# Identification of Sensor Node Localization Anomalies in Parallel Big Data Networks*

SHENG Bo[1*], ZHANG Yuejin[2]
(1.Library and Information Center, Changzhou Art Branch, Jiangsu Union Technical Institute, Changzhou Jiangsu 213000, China;
2.Information Engineering College, East China Jiaotong University, Nanchang Jiangxi 330013, China)

**Abstract**: Parallel big data sensing networks typically have complex topological structures and a large number of nodes, which can affect the reliability of identifying abnormal nodes in the sensing network under big data interference and complexity interference. To this end, a method for identifying sensor node localization anomalies in parallel big data networks is proposed. Boolean model is established to calculate the coverage radius of the node transmission channel, the characteristic parameters in the transmission channel are estimated based on carrier modulation methods, a communication transmission channel model between sensor nodes is estimated, and the node status and distribution are obtained. The distributed sensor sequence sampling model is constructed, the sensor node feature sequence is collected, and the compressed sensing method is used to identify the anomaly of sensor node positioning. The simulation results show that the energy consumption of the proposed method for identifying abnormal nodes is always lower than 3 J, and the average recognition error does not exceed 0.35%, the node positioning accuracy is higher than 95%, and the running time for identifying sensor abnormal nodes is less than 2 ms, which can effectively extend the lifespan of parallel big data networks.

**Key words**: sensing nodes; identification of positioning anomalies; compressed sensing; parallel big data network; boolean model; feature collection
**EEACC**: 7230 **doi**: 10.3969/j.issn.1004-1699.2024.12.022

# 面向并行大数据网络中传感节点定位异常识别*

盛 波[1*], 张跃进[2]
(1.江苏联合职业技术学院常州艺术分院图文信息中心, 江苏 常州 213000; 2.华东交通大学 信息工程学院, 江西 南昌 330013)

**摘 要**: 并行大数据传感网络通常具有复杂的拓扑结构以及较多的节点数量, 在大数据干扰和复杂性干扰下, 影响传感网络异常节点识别的可靠性。为此, 提出面向并行大数据网络中传感节点定位异常识别方法。建立布尔模型计算节点传输信道覆盖半径, 根据载波调制方法估计传输信道中的特征参量, 建立传感节点之间通信传输信道模型, 获取节点状态及分布情况。构建分布式传感序列采样模型, 采集传感节点特征序列, 利用压缩感知方法实现对传感节点定位的异常识别。仿真结果表明, 所提方法识别异常节点的能耗始终低于 3 J, 平均识别误差未超过 0.35%, 节点定位精度高于 95%, 识别传感异常节点的运行时间在 2 ms 以下, 能够有效延长并行大数据网络的寿命。

**关键词**: 传感节点; 定位异常识别; 压缩感知; 并行大数据网络; 布尔模型; 特征采集;
**中图分类号**: TP393 **文献标识码**: A **文章编号**: 1004-1699(2024)12-2159-06

并行大数据传感网络是一种基于并行计算的大数据处理网络, 能够并行处理多个数据源, 具有高可伸缩性、高并行性和高效性等特点。随着并行大数据网络的不断更新发展, 恶意程序、安全漏洞等恶意攻击逐渐产生, 严重影响了传感网络中数据传输的安全性以及可靠性[1]。为此, 需要研究一种并行大数据网络传感节点定位异常识别方法。

陈卡[2]针对传感异常节点检测过程中存在低精度检测的问题, 提出一种基于跳距修正(Hop Distance-Corrected Localization, HDCL)的无线传感网络节点定位方法。该方法利用无偏估计对节点的跳距展开修正处理; 在此基础上通过分析贡献度概念筛选出合适的锚节点, 将该节点作为节点参数指标对比未知节点, 获取异常节点状态; 通过异常节点状态构建定位误差目标函数, 采用 Jaya 算法实现对异常节点的定位识别。朱海龙等[3]提出一种基于置信

