

基于环形缓冲区的 WSN 数据融合技术研究

杨芬婷, 徐震

(武汉轻工大学 电气与电子工程学院, 武汉 430048)

摘要:无线传感器网络监测系统中,环境变化缓慢和节点感知范围重叠所造成的冗余数据会增加节点的数据发送量,降低信息收集效率并导致传感器节点过早死亡.因此,提出一种基于环形缓冲区的簇内数据融合方案.所有节点采用环形缓冲区存储数据.源节点基于环形缓冲区采用二值化相似函数和滑动四分位检测法,在保证数据时间关联性的同时剔除冗余数据和瞬时性异常数据.簇头节点基于加权皮尔逊距离的改进支持度对从源节点接收到的数据进行加权融合.仿真实验表明,所提出的方案在网络剩余节点数、网络剩余能量和网络发送数据包数等3个方面有明显的优势.

关键词:无线传感器网络;数据融合;支持度;环形缓冲区;滑动窗口

中图分类号:TP212.9

文献标志码:A

文章编号:1000-2367(2024)02-0062-10

无线传感器网络凭借节点体积小、成本低和自组网等特点广泛应用于环境监测、灾情警报、健康检测、智能交通等领域^[1-2].传感器节点将感知区域采集的信息进行预处理,经节点单跳传输或多节点转发最终汇聚到汇聚节点或基站.与以往的无线网络不同,无线传感器网络的通信和采集等主要工作需要节点提供足够的电池能量,若传感器节点电池能量极低或耗尽,则该节点将成为故障节点或死亡节点,降低无线传感器网络的寿命.因此,在环境监测的数据收集过程中,通过有效利用资源来降低能耗,在无线传感器网络的研究中占据十分重要的地位^[3-4].

由于传感器节点密集部署,节点采集范围的重叠会产生许多冗余数据,频繁的采集和发送会产生许多不必要的能量消耗并造成网络堵塞.另外,感知环境中的不稳定性如电磁噪声、压力、辐射等外界干扰、信道质量劣化以及传感器节点自身故障都会使传感器节点采集到许多异常数据,从而降低了无线传感器网络的性能^[5-6].数据融合技术结合多个传感器节点信息,多节点协作减少数据冗余和数据传输量的同时提高信息的整体准确性,在降低无线传感器网络能耗方面发挥重要作用^[7-8].

研究表明,将无线传感器网络划分为多个簇,簇内成员节点向簇头节点发送数据,通过簇内融合采集到的数据可以消除冗余数据,从而减少发送到汇聚节点或基站的数据量,降低网络能耗^[9].成员节点将频繁采集的原始信息数据发送至簇头进行融合使得簇头节点信息量不断增加,融合过程中易造成信息丢失.无线传感器网络的监测数据存在时空相关性,节点空间分布使得数据间产生空间相关性,另外,监测数据在时间上也存在相关性.大部分收集的原始数据掺杂着冗余数据,基于数据时空相关性的簇内数据融合技术能够减少无线传感器网络中数据包的传输量,对降低网络能耗起着重要作用^[10-11].

收稿日期:2022-12-14;**修回日期:**2023-01-25.

基金项目:国家自然科学基金(61373091);湖北省自然科学基金(2017CKB893).

作者简介:杨芬婷(1999-),女,内蒙古包头人,武汉轻工大学硕士研究生,主要从事无线传感器网络研究,E-mail:fent99@icloud.com.

通信作者:徐震(1974-),男,河南信阳人,武汉轻工大学副教授,博士,主要从事无线传感器网络以及光纤通信方面的研究,E-mail:xuzhen2046@qq.com.

引用本文:杨芬婷,徐震.基于环形缓冲区的 WSN 数据融合技术研究[J].河南师范大学学报(自然科学版),2024,52(2):62-71.(Yang Fenting, Xu Zhen. Research on WSN data fusion technology based on ring buffer[J]. Journal of Henan Normal University(Natural Science Edition), 2024, 52(2): 62-71. DOI: 10.16366/j.cnki.1000-2367.2022.12.14.0001.)

目前许多学者对无线传感器网络数据融合技术进行了研究.SUN 等^[12]提出基于信任度和改进遗传的多传感器数据融合算法,用三次指数平滑法对数据进行预处理,接着在模糊理论上用指数信任度函数对预处理后的数据进行融合,最后用改进的遗传算法提高算法收敛,优化融合的估计值.YUAN 等^[13]提出一种数据密度相关度的数据融合方法,该空间相关性模型与数据差值相关,以此来度量节点数据与其相邻节点数据间相关性,该算法适合节点部署密集的无线传感器网络.REYANA 等^[14]提出了用于多传感器数据融合算法,将自适应卡尔曼滤波器和决策树算法相结合进行火灾检测,并用模糊优化来提升监测系统决策能力.LI 等^[15]提出一种基于双阈值结合最优中继选择数据聚合方案.当节点的数据量阈值和能量阈值满足要求时才能执行路由,并选择数据包较多或等待时间较长的节点作为传输中继节点,最终完成数据聚合.

还有一些学者对簇内数据融合方案基于分层的角度进行了研究.AGARWAL 等^[16]提出一种基于缓冲区的无线传感器网络数据聚合线性过滤算法.源节点用余弦距离计算采集数据和缓冲区已有数据的关联度并剔除关联度高的数据,簇头节点剔除重复的数据来过滤冗余数据.ALSAFASFEH 等^[17]提出一种基于反向传播神经网络模型的数据融合算法.该算法采用改进的能量有效的阈值感知网络协议(threshold-sensitive energy efficient sensor network protocol, TEEN)进行节点聚类,将簇看作反向传播神经网络对感知数据进行融合,并在簇头提取感知数据的特征值传输到汇聚节点.XIA 等^[18]设计了基于数据融合的智能温室无线温度监测系统.源节点通过改进的无迹卡尔曼滤波器收集和预处理温室的温度数据.簇头用并行逆协方差相交融合算法进行局部融合.汇聚节点采用改进的极值学习算法进行全局融合.DASH 等^[19]提出一种利用传感器时空相关性的数据融合算法,将采集周期分为多个时隙,在源节点采用 Jaccard 相似函数消除冗余,在簇头节点用加权皮尔逊相关系数判断数据间的相关程度,并保留节点间相似度较高的数据.

