

基于时间序列和神经网络的温室传感器节点故障诊断

王俊^{1,2} 刘刚^{1*}

(1. 中国农业大学 信息与电气工程学院/现代精细农业系统集成研究教育部重点实验室,北京 100083;
2. 河南科技大学 车辆与动力工程学院,河南 洛阳 471003)

摘要 温室无线传感器网络中故障节点会产生并传输错误数据,不仅消耗节点的能量和带宽,而且导致错误决策。针对此问题研究一种准确判断节点故障状态的方法。采用时序分析和遗传BP神经网络,建立基于时间序列和神经网络的传感器节点故障诊断系统,通过对传感器样本数据进行时序分析,提取模型参数作为特征向量,并以此对遗传BP神经网络进行网络训练,实现传感器节点故障的诊断。试验结果表明:该方法能够有效地识别传感器节点故障类型,15组测试样本的输出矢量与同类故障基准矢量的欧式距离和为0.007,识别正确率为100%。

关键词 传感器节点; 故障诊断; 时间序列; BP神经网络; 遗传算法

中图分类号 TP 277 文章编号 1007-4333(2011)06-0163-06 文献标志码 A

Fault diagnosis of greenhouse sensors nodes based on time series and neural network

WANG Jun^{1,2}, LIU Gang^{1*}

(1. College of Information and Electrical Engineering/Key Laboratory of Modern Precision Agricultural System Integration of Education Ministry, China Agricultural University, Beijing 100083, China;
2. College of Vehicle & Motive Power Engineering, Henan University of Science and Technology, Luoyang 471003, China)

Abstract The existence of faulty sensor nodes in wireless sensor networks (WSN) causes not only degradation of the network quality of service, but also a huge burden of the limited energy. Therefore, a method of accurate assessment about node failure status needs to be discussed. This paper adopts time series analysis and GA-BP neural network to establish a sensor nodes fault diagnosis system based on the combination of time series analysis and GA-BP neural network. With the aid of analyzing the sensor sample data and extracting model parameters as characteristic vectors, a net training process based on characteristic vector was used to build a GA-BP neural network to detect the sensor nodes faults and improve the diagnosis accuracy. The research result shows that the fault diagnosis method can effectively detect the sensor nodes failure type, the Euclidian Distance Sum of 15 test samples output vector and similar failure vector is 0.007 and the average recognition success rate is 100%.

Key words sensor nodes; fault diagnosis; time series; BP neural network; GA

无线传感器网络是一种全新的信息获取和处理技术,集成了无线通信、传感器和分布式信息处理等技术,将客观上的物理世界和逻辑上的信息世界相融合,具有良好的使用价值和应用前景。在应用过程中,传感器节点受制造工艺的限制和恶劣工作条

件的影响,不可避免地会发生故障。故障节点会产生错误的传感数据,使监控中心无法得到正确的检测信息,从而产生错误的决策,降低整个无线传感器网络的工作效率和服务质量。通过故障诊断准确地对节点状态给出判断,提高传感器的可靠性和数据

收稿日期: 2011-04-06

基金项目: 国家科技重大专项(2009ZX03001-019-02)

第一作者: 王俊,博士研究生, E-mail: wangjun_ujs@yahoo.com.cn

通讯作者: 刘刚,教授,博士生导师,主要从事精细农业研究, E-mail: pac@cau.edu.cn

的有效性,对于保证温室无线传感器网络正常运行具有重要意义。已有研究^[1-6]大多以网络运行状态为征兆,匹配典型特征来识别故障,但对于故障机理及表现形式难以明确辨识的复杂对象,如温室传感器节点,故障诊断比较困难。

本研究拟采用时间序列分析方法对温室传感器节点数据进行建模,提取数据动态特征,利用模型参数作为诊断样本训练 BP 神经网络,确定传感器数据与故障类型之间的非线性映射关系,应用遗传算法提高 BP 神经网络的学习收敛速度和分类效果,结合温度数据进行实证研究。

1 传感器节点故障模型

温室无线传感器网络由基站、簇头节点、传感器节点组成。传感器节点根据应用的需要布置在温室中某些区域,用于采集环境信息,包括温度、湿度、光照度、CO₂ 浓度等。网络拓扑结构见图 1。

