

# 改进型支持度函数的 WSN 水质监测数据融合方法

匡亮<sup>1,2</sup>, 施珮<sup>3,4</sup>, 季云峰<sup>1</sup>, 平震宇<sup>1</sup>, 陈北京<sup>2</sup>

(1. 江苏信息职业技术学院物联网工程学院, 无锡 214153; 2. 南京信息工程大学计算机与软件学院, 南京 210044;  
3. 中国水产科学研究院淡水渔业研究中心, 农业部淡水渔业和种质资源利用重点实验室, 无锡 214081; 4. 江南大学物联网工程学院, 无锡 214122)

**摘要:** 无线传感网络已被广泛应用到水质监测领域中, 针对水质监测中对传感器数据高精度的要求, 该研究提出一种基于支持度函数的数据融合算法。首先, 对各传感器采集的数据进行一致性检测, 保证数据的准确性; 其次, 采用改进的动态时间弯曲距离 (Improved Dynamic Time Warping Distance, IDTW) 对支持度函数 (Support Function, SF) 进行优化, 实现水质参数时间序列数据间的互支持度计算; 最后, 通过加权算法完成数据的融合过程, 实现错误数据的校正, 获得高质量融合数据。该算法在水质监测平台中进行了试验, 结果表明, IDTW-SF 融合算法的平均绝对误差值为 0.279 2%, 融合精度较其他对比算法更高, 且运行速度较快。IDTW-SF 支持度融合算法能够有效弥补现有水质监测系统数据采集准确率低、效率低等缺陷, 基于此方法的水质监测系统提高了溶解氧数据准确率, 在降低水产养殖风险, 提高养殖效率等方面发挥重要作用。

**关键词:** 水质; 监测; 数据融合; 无线传感器网络; 支持度函数; 动态时间弯曲距离

doi: 10.11975/j.issn.1002-6819.2020.16.024

中图分类号: S969.33

文献标志码: A

文章编号: 1002-6819(2020)-16-0192-09

匡亮, 施珮, 季云峰, 等. 改进型支持度函数的 WSN 水质监测数据融合方法 [J]. 农业工程学报, 2020, 36(16): 192-200.

doi: 10.11975/j.issn.1002-6819.2020.16.024 <http://www.tcsae.org>

Kuang Liang, Shi Pei, Ji Yunfeng, et al. Data fusion method for water quality monitoring using WSN based on improved support function[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2020, 36(16): 192-200. (in Chinese with English abstract) doi: 10.11975/j.issn.1002-6819.2020.16.024 <http://www.tcsae.org>

## 0 引言

无线传感器网络 (Wireless Sensor Networks, WSN) 以其灵活的部署、低成本, 小体积等优势被广泛应用于军事侦察、工业监测、农业监测和医疗监控等各个领域<sup>[1-7]</sup>。在水质监测中, 通过部署传感器节点来收集水质数据, 实现监测区域内水质参数的实时监控。然而, 当监测区域较大时, 由于监测区域中水质参数值分布不均匀的问题, 使得单一位置的监测数据不能很好的体现整个监测区域的真实情况<sup>[8]</sup>。同时, 传感器网络中数据的丢失、异常情况时有发生<sup>[9]</sup>, 而水下环境的复杂性增加了异常情况处理的难度。为了实现水质的准确监测, 保障数据的准确性和可靠性, 正确评估水体实际情况, 需要利用数据融合技术对不同位置的传感器数据进行融合, 从而提高水质监测系统的鲁棒性。

数据融合可以有效地减少数据冗余、提高数据质量<sup>[10-12]</sup>。基于不同理论的数据融合方法在使用过程中会呈现各自的局限性<sup>[13-14]</sup>。目前较为常用的数据融合方法

包括统计型算法、人工智能型算法和信息论型算法等。以贝叶斯理论为代表的统计类算法需要在融合多传感器数据之前获取先验知识和概率分布来计算传感器的可靠性<sup>[15]</sup>。以人工神经网络 (Artificial Neural Networks, ANN) 为代表的人工智能型算法可以处理不清楚、不确定的非线性系统的问题, 但复杂的结构和随机参数会导致融合结果不稳定<sup>[16]</sup>。基于支持度函数的融合算法是典型的信息论型算法, 该方法能够获得数据间的关系, 避免不可信数据对融合结果产生的不良影响<sup>[17]</sup>。国内外学者已经对支持度函数的数据融合算法展开了深入地研究。Luo 等<sup>[18]</sup>提出一种支持度函数, 它利用置信概率距离来表征数据间支持度值, 从而构建数据的关联矩阵完成数据融合。然而这种支持度函数需要进行积分运算, 函数计算量较大。罗本成等<sup>[19]</sup>基于肖维涅统计思想构建的支持度函数, 依据设置好的传感器数据偏差的阈值来计算传感器一致性矩阵。该支持度函数在计算时依赖于人为设置的阈值, 数据融合误差不稳定。Yager<sup>[20]</sup>利用指数衰减型支持度函数获取传感器数据间的支持度矩阵, 再通过幂均方加权计算数据融合结果。该方法能够有效地避免支持度函数非 1 即 0 的绝对性问题, 但指数衰减型支持度函数在参数的设置和计算复杂度上还有待改进。熊迎军等<sup>[21]</sup>运用灰色系统接近度思想提出了一种改进指数衰减型支持度函数。该算法不能对时间序列型数据进行分析, 数据融合过程忽略了数据在时间上的特征。段青玲等<sup>[22]</sup>对

收稿日期: 2020-01-14 修订日期: 2020-03-30

基金项目: 中央级公益性科研院所专项基金 (2019JBFM10); 江苏省高等职业教育产教融合集成平台建设计划项目 (苏教职函〔2019〕26 号); 江苏省自然科学基金 (BK20131097); 江苏信息职业技术学院重点课题 (JSITKY201803)

作者简介: 匡亮, 讲师, 主要从事人工智能、无线传感器网络等方向研究。

Email: kuangliang.89@163.com

指数衰减型支持度函数进行了改进,然而该算法在实际应用中的融合精度有待提高。

针对上述支持度函数融合算法存在的问题,本研究提出一种基于改进的动态时间弯曲距离支持度函数(Improved Dynamic Time Warping Distance Optimized Support Function, IDTW-SF)的加权融合算法。利用灰关联分析理论改进传统指数衰减型支持度函数中的指数运算,降低指数运算复杂度。同时,改变单一时刻点支持度的计算为时间段内支持度的计算,并利用动态时间弯曲距离(Dynamic Time Warping Distance, DTW)算法度量时间段内互支持度值。再通过时间分割策略来降低算法复杂度,从而实现 WSN 水质数据的融合。

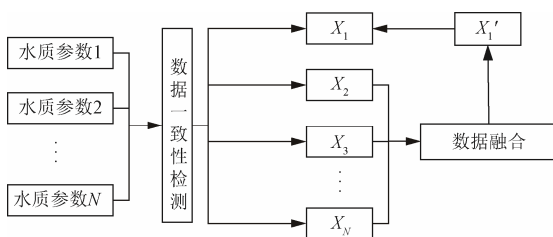
## 1 WSN 数据融合算法

由于同类传感器不同位置感知数据在时间和空间上的相关性,本研究在无线传感网络水质监测系统中,采用数据融合机制来提高数据质量,其主要步骤包括:

1) 通过数据的一致性检测操作对感知的原始数据进行检测,在检测离群值的基础上对丢失数据进行重新构建,获得新的数据集。

2) 当无线传感网络中某个传感器发生故障时,利用基于 IDTW-SF 支持度函数的数据融合算法对故障传感器的异常数据进行融合校正。其融合过程如图 1 所示,  $N$  个传感器节点采集的数据经过一致性检测后构成数据集  $X = \{X_1, X_2, \dots, X_N\}$ 。若传感器 1 为故障传感器,则系统利用融合机制将传感器节点  $X_2 = \{x_{21}, x_{22}, \dots, x_{2t}\}$ ,  $X_3 = \{x_{31}, x_{32}, \dots, x_{3t}\}$  和  $X_N = \{x_{N1}, x_{N2}, \dots, x_{Nt}\}$  进行处理,并获取融合数据  $X_1'$ , 其中  $t$  为采集时刻。

3) 基于融合机制的水质监测系统使用融合结果  $X_1'$  来替换故障传感器 1 中的数据  $X_1$ , 从而提高数据质量,为后续的数据分析和监测提供更可靠的信息。



注:  $\{X_1, X_2, \dots, X_i, \dots, X_N\}$  为  $N$  个传感器节点测量值经一致性检测后形成的数据集;  $X_1'$  为  $N$  个传感器测量值的融合值。

Note:  $\{X_1, X_2, \dots, X_i, \dots, X_N\}$  are  $N$  new sensor datasets constructed by consistency detection of  $N$  sensor measurements;  $X_1'$  is the fused value of  $N$  sensor measurements.

图 1 水质监测系统数据融合流程

Fig.1 Process of data fusion in water quality monitoring

### 1.1 数据的一致性检测

在水质监测中,传感器在水下受到的影响因素远比在空气中多。复杂的影响因素使得传感器更易发生故障和传输错误。对于传输中丢失的数据,采用线性插值法<sup>[23]</sup>对这些数据进行修补。同时,利用自滑动移动平均模型(Model Average Moving Integrated Autoregressive, ARIMA)<sup>[24]</sup>实现感知数据的一致性检测,其主要步骤包括:

1) 首先对水质参数时间序列数据间的相关性进行分

析,测试各感知数据  $X_i$  的稳定性,本研究以水体溶解氧浓度为例。

2) 确定 ARIMA 模型的自回归阶数  $p$  和移动平均阶数  $q$ , 基于设定的参数构建最优 ARIMA 模型。

3) 利用确定的 ARIMA 模型获取判定数据异常的置信区间 PI, 如式 (1) 所示

$$o_i(t) = x_i(t) + C \quad (1)$$

式中  $o_i(t)$  为  $t$  时刻的传感器感知数据,  $x_i(t)$  为 ARIMA 模型的预测值,  $C$  为代价函数<sup>[25]</sup>。式 (2) 为预测值  $x$  的 PI 预测区间。

$$PI = x \pm t_{\alpha/2, n-1} \cdot s \cdot \sqrt{1 + 1/n} \quad (2)$$

式中  $n$  是样本大小,  $s$  为标准偏差,  $t$  分布为遵循自由度为  $(n-1)$  的分布函数。

### 1.2 改进支持度函数

#### 1.2.1 支持度函数

基于支持度函数(Support Function, SF)的融合算法是一种使用较为广泛的数据融合算法。它能够有效地获取数据间的潜在关系,进而获得数据间的支持度。支持度函数即  $\sup(a, b)$ , 用于表征两元素  $a$  和  $b$  之间的接近程度<sup>[26]</sup>。当进行数据融合时,可信度高的数据拥有高的权值,可信度低的数据则拥有低的权值。

目前,较为常用的支持度函数为高斯型支持度函数(Gaussian support function,  $G$ ), 也可称作为指数衰减型支持度函数<sup>[20,27]</sup>, 如式 (3) 所示

$$\begin{aligned} \sup(a, b) &= G(a, b, K, \beta) = K \cdot e^{-\beta \cdot (a-b)^2} \\ K &\in [0, 1], \beta \geq 0 \end{aligned} \quad (3)$$

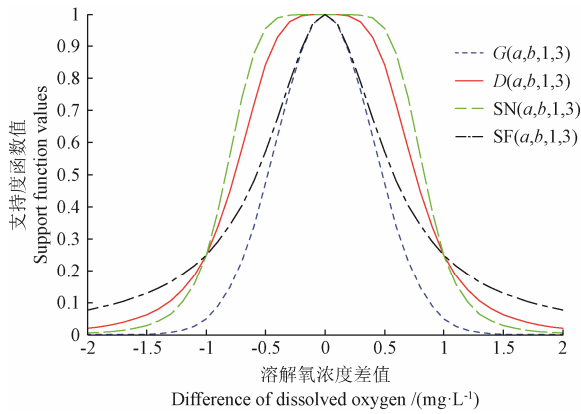
式中  $K$  为函数的振幅,  $\beta$  为函数的衰减系数。当两元素间接近程度越高时,函数的支持度值也越高。当  $a = b$  时,  $\sup(a, b) = K$ 。即,  $a$ 、 $b$  间差距越大,则  $\sup(a, b)$  越趋近于 0。

由于该支持度函数具有对称性,且运算过程是基于指数运算,故  $\sup(a, b)$  的计算也相对复杂和耗时。刘思峰等<sup>[28]</sup>利用灰色关联分析理论表达两元素间的接近程度,即支持度。本研究在这种思想的基础上提出一种支持度函数,对高斯型支持度函数进行改进,从而降低函数的计算复杂度,其函数表达式如式 (4) 所示

$$\begin{aligned} \sup(a, b) &= SF(a, b, K, \beta) = K \cdot (1 + \beta(a-b)^2)^{-1} \\ K &\in [0, 1], \beta \geq 0 \end{aligned} \quad (4)$$

水质监测中,同一深度的溶解氧浓度的数据差异  $< 2 \text{ mg/L}$ , 即  $(a-b) \in [-2, 2]$ 。则本研究所提出的支持度函数  $SF(a, b, K, \beta)$  其特性曲线如图 2 所示。 $(a-b)$  为不同溶解氧传感器感知数据的差值。为了更有效地获取数值间偏差影响,体现各信息权重对支持度的影响,本研究确定参数  $K=1$ ,  $\beta=3$ 。

图 2 显示,  $G(a, b, 1, 3)$ 、 $D(a, b, 1, 3)$ <sup>[21]</sup>、 $SN(a, b, 1, 3)$ <sup>[22]</sup> 和  $SF(a, b, 1, 3)$  分别表示各支持度函数的特性曲线,且当  $(a-b) \in [-1, 1]$  时,  $SF(a, b, 1, 3)$  能够更好的接近  $G(a, b, 1, 3)$ 。事实上,在水质监测中,溶解氧浓度的差异值大多数集中在  $[-1, 1]$ 。在水质监测中使用本研究提出的支持度函数  $SF(a, b, 1, 3)$  能够较好地反映传感网络数据之间的关系。



注:  $G(a,b,1,3)$  为高斯型支持度函数,  $D(a,b,1,3)$  为新型支持度函数,  $SN(a,b,1,3)$  为改进型支持度函数,  $SF(a,b,1,3)$  为本研究提出的支持度函数。

Note:  $G(a,b,1,3)$  is gaussian support function,  $D(a,b,1,3)$  is support function of new type,  $SN(a,b,1,3)$  is improved support function,  $SF(a,b,1,3)$  is the proposed support function.

