Research on Multi-Source Data Feature Fusion Methods for Wireless Sensor Networks*

CHEN Hong[1*], JIANG Wenxian[2], HUANG Liping[1], YU Chongchong[1]
(1.Department of IT Development and Management, Huaqiao University, Xiamen Fujian 361021, China;
2.College of Computer Science and Technology, Huaqiao University, Xiamen Fujian 361021, China)

Abstract: In wireless sensor networks, different sensor nodes may collect overlapping or similar data, resulting in waste of computing, storage, and transmission resources. To this end, a multi-source data feature fusion method for wireless sensor networks is proposed. By combining complementary set empirical mode decomposition and wavelet threshold denoising methods, noise is removed while retaining the main features of the data. The first and second layer features of multi-source data are extracted by using principal component analysis, which are cascaded into the final extracted multi-source data features. The maximum and minimum closeness degree in fuzzy mathematics is used to describe the distance between different features, achieving multi-source data feature fusion. The simulation results show that the node mortality rate after the application of the proposed method is less than 5%, the fusion delay is less than 4 ms, the average node energy consumption remains below 4.5 J, and the feature fusion accuracy in different scenarios is higher than 86.3%, indicating that the proposed method has good multi-source data feature fusion performance and can effectively improve the data transmission performance of wireless sensor networks.

Key words: wireless sensor network; multi source data features; feature fusion; empirical mode decomposition; wavelet threshold denoising; principal component analysis; affinity
EEACC: 7230 **doi**: 10.3969/j.issn.1004-1699.2024.12.017

无线传感网络多源数据特征融合方法研究*

陈 宏[1*], 蒋文贤[2], 黄丽萍[1], 余翀翀[1]
(1.华侨大学信息化建设与管理处, 福建 厦门 361021; 2.华侨大学计算机科学与技术学院, 福建 厦门 361021)

摘 要: 在无线传感网络中, 不同的传感器节点可能会收集到重叠或相似的数据, 造成计算、存储和传输资源的浪费。为此, 提出一种无线传感网络多源数据特征融合方法。结合互补集合经验模态分解和小波阈值去噪方法, 在保留数据主要特征的同时去除噪声。通过主成分分析提取多源数据的第一层和第二层特征, 并将其级联为最终提取的多源数据特征。采用模糊数学中的最大最小贴近度描述不同特征之间的距离, 实现多源数据特征融合。仿真结果表明, 所提方法应用后的节点死亡率低于 5%, 融合延迟小于 4 ms, 平均节点能耗保持在 4.5 J 以下, 不同场景下的特征融合精度高于 86.3%。表明所提方法具有良好的多源数据特征融合性能, 可以有效地提高无线传感网络的数据传输性能。

关键词: 无线传感网络; 多源数据特征; 特征融合; 经验模态分解; 小波阈值去噪; 主成分分析; 亲信度
中图分类号: TP393 **文献标识码**: A **文章编号**: 1004-1699(2024)12-2131-06

随着嵌入式系统以及无线通信技术的不断进步和发展, 无线传感网络逐渐成为一个十分重要的研究领域。在实际运行过程中, 主要通过终端传感节点完成数据采集, 采用逐跳通信的方式将全部数据传输至基站节点。无线传感网络在安全监控以及环境监测等多个不同研究领域内具有十分广泛的应用。但是在全部应用中都面临同样一个难题—传感节点资源受限, 数据融合[1-2]是解决上述问题的重要手段。国内相关专家最近几年也针对传感器数据融合方面的内容展开了大量研究。

Wootton 等[3]在回声状态网络成功处理漏磁数据的基础上, 通过增加回声状态网络展开数据融合。张辉等[4]通过四圆定位法随机选择 2 个检测目标信息的节点, 计算 2 个圆形检测区域边界的交点, 同时

项目来源: 福建省自然科学基金项目(2017J01776)
收稿日期: 2023-08-14 **修改日期**: 2023-09-06

引入卡尔曼滤波增益矩阵实现数据融合处理。莫慧凌等[5]通过张量 Tucker 分解理论组建高阶张量,采用高阶张量提取异构数据的高维特征,最终实现联邦学习中的数据融合。以上几种数据融合方法在使用过程中存在融合延迟较长以及节点死亡率比较高等问题。为此,提出一种无线传感网络多源数据特征融合方法。仿真结果表明,所提方法可以获取比较满意的无线传感网络多源数据特征融合结果。

