

基于数据驱动的 WSN 节点故障诊断算法

常 琳^{1,2}, 黄志清^{1,2}, 张严心³

(1. 北京工业大学 信息学部, 北京 100124; 2. 北京市物联网软件与系统工程技术研究中心, 北京 100124;
3. 北京交通大学 电子信息工程学院, 北京 100044)

摘 要: 节点故障诊断是无线传感器网络持续性监测服务的关键步骤。为准确高效地得到诊断结果, 提出一种基于数据驱动的故障诊断算法。对每个节点获取的信息建立空间高维向量, 引入自身历史数据和邻居节点数据构造十字滑动窗口, 并赋予十字方向上自定义的故障权重, 通过检测异常向量达到故障诊断的目的。实验结果表明, 该算法能够克服计算量大、故障判定条件苛刻等不足, 与分布式故障诊断算法相比, 可使故障诊断正确率提高 15.5%, 故障误警率下降 4.89%。

关键词: 无线传感器网络; 节点故障诊断; 数据驱动; 向量空间模型; 十字滑动窗口; 故障权重

中文引用格式: 常 琳, 黄志清, 张严心. 基于数据驱动的 WSN 节点故障诊断算法[J]. 计算机工程, 2017, 43(9): 105-109.

英文引用格式: CHANG Lin, HUANG Zhiqing, ZHANG Yanxin. Node Fault Diagnosis Algorithm in WSN Based on Data Driven[J]. Computer Engineering, 2017, 43(9): 105-109.

Node Fault Diagnosis Algorithm in WSN Based on Data Driven

CHANG Lin^{1,2}, HUANG Zhiqing^{1,2}, ZHANG Yanxin³

(1. Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China;
2. Beijing Engineering Research Center for IoT Software and Systems, Beijing 100124, China;
3. School of Electronic and Information Engineering, Beijing Jiaotong University, Beijing 100044, China)

[Abstract] Node fault diagnosis is the key step in the continuous monitoring service of Wireless Sensor Network (WSN). In order to implement accurate and efficient node fault diagnosis, this paper proposes a fault diagnosis algorithm based on data driven. It builds high-dimensional vector in space through the information from each node, establishes cross-sliding window by its historical data and neighbor node data, gives customized weights of fault in the cross direction, and ultimately achieves the goal of fault diagnosis by detecting abnormal vector. Experimental results show that this algorithm can overcome defects such as large amount of calculation, harsh conditions of fault judgement, etc. Compared with Distributed Fault Diagnosis (DFD) algorithm, it improves Fault Diagnosis Accuracy (FDA) by 15.5% and reduces False Alarm Rate (FAR) by 4.89%.

[Key words] Wireless Sensor Network (WSN); node fault diagnosis; data driven; Vector Space Model (VSM); cross-sliding window; weight of fault

DOI: 10.3969/j.issn.1000-3428.2017.09.019

0 概述

随着无线传感器网络 (Wireless Sensor Network, WSN) 和传感器硬件技术的发展, 传感器种类不断增多, 无线传感器网络的应用范围也在不断的扩大, 然而部署在实际环境中的传感器节点, 特别是面对恶劣的监测环境时容易发生故障, 导致监测任务失败。因此, 对无线传感器网络中发生的节点故障进行准确诊断, 是无线传感器网络应用能否成功的

关键^[1-3]。

鉴于研究者对于对象的数学模型未知, 即使得知数学模型又无法精确描述, 并且只有大量的监测数据等问题, 针对无线传感器网络中节点故障的诊断, 现有方案通常是基于数据驱动的故障诊断, 主要原理是利用机器学习、信号分析、建模等方法对大量的离/在线过程运行数据分析处理确定故障。文献[4]依据时空相关性将感知信息形成决策树集群, 从而进行能量有效的在线故障诊断。文献[5]提出