1 所提的方案

1.1 簇的构建

无线传感器网络由汇聚节点、簇头节点和源节点三类节点构成.所有节点静止,且通信半径相同.汇聚节点在监测区域外独立设置且能量供应充足.其他节点初始能量相同,且能量有限无法补充.汇聚节点负责收集网络中所有采集的数据.簇头节点负责接收、融合簇内源节点发送的数据,并传输到汇聚节点.源节点负责采集数据并将采集的数据传输到簇头节点.

当节点完成部署后,根据簇头节点选择算法^[17],选择适合的节点当选簇头.担任簇头的节点向邻节点广播簇头信标.邻节点在接收到的簇头消息中选择向距离最近的簇头发送请求消息申请入簇,簇头接收到请求消息后回复确认帧进行确认.当所有节点完成簇的加入,簇头节点为簇内源节点分配标识 ID.网络拓扑如图 1 所示.

1.2 源节点数据处理

在数据驱动的无线传感器网络中,数据融合技术已成为消除冗余数据,减少数据传输量的重要技术之一.在已提出的基于缓冲区的线性滤波算法中,由于节点数据缓冲区较小,数据采集过程中会频繁触发数据的发送.为保证数据完整性和在时间上的前后关联性,本文使用环形缓冲区进行存储,基于环形缓冲区采用相似函数对冗余数据进行过滤,然后使用滑动四分位法进行异常检测,对因电流或电压不稳造成的无效异常进行检测剔除,以及对因天气骤变或突发灾害等有效异常进行上报.

1.2.1 源节点冗余数据过滤

源节点分配一段连续的内存空间,从而构建数组形式的环形缓冲区存储采集数据.环形缓冲区的存储空间在逻辑上首尾相连,在物理存储上为一段一维连续空间.源节点分配好内存空间后设置两个指针:读指针和写指针.初始构建缓冲区时,缓冲区为空,读写指针指向同一位置.当有采集数据存入时,写指针偏移.发送

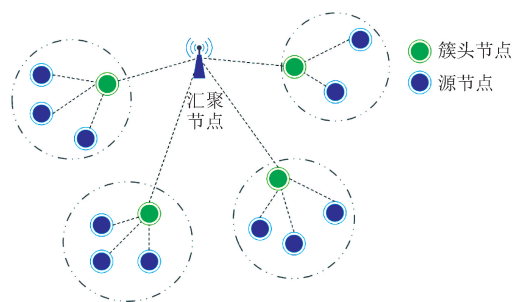


图1 网络拓扑
Fig.1 Network topology

数据时,源节点取出数据,写指针偏移相应数据的长度.以该种方式进行采集数据的存储和取出,既不用频繁分配线性缓冲区,还保证了时间关联性.

每个源节点维护一个环形缓冲区用来存储每个时隙采集的数据.一段时间内,环境信息变化缓慢,节点易存储大量冗余的数据,若将冗余数据全部发送,会产生许多不必要的节点能耗.为将冗余载荷转换为有效载荷进行传输,本文采用简单二值化相似函数进行数据处理.

将采集周期划分为 N 个相同时隙,源节点在一个周期内采集的数据序列为 $\{d_1, d_2, \dots, d_t, \dots, d_N\}$, t 时刻采集的数据为 d_t ,对 d_t 与前一时刻采集数据 d_{t-1} 进行冗余对比,如式(1)所示:

$$M(d_{t-1}, d_t) = \begin{cases} 1, & |d_{t-1} - d_t| \leq \delta, \\ 0, & |d_{t-1} - d_t| > \delta. \end{cases} \quad (1)$$

若小于给定的阈值,则数据相同或极度相似,可判断为冗余数据,不存入缓冲区.为保证数据时间前后关联性,这里仅对相邻时间的数据进行冗余判断.假设节点在当前第 m 时刻保存的数据权重 w_{d_m} 初始化为 1, $m \in [1, k], k \leq N$.若第 $m+1$ 时刻数据判断为冗余,则丢弃,调整第 m 时刻数据权重为 $w_{d_m} + 1$.丢弃的冗余数据越多,表明该数据在该组数据向量中所占的比重越大,该数据权重越大.

经冗余数据过滤后,每个源节点采集的数据会形成一个具有权值的数据集合,即节点 i 的数据集合为 $d_i = \{(d_{i1}, w_{i1}), (d_{i2}, w_{i2}), \dots, (d_{ik}, w_{ik})\}$.权值 $w_{i1}, w_{i2}, \dots, w_{ik}$ 表示当前数据的最终占比.

$$w_{im} = \frac{w_{d_m}}{w_{d_1} + w_{d_2} + \dots + w_{d_k}}, m \in [1, k], \quad (2)$$

其中, $w_{d_1}, w_{d_2}, \dots, w_{d_m}, \dots, w_{d_k}$ 表示存入缓冲区的数据权重.

1.2.2 基于环形缓冲区的滑动四分位检测

考虑到节点因外界噪声或电压电流不稳会产生尖端瞬时性异常,或因节点故障、监测区域产生重大环境变化等突发异常,本文基于环形缓冲区采用滑动窗口的四分位法进行异常值的检测.如图 2 所示.

在基于缓冲区的线性滤波算法^[19]中,节点采用固定帧长的缓冲区存储采集数据,该算法中缓冲区容量较小.本文选用环形缓冲区存储采集数据.环形缓冲区不用频繁分配内存,其低内存利用率非常适合内存有限的节点存储数据^[20],相比固定帧长的缓冲区更适合传感器节点存储采集到的数据.