**项目来源**: 国家自然科学基金重大研究计划培育项目(92159102); 江西省教育厅科学技术研究项目(GJJ2200642)
**收稿日期**: 2023-05-12 **修改日期**: 2023-05-26

规则度(Belief Rule Base,BRB)的传感节点异常检测方法。该方法首先根据传感器网络中节点的工作特性和工作原理分析传感异常节点检测过程;其次将空间和时间两个维度中的所有节点数据提取出来并通过分析空间和时间的相互作用,构建一种新的传感器节点异常检测模型;最后利用 Intel Lab Data 无线传感器数据集对其展开深入的分析,以此获得更准确的结果,实现对传感异常节点的识别。Mishra 等[4]基于模糊的主观聚合概念,建立网络关键节点模糊综合识别模型,确定客观准则权重,对网络中异常节点进行排序,解决多准则决策问题,确定二维标准权重,可以更容易地完成异常节点的识别。

但是,当前方法并不适用于并行大数据传感网络,其通常具有复杂的拓扑结构以及较多的节点数量,在大数据干扰和复杂性干扰下,影响传感网络异常节点识别的可靠性。为了解决上述方法中存在的问题,提出面向并行并行大数据网络中传感节点定位异常识别方法。通过获取传输信道覆盖半径,建立传感节点之间通信传输信道模型,有效获取了传感节点状态及分布情况。利用分布式传感序列采样模型定位异常节点,提高识别效率,通过大数据整理、线性计算、重构和判决快速完成异常节点的识别。

## 1 节点传输信道模型构建

在无线传感网络拓扑结构和传感节点分布模型的基础上,建立了布尔模型,计算传输信道覆盖半径,并以此为基础,利用载波调制方法获取参量特征估计值,构建节点通信传输信道模型,有效获取了传感网络节点状态及分布状况。

### 1.1 获取传输信道覆盖半径

为了识别并行大数据网络下的传感节点定位异常情况,构建网络拓扑结构和传感节点分布模型,并对其展开分析。定义 $\boldsymbol{H}=(\boldsymbol{B},R)$ 为二元有向图的网络拓扑结构[5],其中 $\boldsymbol{B}$ 代表的是在并行大数据网络的监测范围内,无线通信网络的节点集合,$R$ 表示无线通信节点范围,$\boldsymbol{H}$ 表示的是 $R$ 中传感节点密度对集合。

用 $L(t_o)$ 描述区域中通信半径范围内的传感节点分布密度,当节点密度分布处于最优状态时,构建基于并行大数据网络拓扑结构的布尔模型[6]:

$$t_o(c)=L(t_o)+\boldsymbol{H}m(o,c) \tag{1}$$

式中:$m(o,c)$ 表示布尔项向量。通过该模型的构建,能够有效地获取到网络通信节点 $d$ 和中继节点 $a$ 之间的传输信道覆盖半径 $f(d,a)$,计算公式为:

$$f(d,a)=t_o(c)E_d u \tag{2}$$

式中:$u$ 表示传感节点数目,$E_d$ 表示传感节点之间的欧氏距离[7]。

### 1.2 传感节点之间通信传输信道模型

为了达到传感异常节点定位识别的目的,所提方法以传输信道覆盖半径为基础,构建传感节点之间通信传输信道模型[8],由此获得其参量特征值,具体实现过程如下:

首先分析并行大数据网络中传感节点通信的扩展损失 $c_e$ 和衰减损失 $u_e$,得到并行大数据网络通信节点的信道测量公式:

$$\alpha_l=f(d,a)(c_e+u_e)+j\omega_l^o \tag{3}$$

式中:$j$ 为传感节点规律减弱的几何特性,$j=1,2,\cdots,Q$;$\omega_l^o$ 为 $l$ 时刻传感节点 $o$ 的测量向量。因为空间分布中的传感节点的离散特性,所以可以利用不同精度的射线模型估计传感节点的能量传播损失,以此构建信道的检测模型:

$$C=S(l)c(l)+W(l)i_o(l)\alpha_l+F_o(l)u_e\xi(l) \tag{4}$$

式中:$c(l)$ 描述的是传播状态;$S(l)$ 代表的是传感节点的异常特征状态转移偏差;$\xi(l)$ 和 $i_o(l)$ 用于描述模型中的各种干扰项,同时代表均值为零但方差分别为 $W(l)$ 和 $F_o(l)$ 的高斯白噪声[9]。