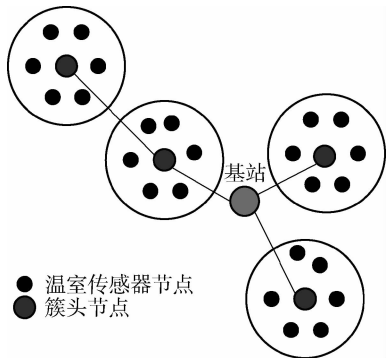


图 1 温室无线传感器网络结构

Fig. 1 Structure of greenhouse WSN

实际应用中,传感器节点是无线传感器网络中最容易出现故障的环节。传感器节点故障分为 2 类:硬故障和软故障。硬故障是指传感器节点的某一模块发生故障,以至不能和其他节点通信(如电池能量耗尽、电路故障、天线损坏和 CPU 模块故障等原因造成的无法通信);软故障是指传感器节点仍可以继续工作并与其它节点通信,但节点所感知的数据不正确,如辐射干扰、电源干扰、传感器故障等原因所引起的传感器误差。硬故障时传感器节点实际已退出无线传感器网络,更换或维修即可。本研究主要针对节点软故障。节点软故障所引起的传感器误差是指无故障时应输出的信号与故障时实际输出信号之间的偏差。按故障的表现可分为卡死故障、

恒增益故障、恒偏差故障、冲击故障^[7]。不失一般性,仅考虑 1 个传感器故障时的模型。

令 $S_{out}i(t)$ 为第 i 个传感器 t 时刻的实际输出, $S_{out}i'(t)$ 为其无故障状态 t 时刻应输出的信号, $i = 1, 2, \dots, m$ 。各故障类型模型如下。

1) 卡死故障。传感器的敏感器件失去作用,传感器的输出为一个常数 A :

$$S_{out}i(t) = A \quad (1)$$

2) 恒增益故障。复杂的工作环境使传感器的增益值发生较大的变化, β_i 为增益变化的比例系数:

$$S_{out}i(t) = \beta_i S_{out}i'(t) \quad (2)$$

3) 恒偏差故障。由于污染的原因,传感器产生慢漂移现象,在某一时间段内,其输出数据表现为恒偏差 Δ :

$$S_{out}i(t) = S_{out}i'(t) + \Delta \quad (3)$$

4) 冲击故障。传感器受外界干扰,在较短的时间内产生突变 D ,其中 $t = p$ 时, θ 值为 1, $t \neq p$ 时, θ 值为 0:

$$S_{out}i(t) = S_{out}i'(t) + D\theta(t) \quad (4)$$

2 故障诊断方法

基于时间序列和神经网络的温室传感器节点故障诊断方法可分为故障样本归纳、特征参数提取、神经网络训练和模式识别 4 个过程。

步骤 1: 由故障发生器产生随机故障,对传感器节点各种不同性质的故障以及不同故障程度的运行状态进行试验、统计归纳^[8],确定各类故障状态的样本。

步骤 2: 利用时间序列模型对故障样本和正常工作状态样本进行分析,提取模型参数。

步骤 3: 将各运行状态的模型参数作为特征矢量,对 BP 神经网络进行训练、计算,使网络实际输出与期望输出误差平方和达到最小,保存网络权值和偏差。

步骤 4: 把需要诊断的传感器节点运行状态的特征矢量输入 BP 神经网络,经计算得到网络输出结果,并将网络输出结果与标准输出向量作比较,判断传感器节点处于正常运行状态或哪种故障状态,从而实现传感器节点故障类型的识别。故障诊断系统结构见图 2。

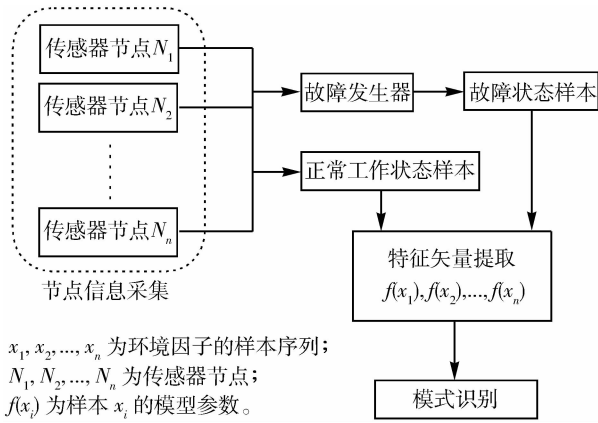


图 2 故障诊断系统结构

Fig. 2 Structure of fault diagnosis system

3 时间序列模型

时间序列建模是温室传感器节点故障诊断的关键,建模内容包括样本序列预处理、模型形式选取、模型参数估计、模型定阶等,其中核心是模型参数估计^[9-10]。

3.1 ARMA 模型表述

传感器节点数据 ARMA 模型可表述为:

$$x_t - \phi_1 x_{t-1} - \dots - \phi_p x_{t-p} = a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q} \quad (5)$$