图2 不同支持度函数特性曲线

Fig.2 Characteristic curve of different support functions

### 1.2.2 动态弯曲距离改进支持度函数

在高斯型支持度函数中,  $t$  时刻 2 个元素间的接近程度为  $\sup(a, b)$ , 其表达式如式 (3) 所示。然而在处理时间序列数据时, 会丢失数据前后之间的关联信息。考虑到水质参数序列数据在时间上的连续性, 本研究对 2 个时间序列  $U$  和  $V$  之间的接近程度进行分析, 完成时间序列间接接近程度  $\sup(U, V)$  的度量, 从而提出基于时间序列的动态弯曲距离改进的支持度函数 (DTW-SF)。DTW 是一种广泛用于度量时间序列之间相似性的算法之一<sup>[29]</sup>。本研究选择 DTW 来度量 2 个序列数据之间的接近程度, 结合 SF 支持度函数, 从而获得 DTW-SF 支持度函数, 其表达式如式 (5) 所示

$$\begin{aligned} \sup(U, V) &= \text{DTW-SF}(U, V, K, \beta) \\ &= K \cdot (1 + \beta \cdot \text{dist}(U, V))^2)^{-1} \\ K &\in [0, 1], \beta \geq 0 \end{aligned} \quad (5)$$

式中  $\text{dist}$  表征时间序列  $U = \{u_1, u_2, \dots, u_p, \dots, u_m\} (1 \leq p \leq m)$  和序列  $V = \{v_1, v_2, \dots, v_q, \dots, v_n\} (1 \leq q \leq n)$  的动态弯曲距离,  $m$  和  $n$  分别为序列的维数。  $S(U, V)$  表征 2 个时间序列的相似度, 即  $\text{dist}$  距离, 其表达式如式 (6) 所示

$$S(U, V) = \text{dist} = \min \left\{ \frac{1}{k} \sqrt{\sum_{k=1}^K w_k} \right\} \quad (6)$$

式中  $w = \{w_1, w_2, \dots, w_k\}$  为 DTW 中的规整路径,  $w_k$  为  $w$  中的第  $k$  个元素;  $d_{pq} = (u_p - v_q)^2$ 。为了获得最小路径距离, 定义  $\text{dist}$  为累加距离,  $\text{dist}(U, V)$  为当前格点  $d_{pq}$  距离的总和<sup>[30]</sup>。

$$\begin{aligned} \text{dist}(U, V) \\ = d_{pq} + \min(\text{dist}(u_{p-1}, v_q), \text{dist}(u_p, v_{q-1}), \text{dist}(u_{p-1}, v_{q-1})) \end{aligned} \quad (7)$$

### 1.2.3 优化的动态弯曲距离改进支持度函数

在传统 DTW 中, 使用较为常见的欧氏距离来测量  $m$  维空间中两点间距离。然而水质数据时间序列的相似度测量具有时间连续性, 单纯采用欧式距离度量各维度数据的相似性往往会忽视各维度之间的内在相关性, 故本研究选择马氏距离替代欧式距离来改进 DTW 算法。同时, DTW

算法的复杂度较高, 故引入时间序列分割策略<sup>[31]</sup>, 从而降低 DTW 算法的时间复杂度, 提高运算效率。通过该分割策略, 将两个时间序列划分为若干子序列, 完成时间序列的分割, 并构建优化的动态弯曲距离改进的支持度函数 (IDTW-SF)。

在 IDTW-SF 支持度函数中, 设置分割时间序列长度  $T=L$ 。则在时间  $T$  内, 由式 (7) 可得时间序列  $U(T)$  和  $V(T)$  之间的支持度函数, 其表达式如式 (8) 所示

$$\begin{aligned} \sup(U(T), V(T)) &= \text{IDTW-SF}(U(T), V(T), K, \beta) \\ &= K \cdot (1 + \beta \cdot \text{dist}(U(T), V(T)))^2)^{-1} \end{aligned} \quad (8)$$

### 1.3 基于优化的动态弯曲距离改进支持度函数的 WSN 数据融合

在水质监测中, 溶解氧传感器的感知数据经过一致性检测后构成  $X_i$ 。在时间间隔  $T$  内, 可得传感器  $i$  和传感器  $j$  的数据分别  $X_i(T)$  和  $X_j(T)$  ( $i, j \in \{1, 2, \dots, N\}$ ), 传感器间的相互支持度  $s_{ij}$  可表达为式 (9) 所示

$$s_{ij} = \sup(X_i(T), X_j(T)) = \text{IDTW-SF}(X_i(T), X_j(T), K, \beta) \quad (9)$$

由此可获得各传感器之间的相互支持度矩阵如式 (10) 所示

$$S = \begin{bmatrix} s_{11} & s_{12} & \dots & s_{1j} & \dots & s_{1N} \\ s_{21} & s_{22} & \dots & s_{2j} & \dots & s_{2N} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ s_{i1} & s_{i2} & \dots & s_{ij} & \dots & s_{iN} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ s_{N1} & s_{N2} & \dots & s_{Nj} & \dots & s_{NN} \end{bmatrix} \quad (10)$$

则在  $N$  个传感器组成的传感器组中,  $T$  时间间隔内的  $N-1$  个传感器对传感器  $i$  的支持度之和如式 (11) 所示

$$\text{sum IDTW-SF}(X_i(T)) = \sum_{i \neq j}^N s_{ij} \quad (11)$$

利用加权融合算法, 则传感器  $i$  的融合估计值如式 (12) 所示

$$X'_i(T) = \sum_{j \neq i}^N \omega_j \cdot X_j \quad (12)$$

式中  $\omega_j = \frac{s_{ij}}{\sum_{i \neq j}^N s_{ij}}$ , 为其他传感器  $j$  的权重值。

## 2 试验与分析

### 2.1 数据来源

本研究以江苏省无锡市南泉养殖基地试验池塘水质监测系统采集的溶解氧数据为试验对象, 获取 2017 年 5 月 24 日至 2017 年 5 月 29 日 6 d 的溶解氧数据。图 3a 所示试验池塘长 80 m, 宽 50 m。根据养殖专家的经验分别在池塘的不同位置部署 5 个溶解氧传感器和增氧机。它们的部署深度相同, 均位于水下 0.6 m 处, 其位置分布 (图 3)。其中, 1 号传感器数据控制 1 号、2 号和 3 号增氧机, 2 号传感器数据控制 4 号增氧机, 5 号传感器数据控制 5 号增氧机。当某一传感器发生故障时, 系统则通