1 多源数据特征融合设计

1.1 无线传感网络多源数据去噪

在无线传感网络中,由于环境干扰、信号衰减等因素的影响,传感器节点采集到的多源数据可能会受到噪声的干扰,导致数据不准确或含有冗余信息。因此,进行多源数据去噪可以消除噪声,减少数据的不确定性,并提高数据的质量。为此,采用小波阈值去噪[6-7]抑制大部分噪声,但也会误删掉部分幅度较小的有效数据。另一方面,如果仅通过互补集合经验模态分解[8](CEEMD)展开去噪分解,虽然可以有效消除高频噪声,但高频范围内的有效信息也可能会被抑制。因此,将 CEEMD 和小波阈值去噪相结合[9],对多源数据展开去噪处理,具体的操作步骤为:

①通过式(1)对原始多源数据展开 CEEMD 分解处理,获取各个 IMF 分量 a_{ij}:

$$x(t)=\sum_{i=1}^{n}\mathrm{IMF}_i(t)+R(t) \tag{1}$$

式中:n 代表原始多源数据集;$\mathrm{IMF}_i(t)$ 代表第 i 个 IMF 分量;$R(t)$ 代表剩余项。

②将全部 IMF 分量按照从小到大的顺序排列,并且选择取值比较大的 IMF 分量,通过确定一个合适的阈值,对原始多源数据进行小波阈值[10-11]去噪处理:

$$\beta=x(t)\,\mathrm{sgn}(e-\tilde{\alpha}) \tag{2}$$

式中:sgn( )代表符号函数;e 代表辅助噪声;$\tilde{\alpha}$ 代表阈值函数。

③将小波去噪后的 IMF 分量以及其他 IMF 分量重构,进而获取去噪后的多源数据 $c(t)$,如式(3)所示:

$$c(t)=nx(t)\beta \tag{3}$$

由此,完成多源数据的去噪处理,这种综合去噪处理方法能有效地去除高频噪声,同时保留幅度较小但有效的信息,从而提高数据的质量和准确性。

1.2 多源数据特征提取

无线传感网络多源数据存在着相互关联和影响。去噪后的多源数据 $c(t)$ 之间因去噪导致关联性降低,无法准确地理解和推断无线传感网络中各个节点之间的关系。因此,为了更好地进行多源数据特征融合,进行多源数据特征提取。

将去噪后的多源数据 $c(t)$ 按照列划分为样本 z,设定每个样本的规格为 $m\times l$。通过式(4)对全部多源数据集展开中心化处理,也就是去均值操作:

$$\tau=\sum_{m=1}^{z}\sum_{l=1}^{z}c(t)_m c(t)_l \tag{4}$$

式中:τ 代表训练样本均值。

将每个多源数据样本 z 减去输入样本的均值后,通过中心化处理得到协方差矩阵 $\boldsymbol{I}$:

$$\boldsymbol{I}=\frac{\tau}{z-1} \tag{5}$$

在协方差矩阵 $\boldsymbol{I}$ 的第一层,用 ZCA 白化矩阵表示为 $\boldsymbol{B}_{(w1)}$,求解公式为:

$$\boldsymbol{B}_{(w1)}=\frac{\boldsymbol{I}}{\sqrt{\tilde{\psi}}} \tag{6}$$

式中:$\tilde{\psi}$ 代表取值为正整数的参数。

当获取 ZCA 白化矩阵后,对于多源数据样本 z,经过 ZCA 白化处理后可以将其表示为式(7)的形式:

$$\tilde{z}=\boldsymbol{B}_{(w1)}z \tag{7}$$

式中:$\tilde{z}$ 代表白化处理后的数据样本。

在完成上述操作后,需要在白化处理后的无线传感网络多源数据集上展开主成分分析(PCA)操作,提取无线传感网络多源数据特征[12]。

对 $\tilde{z}$ 展开 PCA 投影处理,得到多源数据中第一层的特征 S_1:

$$S_1=w_1^t\tilde{z} \tag{8}$$

式中:w_1^t 代表在 t 时间段内的初始权重。通过式(8)获取多源数据中第一层特征,也可以被称为数据的浅层特征,即在表面层的局部特征。

PCA 在数据压缩的同时搜索原始多源数据的最佳描述,确保降维后的数据内冗余信息大幅度降低。为了进一步提升多源数据更深层次的特征,通过 PCA 对多源数据展开第二层特征提取,得到第二层特征 S_2。在完成第一层和第二层的特征提取后,将两层特征展开级联处理,获取最终的多源数据特征 S,如式(9)所示:

$$S=[S_1,S_2]T \tag{9}$$

式中:T 代表时间序列。

1.3 多源数据特征融合

在无线传感网络中,由于传感器节点通常分布广泛,具有多个数据源。为了综合利用不同源的信

息,提供更全面的数据表示,从而提高信息利用率,对多源数据特征 S 进行融合处理。融合过程中,引入采用模糊数学中的最大最小贴近度描述多源数据特征间的距离,各个模糊集之间的差距或者距离可以通过常规的度量理论展开估计。为此,还需要在模糊幂集上构建一个度量,并且对度量展开标准化处理。其中,贴近度和模糊集之间存在密切关联,对应的表达式为:

$$\theta_{(\boldsymbol{a},\boldsymbol{b})}=1-\frac{\sum_{S=1}^{S}\min[\boldsymbol{a}(S),\boldsymbol{b}(S)]}{|S|} \tag{10}$$

式中:$\theta_{(\boldsymbol{a},\boldsymbol{b})}$ 代表模糊集 a、b 之间的贴近度;$\boldsymbol{a}(S)$ 和 $\boldsymbol{b}(S)$ 分别代表模糊集 a、b 在特征值 S 上的隶属度。

由于传感器网络所在位置以及自身部件质量之间的差异性,所以通过亲信度衡量传感网络当前节点测量值和其他节点测量值之间的亲疏程度,进而判断数据是否有效。通过最大最小贴近度概念组建亲信度。根据模糊数据理论,将传感网络中各个节点的测量值设定为一个模糊集,则在 t 时刻模糊集 $\boldsymbol{a}$、$\boldsymbol{b}$ 的亲信度 ϕ 如式(11)所示:

$$\phi=\frac{\min\theta_{(\boldsymbol{a},\boldsymbol{b})}}{\max\theta_{(\boldsymbol{a},\boldsymbol{b})}} \tag{11}$$

式中:$\max\theta_{(\boldsymbol{a},\boldsymbol{b})}$ 和 $\min\theta_{(\boldsymbol{a},\boldsymbol{b})}$ 分别为最大和最小贴近度。

通过式(12)组建亲信度矩阵 $\boldsymbol{\phi}(t)$:

$$\boldsymbol{\phi}(t)=\begin{bmatrix}\phi_{11}(t) & \phi_{12}(t) & \cdots & \phi_{1p}(t)\\ \phi_{21}(t) & \phi_{22}(t) & \cdots & \phi_{2p}(t)\\ \vdots & \vdots & \ddots & \vdots\\ \phi_{q1}(t) & \phi_{q2}(t) & \cdots & \phi_{qp}(t)\end{bmatrix}^{\mathrm{T}} \tag{12}$$

第 α 行亲信度方差 $\nu_{\alpha}^{2}(t)$ 如式(13)所示,方差取值越小,说明其可信度就越高;反之,则说明可信度越低。

$$\nu_{\alpha}^{2}(t)=\frac{\boldsymbol{\phi}(t)^{\mathrm{T}}}{\alpha}-\boldsymbol{\phi}(t)^{\mathrm{T}} \tag{13}$$

综合上述分析,在亲信度方差 $\nu_{\alpha}^{2}(t)$ 较低时,重新提取多源数据特征;在亲信度方差 $\nu_{\alpha}^{2}(t)$ 较高时,进行数据特征融合。在数据特征融合[13]过程中,可以获取基于亲信度方法的多源数据特征融合结果为:

$$S'=\sum_{q=1}^{q}\sum_{p=1}^{p}\tilde{\varpi}_{qp}(t)\nu_{\alpha}^{2}(t) \tag{14}$$