式中: p 和 q 为模型的自回归阶数和移动平均阶数; θ 和 ϕ 是不为 0 的待定模型参数; a_t 为独立的白噪声误差项; x_t 是零均值化后的传感器节点数据时间序列。用 B 表示 k 阶延迟算子,即 $B^k x_t = x_{t-k}$, $B^k a_t = a_{t-k}$, 并令:

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \quad (6)$$

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q \quad (7)$$

则式(5)可表述为 $\phi(B)x_t = \theta(B)a_t$, 称为 p 阶自回归、 q 阶滑动平均模型,简称为 ARMA(p, q)模型。特殊地,若 $p=0$,模型称为滑动平均模型,记为 MA(q);若 $q=0$,模型称为自回归模型,记为 AR(p)。

3.2 样本序列建模及参数估计

1)样本序列预处理。对时间序列 $\{x_t\}$ 进行零均值化处理,得到零均值化后的序列: $\{x'_t\} = \{x_t - \bar{x}\}$, 其中 \bar{x} 是序列 $\{x_t\}$ 的平均值。零均值化后的样本序列往往表现出一定的趋势性,需对数据进行差分处理,形成新的序列: $\{x''_t\} = \{x'_{t+i} - x'_t\}$, 其中 i 为周期长度。

2)定阶。ARMA(p, q)模型建模过程中,首先要解决定阶问题,即估计 p, q 的值,预处理后的样本序列定阶采用 AIC 准则法。AIC 准则函数为:

$$AIC = 2k - 2L(\hat{\beta}) \quad (8)$$

式中: k 为独立参数; \hat{B} 为参数的最大似然估计值; $L(\cdot)$ 为似然函数。ARMA 模型似然函数近似为:

$$\left. \begin{aligned} L(\hat{\beta}) &= -\frac{n}{2} \lg 2\pi - \frac{n}{2} \lg \hat{\sigma}^2 - \frac{S(\hat{\beta})}{2\hat{\sigma}^2} \\ \hat{\sigma}^2 &= \frac{1}{n} S(\hat{\beta}) \\ \hat{\beta} &= (\hat{\varphi}, \theta)^T = (\varphi_1, \varphi_2, \dots, \varphi_p, \theta_1, \theta_2, \dots, \theta_q) \end{aligned} \right\} \quad (9)$$

结合式(8)和(9)得:

$$AIC(p, q) = n \lg \hat{\sigma}^2 + 2(p + q + 1) \quad (10)$$

ARMA(p, q)模型 AIC 定阶准则为:选则 p, q , 使式(10)值最小。

3)参数估计。ARMA 相关矩估计方法有最小二乘估计法、极大似然估计法、最大熵估计法等,本研究采用极大似然估计法计算 θ 和 ϕ 的值。

建模时选取温室温度作为研究对象,实验选用 1 个汇聚节点和 5 个传感器节点,由故障发生器产生随机故障模拟传感器节点 5 种工况(正常运行、卡死故障、恒增益故障、恒偏差故障、冲击故障),采样周期为 8 s,时间序列建模数据量为 50,试验进行 900 个周期,获取每种工况各 18 组时间序列样本。结合以上分析,采用 ARMA(3,1)对样本序列建模,选取自回归参数 ϕ_1, ϕ_2 和 ϕ_3 作为特征矢量。

4 传感器节点故障诊断模型设计

4.1 BP 神经网络模型

BP 神经网络是目前应用最为广泛的神经网络学习算法,可以实现输入和输出间的任意非线性映射,且具有一定的泛化功能,因而比较适合应用于模式识别和故障诊断^[11]。

选择 75 组特征矢量作为训练集用于网络训练,15 组特征矢量作为测试集用于检验分类效果,网络输入层为 3 个节点,分别输入 3 个自回归参数,中间层为 10 个节点,输出层为 5 个节点,分别表示 5 种状态,即网络拓扑结构为 3-10-5。用状态矢量 $\mu_1 = (1, 0, 0, 0, 0), \mu_2 = (0, 1, 0, 0, 0), \mu_3 = (0, 0, 1, 0, 0), \mu_4 = (0, 0, 0, 1, 0), \mu_5 = (0, 0, 0, 0, 1)$ 分别表示正常、卡死、横增益、横偏差、冲击 5 类状态的基准矢量。网络训练采用并行算法,将训练样本输入,调整网络的权重,使得网络输出与理想输出之间的误差

