

0 引言

随着社会经济及城市化的快速发展, 中国不得不面对严重的水环境污染问题。特别是突发水污染事故, 具有较强的不确定性, 且事发突然, 危害性高, 对社会秩序、生态环境, 甚至人的生命财产安全构成严重威胁^[1]。由于突发水污染事故的发生存在巨大的偶然性, 现有技术还很难事先进行有效预测和控制^[2], 因此, 在事故刚发生或刚出现征兆时, 为了最大限度地减少事故造成的影响和损失, 决策者应在第一时间获得水污染监测数据, 并根据污染程度发布准确的预警级别和采取针对性的应急处置措施。其中, 快速准确地获取水质监测数据是对水污染事故进行可靠预警的前提。

目前, 常用的水质数据获取手段主要有人工采样和自动监测 2 种方式。传统的人工方式不能满足突发水污染事故监测预警所需的连续、不间断监测及高时效性的要求, 因此自动监测在水环境监测领域中发挥了越来越重要的作用。而水质自动监测基于传感器技术、计算机技术、自动控制技术和通信技术, 通过搭建有线或无线网络(如无线传感器网络), 实现对水环境水质参数的实时、快速、多源、连续地在线自动监测和远程监控^[3], 从而能够及时掌握监测流域重点断面和水源水体水质状况, 能有效

地对突发水污染事故进行预警, 达到有效防范事故并及时处置的目的^[4]。

传统的人工方式首先对水质参数进行人工采样, 然后通过实验室仪器进行结果分析。与人工方式不同, 自动监测方式完全依赖于前端的水质自动监测传感器, 通过它对水质进行自动采样、结果分析及预报预警。一旦发生污染事故, 传感器监测节点就会将监测数据及预警信息发送给后端供相关部门进行决策, 确保第一时间启动相应的预控措施, 从而采取合理手段减轻污染事故的危害。由于野外监测水域环境的复杂性, 传感器节点可能受到外界干扰或本身出现问题, 从而导致监测的水质数据有较大的偏差, 在极端情况下甚至会导致水污染漏报误报等严重事故。可见, 如何处理异常传感器节点所引起的干扰是突发水污染事故可靠预警必须要解决的关键问题。针对水质传感器监测预警的研究, 仅根据单源水质传感器监测指标是否超标来进行事故预判的传统方法具有明显的局限性, 由于提供的信息量少, 因而常常导致误报率高^[5]。近年来, 国内外学者主要采用时间序列方法、人工智能方法(如机器学习、神经网络等)和聚类方法等^[6-9]对传统方法进行改进, 这些方法均是将历史数据或者最近的水质数据作为原始数据集, 训练得到合适的水质预测模型, 并据此判断水质是否存在异常, 即由过去的数据来预测当前水质是否正常。相较于传统单因子方法, 这些方法预

收稿日期: 2019-11-27

基金项目: 国家自然科学基金(61762063), 江西省自然科学基金(20171BAB202024), 江西省水利厅科技课题(KT201639), 江西省科技厅重点研发课题(20151BBE50077)和江西省教育厅科技课题(GJJ170991, GJJ190958)资助项目。

通信作者: 谭德坤(1973-), 男, 重庆开县人, 副教授, 博士, 主要从事数据融合、智能优化算法研究。E-mail: dktan@nit.edu.cn

测准确率均得到了明显提高. 然而, 上述方法均依赖于历史异常数据, 在没法提供异常特征数据训练集的情况下, 这些方法的使用必然会受到限制; 另外, 野外湖泊、河流水质受水文、气象、污染物等多种复杂因素的综合影响, 其变化规律难以捉摸. 突发事件的发生更是存在巨大的随机性^[10], 利用历史数据训练得到的固有预测模型来对突发事件进行准确预测是困难的, 预警的可靠性难以得到保证.

突发水污染事故的发生是随机的, 传感器节点故障或受干扰会带来监测信息的不完整性和不一致性. 这些不确定性因素会给水污染预警决策带来风险. 而多源数据融合通过对多种数据源自下而上的综合处理, 在最上层得到决策目标的明确描述, 从而降低和消除不确定性对预警决策的影响. 由于 D-S 证据理论不需要先验知识和大量样本, 在处理不确定性问题上具有明显优势^[11]. 因此, 本文采用改进的 D-S 证据理论, 同时考虑到多传感器信息之间的互补性和相关性, 提出了基于多源数据融合的突发水污染事故可靠预警方法. 该方法利用实时的水质自动监测数据, 通过不同水质监测数据之间的相关性及 D-S 融合权重修正策略消除异常传感器监测数据对预警决策的不利影响, 从而降低不确定性的影响, 提高突发水污染事故预警结论的可靠性和时效性.

1 突发水污染预警模型构建

由于水环境监测涉及多种传感器采集的信息, 如水温、pH 值、溶解氧(DO)、电导率、浊度、氨氮、化学需氧量(COD)等. 本文基于多传感器异构信息, 利用融合准则和算法对获取的信息进行融合, 以弥补单源传感信息进行预警所带来的缺点和局限, 达到更准确、客观的效果^[12]. 考虑到突发水污染预警信息准确性要求高的特点, 根据数据融合的分层处理理念, 提出 2 级数据融合突发水污染预警模型. 模型整体框架如图 1 所示. 数据采集层是整个模型的硬件基础, 由水温、氨氮、pH 值及溶解氧等各类传感器对外部水环境数据进行采集, 提供待处理的信息; 数据融合层是整个模型的核心, 基于多源信息融合中的证据理论, 对突发水污染监测信息进行数据和决策 2 级融合处理, 得到需要的决策信息; 决策处理层接收融合结果, 进行可靠性评估与专家决策, 最终得到对突发水污染态势的准确评估, 从而判断是否对突发水污染事故进行预警.