式中:$\tilde{\varpi}$ 代表权重系数。通过式(14)实现了多源数据特征融合处理。

综合上述分析可知,所提方法对无线传感网络多源数据的特征融合过程可以总结为图 1。

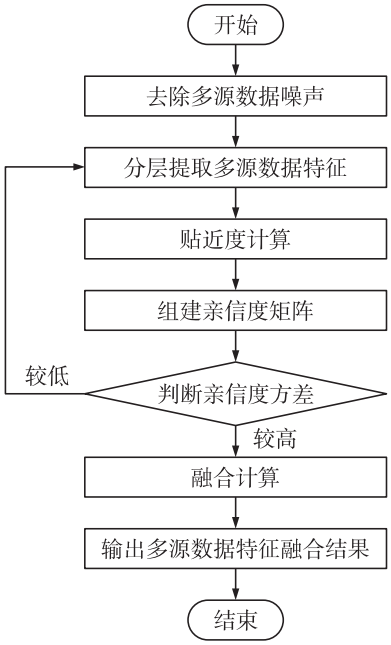

图 1　无线传感网络多源数据融合流程

由图 1 可知,所提方法无线传感网络多源数据融合流程可总结为以下步骤:

步骤 1　结合经验模态分解和小波阈值方法去除多源数据噪声;

步骤 2　通过主成分分析方法分层提取去噪后多源数据的特征;

步骤 3　利用式(10)计算提取到数据特征的贴近度;

步骤 4　依据式(12)组建亲信度矩阵;

步骤 5　依据式(13)判断数据特征之间的亲信度方差,当亲信度方差 $\nu_{\alpha}^{2}(t)$ 较低时,返回步骤 2 重新提取多源数据特征;当亲信度方差 $\nu_{\alpha}^{2}(t)$ 较高时,进行步骤 6。

步骤 6　融合计算,并输出多源数据特征融合结果。

2　仿真与分析

上述过程完成了无线传感网络多源数据特征融合,为了验证所提数据特征融合方法是否能够有效融合多源数据,提高数据处理和分析的效果进行仿真分析。

2.1　仿真设置

2.1.1　环境和参数设置

通过 MATLAB 平台展开仿真分析,仿真环境如图 2 所示。

图 2 中的网络参数设定如表 1 所示。

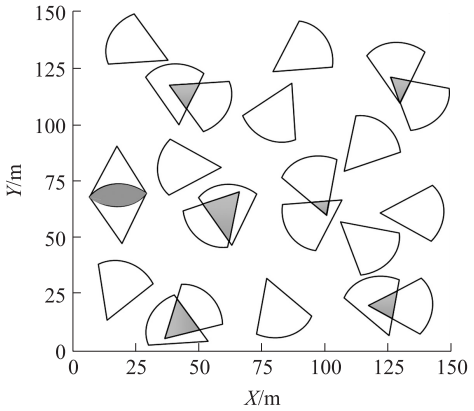

图 2　无线传感网络多源数据随机分布情况

表 1　网络参数设定

仿真参数名称	仿真参数取值
数据长度/bit	2 500
节点数量/个	2 000
区域大小/m	150×150
初始能量/J	1.5
融合单位长度数据能耗/(nJ/bit)	6×10^{-9}
自由空间模型放大器能耗/(pJ/bit/m^2)	14×10^{-16}
BS 位置坐标	(60,60)
无线收发电路能耗/(nJ/bit)	55×10^{-9}

2.1.2　仿真数据集设置

考虑到无线传感网络应用范围较广,为了保证仿真中可以进一步体现多源性,因此,在 150 m×150 m 的监测覆盖范围内,从应用场景和数据类型上多重设置仿真数据集。数据集设置情况如表 2 所示。

表 2　数据集设置

应用场景	数据类型	数据规模/GB
环境监测	PM2.5 和 PM10 浓度、湿度、温度和土壤中的 pH 值、含水量等	5
智能交通系统	车流量、车速、车辆类型等	5
农业监测	土壤湿度、温度和光照等	3
医疗保健	心率、体温和血压等	3
工业自动化	温度、压力和流量等	4
总计	20	

2.2　仿真分析

根据表 1 的仿真参数中的节点数量、区域大小、初始能量等参数,配置无线传感网络的相关参数,并使用文献[3]回波状态网络融合方法和文献[4]卡尔曼滤波增益矩阵融合方法作为对比方法进行仿真,结果如下。