足够小到设定的阈值为止。训练结束以后,故障的特征信息就以神经网络连接权重的方式记录下来。

4.2 遗传 BP 神经网络模型

使用 BP 神经网络进行数据训练和分类验证时发现学习训练收敛速度较慢,分类效果不佳,无法达到期望效果。鉴于遗传算法全局性搜索的特点,采用遗传算法优化 BP 神经网络,寻找最优的网络连接权重。遗传算法优化 BP 网络步骤:

- 1) 初始化种群 p , 包括交叉规模、交叉概率 c 、突变概率 m 以及各层之间的权值的初始化, 权重采用实数编码, 初始种群数取 50。
- 2) 计算每个个体评价函数, 并将其排序。可按式(11)概率值选择网络个体(轮盘赌选择法)。

$$p_s = f_i / \sum_{i=1}^N f_i \quad (11)$$

式中 f_i 为个体 i 的适应度。

- 3) 以概率 c 对个体 G_i 和 G_{i+1} 交叉操作产生新的个体 G'_i 和 G'_{i+1} , 没有进行交叉操作的个体直接复制。
- 4) 利用概率 m 突变产生 G_j 的新个体 G'_j 。
- 5) 将新个体插入到种群 p 中, 并计算新个体的评价函数。
- 6) 如果找到满意的个体, 则结束, 否则转 3)。
- 7) 达到所要求的性能指标后, 将最终群体中的最优个体解码即可达到优化后的网络连接权重系数。

遗传 BP 算法训练网络的误差平方和曲线、适应度曲线见图 3。可见, 遗传 BP 算法在收敛速度和

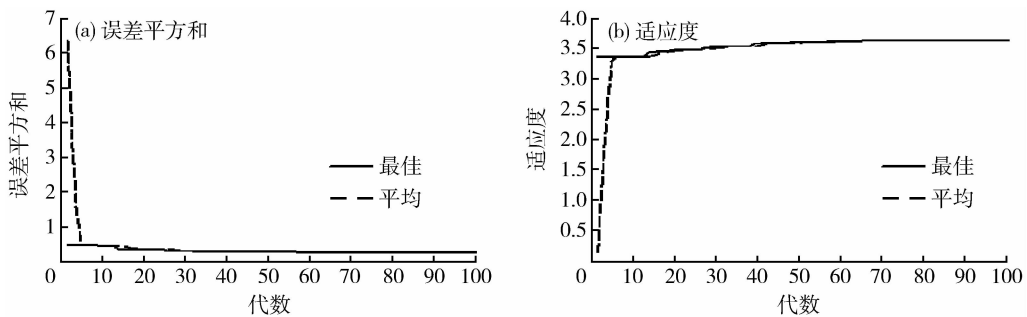


图 3 遗传 BP 算法训练网络的误差平方和与适应度曲线

Fig. 3 Curves of sum squared error and fitness by obtained Genetic BP algorithm

误差精度方面效果明显, 收敛精度非常好; 在进化初期, 适应度提高明显, 经过大约 60 代的搜索后染色体的平均适应度趋于稳定。

5 试验及结果分析

为了比较遗传 BP 网络和 BP 网络的识别性能, 取未参加训练的 15 组状态参数(表 1)进行故障识别。仿真设定: 遗传算法的参数为 $P=50$, 权重初始化空间范围 $[0, 1]$, 交叉率 $c=0.4$, 变异率 $m=0.1$, 最大进化代数 $N=100$; BP 算法参数为动量项系数 $M=0.9$, 学习率 $l=0.01$, 训练误差目标 0.001。训练收敛情况见图 4。训练输出结果见表 2。分析可知:

- 1) 网络经初始寻优后, 初始权值达到最优, 在此基础上网络能准确、全面、快速地表示传感器节点故障诊断的知识, 遗传 BP 网络诊断结果与实际故障状况十分相符, 其准确度明显高于 BP 网络。