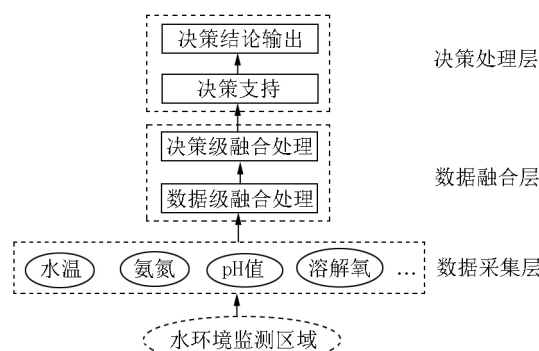


图1 突发水污染预警模型框架图

2 基于改进 D-S 证据理论的多源数据融合方法

经典 D-S 证据理论^[13]是由 A. P. Dempster 提出, G. Shafer^[14]对该理论做了进一步发展, 引入信任函数概念, 形成了一套较好地处理不确定性问题的方法. 与传统概率论方法相比, 该方法不需要先验概率, 能对数据的不确定性进行建模, 较好地处理数据的不确定性和不精确性, 因此, 被广泛应用于解决信息融合问题.

2.1 D-S 证据理论基本概念

定义1 辨识框架 Θ 就是对某个研究对象进行判断的所有可能命题的集合, 是一个互斥的非空有限集, 即

$$\Theta = \{A_1, A_2, \dots, A_n\}.$$

由框架 Θ 所有子集构成的集合被称为 Θ 的幂集, 记为 2^Θ .

定义2 设 Θ 是辨识框架, \emptyset 表示空的命题集合. 若存在映射 $m: 2^\Theta \rightarrow [0, 1]$ 且满足

$$\begin{cases} m(\emptyset) = 0, \\ \sum_{A \subseteq \Theta} m(A) = 1, \end{cases}$$

则 $m(A)$ 被称为辨识框架 Θ 上 A 的基本信任分配函数. $\forall A \subseteq \Theta$, 称 $m(A)$ 为 A 的 mass 函数, 它表示对命题 A 的信任程度大小.

定义3 设 Θ 为辨识框架, $m: 2^\Theta \rightarrow [0, 1]$ 为 Θ 上的基本信任分配函数, $\forall A \subseteq \Theta$, 则称函数 $Bel(A) = \sum_{B \subseteq A} m(B)$ 为辨识框架 Θ 上的信度函数. 相应地, $\forall A \subseteq \Theta$, 称函数 $Pl(A) = \sum_{B \cap A \neq \emptyset} m(B)$ 为辨识框架 Θ 上的似然函数.

$Bel(A)$ 表示 2^Θ 中所有证据对命题 A 的支持度, $Pl(A)$ 表示 2^Θ 中所有证据不反对命题 A 的程度. $Bel(A)$ 和 $Pl(A)$ 的关系为

$$\begin{cases} Pl(A) = 1 - Bel(\bar{A}), \\ Pl(A) \geq Bel(A), \end{cases}$$

其中 \bar{A} 为子集 A 的补集.

定义 4 设 m_1 和 m_2 分别是辨识框架 Θ 上 2 个独立的基本信任分配函数, m_1 对应的焦元为 $\{A_1, A_2, \dots, A_k\}$, m_2 对应的焦元为 $\{B_1, B_2, \dots, B_k\}$. 则定义 D-S 证据合成规则为

$$m(X) = \begin{cases} 0, & X = \emptyset, \\ \frac{\sum_{A_i \cap B_j = X} m_1(A_i) m_2(B_j)}{1 - K}, & X \subseteq \Theta, X \neq \emptyset, \end{cases}$$

其中 $K = \sum_{A_i \cap B_j = \emptyset} m_1(A_i) m_2(B_j)$ 被称为冲突因子,

$0 \leq K \leq 1$, 它表示证据之间冲突程度的大小. K 值越大表示证据之间的冲突越大, 即 m_1 和 m_2 之间越矛盾.

若是多个证据的融合, 则可采用 D-S 证据合成规则进行两两融合.

2.2 改进的 D-S 证据理论融合方法

经典 D-S 证据理论作为目前数据融合中最成熟的方法之一, 对解决不确定性问题具有明显的优势. 但是, 当需要进行处理的证据彼此相互冲突时, 融合结果会存在 L. Zadeh 悖论^[15]、一票否决、信度分配公平性等诸多问题.

在水环境监测实际应用场合中, 传感器节点多部署于野外恶劣环境中, 受周围环境的影响, 多种传感器设备所采集的数据不完全是绝对可靠和正确的, 证据之间的冲突是不可避免的. 若利用经典 D-S 证据理论进行数据融合处理则必然会存在诸多问题. 因此, 在数据融合过程中, 本文从证据源的改进入手, 在 C. K. Murphy^[16] 和 Han Deqiang 等^[17] 提出的证据加权融合策略的基础上, 考虑证据之间的关系以及证据体本身的特性, 对证据源的可靠性及重要性进行量化加权, 并用 D-S 融合规则对改进后的证据源进行融合处理, 从而有效地消除了冲突数据对融合结果的不利影响, 提高了多源数据融合结果的准确性及可靠性.