利用载波调制方法[10]对该信道中的特征参量展开估计。载波调制方法是一种将特征参量通过调制技术嵌入到载波信号中,使得特征参量可以通过载波信号传输的方法,在通信传输中可获取信号的特征参量,从而得到通信传输中传感节点间的初始信道参量状态 $c(0)$,其平均值为 $c_0$,方差为 $a_0$。在传播条件不同的条件下,传感节点特征在 $\xi(l)$ 以及 $i_o(l)$ 中具有独立性,且 $o=1,2,\cdots,M$,在此基础上采用一种自适应量化估计方法对传感节点之间的通信传输信道测量值展开估计,计算方法如下:

$$q_o(l)=u_e w_o(l)+c(l)i_o(l)+Cw_o(l) \tag{5}$$

式中:$w_o(l)$ 表示传输数据的衰减因子,由此,构建传感节点之间通信传输信道模型为:

$$T_w^o(l)=\frac{c_0 C\sum_{o=1} q_o(l)}{f(d,a)a_0[W(l)+F_o(l)]} \tag{6}$$

通过构建上述传感节点通信传输信道模型以及实施参量估计,能够有效地获取到并行大数据网络中传感节点状态及分布情况。

## 2 传感节点定位异常识别方法

构建分布式传感序列采样模型,采集大数据网络传感节点特征序列,通过大数据整理、线性计算及

重构和判决,完成并行大数据网络上所有传感节点的异常定位识别。

### 2.1　分布式传感序列采样

在构建传感节点分布模型及信道模型的基础上,需要根据传感节点的状态及分布情况作进一步的信息融合和自适应特征采样处理[11],因此所提方法利用分布式传感序列采样模型对传感器异常节点的数据展开定位,进而实现传感异常节点数据的智能识别,具体步骤如下:

用 $i$ 表示传感节点,基于并行大数据网络中传感节点信息的特征建立传感节点的分布式采样模型,并设定采集的时间顺序;在此基础上建立基于传感节点数据信息流特征的传感检测特征分布模式:$A_o=(\boldsymbol{a}_{o1},\boldsymbol{a}_{o2},\cdots,\boldsymbol{a}_{oF})$,其中,$\boldsymbol{a}_{oF}$ 表示传感节点感测区域 $V$ 内传感节点与其他节点之间的分布点,并利用该点集构建并行大数据网络传感节点之间的关联规则调度模型:

$$R_t=T_w^o(l)(\beta_1+\beta_2 f_1)+(\beta_1+\beta_2 f_2)+(\beta_1+\beta_3 f_3) \quad (7)$$

式中:$\beta_1$、$\beta_2$ 和 $\beta_3$ 表示不同支持度;$f_1$、$f_2$ 和 $f_3$ 表示不同置信度;采用一种多元信息融合方法[12]对其展开优化调度,根据传感节点检测序列的状态特征识别该模型中模糊信息,以此得到传感节点数据检测的时间窗口函数:

$$\xi(y)=\xi_{\max}-R_t(\xi_{\max}-\xi_{\min})\times y/y_{\max} \quad (8)$$

式中:$y$ 用于描述检测时间;$\xi_{\max}$ 和 $\xi_{\min}$ 代表的是该窗口在取值范围内的最大值与最小值。

在选取的感测范围内,传感节点的总数目用 $I$ 表示,在此基础上求出并行大数据网络传感节点数据的模糊隶属度函数[13],计算方法如下:

$$u(y)=\pi^{-1}\int\frac{\xi(y)}{\chi\upsilon}\mathrm{d}\upsilon+I\chi(\pi y)^{-1} \quad (9)$$

式中:$\upsilon$ 用于描述采集数据时间的延迟状态;$\chi$ 用于描述传感节点数据特征分布长度。

在上述分析的基础上,利用区域信息融合方法获取分布式传感序列采样模型:

$$\chi_f=\frac{\boldsymbol{M}_t-\lambda+1}{u(y)\boldsymbol{M}_t},\quad \lambda\in[2,M_t] \quad (10)$$