2) 由图 4 可以看出, 遗传 BP 网络及 BP 网络的

表 1 待检验传感器状态参数

Table 1 Sensor state parameters to be tested

序号	自回归参数			状态类别
	ϕ_1	ϕ_2	ϕ_3	
1	0.153 0	0.842 5	0.043 6	正常
2	0.406 7	0.900 7	0.355 1	正常
3	0.176 9	0.566 4	0.052 2	正常
4	0.437 0	0.369 3	-0.062 4	卡死
5	0.495 0	0.359 6	-0.034 2	卡死
6	0.837 6	0.065 5	-0.112 0	卡死
7	1.917 0	-0.878 3	-0.044 6	横增益
8	1.917 1	-0.927 3	0.003 5	横增益
9	1.923 9	-0.932 1	0.001 9	横增益
10	0.057 8	0.767 5	-0.055 7	横偏差
11	0.076 1	0.621 0	0.016 4	横偏差
12	0.023 8	0.709 2	-0.042 9	横偏差
13	0.937 1	-0.002 5	-0.061 1	冲击
14	0.925 0	0.001 7	-0.060 3	冲击
15	0.944 9	0.003 0	-0.005 5	冲击

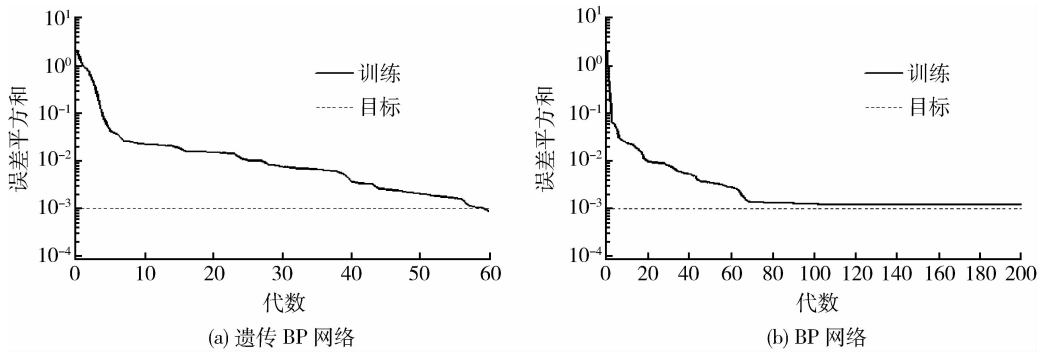


图 4 遗传 BP 网络与 BP 网络的训练误差收敛曲线

Fig. 4 Curves of training error based on GA-BP network and BP network

表 2 遗传 BP 算法与 BP 算法的网络训练识别结果

Table 2 Output result of the network identification by Genetic BP algorithm and BP algorithm

序号	遗传 BP 算法					BP 算法					网络理想输出
	a_1	a_2	a_3	a_4	a_5	a_1	a_2	a_3	a_4	a_5	
1	0.996 3	0.000 0	0.000 0	0.001 8	0.000 0	0.992 0	0.000 0	0.000 0	0.003 6	0.000 0	(1,0,0,0,0)
2	1.000 0	0.000 0	0.000 0	0.000 0	0.000 9	1.000 0	0.000 0	0.000 0	0.011 1	0.002 9	(1,0,0,0,0)
3	0.999 7	0.000 0	0.000 0	0.000 1	0.000 0	1.000 0	0.000 0	0.000 0	0.000 0	0.000 0	(1,0,0,0,0)
4	0.000 0	1.000 0	0.000 0	0.000 0	0.000 0	0.000 1	0.999 9	0.000 0	0.0011	0.000 0	(0,1,0,0,0)
5	0.001 3	0.999 5	0.000 0	0.000 0	0.000 0	0.016 4	0.994 0	0.000 0	0.000 4	0.000 0	(0,1,0,0,0)
6	0.000 0	0.997 5	0.000 0	0.000 0	0.001 9	0.000 0	0.999 5	0.000 0	0.000 0	0.002 6	(0,1,0,0,0)
7	0.000 0	0.000 0	0.999 9	0.000 0	0.000 1	0.000 0	0.000 0	0.999 7	0.000 0	1.000 0	(0,0,1,0,0)
8	0.000 1	0.000 0	0.999 9	0.000 0	0.000 2	0.000 0	0.000 0	0.999 7	0.000 0	1.000 0	(0,0,1,0,0)
9	0.000 1	0.000 0	0.999 9	0.000 0	0.000 2	0.000 0	0.000 0	0.999 7	0.000 0	1.000 0	(0,0,1,0,0)
10	0.000 0	0.000 0	0.000 0	1.000 0	0.000 0	0.000 0	0.000 0	0.000 0	1.000 0	0.000 0	(0,0,0,1,0)
11	0.000 0	0.000 0	0.000 0	0.999 9	0.000 0	0.000 2	0.000 0	0.000 0	0.992 5	0.000 0	(0,0,0,1,0)
12	0.000 0	0.000 0	0.000 0	1.000 0	0.000 0	0.000 0	0.000 0	0.000 0	1.000 0	0.000 0	(0,0,0,1,0)
13	0.000 0	0.000 3	0.000 3	0.000 0	0.998 9	0.000 0	0.004 7	0.000 1	0.000 0	0.996 2	(0,0,0,0,1)
14	0.000 0	0.000 3	0.000 2	0.000 0	0.998 4	0.000 0	0.007 1	0.000 1	0.000 0	0.995 0	(0,0,0,0,1)
15	0.000 2	0.000 0	0.000 4	0.000 0	0.999 9	0.000 0	0.000 0	0.000 6	0.000 0	1.000 0	(0,0,0,0,1)