基金项目: 国家发改委项目“森林监测无线传感器网络管理与事件检测技术研究”(Q5025001201502)。

作者简介: 常 琳 (1991—), 女, 硕士研究生, 主研方向为无线传感器网络故障诊断; 黄志清 (通信作者), 张严心, 副教授、博士。

收稿日期: 2016-07-12 **修回日期:** 2016-08-29 **E-mail:** zqhuang@bjut.edu.cn

基于簇方式的传感器故障诊断算法。文献[6]提出通过部署节点提供的故障信息构建故障证据融合树的诊断故障算法。文献[7]提出一种经典的节点故障诊断算法,即分布式故障检测(Distributed Fault Diagnosis, DFD)算法。文献[8]结合增量探测方案和动态概率推理模型,提出在线故障检测方案来确认各种网络异常的来源。文献[9]利用节点感知传感数据的空间相似性诊断节点的状态。文献[10]提出通过加权中值的算法,用于无线传感器网络中故障的诊断。文献[11-12]提出向量空间模型(Vector Space Model, VSM)概念。文献[13]提出滑动窗口处理数据信息,能够高质量、小内存、速度快地解决问题。

本文在上述研究的基础上,基于数据驱动技术,提出一种结合向量空间模型和十字滑动窗口的故障诊断算法,以解决计算量大、故障判定条件苛刻等问题,实现无线传感器节点故障的诊断以及故障类型的确定。

1 算法原理

本文提出的故障诊断算法总体分为3个阶段:数据预处理阶段,向量空间模型建立阶段,十字滑动窗口诊断阶段。

1.1 数据预处理

小波变换是一种在时间频率基础上分析信号的方法,能够聚焦到信号的任意微小细节,可以去除信号中存在的噪声,是分析和处理信号的理想工具。

在无线传感器网络的应用中,通常是由多个传感器组合而成的感知节点对环境进行监测,由于设备自身配置和受环境中电磁波的影响,信号中会产生噪声。由于小波变换去噪效果较好,因此采用小波变换对输出多维信号进行去噪处理,去除原始信号中的噪声 $e(x)$ 。原始信号表示如下:

$$S(x) = \begin{cases} s(i) = f(i) + g(i) + e(i) \\ s(i+1) = f(i+1) + g(i+1) + e(i+1) \\ \vdots \\ s(j-1) = f(j-1) + g(j-1) + e(j-1) \\ s(j) = f(j) + g(j) + e(j) \end{cases} \quad (1)$$

其中, i, j 为获取的感知信号的时间序列,且 $i, j \in 1, 2, \dots, n, n$ 为整数; $S(x)$ 表示含有噪声的信号; $f(x)$ 为真实信号; $g(x)$ 为故障; $e(x)$ 为噪声且 $x \in i, i+1, \dots, j$ 。

在去除噪声的过程中,遇到数据剧烈变化时,过多分解有可能把数据峰值变化作为信号噪声消除,因此,需要综合考虑去除噪声的分解过程。

1.2 向量空间模型建立

向量空间模型是一种简便、高效的文本表示模

型,已被成功地应用于著名的 SMART 文本检索系统^[11-12]。

空间向量具有成熟的相似性判定原理,并且可以很方便地将向量之间的差异度量化为向量大小和方向上的差异,因此,其被引入节点故障诊断过程。

在一般的使用环境中,可利用欧氏距离或余弦相似度判定 2 个体是否相似,但在节点的故障诊断过程中,存在以下问题:

1) 单独采用欧氏距离判定向量个体是否相似,存在以下弊端:若 n 维向量中存在某一维数值改变时,将直接影响空间中映射点的位置,间接影响欧氏距离大小,个体差异性表现不完全。

2) 单独采用余弦相似度判定向量个体是否相似,存在以下弊端:若 n 维向量中存在某一维数值改变时,将会影响向量方向的变化,导致向量个体之间相似性判定不准确。

因此,综合向量的方向和大小 2 个因素,本文引入故障贡献率作为故障诊断的判定依据。

将故障贡献率 λ 定义为 2 个体余弦相似度值的倒数与欧式距离乘积的百分数的值。 λ 的值越大,说明向量个体之间差异越大,出现故障的可能性越大。 λ 计算公式如下:

$$\lambda = \left| \frac{|S_i| \times |S_j|}{S_i \times S_j} \right| \times \sqrt{\sum_{i,j=1}^n (S_i - S_j)^2} \times 100\% \times \lambda \quad (2)$$

其中,向量 S_i 和向量 S_j 分别表示不同的感知节点构建的向量空间模型, i 与 j 为正整数。

在无线传感器网络中,节点在感知区域内能够感知的信息是由不同传感模块对所监测环境的监测数据,针对单独的某个节点,可以被抽象为一个 n 维的空间向量模型:

$$S = s(s_1, s_2, \dots, s_n) \quad (3)$$

其中, n 为正整数,代表不同种类的感知模块; s_n 为处理后的感知模块采集的数据。

举例说明向量空间模型的建立,如图 1 所示。其中,节点 S_A 和 S_B 为部署在监测区域内的感知节点,通过建立向量空间模型,将其感知信息映射到高维空间中的两点, A 和 B 分别为对应的映射到空间的向量。