滑动窗口策略常使用于时间相关性强的数据.随着时间进行,历史数据的参考意义小于最近时间数据的参考意义^[21],因此,当使用滑动窗口进行检测时,如果采集数据充满滑动窗,便进行一次检测.

基于滑动窗口的异常检测采用滑动四分位异常检测机制实现.选择四分位检测法进行异常值判别能够减小节点计算复杂度.滑动四分位法异常检测过程如下:

1)假设进入滑动窗内的数据序列为 $d_1, d_2, \dots, d_p, \dots, d_q$,对序列进行从小到大的排序后,选取当前数据序列的 $1/4$ 处序列值(Q_1), $3/4$ 处序列值(Q_3).则四分位 $IQR = Q_3 - Q_1$.

2)计算当前序列对应的上下限.考虑到不同时间段窗口内的数据变化会产生不同程度的波动,所以需构建窗口的动态阈值.设定宽容度常数为 α_1 和 α_2 ($\alpha_2 = 1/\alpha_1$),其大小可以动态调整.则上下限计算式为:

$$\begin{cases} \epsilon_a = Q_1 - \alpha_1 IQR, \\ \epsilon_b = Q_3 - \alpha_2 IQR. \end{cases} \quad (3)$$

3)异常值判定:若 d_p 在区间 $[\epsilon_a, \epsilon_b]$ 内,则为非异常值.若超出该区间,则为异常值.

4)源节点设置异常计数标志记录检测出的异常数据个数.当异常值个数累计大于缓冲区长度 l 的二分之一,源节点唤醒簇头节点,并保留该部分连续异常值,将当前缓冲区数据全部发送给簇头节点.

若异常值个数并未超过缓冲区长度的二分之一,则滑动窗口后移,等待下一个时间段所采集的数据填满缓冲区再进行异常检测.环形缓冲区滑动窗口上的数据检测完成后,源节点采用二值化判断是否进行该部分数据的发送.判断公式如下:

$$S = \begin{cases} 1, & f_{abn} \geq 0.5 \cdot l, \\ 0, & f_{abn} < 0.5 \cdot l, \end{cases} \quad (4)$$

其中, S 为强制发送标志, f_{abn} 表示异常计数标志, l 为节点缓冲区长度.正常情况下,节点不会强制发送.一

个周期即将结束时,唤醒簇头节点等待接收源节点异常检测和冗余数据处理后的数据.周期时间到达时,源节点剔除检测出的瞬时性异常数据,并发送给簇头节点,其过程如图 3 所示.

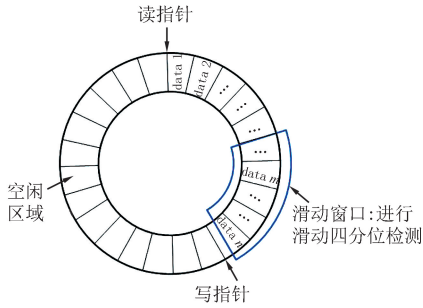


图2 基于环形缓冲区的滑动窗口

Fig.2 A sliding window based on a ring buffer

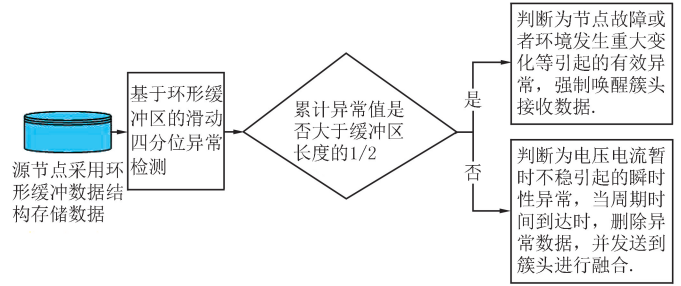


图3 源节点异常检测机制

Fig.3 Source node anomaly detection mechanism

1.3 基于支持度函数的簇头节点融合

1.3.1 支持度函数

支持度函数^[22]的提出是在数据融合的过程中将数值间的关联信息加入,进一步优化数据融合方案. $Sup(a, b)$ 表示 b 对 a 的支持程度,即数值 a 和 b 的接近程度. 支持度函数需满足以下 3 个性质:(1) $Sup(a, b) \in [0, 1]$; (2) $Sup(a, b) = Sup(b, a)$; (3) $Sup(a, b) \geq Sup(x, y)$, 若 $|a - b| < |x - y|$.

目前,常用的支持度函数为高斯支持度函数^[23],其函数形式为:

$$Sup(a, b) = G(a, b, K, \beta) = K \times e^{-\beta \cdot (a-b)^2}, \quad (5)$$

其中, K 表示幅度, $K \in [0, 1]$; β 表示函数衰减因子, β 越大,数值 a 和 b 的支持度越小.当数值 a 和 b 相同时,支持度为 1.

指数形式的高斯支持度函数计算较为复杂,为降低计算复杂度,刘思峰等^[24]基于灰色接近关联度来描述两数值接近程度,提出一种无需指数运算的新型支持度函数.其函数表达式为:

$$Sup(a, b) = SF(a, b, K, \beta) = K(1 + \beta(a - b)^2)^{-1}. \quad (6)$$

考虑到序列的时间前后关联,匡亮等^[25]提出了基于优化动态弯曲距离的支持度函数 IDTW-SF.其表达式为:

$$Sup(U(T), V(T)) = IDTW-SF(U(T), V(T), K, \beta) = K(1 + \beta \times \text{dist}(U(T), V(T)))^{-1}. \quad (7)$$

1.3.2 基于加权皮尔逊距离的改进支持度函数

基于优化动态弯曲距离的支持度函数虽然考虑了采集信息的时间关联性,但通过寻找两序列在时间轴上的对齐方式来计算最短距离,需要花费较高的计算成本.