式中:$\boldsymbol{M}_t$ 指的是传感节点数据信息流的特征向量;$\lambda$ 代表的是特征向量参数。

通过分布式传感序列采样模型的构建,实现对并行大数据网络传感节点特征序列的采集,将其作为下属传感节点异常识别方法的参数指标,以此提高异常节点检测效率。

### 2.2　传感节点异常识别方法

在采集到并行大数据网络中传感节点的特征序列后,在大数据干扰下根据其特性选取一种基于压缩感知的传感节点定位异常识别方法,主要分为三个步骤:

①大数据整理

在并行大数据网络运行过程中,其正常节点和异常节点均会产生不同程度的电量损耗[14],因此很难分辨所有节点在时域中产生电量损耗向量 $\boldsymbol{B}$ 的稀疏性,从而影响数据重构的成功率以及对异常节点的判别,需要通过阈值计算将二者的电能损耗向量 $\boldsymbol{B}$ 做更深入的整理:

传感节点在距离 $f$ 内发送 1-bit 信号的能量消耗情况如下式:

$$R(f)=\begin{cases}\chi_f R_e+\varepsilon_g f^2 & f<f_0\\ R_e+\varepsilon_q f^4 & f\geqslant f_0\end{cases} \quad (11)$$

式中:$R_e$ 表示接收器接收和发射器发射信号为 1-bit 时消耗的能量,也是传感节点能量损耗的主要来源;$\varepsilon_g$ 和 $\varepsilon_q$ 分别代表的是空间损耗能量和多路径损耗能量。当距离 $f$ 大于距离 $f_0$ 时,能量损耗较大,反之较小。

在此基础上求出在第 $t$ 个检测周期内,并行大数据网络以及其传感节点的平均能耗 $\overline{H}(t)$:

$$\overline{H}(t)=M^{-1}\sum_{o=1}^{M^2}R_o(t)-R(f)\sum_{o=1}^{M^2}R_o(t-1) \quad (12)$$

式中:$R_o(t)$ 为初始能量,若传感节点的电量损耗值不超过 $s\times\overline{H}(t)$,则输出 1,反之输出 0,其中 $s$ 代表的是门限调整系数。通过上述计算可获得传感节点在检测周期内的电量损耗状态向量 $\boldsymbol{C}_M$,将其转换为下式:

$$\boldsymbol{C}_M=\begin{cases}0 & c_o<s\times\overline{H}(t)\\ 1 & c_o\geqslant s\times\overline{H}(t)\end{cases}\quad 1\leqslant o\leqslant M^2 \quad (13)$$

最后对门限系数 $s$ 展开合理的取值,以此减少 $C_M$ 中输出为 1 的元素个数,传感节点电量损耗状态向量在通过上述数据整理后自动变为稀疏的 01-向量。

②线性计算

利用具有较好稀疏特性的 01-向量能够稳定实现对向量 $\boldsymbol{C}_M$ 的压缩感知。针对并行大数据网络的实际应用方向,所提方法选取一种基于 LDPC 码[15]的校验矩阵,因其元素中只包含 0 和 1 两种类型,可以作为测量矩阵对传感节点数据展开压缩感知,表达式如下:

$$\boldsymbol{U}=\boldsymbol{\phi}\boldsymbol{\Psi}\boldsymbol{C}_M \quad (14)$$

式中:$\boldsymbol{\Psi}$ 为稀疏基地矩阵;$\boldsymbol{\phi}$ 为低密度奇偶校验码的稀疏校验矩阵,主要运算方式为线性计算。

通过上述计算能够看出，$\boldsymbol{C}_M$ 中输出为 1 的元素一直未超过向量元素的个数，因此可以看出时域中具有稀疏特性，在观测端 $Q_o$ 对收集到的数据展开线性计算，以此获得观测结果 $\boldsymbol{U}$。

③重构和判决

基于在前一阶段所得到的观测结果以及测试矩阵，可以较好地实现对传感节点电量消耗状态向量的重建，由于 LDPC 码的校验矩阵代替测试矩阵，可选为合理的重建方法。置信传播方法[16]（Belief Propagation，BP）是最为典型的低密度奇偶校验码（Low-Density Parity-Check，LDPC）译码方法，该方法将 LDPC 编码转化为一种 Tanner 图像，利用图中信息的持续迭代完成收敛。两者之间的对应关系如图 1 所示。