过数据融合机制对其余传感器的数据进行融合，获取高质量的融合数据。各传感器节点每 10 min 采样数据，共计 864 个数据点。所有数据经网关发送至上位机存储，并供后台数据处理软件进行分析处理。

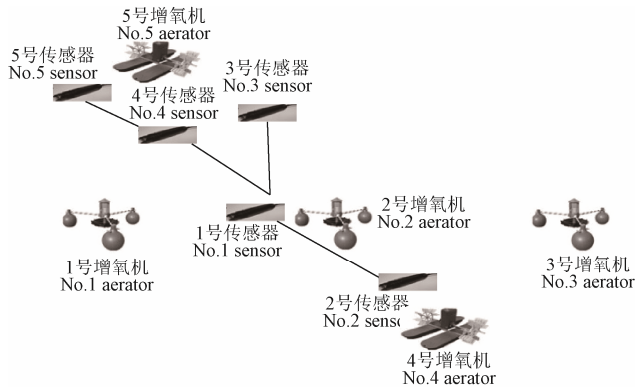


Fig.3 养殖池塘的水质监测设备部署图

Fig.3 Deployment diagram of water quality monitoring devices in the aquaculture pond

## 2.2 一致性检测分析与时间序列分割

本研究使用 ARIMA 完成 864 个样本数据的一致性检测，其中异常数据共 35 个，丢失数据 16 个。通过仿真试验对样本数据进行测试，设置置信区间为 95%，获得样本数据的异常检测率（Detection Rate, DR, %）为 94.29%。其中，异常检测率  $DR = TP \cdot 100\% / (TP + FN)$ ；正确率（True Positive, TP, %）为异常数据被正确检测出的个数；准确率（False Negative, FN, %）为异常数据被错误检测的个数。本研究所有试验程序均在 MATLAB 2014a 环境下上运行和实现，计算机操作系统为 Windows 10 (64 bit)，运行内存为 16 GB，CPU 频率为 2.7 GHz。

本研究选择平均绝对误差（Mean Absolute Error, MAE）和运行时间（Time, s）作为数据融合算法性能评估的指标。6 d 内基于 IDTWS-SF 支持度函数的融合算法的性能结果（图 4）。

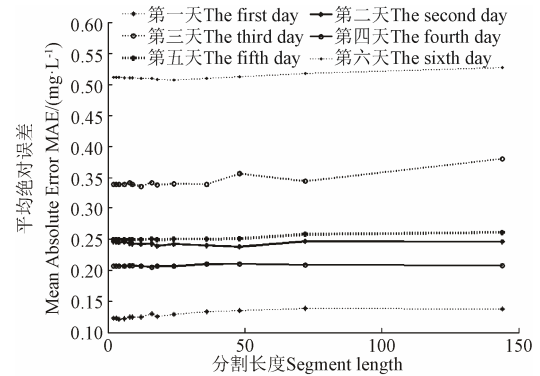
如图 4 所示，6 d 内 IDTWS-SF 的融合精度 MAE 值随着分割长度的不断增大无明显变化，处于较稳定状态。而运行时间参数随着分割长度的不断增大几乎呈线性变化趋势。由此可见，分割长度对融合精度 MAE 值的影响较小，对算法运行时间有较大的影响。即分割长度越小，IDTWS-SF 支持度函数的综合性能越好。综上所述，本研究确定时间序列的分割长度  $L=2$ 。

## 2.3 结果与分析

### 2.3.1 支持度函数性能分析

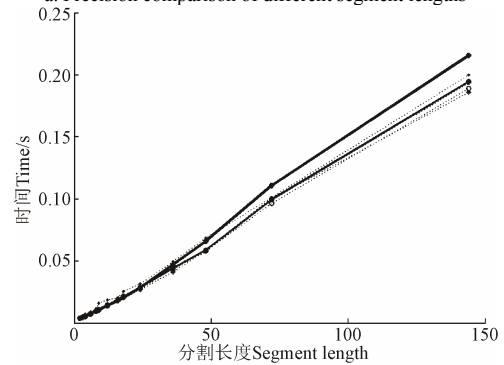
基于上述参数设置情况，使用基于 IDTWS-SF 支持度函数的加权融合算法对水质参数数据进行融合，本研究以溶解氧浓度为例进行试验。6 d 内溶解氧浓度的融合输出曲线如图 5 所示，IDTWS-SF 支持度函数融合输出值能较好的拟合观测值，6 d 内融合值曲线与观测值曲线较为一致，无明显的融合异常情况。图 6 为 IDTWS-SF 支持度函数加权融合算法的融合误差率，图中所有测试样本的融合误差率均低于 0.25，且大部分样本点的融合误

差率低于 0.05。试验结果证明，该算法有较好的融合效果，能有效地实现水质参数溶解氧浓度的多传感器数据融合。



a. 不同分割长度融合精度对比

a. Precision comparison of different segment lengths



b. 不同分割长度融合时间对比

b. Time comparison of different segment lengths

注：横坐标表征时间序列的分割长度，范围为（1，144）。溶解氧时间序列一天的序列长度为 144。单位长度表征间隔 10 min 采集的数据。  
Note: Horizontal axis shows the segment length range of time series (1,144). The length of dissolved oxygen time series is 144. The unit length represents the data collected in 10 min interval.

图 4 优化的动态弯曲距离改进支持度函数六天内不同分割长度融合性能对比图

Fig.4 Performance comparison of IDTWS-SF with different segment lengths in six days

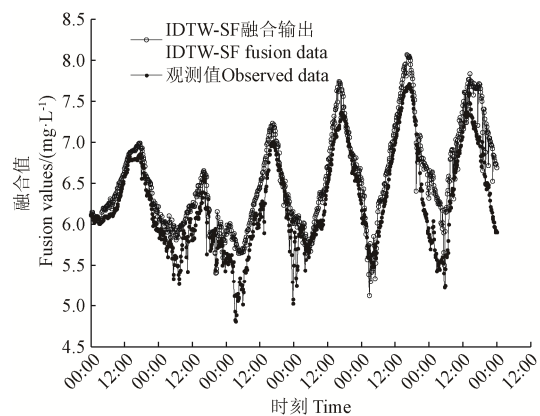


图 5 优化的动态弯曲距离改进支持度函数加权融合结果  
Fig.5 Results of IDTWS-SF support degree function weighted fusion

为了有效地评估 IDTW 算法对支持度函数 SF 的改进效果，分别选择 SF 支持度函数、余弦夹角<sup>[32]</sup>改进 SF 支持度函数（Cosine Angle Improved SF Function, Cos-SF）、传统 DTW 改进 SF 支持度函数进行对比试验。对 6 d 内的水质参数溶解氧浓度的融合残差值进行对比，融合残差结果（图 7）。