加权皮尔逊距离在强调时间序列变化趋势相似程度的同时,通过调整不同维度的权重来控制每个维度的贡献.加权皮尔逊距离具有平移不变性.这种平移不变的特性对时间序列到达同一簇头的前后延迟具有包容性.本文采用加权皮尔逊距离^[26]计算支持度函数中两数据序列的距离,并在簇头采用加权皮尔逊距离的改进支持度函数(weighted Pearson distance improvement support function, wPd-SF)进行加权融合.

由于成员节点发送数据之前会过滤冗余数据并剔除异常数据,所以,到达簇头的数据时间序列长度可能会不同.在计算两序列间加权皮尔逊距离时,需要将两序列拓展到相同的长度.簇头节点用源节点 i 生成的带权值的数据集合 $d_i = \{(d_{i1}, \omega_{i1}), (d_{i2}, \omega_{i2}), \dots, (d_{ik}, \omega_{ik})\}$, 计算加权平均数 x_i .并用 x_i 来补齐序列. x_i 计算如下: $x_i = (d_{i1}\omega_{i1} + d_{i2}\omega_{i2} + \dots + d_{ik}\omega_{ik})/k$.

假设同一簇内的两个不同节点在相同周期时间发送至簇头节点的数据序列为 $X_i = (x_{i1}, x_{i2}, \dots, x_{in})$ 和 $X_j = (x_{j1}, x_{j2}, \dots, x_{jn})$.则他们的加权皮尔逊相关系数 $\rho(X_i, X_j)$ 和加权皮尔逊距离 $D_w(X_i, X_j)$ 分别

$$\text{为: } m(X_i) = \frac{\sum_{k=1}^n \omega_k x_{ik}}{\sum_{k=1}^n \omega_k}, \text{cov}(X_i, X_j, W) = \frac{\sum_{k=1}^n \omega_k (x_{ik} - m(X_i))(y_{jk} - m(X_j))}{\sum_{k=1}^n \omega_k}, \rho(X_i, X_j) =$$

$$\frac{\text{cov}(X_i, X_j, \mathbf{W})}{\sqrt{\text{cov}(X_i, X_i, \mathbf{W}) \text{cov}(X_j, X_j, \mathbf{W})}}, D_w(X_i, X_j) = 1 - \rho(X_i, X_j),$$

其中, \mathbf{W} 表示权重矩阵. 通过权重来体现各维度上采集数值对所在时间序列的重要程度. 维度上的权重越高, 表示采集数据时与该维度上的数值相同或相似的值越多, 可以将其看作当前时间序列上的重要采集点. 若两时间序列的重要采集点出现在同一时间维度, 则这两个时间序列变化趋势相似. 在消除冗余数值的同时, 通过时间序列上的重要采集点以增强对两时间序列变化趋势的描述. 利用加权皮尔逊距离来计算两时间序列距离, 更准确把握时间序列间相关程度.

基于加权皮尔逊距离的改进支持度函数定义为: $\text{Sup}(X_i, X_j) = \text{wPd} - \text{SF}(X_i, X_j, K, \beta) = K \times (1 + \beta \times D_w(X_i, X_j))^2)^{-1}$, 即簇内节点 i 和 j 间的支持度为 $S_{ij} = \text{Sup}(X_i, X_j)$. 节点相互支持度矩阵可定义如下:

$$\begin{bmatrix} S_{11} & S_{12} & \cdots & S_{1j} & \cdots & S_{1n} \\ S_{21} & S_{22} & \cdots & S_{2j} & \cdots & S_{2n} \\ \vdots & & & & & \vdots \\ S_{i1} & S_{i2} & \cdots & S_{ij} & \cdots & S_{in} \\ \vdots & & & & & \vdots \\ S_{n1} & S_{n2} & \cdots & S_{nj} & \cdots & S_{nm} \end{bmatrix},$$

其中, n 为簇内传感器节点数. 簇内传感器节点对传感器节点的支持度之和为: $\text{sum}(\text{wPd} - \text{SF}(X_i, X_j)) = \sum_{j=1}^n S_{ij}$.

传感器节点的融合权值 w_i 为: $w_i = \text{sum}(\text{wPd} - \text{SF}(X_i, X_j))$.

在簇内节点均无故障的情况下, 簇头节点将簇内成员节点发送的采集信息经加权融合成一组最优融合值, 其表达式如下:

$$X(t) = \sum_{i=1}^n (w_i \times X_i(t)) / \sum_{i=1}^n w_i,$$

其中, t 指周期内第 t 时刻, $X_i(t)$ 指 t 时刻采集的数据.

若簇内存在故障节点, 簇头节点丢弃故障节点数据并用融合的估计值代替故障节点采集的数据值. 故障节点的融合估计值计算如下: $X'_i(T) = \sum_{j=1}^n w_j X_j(T) (j \neq i)$, $w_j = \frac{S_{ij}}{\sum_{j=1}^n S_{ij}} (j \neq i)$, 其中, T 为周期长度.

最后, 如图 4 所示, 簇头节点对簇内节点数据进行最终融合, 将多个传感器节点数据融合成一组数据并传到汇聚节点, 从而减少数据传输量, 降低节点的能耗.

2 仿真和结果分析

为评估本文所提方案, 对其进行仿真测试. 本文实验能耗参照一阶能耗模型公式^[27]. 实验环境基于 Windows 10(64 bit), 运行内存 16 GB, 处理器为 Intel(R)Core(TM)i5-7300HQ CPU @ 2.50 GHz 2.50 GHz. 仿真和测试实验在 MATLAB R2021a 中进行. 实验数据集采用英特尔伯克利研究实验室的公开数据集^[28], 其中包含了不同传感器信息, 包括温度、湿度、光照和电压. 该真实数据集还包含了节点 ID 以及时间等信息. 在本文中, 选取环境监测最常见的温度属性进行实验, 其传感器节点分布图如图 5 所示.