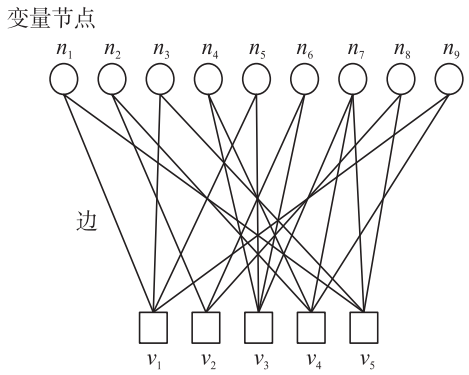


**图 1　低密度奇偶校验矩阵和 Tanner 图像的对应**

通过该方法完成对 Tanner 图中所有数据信息传递流程的持续迭代，当经过校验节点和变量节点之间的交换和迭代过程获得了一定意义上的收敛速度效应后，能够有效地获取到传感节点电量损耗状态向量 $\boldsymbol{C}_M$ 的估计值。该检测周期内的电量输出峰值等于 1 的元素所对应的传感节点将被判定为电能消耗过大的节点，在下一个检测周期也将尽量避免该节点[17]；若在连续 $N$ 个检测周期中，某一传感节点仍持续产生较大的电量消耗值，即被判定为异常传感节点。

由上述步骤可知，本次研究通过建立节点传输信道模型对传感节点进行初步筛选，采集大数据网络传感节点特征序列后，采用压缩感知的传感节点定位异常识别方法，快速完成对并行大数据网络上所有传感节点的异常定位识别。

## 3　仿真与分析

为了验证面向并行大数据网络中传感节点定位异常识别方法的应用有效性，对其进行仿真。仿真参数如表 1 所示。

**表 1　仿真参数设置**

| 配置 | 参数 |
|---|---|
| 操作系统 | Windows 10 |
| 仿真软件 | MATLAB R2014b |
| CPU | Intel Core i5-4200U |
| GPU | FP32 |
| 位数 | 64 bit |
| 传感节点 | 300 个 |
| 迭代次数 | 6 000 次 |

### 3.1　识别能耗仿真

将并行大数据网络中某区域的 300 个未知传感节点作为仿真条件并实施 6 000 次迭代，采用所提方法和文献[2]方法（基于跳距修正的节点识别方法）、文献[3]方法（基于置信规则库的节点识别方法）、文献[4]方法（基于对数概念的节点识别方法）识别异常节点定位所耗能量，对比结果如图 2 所示。

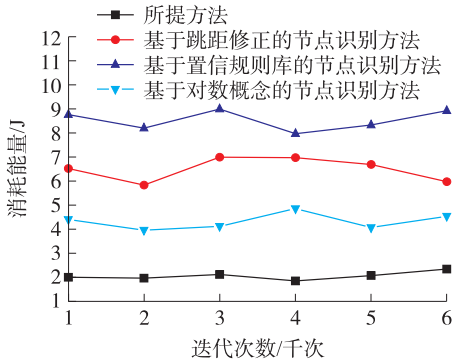


**图 2　各方法识别所耗能量比较**

分析图 2 可知，文献[3]方法所消耗的能量相对较高，产生大量的计算过程，浪费能源；文献[2]方法虽比文献[3]方法好一些，但总体波动较大，影响对传感节点的异常识别效率。而所提方法在达到 6 000 次迭代时，消耗能量仍低于 3 J。因为所提方法采用基于 LDPC 码的校验矩阵，对传感节点数据展开压缩感知，完成对 Tanner 图中所有数据信息传递流程的持续迭代，获取传感节点电量损耗估计值，可以更好地了解节点的能量消耗情况，从而优化节点能量消耗的策略，延长节点寿命。