①传感网络剩余总能量

分析各个方法在运行轮数持续增加情况下传感器网络内剩余总能量的变化情况。分析过程中,在表 1 的参数设定下,基于融合单位长度数据能耗为 6×10^{-9} nJ/bit 来计算每个节点的能量消耗和剩余能量。结合所有节点的能量消耗和剩余能量,计算传感网络的剩余总能量。结果如图 3 所示。

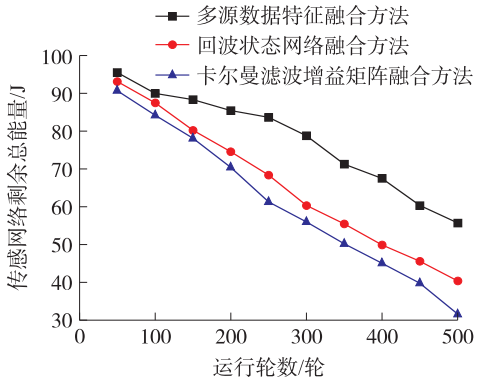

图 3　不同运行轮数下各个方法的传感网络剩余总能量比较

分析图 3 可知,在运行轮数持续增加的情况下,各个方法对应的传感网络剩余总能量呈现直线下降趋势。经过 500 轮运转后,多源数据特征融合方法的传感网络剩余总能量分别高于另外两种方法 15.3 J 和 24.1 J,充分验证了多源数据特征融合方法的优越性。综合上述数据可知,多源数据特征融合方法的剩余总能量高于两种对比方法,所提方法应用后,无线传感网络中的节点能够有效地利用能量进行数据传输和处理,因为所提方法通过集合了经验模态分解和小波阈值去噪方法的优势,去除了多源数据中的噪声,避免了无线传感网络节点通信时噪声数据的干扰,降低了节点在数据融合过程中的能耗,从侧面印证了所提方法对多源数据的融合效果较好。

②节点死亡率

在无线传感网络中,功效会随着节点的死亡而降低直至消失,所以节点的死亡率也是衡量方法是否优越的一项重要指标。在传输过程中,在表 1 的参数设定下,基于融合单位长度数据能耗为 6×10^{-9} nJ/bit 来计算每个节点的能量消耗和剩余能量,根据设定的初始能量 1.5 J 和运行轮数,判断节点是否将要耗尽能量或已经死亡,从而得出节点死亡率的评估结果。仿真测试结果如图 4 所示。

分析图 4 可知,各个方法对应的节点死亡率会随着运行轮数的增加呈现上升趋势。多源数据特征融合方法对应的节点死亡率保持在 5%以下;回波状态网络融合方法对应的节点死亡率保持在 11%以下;卡尔曼滤波增益矩阵融合方法对应的节点死亡率保持在 12%以下。因为多源数据特征融合方法利用主成分分析方法提取了多源数据的第一层和第二层特征,

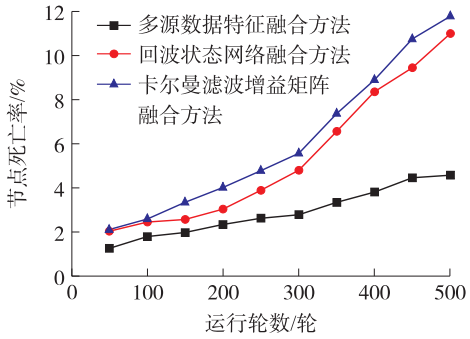

图 4　不同运行轮数下各个方法的节点死亡率比较

并将其进行级联保证了每一个节点传输的数据都是有效的,避免节点因为无效或冗余数据的传输而浪费能量资源。同时,所提方法从不同的角度收集和汇总各种数据,更全面地了解节点的情况和趋势,从而更准确地评估节点的健康状态和风险程度,延长节点的寿命,降低了节点死亡率。

③无线传感网络多源数据特征融合延迟

融合延迟分析时,在表 1 的参数设定下,在 500 轮次下,记录 2 000 个节点传输多源数据从采集节点传输到基站的时间戳,并计算无线传感网络多源数据特征融合的延迟。通过图 5 分析不同方法的无线传感网络多源数据特征融合延迟变化情况。