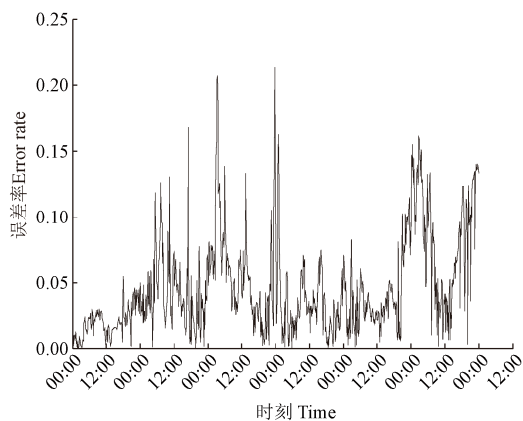


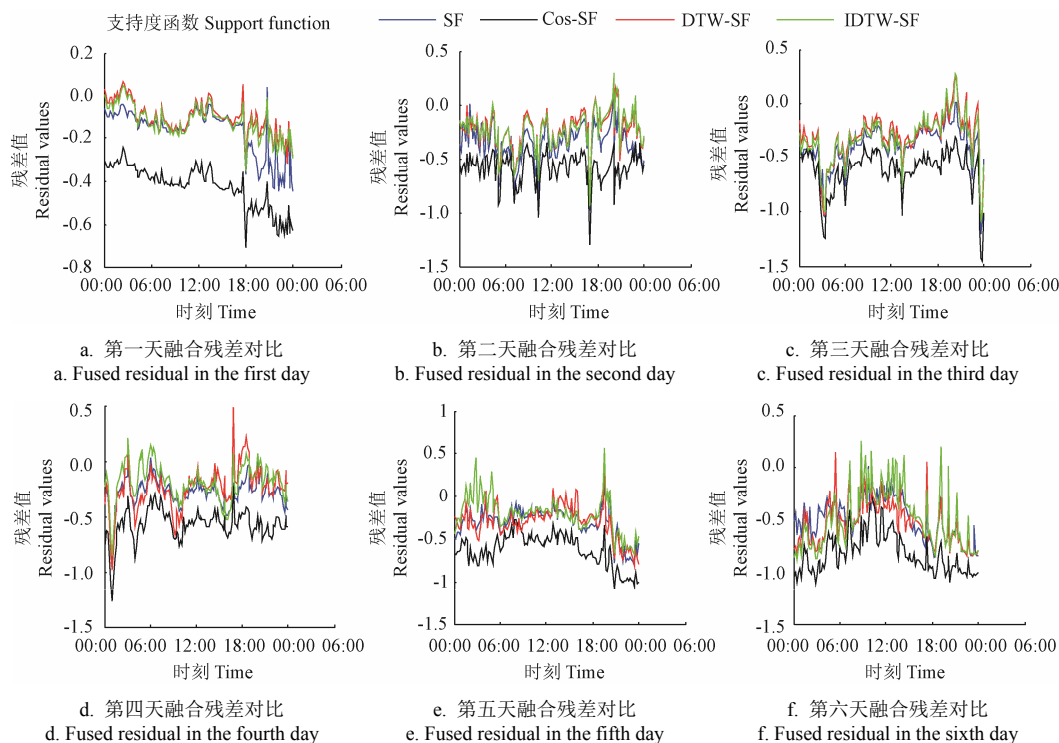
图6 优化的动态弯曲距离改进支持度函数融合误差率

Fig.6 Fusion error rate of IDTWS-SF

图7 为不同的改进 SF 支持度函数的加权融合算法的

融合残差结果对比图。通过观察这 4 种支持度函数融合算法的残差曲线的浮动范围可以发现, 6 d 内测试数据集中 Cos-SF 支持度函数融合算法的残差浮动范围明显大于其他 3 个支持度函数。SF、DTW-SF 和 IDTW-SF 3 种支持度函数的融合残差浮动情况较为接近。为了更好地对 4 种支持度算法的性能进行对比分析, 对支持度函数融合精度指标 MAE 和时间进行计算, 各支持度函数融合结果(表 1)。

表 1 显示 IDTW-SF 支持度函数加权融合算法的性能明显优于其他 3 个支持度函数融合算法。就 MAE 指标而言, IDTW-SF 的 MAE 值相较于 DTW-SF、Cos-SF 和 SF 分别降低了 6.308 7%、54.214 5%和 16.306 9%。由此可见, 改进的 DTW 方法能够有效地提高支持度函数的融合精度。在运行时间上, IDTW-SF 比 DTW-SF 快 3.576 s, 相较于 Cos-SF 和 SF 支持度函数融合算法仅慢 0.001 1 s 和 0.003 5 s, 运行速度相差很小。



注: SF 为支持度函数, Cos-SF 为余弦夹角改进支持度函数, DTW-SF 为动态弯曲距离改进支持度函数, IDTW-SF 为优化的动态弯曲距离改进支持度函数。

Note: SF is a support function; Cos-SF is Cosine angle improved support function; DTW-SF is dynamic warping distance improved support function; IDTW-SF is optimized dynamic warping distance improved support function.

图7 不同的改进支持度函数 SF 的融合残差对比图

Fig.7 Comparison of different optimized SF support degree functions for fused residual

表 1 不同的改进 SF 支持度函数算法的融合结果

Table 1 Weighted fusion results of different optimized support degree functions

| 指标 Indicator                             | SF      | Cos-SF  | DTW-SF | IDTW-SF |
|--|---------|---------|--------|---------|
| 平均绝对误差<br>Mean Absolute Error<br>MAE / % | 0.333 6 | 0.609 8 | 0.298  | 0.279 2 |
| 运行时间<br>Run time/s                       | 0.019 2 | 0.021 6 | 3.599  | 0.022 7 |

DTW-SF 与 Cos-SF、SF 支持度函数融合结果的对比试验发现, DTW-SF 支持度函数有较高的 MAE 值。

这一结果表明, DTW 比余弦距离和欧式距离更适用于度量时间序列间的支持度, 它与 SF 支持度函数的结合能够有效的提高融合精度。而马氏距离和时间分割策略改进的 DTW 算法, 不仅提高算法的融合精度, 而且保障了算法的运行效率。综合融合精度和效率两方面考虑, IDTW-SF 支持度函数更适用于水质监测数据的融合。