### 3.2　节点定位精度仿真

节点的定位精准度是指传感节点在空间位置上的测量精度，是影响节点测量数据准确性和可信度的重要因素。为了验证所提方法的节点定位精度，在上述条件下，利用四种方法测试节点定位精度，结果如图 3 所示。

由图 3 可知，随着节点数量的增加，四种方法的定位精度均有所下降，但所提方法的定位精度始终高于 95%，且在四种方法中的定位精度最高。这是

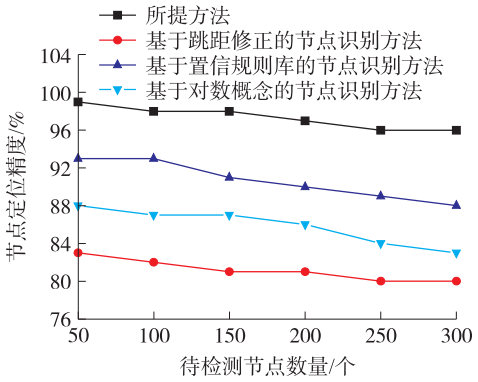


**图 3　各方法节点定位精度对比**

因为所提方法根据建立的节点通信传输信道模型,获取了节点状态及分布情况,提高了定位精准性。

### 3.3　识别精度仿真

识别精度是并行大数据网络中传感节点异常识别的一个重要影响因素,为了体现所提方法识别精度的优劣,在仿真条件相同的情况下,利用上述四种方法分别识别网络中待检测的节点,结果如图 4 所示。

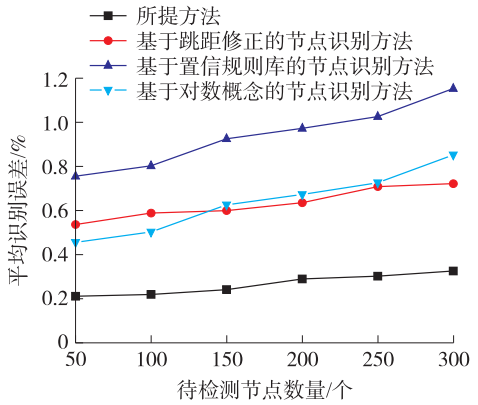


**图 4　各方法误差对比**

由图 4 可以看出,随着传感节点数量的增加,四种方法的平均识别误差均有不同程度的差异,在相同节点数量中,所提方法的误差均值最小,未超过 0.35%,对异常传感节点的识别精度较高。这是因为所提方法对大数据进行了整理,能够有选择性地选取传感节点信息,进而保证异常传感节点的识别精度。

### 3.4　识别时间仿真

为了进一步验证所提方法的可行性,仿真在 300 个传感节点选取 100 个传感异常节点,并按递增规律随机分布在含有大量未知节点的并行大数据网络中,对比上述四种方法识别异常节点的运行时间,结果如图 5 所示。

由图 5 得出,当传感异常节点增加时,文献[3]方法需要对包括原始节点的整个新节点序列重新展

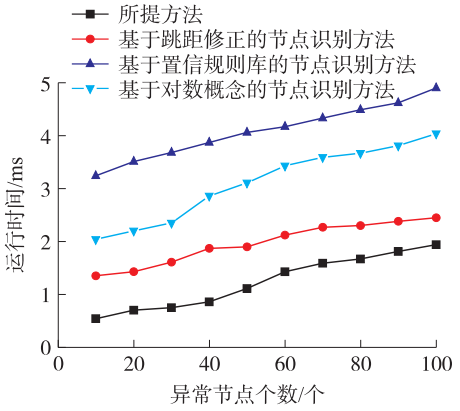


**图 5　各方法运行时间对比**

开计算识别,因此运行时间相对较长;文献[2]方法虽然减少了重复计算的过程,但仍存在一些误识别。而所提方法在两种对比方法的基础上,在保留原始节点检测数据的同时能够快速识别出并行大数据网络中的传感异常节点,运行时间在 2 ms 以下。这是因为该方法构建了分布式传感序列采样模型,运用该模型完成分布式传感序列的采样,并将其作为异常识别的参数指标,有效提高了异常节点的识别效率。

## 4　结束语

针对并行大数据网络中传感节点定位异常识别效率较低和精度较低等问题,提出面向并行大数据网络中传感节点定位异常识别方法。该方法在获取传感节点的传播衰减系数和相对时延的基础上,建立分布式传感序列采样模型,获得传感节点特征序列,利用基于压缩感知的传感节点定位异常识别检测方法,完成对传感节点定位异常的有效识别。经验证,所提方法可以有效地延长并行大数据网络的寿命,具有较高的识别效率和识别精度。

**参考文献:**

[1]　唐新梅,覃开贤.大数据移动通信网络数据交换隐私隐写加密[J]. 计算机仿真,2022,39(6):403-407.