仿真参数如表 1 所示.

2.1 源节点数据异常检测

本文所提方案采用滑动四分位法对温度数据集进行异常检测, 滑动四分位法无需用到早期的历史数据, 适用于存储资源有限的传感器节点. 传感器节点采集过程中受到环境和噪声等干扰会产生两种类型的异常数据: 一种是瞬时性的尖端异常, 另外一种连续多个偏离正常值. 这里选用同一传感器节点连续采集的 1 500 个温度数据, 其中分别设置 60 个加入噪声干扰的瞬时性和连续性异常数据. 这里, 衡量异常值检测方法优劣的指标选用 F_1 分数^[29]. 其表达式如下所示:

$$F_1 = \frac{2}{\frac{1}{AR} + \frac{1}{CR}} = 2 \times \frac{AR \times CR}{AR + CR},$$

式中,准确率 AR 指检测出的异常值中,真实异常值个数占所有检测出的异常值个数的比值.覆盖率 CR 指检测出的异常值中,真实异常值个数占总的真实异常值个数的比值.对温度数据集,测试不同滑动窗口宽度和宽容度常数组合情况下的异常检测 F_1 值. F_1 分数值越大表示检测效果越好.如图 6 和图 7 所示,实验表明,在合适的滑动窗宽和宽容度常数下,两种不同异常值检测的最优 F_1 值可以达到 91.47% 和 96.77%.

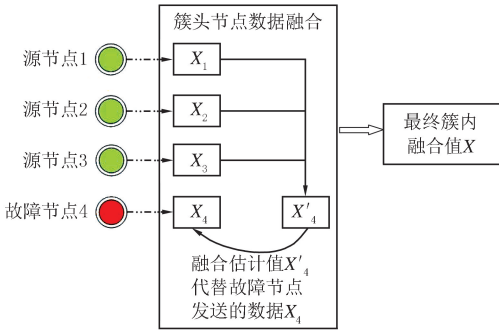


图4 簇头节点数据融合

Fig. 4 Data fusion of cluster head nodes

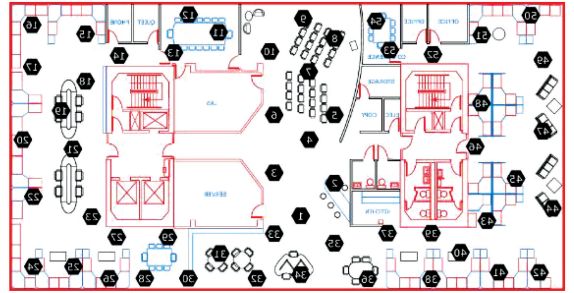


图5 英特尔伯克利实验室传感器节点分布

Fig. 5 Distribution of sensor nodes at Intel Berkeley Lab

表 1 仿真参数

Tab. 1 Simulation parameter

仿真参数	仿真区域	节点个数	$E_{tx}/$ ($nJ \cdot bit^{-1}$)	$E_{rx}/$ ($nJ \cdot bit^{-1}$)	E_{fs} (free space)/ ($pJ \cdot bit^{-1} \cdot m^{-2}$)	E_{amp} (Multipath fading)/ ($pJ \cdot bit^{-1} \cdot m^{-4}$)	节点初能 量 E_n/J	轮数	感知范 围/m
参数值	100 m×100 m	50	50	50	10	0.001 3	0.5	8 000	15

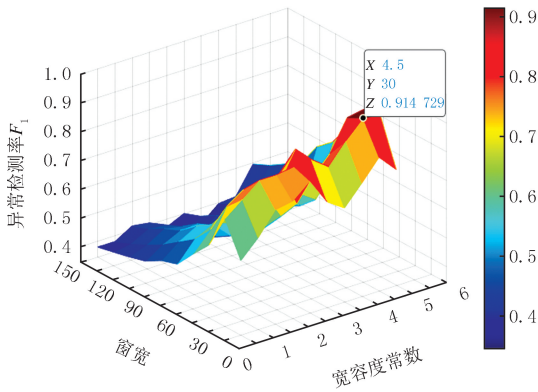


图6 瞬时性异常数据的检测率

Fig. 6 Detection rate of transient anomaly data

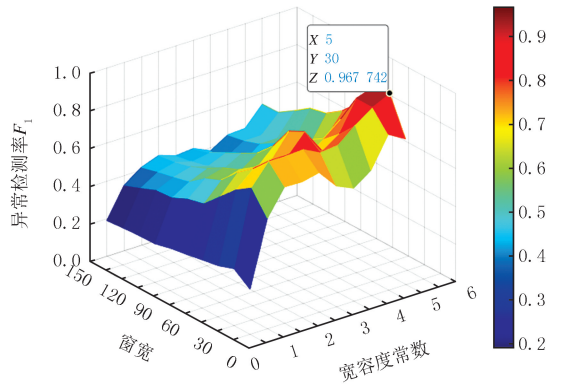


图7 连续性异常数据的检测率

Fig. 7 Detection rate of continuous anomalous data

2.2 簇头节点数据融合测试

为测试本文提出的加权皮尔逊距离改进支持度函数(wPd-SF)的融合效果,对融合温度传感器节点组采用方差来评估融合结果,对故障节点的融合估计值采用平均绝对误差(mean absolute error, MAE)进行评估.这里,选取位置邻近的 5 个传感器节点在同一天内每 5 min 采集 1 次温度,每小时进行 1 次温度数据融合测试.所选取的传感器组温度监测值如图 8 所示.

在节点无故障的情况下,将本文提出 wPd-SF 与新型支持度函数 D-SF^[30]、改进型支持度函数 SN-SF^[31]、动态弯曲距离支持度函数 DTW-SF^[22] 和优化动态弯曲距离支持度函数 IDTW-SF^[25] 进行融合结果方差对比.其结果如图 9 所示.