2.2.1 融合数据的可靠性 多源数据融合的可靠性是对各个传感器所采集数据的可靠性估计, 可靠的数据之间的融合其最终结果才是可靠的.

基于数理统计理论, 大量数据出现异常的多个随机事件的联合概率往往非常低. 因此, 可从数据之间的相互关系分析数据的可靠性. 若一个数据与其他大部分数据较接近, 这反映了该数据被其他大部分数据所支持, 则该数据可靠性较高; 若一个数据与

其他大部分数据冲突, 则该数据可靠性较低, 表现为异常数据.

对各种传感器采集的水环境监测数据, 数据的可靠性由传感器测量数据之间的相互支持度进行表征. 若某传感器所测数据被较多的传感器支持, 则其可靠性也较高. 反之, 若某传感器所测数据的支持度偏低, 则其可靠性也较低. 根据统计学视角, 数据之间的相对距离能够表示证据之间的相互支持度或者相似度, 本文采用目前最常用的类欧氏距离 Jousselme 距离^[18] 来度量待融合证据间的相互支持度, 其定义如下:

设 m_1 和 m_2 是定义在辨识框架 Θ 上的 2 条证据体, 则它们之间的 Jousselme 距离为

$$d(m_1, m_2) = \sqrt{(m_1 - m_2)^T D (m_1 - m_2) / 2},$$

其中 D 为 Jaccard 系数矩阵, 其元素定义为

$$D(A, B) = |A \cap B| / |A \cup B|, A, B \in 2^\Theta.$$

若 2 条证据体之间的距离越大, 则它们之间的相互支持度越低. 设任意 2 条证据体之间的距离为 $d(m_i, m_j)$, 则 2 条证据体之间的相似度为

$$Sim(m_i, m_j) = 1 - d(m_i, m_j).$$

在计算出证据体之间的相似度后, 证据体的支持度计算公式为

$$Sup(m_i) = \sum_{j=1, j \neq i}^n Sim(m_i, m_j).$$

若一个证据体的支持度越高, 则其可靠性也越高, 证据体的可靠性大小用可靠度权重 W_R 进行表征. 某个证据体 m_i 的可靠度权重计算公式为

$$W_{R_i} = Sup(m_i) / \sum_{i=1}^n Sup(m_i). \quad (1)$$

2.2.2 融合数据的重要性 在突发水污染事故预测预警中, 对传感器所采集的数据, 要考虑其可靠性. 由于不同水质参数超标所引起的水污染事故严重程度不同, 如有毒物质(如剧毒氰化物)所引起的水污染其危害程度远超一般性的生活污水排污所引起的污染事故. 因此, 数据的重要性也是在水污染事故预警决策时必须考虑的一个重要方面.

根据 GB 3838—2002《地表水环境质量标准》^[19], 用于水环境监测的项目种类比较多, 共有 27 项水质指标, 主要分为固体污染物、需氧污染物、有毒污染物、生物污染物等 4 大类监测指标. 各种传感器所监测的水质参数在水污染预警决策评价中的影响轻重程度不同, 本文用重要性权重 W_I 进行表征, 其计算方法为

$$W_{I_i} = \alpha_i / \sum_{j=1}^n \alpha_j, \quad (2)$$

其中 α_i 为所监测污染物的生物毒性权重系数,该系数用以评价污染物的毒性级别大小,其计算公式^[20]为

$$\alpha_i = 1/E_{C_{50}}^i, \quad (3)$$

其中 $E_{C_{50}}^i$ 为第 i 种水质污染物样品的半数效应浓度值.生物毒性测试是评判水污染物综合毒性效应的基础,在突发水污染预警中,常采用发光细菌急性毒性试验法测量各种有毒污染物的生物毒性.

2.2.3 融合数据的综合权重 可靠度权重体现了某个传感器所监测数据被其他有相关性的同种或异种传感器所支持的程度,而重要性权重体现了不同监测指标在水污染评价体系中的重要程度.综合权重为可靠度权重与重要性权重的线性组合,其计算公式为

$$W_i = \lambda W_{R_i} + (1 - \lambda) W_{R_i}, \quad 0 \leq \lambda \leq 1, \quad (4)$$

其中 W_i 为综合权重; λ 为调节因子,其反映决策者对可靠度权重 W_{R_i} 及重要性权重 W_{I_i} 在综合权重中占比的偏好程度,当数据的可靠性与其重要性同等看待时,取 $\lambda = 0.5$.

2.3 多源数据融合策略

在水环境实时监测预警系统中,需采用多种传感器对水质参数进行监测,如水温、pH 值、溶解氧、化学需氧量、氨氮、重金属等.对传感器监测获得的多源异构复杂信息,本文分别在传感器节点端、数据层和决策层中进行融合处理,最后针对融合结果进行预警评估决策.