注： a_1, a_2, \dots, a_5 为网络输出向量的元素。

前期收敛效果相近；在后期 BP 网络收敛很慢，几乎停滞不前，需经较长步数后才能达到收敛精度；而遗传 BP 网络能 60 步时迅速得到最优解，比 BP 网络的收敛速度快得多。

3) 从表 2 的训练结果可以看出，测试样本经遗传 BP 网络识别的输出矢量与同类故障基准矢量的

距离都接近于 0，而与不同类故障的基准矢量的距离都比较大；遗传 BP 网络和 BP 网络与标准矢量的欧式距离的和分别为 0.007 0 和 3.042 6，表明遗传 BP 网络能很好的进行传感器节点故障诊断，识别精度高，而传统 BP 网络识别精度较低，尤其无法识别故障状态 μ_3 。

6 结 论

本研究利用时间序列 ARMA(3,1)建模,将传感器节点数据的自回归参数 ϕ_1 、 ϕ_2 和 ϕ_3 作为特征矢量训练 BP 网络,利用实数编码的遗传算法优化 BP 神经网络,为利用时间序列和神经网络诊断传感器节点故障提供了理论依据,主要结论如下:

1)时间序列用于传感器节点故障诊断,可充分利用时间序列分析的优越性,提取传感器节点各种状态下的特征参数,为多参量融合的传感器节点故障诊断提供条件。

2)遗传算法具有很强的宏观搜索能力,且能以最大的概率找到全局最优解。仿真表明,将遗传算法和 BP 网络两者结合,能达到优化网络的目的,训练误差下降速度很快,60 步时就达到设定的网络误差值。

3)将时间序列和遗传 BP 网络用于传感器节点故障诊断,通过对 75 组样本进行训练、15 组样本识别与分析,诊断结果区分明显,表明该方法能深入挖掘数据中有关故障信息,在故障模式识别中具有较高的辨识度。

参 考 文 献

[1] 季塞,袁慎芳,马廷准,等.无线传感器网络中节点故障诊断方

法的研究[J].计算机工程与应用,2010,46(23):95-97

- [2] 雷霖,代传龙,王厚军.基于 Rough set 理论的无线传感器网络节点故障诊断[J].北京邮电大学学报,2007,30(4):69-73
- [3] 蒋鹏.一种改进的 DFD 无线传感器网络节点故障诊断算法研究[J].传感技术学报,2008,21(8):1417-1421
- [4] Chessa S,Santi P. Crash faults identification in wireless sensor networks[J]. Computer Communications,2002,25:1273-1282
- [5] Sheth A,Hartung C,Han R. A decentralized fault diagnosis system for wireless sensor networks[C]//IEEE International Conference on Mobile Adhoc and Sensor Systems Conference, Washington,DC,2005:192-194
- [6] Chatzigiannakis V,Papavassiliou S. Diagnosing anomalies and identifying faulty nodes in sensor networks [J]. IEEE Sensor Journal,2007,7(5):637-645
- [7] 张荣标,蔡兰,王贵成.基于时间序列和专家系统的 pH 值传感器故障诊断的研究[J].电子测量与仪器学报,2001,15(3):30-35
- [8] 白斌.基于时空序列分析的温室测控系统 WSN 故障诊断[D].镇江:江苏大学,2009
- [9] 朱建元.基于 BP 神经网络与时间序列分析的柴油机故障诊断[J].上海海事大学学报,2006,27(4):22-27
- [10] 张树京,齐立心.时间序列分析简明教程[M].北京:清华大学出版社,2003
- [11] 周开利,康耀红.神经网络模型及其 MATLAB 仿真程序设计[M].北京:清华大学出版社,2003

(责任编辑:刘迎春)