由图 9 可以看出,本文提出的支持度函数 wPd-SF 对簇内传感器节点的融合方差小于其他支持度函数.

在计算故障节点的融合估计值过程中,需要传感器节点采集时间序列间的相互支持度矩阵来进行估计值计算.对于 D-SF 以及 SN-SF 而言,它们在每个时刻都要计算出一次当前时刻的支持度矩阵,无法顾及时

间前后的关联性,在融合阶段不便计算故障节点估计值.所以,该部分实验选择能够计算两序列间距离 DTW-SF 和 IDTW-SF 进行 MAE 值对比,结果如图 10 所示.

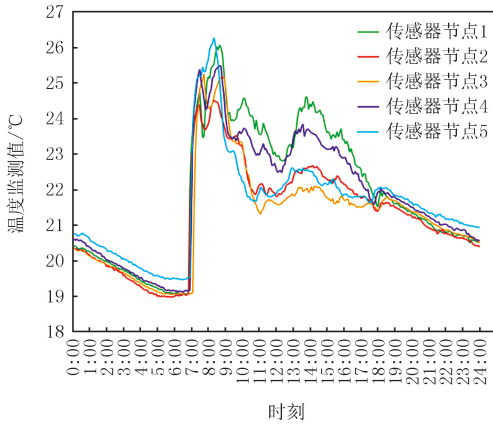


图8 温度传感器节点监测值

Fig.8 Temperature sensor node monitoring values

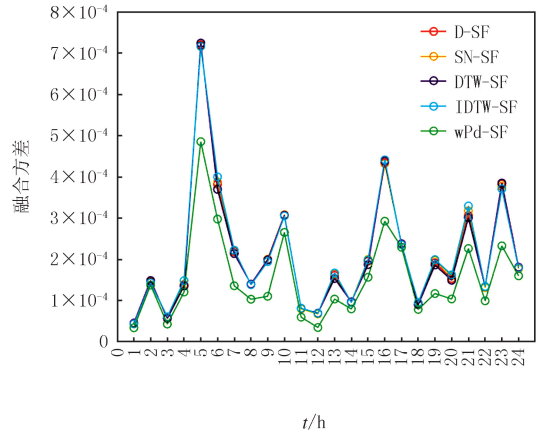


图9 融合方差对比

Fig.9 Fusion variance comparison

由图 10 可以看出,对于故障节点融合估计值的计算,DTW-SF 在环境信息较平稳时 MAE 值较低,但仍高于 IDTW-SF 和本文所提出的 wPd-SF;在环境信息变化波动较大时,DTW-SF 的估计值出现较大误差,稳定性远不如 IDTW-SF 和 wPd-SF.使用 wPd-SF 对故障节点的融合估计值与该故障节点正常工作监测值的 MAE 值小于其他 2 个支持度函数.

表 2 和表 3 给出了各支持度函数计算时间的对比.在无故障节点情况下,本文提出的 wPd-SF 的计算时间明显少于 D-SF、SN-SF 和 DTW-SF.wPd-SF 的计算时间和 IDTW-SF 的计算时间接近,但仍低于 IDTW-SF 的计算时间.在计算故障节点的融合估计值时,wPd-SF 的计算时间明显低于 DTW-SF 和 IDTW-SF.

表 2 各支持度函数融合整组采集数据所用时间

Tab. 2 The time it takes for each support function to collect data from the whole set of data

支持度函数	D-SF	SN-SF	DTW-SF	IDTW-SF	wPd-SF
时间/s	0.033 6	0.070 6	0.126 4	0.017 1	0.012 9

2.3 整体网络性能评估

为更准确地评估本文所提出的数据融合算法的性能,本文方案从网络剩余节点、网络剩余能量、网络数据包发送量 3 个方面,与分层传输缩减 ETDTR 算法^[32]、基于数据时空相关性的数据聚合 STCDRR 方案^[19]和基于缓冲区的数据聚合线性滤波 BFL 算法^[16]进行对比.

2.3.1 网络剩余节点数分析

网络剩余节点数量多少决定了传感器网络的寿命.如图 11 所示,在前 1 500 轮之前,4 种算法都没有节点死亡,在 2 000 轮时,BFL 算法仍未有死亡节点.但随着轮次的继续进行,BFL 算法死亡节点数增加变快.这是由于 BFL 算法中节点缓冲区较小,当缓冲区满的时候易触发被替换数据的发送,所以节点能耗变快.随着轮次增加可以看出,本文所提出方案的网络剩余节点数多于其他 3 个算法.

2.3.2 网络剩余能量分析

如图 12 所示,该图为 4 种算法的网络剩余能量.随着轮次进行至 8 000 轮,本文所提方案和 ETDTR 算法、BFL 算法和 STCDRR 算法的网络剩余能量均低于网络初始总能量的 50%.ETDTR 算法、BFL 算法和 STCDRR 算法的网络剩余能量仅剩 16.12%、22.95% 和 7.72%.而本文所提方案的网络剩余能量占网络初始总能量的 39.33%.仿真实验表明,本文所提算法在网络剩余能量方面优于其他 3 个算法.