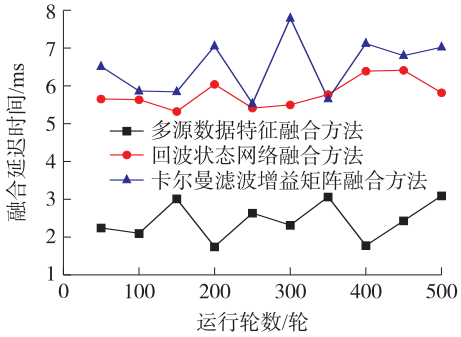

图 5　不同运行轮数下各个方法的无线传感网络多源数据特征融合延迟比较

分析图 5 可知,和另外 2 种数据特征融合方法相比,多源数据特征融合方法在展开无线传感网络多源数据特征融合过程中,出现的延迟要低于其他方法,保持在 4 ms 以下。主要是因为多源数据特征融合方法利用经验模态分解和小波阈值对全部无线传感网络多源数据展开了去噪处理,有效剔除了数据内带有噪声的数据,最大程度降低无线传感网络多源数据特征融合延迟。

④平均节点能耗

在分析平均节点能耗时,根据表 1 的仿真参数中的节点数量、区域大小、初始能量等参数,配置无线传感网络的相关参数,并在 2 000 个节点下,基于 6×10^{-9} nJ/bit 的融合单位长度数据能耗,得到各个方法的平均节点能耗变化情况,结果如图 6 所示。

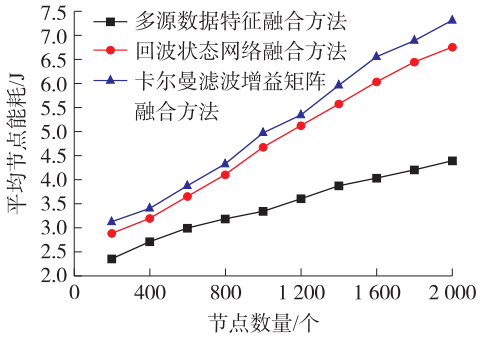

图 6　不同节点数量下各个方法的平均节点能耗比较

分析图 6 可知,各个方法的平均节点能耗会随着节点数量的增加而增加。但是和另外两种方法相比,多源数据特征融合方法平均节点能耗上升幅度比较平缓,且保持在 4.5 J 以下。因为多源数据特征融合方法采用模糊数学融合了多源数据特征,这一过程中通过综合考虑点自身的能耗信息、环境参数和外部影响因素等多个数据源,提高了节点能耗的预测能力,并根据实际情况对节点能耗进行动态调整和优化,通过上述过程优化了特征融合效果,进一步提高了能量利用效率,降低了平均节点能耗,更加适用于解决无线传感网络多源数据特征融合问题。

⑤多场景下的特征融合精度

在多场景下的无线传感网络多源数据特征融合中,特征融合精度是评估不同方法在整合来自不同场景的特征时的准确性和可靠性的指标。它反映了不同方法对多个场景中多源数据不同特征的融合程度。因此,在表 2 应用场景的 20 GB 的数据规模下,分别分析多源数据特征融合方法、回波状态网络融合方法和卡尔曼滤波增益矩阵融合方法对多源数据的特征融合精度,结果如图 7 所示。

分析图 7 可知,在多场景下的无线传感网络多源数据特征融合中,多源数据特征融合方法的特征融合精度高于 86.3%;相较于回波状态网络融合方法和卡尔曼滤波增益矩阵融合方法的特征融合精度 74.5%