2.3.1 传感器节点端 在节点端,采用预设阈值并越限报警的方法来进行第 1 层融合预处理,若污染物浓度监测值超过给定阈值,则判断可能存在水污染事故,预警节点便将监测值发给后端,同时也触发系统采集其他节点的监测数据以便进行更高层级的融合处理.节点阈值根据 GB 3838—2002《地表水环境质量标准》^[19] 某类水域某种监测项目的标准限值进行确定,为了防患于未然,节点端预警阈值根据文献[21]提供的频次分析法进行确定,即以 5% 频次的浓度线为基准,将其设为浓度不正常状态的阈值,当某水质监测指标达到该限值时,节点端就会发送预警信息给后端,从而让系统的预警机制提前介入工作.这里需特别说明的是节点端预警信息仅触发系统多源数据融合预警机制运行工作,并不作为最终预警结论供相关部门决策使用.

2.3.2 数据层 对节点端传送过来的带预警信息的水污染监测数据,仅靠单个传感器节点的信息对水污染事故进行决策判断具有明显的局限性.无论是同种还是异种传感器,当污染事故发生时,各个监测项目都会呈现出同步变化的规律性,如电导率与

铜、铅、镉等重金属污染物浓度呈正相关,若水体中重金属浓度偏高,而电导率也同步偏高,则多为重金属污染.溶解氧和氨氮、总磷呈负相关,若溶解氧下降,而氨氮或总磷浓度同步上升,则较大可能是氨氮或者总磷超标发生污染.因此,利用水质监测传感器之间的这种相关性,可分析监测传感器所发送预警信息的可靠性.若某个传感器发送了污染预警信息,但是与其有相关性的其他同种或异种传感器的监测信息未同步变化,则可判断该传感器发生了异常,发生水污染事故的可能性较低;反之,若某个传感器发出污染预警信息,而与之相关的其他传感器也同步发出预警,则发生水污染事故的可能性就较高.

在数据层,通过分析多源(同种或异种)传感器信息之间的相关性,利用本文提出的改进 D-S 证据理论分析监测传感器所发送预警信息的可靠性,并计算其可靠度权重.通过该层融合处理,系统对单个传感器发出的预警信息的可靠性有一个初步的判断,并将判断结果发送给更高层级的决策层进行进一步融合处理.

2.3.3 决策层 在数据层,得到了发出预警信息数据可靠性的判断结果.当有多种传感器同时发出了预警信息时,哪种预警信息需要优先考虑?在决策层,采用改进的 D-S 证据理论对多个决策结果进行融合处理,在融合处理过程中,不但要考虑证据的可靠性,而且也要考虑证据的重要性,即引入综合权重对待融合证据进行加权平均,从而得到可靠性较高的最终预警决策结论.

2.4 融合算法描述

根据以上多源数据融合策略,本文引入数据的可靠度及重要性权重,对不同证据进行加权平均,随后采用 D-S 融合规则进行融合处理,最后得到突发水污染预警决策结果.算法的主要步骤如下:

(i) 首先建立决策系统辨识框架.根据决策预警要求,设辨识框架 $\Theta = \{A, B, C\} = \{\text{正常}, \text{黄色预警}, \text{红色预警}\}$.

(ii) 从水质监测传感器采集获取水污染决策数据.在已确定的辨识框架内,利用传感器对水环境进行实时监测,采集水质参数.若某个节点端污染物浓度超过预设阈值,则触发系统预警处理机制,将预警节点及其他节点的监测值均发给后端进行数据层融合处理;否则,节点仅处于实时监测状态而不向后端发送数据.

(iii) 确定数据基本信任分配函数.在数据层,根据多源(同种或异种)传感器信息之间的相关性,针对传感器节点发送过来的水污染决策数据,确定

每个独立传感器数据对辨识框架内决策命题的基本信任分配函数,在 D-S 证据理论中被称为证据。

(iv) 计算证据体的可靠度权重. 首先计算出证据体之间的 Jousselme 距离,然后计算相似度和支持度,最后根据(1)式计算可靠度权重。

(v) 计算证据体的重要性权重. 根据(2)式计算数据的重要性权重。

(vi) 计算证据体的综合权重. 由可靠度权重及重要性权重,根据(4)式计算证据体的综合权重。

(vii) 对证据做加权平均. 考虑到证据体的可靠度及重要性,通过证据体的综合权重 W_i 对原来的证据体进行加权修正,修正公式为 $m_w(A) = \sum_{i=1}^n W_i m_i(A)$ 。

(viii) 多源数据融合处理. 在决策层,将加权修正后的证据采用 D-S 融合规则进行两两融合。

(ix) 决策处理. 根据相应的决策模型进行决策,选择基本信任分配函数值超过某个预设阈值的证据融合结果作为预警结论输出。

3 实验结果及分析

设研究区域为某湖泊饮用水源地,根据GB 3838—

2002《地表水环境质量标准》要求,湖泊水体需满足Ⅲ类水质标准. 针对实际应用需求,系统可部署多种传感器对水源地进行实时监测,对同种水质参数也可以部署多个传感器节点以增强系统的可靠性. 本文取其中9种常规水质指标测量值作为多源数据融合实验的基本数据,每种水质指标取2个监测节点值. 为了验证本文方法的有效性和可靠性,进行了大量仿真实验,并抽取了其中3组典型实验数据进行分析. 这3组实验数据分别模拟突发水污染事故实时监测中可能出现的3种极端案例。