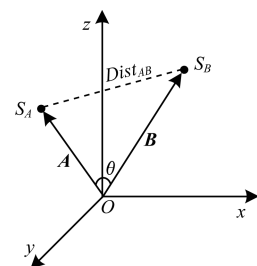


图 1 向量空间模型

本文通过建立向量空间模型, 克服了传统故障诊断算法计算量大的不足, 原因在于通过向量空间模型的建立, 将节点获取的信息映射到空间中的某一个点, 不同的感知节点在空间中将会被映射为不同的点, 可通过计算故障贡献率进行故障判定。

1.3 十字滑动窗口诊断

本文还引入了十字滑动窗口, 滑动窗口是一种流量控制技术^[13], 十字窗口则是在普通滑动窗口基础上增加了横向的滑动窗口。采用十字滑动窗口可以很好地克服过于依赖邻居节点数据、过于依赖自身历史数据、过于依赖大量数据集的缺点, 因此, 本文引入十字滑动窗口进行节点故障诊断。

从无线传感器网络所监测区域的信息可知, 节点获取的感知信息一般具有时空相关性, 若采用纵向滑动窗口对某一节点依据其历史数据进行故障判定, 弊端在于若感知数据为一固定值, 容易将故障节点误判为正常节点。若采用横向滑动窗口对节点以及邻居节点进行相似性比对, 判定故障与否, 其弊端在于当遇到某一节点与其邻居节点完全相似时, 不能准确判定节点是否一定为故障, 容易造成误判。

因此, 将横向与纵向相结合形成的滑动窗口, 称之为十字滑动窗口, 可以克服传统诊断算法以下缺点: 过于依赖邻居节点数据, 依赖自身历史数据, 以及依赖大量数据集。

本文将节点自身历史数据作为十字滑动窗口的纵向窗口, 邻居节点作为十字滑动窗口的横向窗口, 建立的十字滑动窗口示意图如图 2 所示。其中, * 表示感知环境中部署的感知节点; L, R 分别表示以某感知节点 i 为中心、左侧和右侧的邻居节点序列, 窗口的横向大小为邻居节点的数目, U, D 分别表示以某感知节点 i 为中心, 窗口的纵向大小为节点 i 在从 t_1 时刻到 t_2 时刻到来的历史数据的数目。

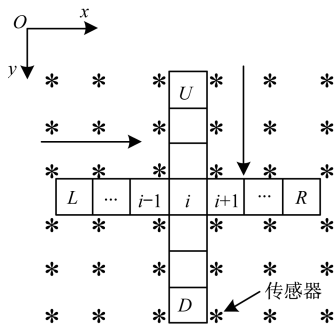


图 2 十字滑动窗口

十字滑动窗口移动顺序: 以某感知节点 i 为中心, 以队列的形式进行, 横向从左至右, 纵向从上至

下进行扫描, 从原队列中删除原始向量, 插入新向量。经过出队和进队的操作, 完成队列更新的同时进行向量更新。

十字滑动窗口具体操作: 将十字滑动窗口的中心向量, 依次与 4 个方向上的所有向量做故障贡献率的计算。

将十字滑动窗口 4 个方向 (L, R, U, D) 上的故障权重分别自定义为 25%, 表示如下:

$$Weight_{\text{fault}} = \begin{cases} 25\%, Dir_L \\ 25\%, Dir_R \\ 25\%, Dir_U \\ 25\%, Dir_D \end{cases} \quad (4)$$

根据 4 个方向上的故障权重, 累加各个方向上故障贡献率的值, 若超过阈值, 表明十字滑动窗口的中心向量异常, 如果向量的每一维都出现异常, 则表明该节点故障; 如果向量的某一维异常, 则表明产生该维数据的传感器异常, 然后根据不同故障类型对应的阈值范围表给出故障类型诊断结果。

2 实验设计

2.1 实验数据及实验环境

实验采用的感知数据集来源为英特尔伯克利实验室, 该实验室将 54 个 Mica2Dot 传感器节点部署在不同房间以及不同位置, 采集感知数据^[14], 如图 3 所示。