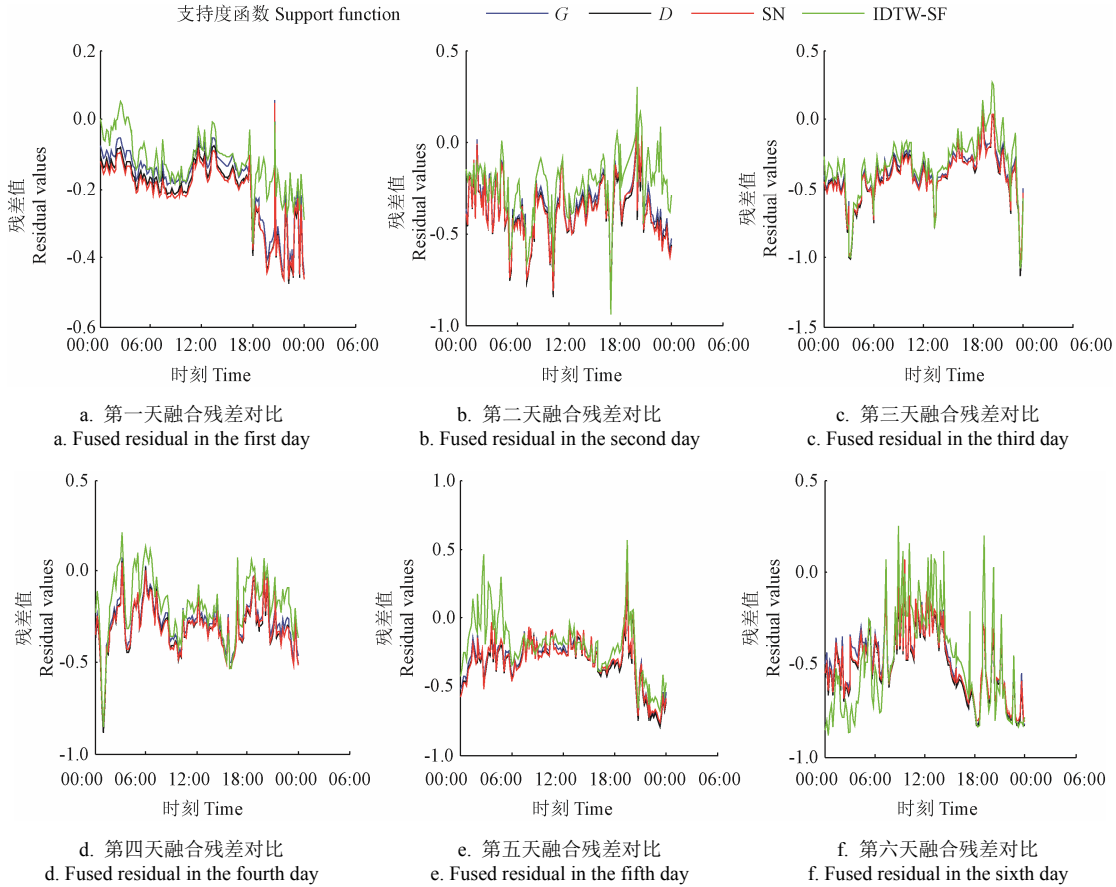
### 2.3.2 不同的支持度函数融合结果分析

为了深入评估本研究提出的 IDTW-SF 支持度融合算法的效果, 选择高斯型支持度函数  $G$  函数<sup>[20]</sup>、改进支持

度函数  $D$  函数<sup>[21]</sup>、改进支持度函数  $SN$  函数<sup>[22]</sup>3 种支持度函数作为对比函数。将 6 d 内 4 种支持度函数的溶解氧融合结果残差值进行对比分析（图 8）。

图 8 为 4 种支持度函数的加权融合结果的残差对比图。通过观察 4 种支持度函数融合算法的残差曲线浮动范围可以发现，尽管 IDTW-SF 支持度函数的融合残差值

的浮动范围比其他 3 个支持度函数大，但其残差值更接近 0。结果表明，该算法的融合值与实际值差异更小。其他 3 个支持度函数融合残差浮动范围则较为接近，无明显差异度。为了深入分析各支持度函数性能，将指标 MAE 和运行时间作为评价指标，获得如表 2 所示的支持度函数融合性能结果。



注：G 为高斯型支持度函数；D 为新型支持度函数；SN 为改进型支持度函数；IDTW-SF 为优化的动态弯曲距离改进支持度函数。  
Note:  $G$  is gaussian support function;  $D$  is support function of new type;  $SN$  is improved support function; IDTW-SF is improved dynamic warping distance improved support function.

图 8 4 种支持度函数融合残差对比图  
Fig.8 Comparison of four support degree functions for fused residual

| 表 2 四种支持度函数的融合性能对比   |         |         |         |          |
|--|---------|---------|---------|----------|
| Table 2 Weighted fusion results of four support degree functions |         |         |         |          |
| 指标 Indicator   | $G$     | $D$     | $SN$    | IDTWS-SF |
| 平均绝对误差 Mean Absolute Error MAE/%                                 | 0.335 0 | 0.360 9 | 0.356 4 | 0.279 2  |
| 运行时间 Run time/s  | 0.021 2 | 0.019 6 | 0.018 9 | 0.021 7  |

由表 2 可以清楚地发现，相较于其他 3 种支持度函数，IDTWS-SF 的融合性能有明显的优势。就 MAE 值而言，IDTWS-SF 相较于  $G$  函数、 $D$  函数和  $SN$  函数分别降低了 16.656 7%、22.637 9%和 21.661 1%。在算法运行速度上，IDTWS-SF 比  $G$  函数、 $D$  函数和  $SN$  函数分别多耗时 0.000 5、0.002 1 和 0.002 8 s，运行时间差异较小。 $D$  函数和  $SN$  函数均为高斯支持度函数改进后的支持度函数。2 种支持度函数的融合运行时间比高斯函数快 0.001 6 和 0.002 3 s，在融合精度上差异较小。

上述结果表明，将改进的 DTW 算法引入到 SF 支持度函数中可以有效地提高加权融合算法的融合性能。基于型支持度函数 IDTWS-SF 的数据融合算法拥有较好的融合性能，更适用于实际的水质监测。

3 结 论

针对水质监测系统对多传感器数据融合的高精度、快速响应要求，本研究提出一种采用 IDTW-SF（Improved Dynamic Time Warping Distance Optimized Support Function）支持度函数的数据融合算法。该支持度函数在改进传统指数衰减支持度函数的基础上，对时间段内传感数据间互支持度进行计算。IDTW-SF 无需自行设定阈值，能够有效的避免传统支持度函数非 1 即 0 的问题。在 IDTW-SF 中，DTW 算法对 SF 函数的改进能够有效地获取时间序列数据在时间上的关联性。同时，引入的马氏距离和时间序列分割策略能够降低 DTW 算法的复杂

度,提高算法的融合精度。基于 IDTW-SF 支持度函数融合算法有较高的融合精度,其融合平均绝对误差值为 0.279 2%。算法的运算效率较高,运行时间仅 0.021 2 s。本研究以江苏省无锡市南泉养殖基地试验池塘水质参数为试验对象,对引起鱼类异常死亡重要因素之一的溶解氧数据进行融合,结果表明 IDTW-SF 支持度函数的融合效果相较于指数衰减型支持度函数和其他对比的支持度函数有较大程度的提高,能够应用于无线传感网络水质监测系统中,对提高鱼类存活效率,降低渔民养殖风险有重要意义。