案例1 某个传感器节点发生了故障,从而在节点端误报了预警信息。

从表1中数据可以看出,氨氮监测传感器节点2的监测值超过了Ⅲ类水质标准最大限值,从而将预警信息发送给后端进行融合处理,而另一传感器节点1则表明监测指标正常. 这表明传感器数据之间发生了明显的冲突. 针对此种同类传感器监测数据发生冲突的情况,仅凭这2个节点的监测信息很难排除异常传感器监测节点所采集的数据. 针对矛盾的信息,决策者也无法作出可靠的预警结论。

表1 水环境监测预警实验数据1

监测项目	水质指标监测节点采集数据		Ⅲ类水质标准
	节点1	节点2	
温度/℃	18.2	18.1	温升≤1℃ 温降≤2℃
pH值	7.25	7.27	[6.00,9.00]
溶解氧/(mg·L ⁻¹)	5.82	5.85	[5.00,+∞)
氨氮/(mg·L ⁻¹)	0.42	2.13	[0,1.00]
化学需氧量/(mg·L ⁻¹)	13.93	13.78	[0,20.00]
总氮/(mg·L ⁻¹)	0.58	0.57	[0,1.00]
总磷/(mg·L ⁻¹)	0.031	0.032	[0,0.05]
铜/(mg·L ⁻¹)	0.245	0.244	[0,1.00]
铅/(mg·L ⁻¹)	0.023	0.023	[0,0.05]

本文根据水质特征指标之间的相关关系,分析水质监测传感器之间的关联度,通过紧密关联的传感器节点的监测信息分析预警节点的可靠性,从而进一步排除异常数据对预警结论的干扰. 水质指标之间相关程度的高低用相关系数 r ($-1 \leq r \leq 1$) 进行表征. 本文设相关系数 $|r| \geq 0.5$ 的水质特征参数为紧密相关性指标. 由文献[22-24]可知,氨氮与总氮、总磷、溶解氧等水质指标之间呈现出显著相关性(见表2)。

表2 氨氮水质指标相关系数矩阵($|r| \geq 0.5$)

指标	氨氮	总氮	总磷	溶解氧
氨氮	1.000 0	0.937 8	0.806 0	-0.744 5

从表2可以看出,氨氮与总氮、总磷以及溶解氧等水质指标具有较强的相关性. 当污染事故发生时,这些紧密相关水质监测数据会与氨氮监测数据呈现同步变化的规律性. 因此,通过水质监测数据可以间接分析氨氮监测传感器节点所采集数据的可靠性和准确性。

针对辨识框架内的决策命题,如何确定传感器对命题的基本信任分配函数(即 $mass$ 函数)是融合判断的关键. 本文利用模糊集合理论,将模糊隶属度函数作为基本信任分配函数. 传感器节点所监测数据按照某种模糊隶属度分别属于3类评价标准等级,即监测指标正常、黄色预警态和红色预警态,由

三角模糊数构造 3 个模糊子集 T_1 、 T_2 和 T_3 , 对应隶属度函数分别为 $f_1(t)$ 、 $f_2(t)$ 及 $f_3(t)$ 且 $f_1(t) + f_2(t) + f_3(t) = 1$ (见图 2)。

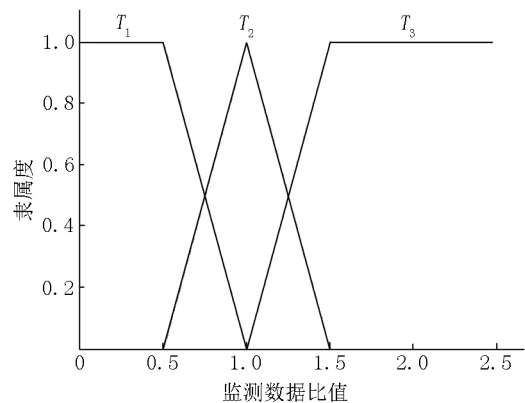


图 2 节点监测值的隶属度函数

在图 2 中, 监测数据比值为传感器节点监测值与水质标准限值之比, 其大小可反映水质污染的严重程度; 模糊隶属度表示传感器节点监测数据隶属于各评价标准等级的大小. 对于溶解氧这种含量越多越好的指标, 标准限值为最低评分值, 且模糊子集 T_1 、 T_2 和 T_3 分别对应红色预警态、黄色预警态和监测指标正常.

利用前述三角模糊隶属度函数, 可以求得与氨氮紧密相关的水质监测传感器节点对决策命题的基本信度, 各节点的 $mass$ 函数值如表 3 所示.

表 3 与氨氮紧密相关节点的基本信度分配函数

监测项目	节点 1 $mass$ 值			节点 2 $mass$ 值		
	$m(A)$	$m(B)$	$m(C)$	$m(A)$	$m(B)$	$m(C)$
溶解氧	0.328	0.672	0	0.340	0.660	0
氨氮	1	0	0	0	0	1
总氮	0.840	0.160	0	0.860	0.140	0
总磷	0.760	0.240	0	0.720	0.280	0

从表 3 可以看出, 氨氮节点 2 的证据与其他节点的证据明显冲突. 根据 (1) 式, 计算出了各节点数据的可靠度权重(见表 4). 从表 4 可看出, 氨氮节点 2 的可靠度权重极低, 与其他节点的可靠度权重存在显著差异, 应视为异常数据, 该异常数据在决策融合中对最终结果的影响也应该最小.