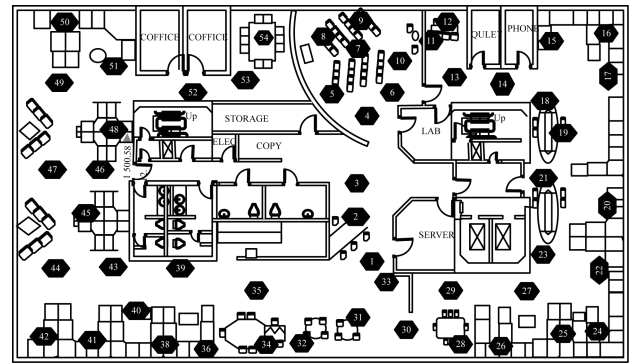


图 3 实验室节点部署分布图

该节点的感知模块包括温度传感器、湿度传感器、光照传感器。每 31 s 有一次湿度、温度、光和电压值的更新变化, 实验运行时间为 2004 年 2 月 28 日—2004 年 4 月 5 日, 共计 37 d。

实验算法实现语言为 C++, 开发环境为 Visual Studio 2013。

2.2 衡量指标

数据预处理中去除噪声效果评价指标分为信噪比 SNR 和平滑度 SR ^[15]。

信噪比为测量信号中原始信号与去噪后的信号

之间的方差的平方根,计算公式如下:

$$SNR = 10 \times \lg \left[\frac{\sum_{i=1}^N S_i'^2}{\sum_{i=1}^N (S_i' - S_i)^2} \right] \quad (5)$$

平滑度为去噪后信号差分数的方差根与原始信号差分数的方差根之比,计算公式如下:

$$SR = \frac{\sum_{i=1}^{N-1} (S_{i+1} - S_i)^2}{\sum_{i=1}^{N-1} (S_{i+1}' - S_i')^2} \quad (6)$$

其中, S_i 为原始信号; S_i' 为去噪之后的信号。

故障诊断正确率定义为诊断出故障节点的数目与网络中实际存在故障的节点数目的比值的百分数,计算公式为:

$$f_{FDA} = \frac{N_{\text{fault}}}{N_{\text{total fault}}} \times 100\% \quad (7)$$

故障诊断误警率指误诊为故障节点的数量与正常的传感器节点数量的比值的百分数,计算公式如下:

$$f_{FAR} = \frac{N_{\text{normal_fault}}}{N_{\text{normal}}} \times 100\% \quad (8)$$

2.3 实验流程

根据算法原理,故障诊断实验流程如图 4 所示。

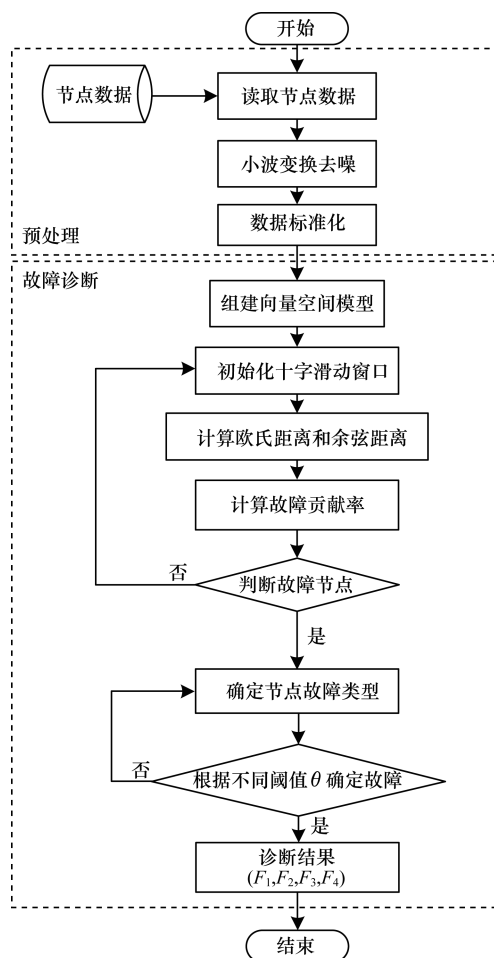


图 4 故障诊断流程

实验设计分为 2 个模块:数据预处理模块和故障诊断模块,故障诊断模块包括向量空间模型建立阶段、十字滑动窗口诊断阶段。

在数据预处理阶段,会将获取的节点信息进行数据噪声处理,对感知信号 S 采用两层分解,避免了过多分解把数据峰值变化作为信号噪声消除。

在向量空间模型建立阶段和十字滑动窗口诊断阶段,会将预处理后的信息先建立向量模型,依据伯克利实验室的数据源,建立向量空间模型。以感知节点序号为 1 为例,建立向量空间模型如下:

$$S_1 = s(s_1, s_2, s_3, s_4) \quad (9)$$

其中, s_i 表示温度传感器、湿度传感器、光照传感器以及电源电压的数值, $i \in (1, 2, 3, 4)$ 。

因此,采用 54 个 Mica2Dot 传感器节点,构建的向量空间模型可表示为:

$$(S_1, S_2, \dots, S_{54})^T \quad (10)$$

以图 3 中 7 号感知节点为例,建立十字滑动窗口,其邻居节点分别为 4, 5, 6, 8, 9, 10, 在某时间段 (t_1, t_2) 内的历史数据作为纵向窗口,邻居节点序列作为横向窗口,建立十字滑动窗口,根据故障贡献率以及自定义故障权重,判定 7 号节点的状态。

2.4 节点故障类型判定

经分析大量感知历史数据存在的故障特征,依照数据准确性将节点故障分为 4 种^[16]:传感器失效故障,偏移故障,恒偏差故障以及因传感器测量灵敏度下降和精度下降故障,表示为 $F_1 \sim F_4$ 。

本文结合以上分类方法,根据不同故障类型对应的阈值 θ 的范围进行故障类型确定,阈值 θ 的范围是由 2.2 节中提出的故障贡献率计算得出,如表 1 所示。

表 1 阈值的取值范围

阈值	取值范围
θ_1	$(2.047E-01, +\infty)$
θ_2	$(7.42E-02, 1.14E-01)$
θ_3	$(1.09E-02, 6.09E-02)$
θ_4	$(6.21E-02, +\infty)$

由故障类型的诊断结果得知,节点故障类型多为节点失效故障,其他故障类型出现的情况较少。

2.5 实验结果对比分析

在不同的节点故障概率下,将本文提出的基于数据驱动的 CW-VSM 算法与同属于基于数据驱动的经典 DFD 算法两者做了对比实验,对比结果如图 5 所示。

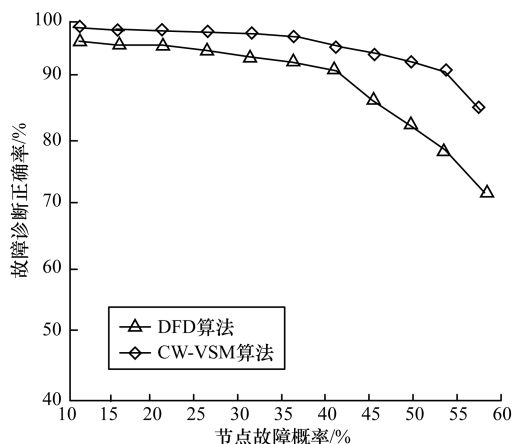


图5 2种算法的故障诊断正确率对比

由图5可知,DFD算法在节点故障概率为30%左右时,诊断精确度降幅较大,其原因可能为过度依赖邻居节点的状态,并且在确定是否为故障时,判定条件比较苛刻引起的。本文提出的CW-VSM算法,诊断精度变化较平稳,原因在于利用十字滑动窗口结合邻居节点比较的同时,将节点自身历史数据参与了比较。

在不同的节点故障概率下,将本文提出的基于数据驱动的CW-VSM算法与同属于基于数据驱动的经典DFD算法,两者做了故障误警率高低的对比实验,对比结果如图6所示。

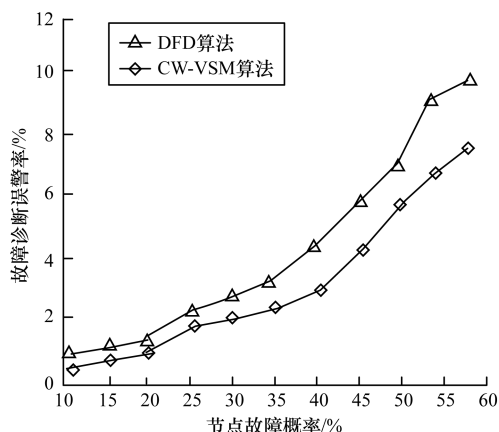


图6 2种算法的故障诊断误警率对比

由图6可知,DFD算法和本文的CW-VSM算法误警率均呈上升趋势,在节点故障概率为35%左右时,DFD算法对应的故障误警率变化幅度较大,其原因可能为诊断过程中忽略了节点本身历史状态。但是,CW-VSM算法对应的故障误警率较前者均表现为较低的趋势,原因在于通过节点历史数据和邻居节点数据综合判定,得到了较为准确的判定结果。