表 3 各支持度函数计算故障节点融合估计值所用时间

Tab. 3 The time taken by each support function to calculate the fusion estimate of fault nodes

度量	DTW-SF	IDTW-SF	wPd-SF
时间/s	0.073 8	0.012 4	0.004 3

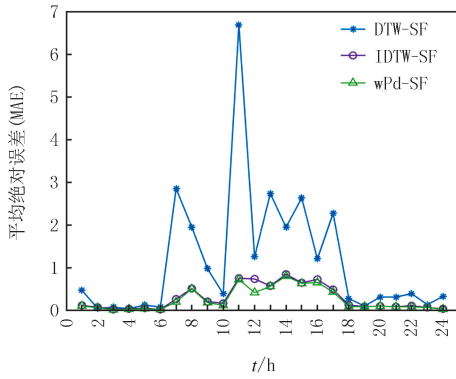


图10 故障节点融合估计值的MAE值

Fig.10 MAE of the failed node fusion estimate

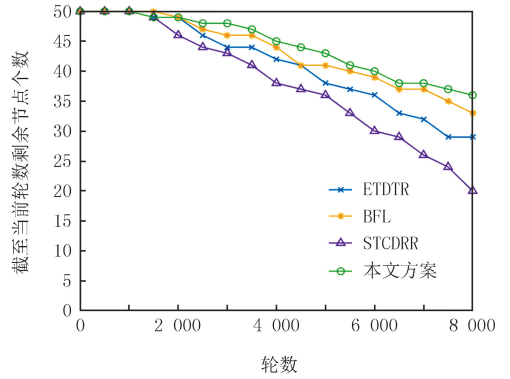


图11 网络剩余节点个数

Fig.11 Number of remaining nodes in the network

2.3.3 网络发送数据包数分析

随着无线传感器网络生存周期的延长,网络发送数据包数也会逐渐累加.图 13 结果显示每个算法的网络发送数据包数情况.相对 ETDTR 算法、BFL 算法及 STCDRR 算法而言,本文所提方案分别大约减少了 26.19%、14.58% 和 33.13% 的数据包发送量.本文所提方案传输的数据包数少于其他 3 个算法.

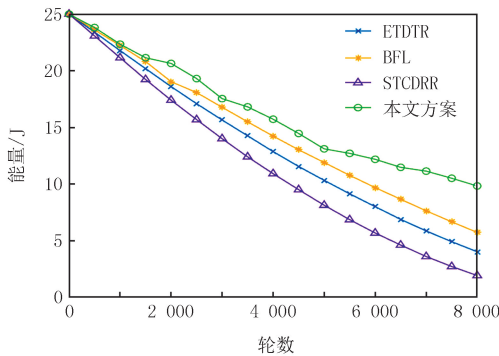


图12 网络剩余能量

Fig.12 Residual energy of the network

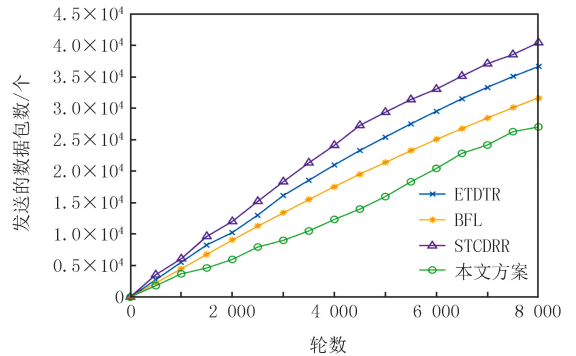


图13 网络发送数据包数

Fig.13 The number of packets sent by the network

3 总 结

本文提出一种基于环形缓冲区的无线传感器网络簇内数据融合算法.源节点采用环形缓冲区存储采集数据.源节点在数据存入缓冲区前采用相似函数判断数据冗余,并丢弃冗余的数据.且源节点基于缓冲区的滑动窗口采用滑动四分位法进行异常数据的检测.若异常值个数累计超过预设范围,则强制唤醒簇头去接收数据.若没有超预设范围,则在周期时间到达后,剔除异常值并发送到簇头进行融合.簇头使用基于加权皮尔逊距离的改进支持度函数对簇内采集的数据进行加权数据融合.在没有故障节点的情况下,簇头对整组传感器节点发来的数据进行融合.若存在故障节点,为保证数据融合的准确性,通过融合其他正常节点数据对故障节点进行估计值计算,并用估计值代替故障节点数据.

仿真结果表明,本文所提方案融合误差小于其他支持度函数融合误差,计算时间小于其他支持度函数.且对不同类型异常数据异常检测率均在 91% 以上.整体性能评估上,本文所提出的算法在网络剩余节点个数、网络剩余能量和网络发送数据包数等 3 个方面的评估均优于 ETDTR 算法、BFL 算法和 STCDRR 算法.

参 考 文 献

[1] WAN R Z, XIONG N X, HU Q H, et al. Similarity-aware data aggregation using fuzzy c-means approach for wireless sensor networks[J]. EURASIP Journal on Wireless Communications and Networking, 2019, 2019(1): 59.

[2] YANG C Y, LIN C Y, GALSANBADAM S, et al. Multivariable support vector regression with multi-sensor network data fusion[C]// 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC). [s.l.]: IEEE, 2019: 4029-4034.