表 4 与氨氮紧密相关节点的可靠度权重

监测项目	可靠度权重	
	节点 1	节点 2
溶解氧	0.117 13	0.119 18
氨氮	0.126 43	0.018 74
总氮	0.152 68	0.151 47
总磷	0.157 17	0.156 21

求得了各节点数据的可靠度权重后, 可采用本文提出的 D-S 融合规则进行多源数据融合处理. 由于氨氮、总磷、总氮属于营养污染物, 当其浓度较低时对发光细菌生物毒性效应影响不明显^[25], 故本案例在决策融合过程中不考虑重要性权重的影响. 为了验证本文方法的优势, 将最终融合结果与 GB 3838—2002《地表水环境质量标准》规定的单因子法、加权平均融合法、经典 D-S 方法及 Murphy 法进行了比较, 结果如表 5 所示.

表 5 在案例 1 中 5 种融合方法预警结果比较

方法	融合决策数据	预警结论
单因子法	氨氮 2.13 $\text{mg} \cdot \text{L}^{-1}$	氨氮超标, 红色预警
加权平均融合法	氨氮 1.848 $\text{mg} \cdot \text{L}^{-1}$	氨氮超标, 红色预警
经典 D-S 方法	$m(A) = 0$	无
	$m(B) = 0$	
	$m(C) = 0$	
Murphy 方法	$m(A) = 0.998\ 5$	正常
	$m(B) = 0.001\ 5$	
	$m(C) = 0$	
本文方法	$m(A) = 0.999\ 2$	正常
	$m(B) = 0.000\ 8$	
	$m(C) = 0$	

从表 5 可看出, 单因子法及加权平均法均不能排除异常节点对最终融合结果的影响, 从而发出错误警报. 而经典 D-S 方法在节点证据显著冲突的情况下, 根本不能得出有效的预警结论. Murphy 方法及本文方法均能排除异常证据的不利影响, 从而作出正确的预警结论, 但是本文方法考虑了紧密相关节点证据之间的关系, 最终决策数据 $m(A)$ 取值最大, 即正确结论的可靠度更高, 不确定性降低, 本文方法的聚焦效果更好.

案例 2 监测区域发生了铜和铅 2 种重金属污染, 监测节点端均发出了预警信息.

从表 6 可看出, 铜和铅 2 种重金属水质传感器监测值均超标, 从而发出预警信息给决策层进行融合处理. 根据前述理论, 在求得铜和铅 2 种监测节点的可靠度权重后, 需考虑 2 种重金属离子的生物毒性效应. 基于发光细菌法, 测得重金属铅和铜的生物毒性效应值 $E_{C_{50}}$ 分别为 0.427 $\text{mg} \cdot \text{L}^{-1}$ 、2.74 $\text{mg} \cdot \text{L}^{-1}$ ^[26], 由 (2) 式和 (3) 式求得铅和铜监测节点的重要度权重(见表 7).

表 6 水环境监测预警实验数据 2

监测项目	水质指标监测节点采集数据		III 类水质标准
	节点 1	节点 2	
温度/℃	28.2	28.3	温升≤1℃ 温降≤2℃
pH 值	7.43	7.41	[6.00 9.00]
溶解氧/(mg · L ⁻¹)	5.82	5.85	[5.00 ,+ ∞)
氨氮/(mg · L ⁻¹)	0.58	0.58	[0 1.00]
化学需氧量/(mg · L ⁻¹)	21.77	21.78	[0 20.00]
总氮/(mg · L ⁻¹)	0.71	0.72	[0 1.00]
总磷/(mg · L ⁻¹)	0.032	0.031	[0 0.05]
铜/(mg · L ⁻¹)	1.645	1.647	[0 1.00]
铅/(mg · L ⁻¹)	0.068	0.069	[0 0.05]

表 7 监测节点重要度权重

监测项目	监测节点的重要度权重
铅	0.865 2
铜	0.134 8

根据节点的可靠性权重及重要度权重,可计算得到其综合权重,然后对原决策证据进行加权修正,并用 D-S 融合规则两两融合得到了最终结果.在案例 2 中各方法最终预警结果比较如表 8 所示.

表 8 在案例 2 中 5 种融合方法预警结果比较

方法	融合决策数据	预警结论
单因子法	铜	铜超标,红色
	1.647 mg · L ⁻¹	预警
加权平均融合法	铜	铜超标,红色
	1.646 mg · L ⁻¹	预警
经典 D-S 方法	$m(A)=0$	红色预警,预警原因是铜超标
	$m(B)=0$	
	$m(C)=1$	
Murphy 方法	$m(A)=0$	红色预警,主要原因是铜超标
	$m(B)=0.031 4$	
	$m(C)=0.968 6$	
本文方法	$m(A)=0$	红色预警,主要原因是铅超标
	$m(B)=0.001 7$	
	$m(C)=0.998 3$	

在表 6 中,存在铅、铜和化学需氧量 3 种水质参数均超标的情况,由于单因子法及加权平均融合法

均以最差水质指标数据进行分析判断,而铜超标最严重,因此表 8 中的这 2 种方法预警结论均是铜超标红色预警;由于经典 D-S 方法的组合规则倾向于放大证据之间的共识,而忽视证据之间的冲突,所以融合决策数据中 $m(A)=0$ $m(B)=0$ $m(C)=1$,且得出 $m(C)=1$ 的原因是由铜监测节点的基本信度分配函数值所决定的,从而在融合过程中忽略了其他证据节点数据的影响;而 Murphy 方法为了减轻证据之间的冲突,在融合过程中对各证据进行平均计算,从而融合计算得到 $m(C)=0.968 6$,即红色预警结论,但是该结论 52.15% 归功于铜监测节点的证据贡献;由于本文方法既考虑了监测节点之间的可靠性,又根据水污染程度的严重性考虑了监测节点的重要性,在案例 2 中,铅与铜相比,虽然超标倍数不高,但是铅的生物毒性更高,对水体污染的严重程度更大.因此,本文融合计算得到 $m(C)=0.998 3$,发出红色预警结论,而且得出该结论 70.60% 归功于铅监测节点的证据贡献,这说明:由于铅监测节点重要性更高,所以在做决策预警结论时对其应优先考虑.