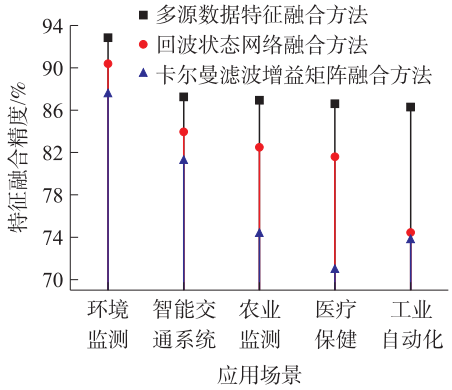

图 7　各个方法在多场景下的特征融合精度比较

和 73.6%更高。由此可知,所提无线传感网络多源数据特征融合方法可以适应不同的多源数据类型和无线传感网络应用场景,具有较好的特征融合效果。

3 结束语

为了更好地融合多源数据特征,优化无线传感网络传输速率,提出一种无线传感网络多源数据特征融合方法。通过仿真证明,采用多源数据特征融合方法可以有效提升传感网络剩余总能量,同时还可以降低节点死亡率、多源数据特征融合延迟以及平均节点能耗,更好地实现无线传感网络多源数据特征融合,提高了无线传感网络的数据传输性能。由于受到研究周期以及条件等多方面因素限制,多源数据特征融合方法获取的研究成果没有达到深入和仔细的阶段,未来将对多源数据特征融合方法展开进一步完善,确保其可以得到更加广泛地应用。

参考文献:

[1] 刘航,吕国林. 多源数据融合的无人机性能参数智能感知研究[J]. 计算机仿真,2022,39(1):40-43,48.

[2] 付翔,何宗权,黄斌,等. 基于多信息数据融合滤波的坡度识别算法[J]. 江苏大学学报(自然科学版),2021,42(2):173-179.

[3] Wootton A J, Day C R, Haycock P W, et al. Heterogeneous Data Fusion for the Improved Non-Destructive Detection of Steel-Reinforcement Defects Using ECHO State Networks[J]. Structural Health Monitoring, 2022,21(6):2910-2921.

[4] 张辉,黄向生. 基于 UKF 的无线传感器异步数据融合优化算法[J]. 重庆大学学报,2021,44(5):115-123.

[5] 莫慧凌,郑海峰,高敏,等. 基于联邦学习的多源异构数据融合算法[J]. 计算机研究与发展,2022,59(2):478-487.

[6] Cesaire Y,Trujillo R. Local Adaptive Bivariate Shrinkage Function for Seisogram Wavelet Based Denoising[J]. Latin America Transactions,2021,19(2):342-348.

[7] 丁明宽,石志勇,韩兰懿,等. 基于 EMD-DFA-小波阈值的 MEMS 陀螺信号去噪方法[J]. 火炮发射与控制学报,2021,42(2):50-56.

[8] Mohammadi A K,Mohebian R,Moradzadeh A. High-Resolution Seismic Impedance Inversion Using Improved Ceemd with Adaptive Noise[J]. Journal of Seismic Exploration,2021,30(5):481-504.

[9] 王民顿,尚俊娜. 基于 CEEMD 和改进小波阈值法的钢架结构沉降数据去噪方法[J]. 大地测量与地球动力学,2022,42(11):1191-1195.

[10] 赵永梅. VMD 和小波阈值重构的电力电缆局部放电信号去噪法[J]. 西安科技大学学报,2021,41(4):739-746.

[11] 周建新,周凤祺. 基于改进协同量子粒子群的小波去噪分析研究[J]. 电光与控制,2022,29(1):47-50.

[12] Ahmed M Z,Mahesh C. An Efficient Image Based Feature Extraction and Feature Selection Model for Medical Data Clustering Using Deep Neural Networks[J]. Traitement du Signal,2021,38(4):1141-1148.

[13] 廖亚桢,刘昱,张立强,等. 面向 PDR 定位系统的多层感知机 IMU 阵列数据融合方法[J]. 电子测量与仪器学报,2022,36(8):35-42.

陈 宏(1980—),女,汉族,福建三明人,硕士,工程师,信息系统项目管理师,研究方向:无线传感器网络、信息网络安全,主持或参与多个福建省中青年教师教育科研项目、国家重点实验室开放课题、及企业合作项目,累计发表论文 5 篇,参与发明专利一项,chenhong99611@126.com;

黄丽萍(1981—),女,汉族,湖南常德人,硕士,工程师,研究方向:无线传感器网络、信息网络安全;

蒋文贤(1974—),男,汉族,福建漳州人,博士研究生,副教授,研究方向:无线传感器网络、信息网络安全;

余翀翀(1980—),男,汉族,福建泉州人,硕士,高级实验师,研究方向:无线传感器网络、信息网络安全。