- [3] DANANJAYAN S, ZHUANG J J, TANG Y, et al. Wireless sensor deployment scheme for cost-effective smart farming using the ABC-TEEM algorithm[J]. *Evolving Systems*, 2023, 14(4): 567-579.
- [4] ZHANG J, LIN Z W, TSAI P W, et al. Entropy-driven data aggregation method for energy-efficient wireless sensor networks[J]. *Information Fusion*, 2020, 56: 103-113.
- [5] IZADI D, ABAWAJY J H, GHANAVATI S, et al. A data fusion method in wireless sensor networks[J]. *Sensors*, 2015, 15(2): 2964-2979.
- [6] ULLAH I, YOUN H Y. A novel data aggregation scheme based on self-organized map for WSN[J]. *The Journal of Supercomputing*, 2019, 75(7): 3975-3996.
- [7] SOLTANI M, HEMPEL M, SHARIF H. Data fusion utilization for optimizing large-scale Wireless Sensor Networks[C]//2014 IEEE International Conference on Communications(ICC). [s.l.]: IEEE, 2014: 367-372.
- [8] SONG Y, LIU Z, HE X, et al. Research on data fusion scheme for wireless sensor networks with combined improved LEACH and compressed sensing[J]. *Sensors*, 2019, 19(21): 4704-4733.
- [9] XIAO X X, HUANG H N, WANG W. Underwater wireless sensor networks: an energy-efficient clustering routing protocol based on data fusion and genetic algorithms[J]. *Applied Sciences*, 2020, 11(1): 312.
- [10] CAO L, CAI Y, YUE Y G. Data fusion algorithm for heterogeneous wireless sensor networks based on extreme learning machine optimized by particle swarm optimization[J]. *Journal of Sensors*, 2020, 2020: 1-17.
- [11] JARWAN A, SABBAH A, IBNKAHLA M. Data transmission reduction schemes in WSNs for efficient IoT systems[J]. *IEEE Journal on Selected Areas in Communications*, 2019, 37(6): 1307-1324.
- [12] SUN G L, ZHANG Z Y, ZHENG B W, et al. Multi-sensor data fusion algorithm based on trust degree and improved genetics[J]. *Sensors*, 2019, 19(9): 2139.
- [13] YUAN F, ZHAN Y J, WANG Y H. Data density correlation degree clustering method for data aggregation in WSN[J]. *IEEE Sensors Journal*, 2014, 14(4): 1089-1098.
- [14] REYANA A, VIJAYALAKSHMI P. Multisensor data fusion technique for energy conservation in the wireless sensor network application "condition-based environment monitoring"[J]. *Journal of Ambient Intelligence and Humanized Computing*, 2021, 13: 1-10.
- [15] LI A, LIU W, ZENG L J, et al. An efficient data aggregation scheme based on differentiated threshold configuring joint optimal relay selection in WSNs[J]. *IEEE Access*, 2021, 9: 19254-19269.
- [16] AGARWAL A, JAIN K, DEV A. BFL: a buffer based linear filtration method for data aggregation in wireless sensor networks[J]. *International Journal of Information Technology*, 2022, 14(3): 1445-1454.
- [17] ALSAFASFEH M, ARIDA Z A, SARAEREH O A, et al. An optimized data fusion paradigm for WSN based on neural networks[J]. *Computers, Materials & Continua*, 2021, 69(1): 1097-1108.
- [18] XIA S B, NAN X Y, CAI X, et al. Data fusion based wireless temperature monitoring system applied to intelligent greenhouse[J]. *Computers and Electronics in Agriculture*, 2022, 192: 106576.
- [19] DASH L, PATTANAYAK B K, MISHRA S K, et al. A data aggregation approach exploiting spatial and temporal correlation among sensor data in wireless sensor networks[J]. *Electronics*, 2022, 11(7): 989.
- [20] FELDMAN S, DECHEV D. A wait-free multi-producer multi-consumer ring buffer[J]. *ACM SIGAPP Applied Computing Review*, 2015, 15(3): 59-71.
- [21] YIN C Y, ZHANG S, WANG J, et al. Anomaly detection based on convolutional recurrent autoencoder for IoT time series[J]. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2022, 52(1): 112-122.
- [22] SHI P, LI G H, YUAN Y M, et al. Data fusion using improved support degree function in aquaculture wireless sensor networks[J]. *Sensors*, 2018, 18(11): 3851.
- [23] YAGER R R. The power average operator[J]. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 2001, 31(6): 724-731.
- [24] 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 S F, XIE N M, 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.
- [25] 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.
KUANG L, SHI P, JI Y F, 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*, 2020, 36(16): 192-200.
- [26] LIN R H, WU B D, SU Y. An adaptive weighted Pearson similarity measurement method for load curve clustering[J]. *Energies*, 2018, 11(9): 2466.
- [27] HEINZELMAN W B, CHANDRAKASAN A P, BALAKRISHNAN H. An application-specific protocol architecture for wireless microsensor networks[J]. *IEEE Transactions on Wireless Communications*, 2002, 1(4): 660-670.
- [28] MADDEN S. Intel Berkeley Research Lab[EB/OL]. [2022-12-09]. <http://db.csail.mit.edu/labdata/labdata.html>.

- [29] 李艺,华静,刘保双,等.大气污染物监测数据异常值判别方法研究[J].环境科学学报,2022,42(12):341-352.
LI Y,HUA J,LIU B S,et al.Study on the outlier identification approaches for atmospheric pollutant monitoring data[J].Acta Scientiae Circumstantiae,2022,42(12):341-352.
- [30] 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 Y J,SHEN M X,LU M Z,et al.Algorithm of real time data fusion for greenhouse WSN system[J].Transactions of the Chinese Society of Agricultural Engineering,2012,28(23):160-166.
- [31] 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 Q L,XIAO X Y,LIU Y R,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,2017,33(S1):239-245.
- [32] IDREES A K,ALHUSSAINI R,ABED SALMAN M.Energy-efficient two-layer data transmission reduction protocol in periodic sensor networks of IoTs[J].Personal and Ubiquitous Computing,2023,27(2):139-158.

Research on WSN data fusion technology based on ring buffer

Yang Fenting, Xu Zhen

(School of Electrical and Electronic Engineering, Wuhan Polytechnic University, Wuhan 430048, China)

Abstract: In the wireless sensor network monitoring system, redundant data caused by slow environmental changing and overlapping sensing range of nodes will increase the amount of data sent by sensor nodes, reduce the efficiency of information collection, and lead to premature death of sensor nodes. Therefore, this paper proposes an intracluster data fusion scheme based on ring buffer. All sensor nodes use ring buffers to store data. Based on the ring buffer, the source node adopts the binarized similarity function and sliding quartile detection method to eliminate redundant data and transient abnormal data while ensuring data time correlation. Based on the improved support of weighted Pearson distance, the cluster head node carries out weighted fusion of the data received from the source node. Simulation experiments show that the proposed scheme has obvious advantages in the number of remaining nodes in the network, the remaining energy of the network and the number of packets sent by the network.

Keywords: wireless sensor network; data fusion; support function; ring buffer; sliding window

[责任编辑 陈留院 赵晓华]