#### [参 考 文 献]

- [1] 杨传顺. 基于 WSN 的目标定位跟踪在 C4ISRT 中的应用[J]. 北京联合大学学报, 2012, 26(4): 23-27.  
Yang Chuanshun. Applications of target location and tracking based on WSN on C4ISRT[J]. Journal of Beijing Union University, 2012, 26(4): 23-27. (in Chinese with English abstract)
- [2] 赵小敏, 梁学利, 蒋双双, 等. 安全的 WSN 数据融合隐私保护方案设计[J]. 通信学报, 2014, 35(11): 154-161.  
Zhao Xiaomin, Liang Xueli, Jiang Shuangshuang, et al. Design of secure privacy-preserving data aggregation scheme for wireless sensor network[J]. Journal on Communications, 2014, 35(11): 154-161. (in Chinese with English abstract)
- [3] Sahoo C. A cooperative strategy for collision detection and prevention for unmanned ground vehicles in military applications using WSN based VANET[J]. International Journal of Computer Applications, 2014, 102(1): 20-26.
- [4] Dey N, Ashour A S, Shi Fuqian, et al. Developing residential wireless sensor networks for ECG healthcare monitoring[J]. IEEE Transactions on Consumer Electronics, 2017, 63(4): 442-449.
- [5] Bapat V, Kale P, Shinde V, et al. WSN application for crop protection to divert animal intrusions in the agricultural land[J]. Computers and Electronics in Agriculture, 2017, 133: 88-96.
- [6] 王纪章, 彭玉礼, 李萍萍. 基于事件驱动与数据融合的温室 WSN 节能传输模型[J]. 农业机械学报, 2013, 44(12): 258-261.  
Wang Jizhang, Peng Yuli, Li Pingping. Energy transmission model of WSN in greenhouse based on event-driven and data fusion[J]. Transactions of the Chinese Society for Agricultural Machinery, 2013, 44(12): 258-261. (in Chinese with English abstract)
- [7] 张小栓, 刘贺, 崔衍, 等. 钾肥生产原卤井无线传感器网络监测系统[J]. 农业工程学报, 2017, 33(S1): 199-205.  
Zhang Xiaoshuan, Liu He, Cui Yan, et al. Monitoring system for brine well in production of potash fertilizer based on wireless sensor network[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2017, 33(S1): 199-205. (in Chinese with English abstract)
- [8] Han Guangjie, Liu Li, Jiang Jinfang, et al. Analysis of energy-efficient connected target coverage algorithms for industrial wireless sensor networks[J]. IEEE Transactions on Industrial Informatics, 2015, 13(1): 135-143.
- [9] 黄鹤, 刘一恒, 赵熙, 等. 多层多源信息融合旋翼无人机测高算法[J]. 中国惯性技术学报, 2018, 26(3): 42-48, 55.  
Huang He, Liu Yiheng, Zhao Xi, et al. Multi-layer multi-source information fusion algorithm for hovering UAV height measurement[J]. Journal of Chinese Inertial Technology, 2018, 26(3): 42-48, 55. (in Chinese with English abstract)
- [10] 焉晓贞. 无线传感器网络中传感数据估计方法研究[D]. 哈尔滨: 哈尔滨工程大学, 2012.  
Yan Xiaozhen. Sensor Data Estimation Method for Wireless Sensor Networks[D]. Harbin: Harbin Engineering University, 2012. (in Chinese with English abstract)
- [11] 杨东华, 李苏宁, 王宏志, 等. 基于任务合并的并行大数据清洗过程优化[J]. 计算机学报, 2016, 39(1): 97-108.  
Yang Donghua, Li Ningning, Wang Hongzhi, et al. The optimization of the big data cleaning based on task merging[J]. Chinese Journal of Computers, 2016, 39(1): 97-108. (in Chinese with English abstract)
- [12] 王玲. 无线传感器网络时空相关性数据融合算法研究[D]. 重庆: 重庆大学, 2014.  
Wang Ling. Study of Temporal-Spatial Correlation Based Data Fusion Algorithm in Wireless Sensor Networks[D]. Chongqing: Chongqing University, 2014. (in Chinese with English abstract)
- [13] 周艳青, 薛河儒, 姜新华, 等. 基于改进的卡尔曼滤波算法的气象数据融合[J]. 计算机系统应用, 2018, 27(4): 184-189.  
Zhou Yanqing, Xue Heru, Jiang Xinhua, et al. Meteorological data fusion based on proposed Kalman filter method[J]. Computer Systems & Applications, 2018, 27(4): 184-189. (in Chinese with English abstract)
- [14] 沙敏, 宋超, 张正勇, 等. 谱图数据融合结合模式识别算法鉴别苹果香精[J]. 食品科学, 2016, 37(22): 192-197.  
Sha Min, Song Chao, Zhang Zhengyong, et al. Discrimination of apple essences based on spectral data fusion combined with pattern recognition algorithm[J]. Food Science, 2016, 37(22): 192-197. (in Chinese with English abstract)