案例 3 监测区域氨氮、总氮、总磷 3 种水质指标浓度超过了预警阈值,此 3 种水质指标监测节点端均发出了预警信息.

根据表 9 中提供的案例 3 实验数据,采用 5 种方法分别进行融合处理,所得最终预警结果如表 10 所示.

表 9 水环境监测预警实验数据 3

监测项目	水质指标监测节点采集数据		III 类水质标准
	节点 1	节点 2	
温度/℃	28.2	28.3	温升≤1℃ 温降≤2℃
pH 值	7.43	7.41	[6.00 9.00]
溶解氧/(mg · L ⁻¹)	5.82	5.85	[5.00 ,+ ∞)
氨氮/(mg · L ⁻¹)	0.862	0.864	[0 1.00]
化学需氧量/(mg · L ⁻¹)	13.56	13.58	[0 20.00]
总氮/(mg · L ⁻¹)	0.972	0.973	[0 1.00]
总磷/(mg · L ⁻¹)	0.049	0.048	[0 0.05]
铜/(mg · L ⁻¹)	0.213	0.214	[0 1.00]
铅/(mg · L ⁻¹)	0.023	0.024	[0 0.05]

表 10 在案例 3 中 5 种融合方法预警结果比较

方法	融合决策数据	预警结论
单因子法	总磷	正常
	$0.049 \text{ mg} \cdot \text{L}^{-1}$	
加权平均融合法	总氮	正常
	$0.9725 \text{ mg} \cdot \text{L}^{-1}$	
经典 D-S 方法	$m(A) = 0$	黄色预警
	$m(B) = 1$	
	$m(C) = 0$	
Murphy 方法	$m(A) = 0$	黄色预警
	$m(B) = 1$	
	$m(C) = 0$	
本文方法	$m(A) = 0$	黄色预警
	$m(B) = 1$	
	$m(C) = 0$	

从表 10 融合结果可以看出,当在监测区域内仅受到轻微污染而未超标的情况时,传统的单因子法及加权平均融合法均指示监测指标正常,系统不会向外界发出预警信息.而经典 D-S 方法、Murphy 方法和本文方法融合了多源数据信息,从而发出黄色预警信息,这有利于提醒决策层及时启动相应级别的应急预案,防患于未然.后 3 种方法所得到的融合决策数据也是相同的,实验结果说明当传感器各节点所提供证据趋于一致时,即当证据之间冲突很小或者无冲突时,基于证据理论的各种方法均能得到正确的决策结论.

4 结论

突发水污染事故具有很强的不确定性,而且会对人类的生产生活造成巨大危害,因此对突发水污染事故进行可靠预警具有重要意义.本文采用改进的 D-S 证据理论对传感器节点所采集的水质自动监测数据进行多源数据融合处理,从而降低不确定性的影响,得到可靠性较高的最终预警结果.本文主要结论如下:

1) 考虑了多源(同种或异种)传感器监测信息之间的相关关系,采用一元线性回归法对水质监测指标之间的相关性进行分析,建立紧密相关节点之间的关联矩阵.基于模糊三角函数,构建紧密相关节点的基本信度分配函数,采用 Jousselme 距离来度量紧密相关节点证据之间的相互支持度,通过观测节点被其他紧密相关节点的支持程度来反映其可靠性.典型案例的实验结果表明:与传统方法相比,本文方法能有效解决证据冲突问题,排除异常节点对融合结果的不利影响,从而得到可靠度更高、聚焦性

更好的预警结论.

2) 根据传感器节点所监测水质指标的生物毒性大小来定义其在融合决策中的重要性权重,从而反映水污染物所造成水污染事故的严重程度.本文方法既考虑了节点数据的可靠性权重,又考虑了重要度权重,即在 D-S 融合过程中引入综合权重进行加权修正.相较于传统以污染物浓度超标进行预警的方法,本文方法不仅考虑了污染物浓度超标的影响,而且也考虑了其生物毒性效应对环境影响的严重性,即优先考虑有毒污染物在预警决策时的贡献,所得预警结论更符合实际情况.

3) 相较于传统的单因子法及加权平均融合方法,本文方法考虑了多源(同种或异种)传感器监测数据信息.当监测区域仅受到轻微污染而未超标时,本文方法能充分考虑各水质指标对整体综合评价的作用,即通过多源数据融合处理,改变单一评价指标信息量少的局限性,适时发出相应级别的预警信息,从而提醒决策层提早准备,防患于未然.

5 参考文献

[1] 姜继平,王鹏,刘洁,等.突发水污染预警应急响应研究与实践的方法学辨析[J].环境科学学报,2017,37(9):3621-3628.

[2] 彭雪非.突发水污染事故的预警与应急监测技术进展[J].安徽农业科学,2014,42(20):6653-6655.