[2]　陈卡.基于跳距修正的 WSNs 的无线传感网络节点定位算法[J]. 组合机床与自动化加工技术,2021(12):23-26.

[3]　朱海龙,耿文强,韩劲松,等. 利用置信规则库构建 WSN 节点故障检测模型[J]. 智能系统学报,2021,16(3):511-517.

[4]　Mishra A K,Joshi N,Mathur I. A fuzzy based integrated model for identification of vital node in terrorist network using logarithmic concept[J]. Journal of Intelligent & Fuzzy Systems,2020,39(3):3617-3631.

[5] 张庆霞,李军,刘坤,等. 基于增材制造技术的电枢结构优化与试验研究[J]. 兵器材料科学与工程,2023,46(1):13-18.

[6] 刘开元,谢田法.右删失数据下带协变量威布尔分布模型参数的估计[J]. 系统科学与数学,2022,42(9):2497-2507.

[7] Zhou R,Wang X,Wan J,et al. EDM-Fuzzy:An Euclidean Distance Based Multiscale Fuzzy Entropy Technology for Diagnosing Faults of Industrial Systems[J]. IEEE Transactions on Industrial Informatics, 2021, 17 (6): 4046-4054.

[8] 谢文平,尹禄高,陈小敏,等. 塔架环境通信网络信道模型及性能研究[J]. 信号处理,2021,37(11):2077-2083.

[9] Mukherjee A,Mandal S,Ghosh D,et al. Influence of Additive White Gaussian Noise on the OEO Output[J]. IEEE Journal of Quantum Electronics, 2021, 57(1): 5000110.

[10] 黄章斌,杨荣杰.OFDM 雷达信号子载波调制方式识别方法[J]. 火力与指挥控制,2022,47(3):111-115,119.

[11] 曲仝,孙华东,郑超,等. 计及量测噪声的特征节点对联合判稳与紧急控制[J]. 中国电机工程学报,2022,42(23):8496-8509.

[12] 孟月波,陈宣润,刘光辉,等. 高低密度多维视角多元信息融合人群计数方法[J]. 控制与决策,2023,38(1):181-189.

[13] 梁蓉蓉,潘小东.模糊集合隶属函数的确定方法及实验[J]. 四川师范大学学报(自然科学版),2021,44(4):479-486.

[14] 孙璇.无线传感器网络电量损耗异常节点识别[J]. 计算机工程与设计,2020,41(10):2724-2728.

[15] Yu Q,Wu F,Zhang M,e t al. Improving reliability using phase distribution aware LDPC code for holographic data storage[J]. Applied Optics,2022,61(21):6119-6127.

[16] 牛进,王晓峰,左逢源,等. 基于二维结构熵的置信传播算法收敛性分析[J]. 计算机应用研究,2021,38(7):2032-2036,2043.

[17] 桂易琪,叶子鉴,陈永康.一种基于区域划分和内容流行度的 NDN 缓存策略[J]. 江苏大学学报(自然科学版),2023,44(2):186-193.

**盛　波**(1979—),女,汉族,江苏常州人,工程硕士,副教授,研究方向:计算机科学与技术,就职于江苏联合职业技术学院常州艺术分院,长期从事校园网络建设管理、信息化建设管理、教学工作等。以第一作者在省级以上刊物上发表论文多篇,参与多项省、市级课题研究,参编教材 3 部,Shengb18542@163.com;

**张跃进**(1978—),男,汉族,湖北钟祥人,博士,教授,研究方向:算法设计与分析、生物医学工程。