- abstract)
- [15] D'Addabbo A, Refice A, Lovergine F P, et al. DAFNE: A Matlab toolbox for Bayesian multi-source remote sensing and ancillary data fusion, with application to flood mapping[J]. *Computers & Geosciences*, 2018, 112: 64-75.
  - [16] Jha S K, Josheski F, Marina N, et al. GC-MS characterization of body odor for identification using artificial neural network classifiers fusion[J]. *International Journal of Mass Spectrometry*, 2016, 406, 35-47.
  - [17] Bai Xingzhen, Wang Zidong, Sheng Li, et al. Reliable data fusion of hierarchical wireless sensor networks with asynchronous measurement for greenhouse monitoring[J]. *IEEE Transactions on Control Systems Technology*, 2019, 27(3): 1036-1046.
  - [18] Luo R C, Lin M H. Dynamic multi-sensor data fusion system for intelligent robots[J]. *IEEE Journal of Robotics and Automation*, 1988, 4(4): 386-396.
  - [19] 罗本成, 原魁, 陈晋龙, 等. 一种基于不确定分析的多传感器信息动态融合方法[J]. *自动化学报*, 2004, 30(3): 89-97.  
Luo Bencheng, Yuan Kui, Chen Jinlong, et al. Uncertainty analysis based dynamic multi-sensor data fusion[J]. *Acta Automatica Sinica*, 2004, 30(3): 89-97. (in Chinese with English abstract)
  - [20] Yager R R. The power average operator[J]. *IEEE Transactions on Systems, Man, and Cybernetics*, 2001, 31(6): 724-731.
  - [21] 熊迎军, 沈明霞, 陆明洲, 等. 温室无线传感器网络系统实时数据融合算法[J]. *农业工程学报*, 2012, 28(23): 160-166.  
Xiong Yingjun, Shen Mingxia, Lu Mingzhou, et al. Algorithm of real time data fusion for greenhouse WSN system[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2012, 28(23): 160-166. (in Chinese with English abstract)
  - [22] 段青玲, 肖晓琰, 刘怡然, 等. 基于改进型支持度函数的畜禽养殖物联网数据融合方法[J]. *农业工程学报*, 2017, 33(S1): 239-245.  
Duan Qingling, Xiao Xiaoyan, Liu Yiran, et al. Data fusion method of livestock and poultry breeding internet of things based on improved support function[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2017, 33(S1): 239-245. (in Chinese with English abstract)
  - [23] Aliaga R J. Real-time estimation of zero crossings of sampled signals for timing using cubic spline interpolation[J]. *IEEE Transactions on Nuclear Science*, 2017, 64(8): 2414-2422.
  - [24] 麦鸿坤, 肖坚红, 吴熙辰, 等. 基于 R 语言的负荷预测 ARIMA 模型并行化研究[J]. *电网技术*, 2015, 39(11): 3216-3220.  
Mai Hongkun, Xiao Jianhong, Wu Xichen, et al. Research on ARIMA model parallelization in load prediction based on R language[J]. *Power System Technology*, 2015, 39(11): 3216-3220. (in Chinese with English abstract)
  - [25] Hill D J, Minsker B S. Anomaly detection in streaming environmental sensor data: A data-driven modeling approach[J]. *Environmental Modelling & Software*, 2010, 25(9): 1014-1022.
  - [26] Shi Zhenyun, Tian Jiawei, Luo Ruidong, et al. Multifeedback control of a shape memory alloy actuator and a trial application[J]. *IEEE Transactions on Systems Man & Cybernetics Systems*, 2018, 48(7): 1106-1119.
  - [27] 孙智国. 一种支持度函数的改进及其 WSN 数据融合应用研究[D]. 重庆: 重庆大学, 2015.  
Sun Zhiguo. Research on a Support Function Improvement and Its Application for WSN Data Fusion[D]. Chongqing: Chongqing University, 2015. (in Chinese with English abstract)
  - [28] 刘思峰, 谢乃明, Jeffery F. 基于相似性和接近性视角的新型灰色关联分析模型[J]. *系统工程理论与实践*, 2010, 30(5): 881-887.  
Liu Sifeng, Xie Naiming, Jeffery F. On new models of grey incidence analysis based on visual angle of similarity and nearness[J]. *Systems Engineering-Theory & Practice*, 2010, 30(5): 881-887. (in Chinese with English abstract)
  - [29] 李正欣, 张凤鸣, 李克武, 等. 一种支持 DTW 距离的多元时间序列索引结构[J]. *软件学报*, 2014, 25(3): 560-575.  
Li Zhengxin, Zhang Fengming, Li Kewu, et al. Index structure for multivariate time series under DTW distance metric[J]. *Journal of Software*, 2014, 25(3): 560-575. (in Chinese with English abstract)
  - [30] Schultz D, Jain B. Nonsmooth analysis and subgradient methods for averaging in dynamic time warping spaces[J]. *Pattern Recognition*, 2018, 74: 340-358.
  - [31] 汪竹蓉, 李伟, 朱碧磊, 等. 基于音乐内容分析的音频认证算法[J]. *计算机研究与发展*, 2012, 49(1): 158-166.  
Wang Zhurong, Li Wei, Zhu Bilei, et al. Audio authentication based on music content analysis[J]. *Journal of Computer Research and Development*, 2012, 49(1): 158-166. (in Chinese with English abstract)
  - [32] George K K, Kumar C S, Ramachandran K I, et al. Cosine distance features for improved speaker verification[J]. *Electronics Letters*, 2015, 51(12): 939-941.



## Data fusion method for water quality monitoring using WSN based on improved support function

Kuang Liang<sup>1,2</sup>, Shi Pei<sup>3,4</sup>, Ji Yunfeng<sup>1</sup>, Ping Zhenyu<sup>1</sup>, Chen Beijing<sup>2</sup>

(1. School of IoT Engineering, Jiangsu Vocational College of Information Technology, Wuxi 214153, China; 2. School of Computer and Software, Nanjing University of Information Science & Technology, Nanjing 210044, China; 3. Tilapia Industrial Research Center of China Agricultural Research System, Key Laboratory of Freshwater Fisheries and Germplasm Resources Utilization, Ministry of Agriculture Freshwater Fisheries Research Center, Chinese Academy of Fishery Sciences, Wuxi 214081, China; 4. School of IoT Engineering, Jiangnan University, Wuxi 214122, China)

**Abstract:** Wireless sensor network has been widely used in various types of industries, such as water quality detection. Due to all kinds of device faults and transmission faults, there are some outliers during water quality monitoring. In a large monitoring area, the difference between monitoring parameter values exists due to uneven distribution. The monitoring data in a single location is unsuitable to represent the real situation of the whole monitoring area. The data fusion method is used to fuse data in multiple locations. Traditional methods in water quality data fusion have problems of low accuracy and efficiency for limited to poor generalization and complex calculations. In response to requirements of high-precision for sensor data in water quality detection, a novel data fusion method based on a new support function IDTW-SF (Improved Dynamic Time Warping Distance Optimized Support Function) was proposed in this study. Based on the importance of dissolved oxygen in various water quality parameters, it was used as an example to study this research in this study. The purpose of data fusion was for correcting outliers to obtain high-quality data. Firstly, the consistency detection of sensor data improved the quality of the fusion data. With high computing complexity, the traditional Gaussian support function was a defective method in data fusion. The dissolved oxygen content was used for example to study the new data fusion method. An improved dynamic time warping distance IDTW (Improved Dynamic Time Warping Distance) was used to optimize a new support function SF (Support Function), thus calculated the support degree value between water quality time series data. Unlike the Gaussian support function, the SF function obtained mutual support degree of sensors without the exponent calculation. The weighted algorithm was used to complete the data fusion process. Based on Grey correlation analysis, the IDTW-SF combined the dynamic time warping distance, time segment strategy and Mahalanobis distance together. DTW algorithm was applied to replace the Euclidean distance and compute the distance between time series. Time series segmentation strategy was utilized to reduce the computation dimension of the DTW algorithm. To prove the validity of this fusion algorithm, various experiments were carried out on a water quality monitoring platform of aquaculture pond from May 24<sup>th</sup>, 2017 to May 29<sup>th</sup>, 2017 in the Freshwater Fisheries Research Center, Jiangsu Province. Different distance measures were applied to optimized the SF function and construct new support functions, such as Cos-SF (Cosine angle optimized SF) and DTW-SF. Meanwhile, existed support functions were realized in this study, such as the G function (Gaussian function). These functions were all used as a comparison. The experimental results showed that the mean absolute error value of the proposed fusion method was 0.279 2%. Compared with DTW-SF, Cos-SF, and SF, the mean absolute error value was reduced by 6.308 7%, 54.214 5%, and 16.306 9%, respectively. And the fusion effect of dissolved oxygen was improved obviously. The run time of the IDTW-SF was just 0.021 7 s. The fusion accuracy of the IDTW-SF support degree function was higher than the counterpart existed functions, and its efficiency was also high. Meanwhile, the residual value of IDTW-SF also had a distinct advantage over the other contrast functions. It was obvious that the combination of the IDTW algorithm and the SF method was reasonable and effective. To summarize, the proposed fusion algorithm in this study fully obtained the correlation of monitoring data and overcame shortcomings of traditional function, thus improved the accuracy and efficiency of fusion results. High fusion accuracy could meet the needs of water quality monitoring in the real world.

**Keywords:** water quality; monitoring; data fusion; wireless sensor network; support function; dynamic time warping distance