3 结束语

在无线传感器网络中,如何准确高效地诊断节点故障是近年来研究的热点,本文针对传统算法过于依

赖节点自身数据和计算量大等不足,提出一种基于数据驱动的故障诊断算法CW-VSM,通过建立高维空间向量、引入自身历史数据和邻居节点数据构造十字滑动窗口、窗口内计算向量相似度检测异常向量,达到准确诊断无线传感器节点故障的目的。实验结果表明,与DFD算法相比,本文算法在提高故障诊断正确率的同时,可将误警率控制在较低的范围。由于本文主要针对节点故障进行诊断,因此下一步将对网络故障的诊断进行研究。

参考文献

- [1] 孙利民. 无线传感器网络[M]. 北京:清华大学出版社,2005.
- [2] 任丰原,黄海宁,林 闯. 无线传感器网络[J]. 软件学报,2003,14(7):1282-1291.
- [3] MAO Xufei, MIAO Xin, HE Yuan, et al. CitySee: Urban CO2 Monitoring with Sensors [C]//Proceedings of IEEE INFOCOM'12. Washington D. C., USA: IEEE Press, 2012:1611-1619.
- [4] MAHAPATRO A, KHILAR P M. Online Distributed Fault Diagnosis in Wireless Sensor Networks [J]. Wireless Personal Communications, 2013, 71(3):1931-1960.
- [5] 杜 莹,程 普. 基于簇的分布式传感器故障检测算法[J]. 计算机工程, 2014, 40(2):110-113, 118.
- [6] MA Qiang, LIU Kebin, MIAO Xin, et al. Sherlock is Around: Detecting Network Failures with Local Evidence Fusion [J]. IEEE Transactions on Parallel and Distributed Systems, 2015, 26(5):1430-1440.
- [7] CHEN Jinran, KHER S, SOMANI A. Distributed Fault Detection of Wireless Sensor Networks [C]//Proceedings of 2006 Workshop on Dependability Issues in Wireless Ad Hoc Networks and Sensor Networks. New York, USA: ACM Press, 2006:65-72.
- [8] GONG Wei, LIU Kebin, LIU Yunhao. Directional Diagnosis for Wireless Sensor Networks [J]. IEEE Transactions on Parallel and Distributed Systems, 2015, 26(5):1290-1300.
- [9] 季 赛,袁慎芳,李含光. WSN中故障诊断性能与平均节点度研究[J]. 计算机工程, 2010, 36(7):14-16.
- [10] 高建良,徐勇军,李晓维. 基于加权中值的分布式传感器网络故障检测[J]. 软件学报, 2007, 18(5):1208-1217.
- [11] MIKOLOV T, CHEN Kai, CORRADO G, et al. Efficient Estimation of Word Representations in Vector Space [J/OL]. (2013-09-07). <https://arxiv.org/pdf/1301.3781v3.pdf>.
- [12] TURNEY P D, PANTEL P. From Frequency to Meaning: Vector Space Models of Semantics [J]. Journal of Artificial Intelligence Research, 2010, 37(1):141-188.
- [13] 常建龙,曹 锋,周傲英. 基于滑动窗口的进化数据流聚类[J]. 软件学报, 2007, 18(4):905-918.
- [14] Intel Lab. Berkeley Data Set [EB/OL]. [2016-04-05]. <http://berkeley.intelresearch.net/labdata/>.
- [15] 张德干,康学净,王 冬. 一种球坐标系下小波收缩去噪新算法[J]. 光电子·激光, 2012, 23(1):180-187.
- [16] NI K, RAMANATHAN N, CHEHADE M N H, et al. Sensor Network Data Fault Types [J]. ACM Transactions on Sensor Networks, 2009, 5(3):1-29.