[3] 施建强.水环境实时自动监测系统的设计与实现[J].计算机工程与应用,2004,40(13):222-226.

[4] 袁维林.水质自动监测技术在水环境保护中的应用[J].低碳世界,2016(5):2-3.

[5] 蔡世清,周杰.基于支持向量机的多传感器数据融合算法[J].计算机工程与设计,2016,37(5):1352-1356.

[6] 彭能松,张维纬,张育钊,等.基于时间序列数据的无线传感器网络的异常检测方法[J].传感技术学报,2018,31(4):595-601.

[7] Cauteruccio F, Fortino G, Guerrieri A, et al. Short-long term anomaly detection in wireless sensor networks based on machine learning and multi-parameterized edit distance[J]. Information Fusion, 2019, 52: 13-30.

[8] 胡石,李兆辉,卢文伟,等.基于神经网络的无线传感器网络异常数据检测方法[J].计算机科学,2014,41(S2):208-211.

[9] 史斌,姜继平,王鹏.基于高频在线水质数据异常的突发污染预警[J].中国环境科学,2017,37(11):4394-4400.

[10] 靳春玲,王运鑫,贡力.基于模糊层次评价法的黄河兰州段突发水污染风险分析[J].安全与环境学报,

- 2018, 18(1): 363-368.
- [11] 严志军, 陶洋. 基于证据距离和不确定度的冲突数据融合算法 [J]. 计算机工程与应用, 2018, 54(10): 54-58, 104.
- [12] 李涛, 沈江. 基于多传感器信息融合的自然灾害预警模型研究 [J]. 电子科技大学学报: 社科版, 2015, 17(1): 34-38.
- [13] Dempster A P. Upper and lower probabilities induced by a multi-valued mapping [J]. Annual Math Statist, 1967, 38(4): 325-339.
- [14] Shafer G. A mathematical theory of evidence [M]. Princeton: Princeton University Press, 1976.
- [15] Zadeh L. A simple view of the Dempster-Shafer theory of evidence and its implication for the rule of combination [J]. AI Magazine, 1986, 7(1): 85-90.
- [16] Murphy C K. Combining belief function when evidence conflicts [J]. Decision Support Systems, 2000, 29: 1-9.
- [17] Han Deqiang, Deng Yong, Liu Qi. Combining belief functions based on distance of evidence [J]. Decision Support Systems, 2004, 38(3): 489-493.
- [18] Jousselme A L, Grenier D, Bossé E. A new distance between two bodies of evidence [J]. Information Fusion, 2001, 2(2): 91-101.
- [19] 国家环境保护总局科技标准司. 地表水环境质量标准: GB 3838—2002 [S]. 北京: 中国环境科学出版社, 2002.
- [20] 游猛. 典型城市污水处理工艺中多种生物毒性检测及水质安全评价 [D]. 西安: 西安建筑科技大学, 2017.
- [21] 曹伟, 秦延文, 张雷, 等. 化工企业废水污染源在线监控预警阈值确定方法研究 [J]. 环境科学研究, 2019, 32(3): 1-15.
- [22] 马自伟, 吴佳宁, 陈明, 等. 地表水中氨氮、总氮和总磷的污染程度及相关性分析 [J]. 广州化学, 2017, 42(5): 7-11, 31.
- [23] 尚小龙, 余厚平, 陈朴青, 等. 乐安河-鄱阳湖段水环境因子对重金属 Cu、Pb、Cd 迁移转化的影响 [J]. 江西师范大学学报: 自然科学版, 2014, 38(6): 650-655.
- [24] 熊宇奇. 典型村镇水污染指标相关性分析 [D]. 长沙: 湖南大学, 2014.
- [25] 李真. 海洋疏浚物生物毒性检测技术的研究 [D]. 青岛: 中国海洋大学, 2014.
- [26] Mowat F S, Bundy K J. Experimental and mathematical computational assessment of the acute toxicity of chemical mixtures from the Microtox assay [J]. Advances in Environmental Research, 2002, 6(4): 547-558.

The Reliable Warning Method for Sudden Water Pollution Based on Multi-Source Data Fusion

FU Tao¹, TAN Dekun^{2*}, FU Xuefeng², TU Zhenyu², WANG Hui²

(1. Jiangxi Provincial Flood Control and Information Center, Nanchang Jiangxi 330009, China; 2. Jiangxi Province Key Laboratory of Water Information Cooperative Sensing and Intelligent Processing, Nanchang Institute of Technology, Nanchang Jiangxi 330099, China)

Abstract: In the field of automatic monitoring of sudden water pollution accidents, the abnormal data caused by sensor nodes is an important reason to affect the warning reliability for automatic monitoring system. Considering the complementarity and correlation between multi-sensor information, a new reliable warning method for accidental water pollution based on multi-source data fusion is presented in this paper. The comprehensive weight is used to modify the original evidences based on the improved Dempster-Shafer (D-S) evidence theory, then the multi-source evidences are fused by utilizing the combination rule of D-S theory, and finally the early warning decision for sudden water pollution accidents can be made according to the fusion result. Compared with the traditional methods, the case analysis and experimental results show that the proposed method can make the warning decisions with higher credibility and better focus.

Key words: sudden water pollution; abnormal data; D-S theory; multi-source data fusion; forecasting and warning

(责任编辑: 冉小晓)