

无线传感器节点的故障诊断研究

尚兴宏^{1,2} 尚曦乐³ 章静³ 钱焕延¹

(南京理工大学计算机科学与技术学院 南京 210094)¹ (上海华为技术有限公司 上海 201206)²
(上海金洲小学 上海 200333)³

摘要 无线传感器网络节点发生故障不仅消耗节点的能量和网络带宽,甚至会造成网络瘫痪。在分析无线传感器网络节点故障类别的基础上,分别使用相关向量机、支持向量机等算法对其进行研究,并用节点的特征值及相应的故障类型训练相关向量机及支持向量机的分类器。仿真结果表明,相关向量机比支持向量机和人工神经网络有更高的诊断精度。

关键词 无线传感器网络,故障诊断,相关向量机,支持向量机,人工神经网络

中图分类号 TP393 文献标识码 A

Research on Fault Diagnosis of Wireless Sensor Nodes

SHANG Xing-hong^{1,2} SHANG Xi-yue³ ZHANG Jing³ QIAN Huan-yan¹

(School of Computer Science and Technology, Nanjing University of Science and Technology, Nanjing 210094, China)¹

(Huawei Technologies Co., Ltd., Shanghai 201206, China)² (Jinzhou Primary School, Shanghai 200333, China)³

Abstract In Wireless Sensor Networks(WSNs), faulty sensor nodes may consume the limited energy and bandwidth of network. Furthermore, they will produce the paralysis of entire systems. In this paper, RVM and SVM algorithm was applied to fault diagnosis for sensor nodes based on the analysis of fault type of WSNs nodes, and then the values of the features and the corresponding fault types of wireless sensor nodes were used to train RVM and SVM classifier. Simulation results show that the diagnosis results of the RVM model for wireless sensor are better than those of SVM and ANN.

Keywords Wireless sensor networks(WSNs), Fault diagnosis, Relevance vector machine(RVM), Support vector machine(SVM), Artificial neural networks(ANN)

1 引言

无线传感器网络是一个能量、计算能力、资源有限的自组织网络,WSN节点常常被部署在一些不可控甚至十分危险的环境中^[1]。一旦节点被部署后,往往很难再对节点补充能量,或者进行维修管理等工作。节点一旦失效或者出现故障,可能会导致测量错误,使节点所具有的功能降低或消失,甚至将会导致整个网络的瘫痪。为了保证网络的服务质量并延长网络的生命周期,进行无线传感器网络的节点故障诊断显得尤为重要。近年来,无线传感器的故障诊断研究已成为一个新的热门课题。

相关向量机(relevance vector machine, RVM)^[2]是Tipping于2001年提出的一种与支持向量机(support vector machine, SVM)^[3]相似的稀疏概率模型,它的训练过程是在贝叶斯框架下进行的^[4,5]。RVM可用于回归估计预测,从而获得预测值的分布,不仅具有出色的非线性拟合和泛化能力,而且弥补了SVM需要估计正则化参数、无法得到基于概率的预测结果的缺陷^[6]。RVM在很多领域都得到了成功的应用,

现正逐渐成为机器学习界的研究热点。小波分析具有良好的局部化和多尺度特性,并且在函数逼近、回归分析和数据去噪等领域受到了越来越多的关注^[7]。

本文在分析无线传感器节点故障类别的基础上,分别选取60、120、180及240个节点,并将相关向量机和支持向量机应用于其故障诊断中,同时与人工神经网络算法进行了比较。结果显示,相关向量机比支持向量机和人工神经网络具有更高的诊断精度。

2 节点常用故障

一个无线传感器网络由很多随机地分布在检测区域的传感器节点构成,当WSN发生故障时,一般有两种情况,其一是用户无法接收到数据,其二是用户接收到的不是真实的测量数据。对于第一种情况,发生故障的原因往往是一个或几个节点的模块部件如电源单元模块、通信模块等发生了故障;而对于第二种情况,往往是因为某个节点的传感部件发生故障导致。所以,WSN的故障总是由其相关的节点的故障造成的,而节点的故障也是由其各个功能部件构成的。因此对节

本文受国防科工委应用基础资金资助。

尚兴宏(1975—),男,博士,高级工程师,主要研究方向为无线传感器网络技术,E-mail:shang_xh@hotmail.com;钱焕延 男,教授,博士生导师,主要研究方向为现代通信技术、网络技术。

点进行故障诊断的本质就是对节点的各个部件进行故障诊断。

目前在实验中出现较多的主要是冲击故障、偏置故障、短路故障、漂移故障^[8]。从信号角度来看,其主要表现形式为:

- (1)冲击故障,传感部件的随机干扰等会造成短时间的脉冲波;
- (2)偏置故障,会造成信号与原信号有一固定的偏差;
- (3)短路故障,内部短路,信号接近于0;
- (4)漂移故障,以一定的速率偏移原信号。

3 算法理论

3.1 支持向量机

支持向量机(Support Vector Machine, SVM)方法是一种强有力的机器学习方法,现已广泛应用于模式识别等方面。其建立在统计学习理论的 VC 维理论和结构风险最小化(structural risk minimization, SRM)准则基础上,根据有限的样本信息,在模型的复杂性(即对特定训练样本的学习精度)和学习能力(即无错误地识别任意样本的能力)之间寻求最佳折衷,以期获得最好的推广能力^[9]。

假设大小为 n 的训练样本集为 $\{x_i, y_i\}_{i=1}^n, \{x_i\} \subset R^n$ 。这里 x_i 表示样本输入数据, y_i 为相应的预测值。支持向量机即寻找最优的分类超平面 $wx + b = 0$, w 为权重矢量, b 为偏差。

最优分类超平面满足下列等式(1)的条件:

$$\begin{cases} \min \frac{\|w\|^2}{2} \\ \text{subject to } y_i(wx_i + b) \geq 1, \quad i=1, 2, \dots, n \end{cases} \quad (1)$$

该条件可以通过求解拉格朗日泛函的鞍点给出:

$$L(w, b, a_i) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^n a_i [(wx_i + b)y_i - 1] \quad (2)$$

式中, a_i 为拉格朗日乘子。依据拉格朗日对偶理论将其转化为对偶问题:

$$\begin{cases} \max W(a) = \sum_{i=1}^n a_i - \frac{1}{2} \sum_{i,j=1}^n a_i a_j y_i y_j (x_i x_j) \\ \text{subject to } \sum_{i=1}^n a_i y_i = 0, a_i \geq 0 \end{cases} \quad (3)$$

式(3)通过二次规划求解,得到最优的 w^* 和 b^* 为:

$$\begin{cases} w^* = \sum_{i=1}^n a_i^* x_i y_i \\ b^* = -\frac{1}{2} w^* (x_r + x_s) \end{cases} \quad (4)$$

式中, x_r 和 x_s 为两个类别中任意的一对支持向量。

最终得到的最优分类函数为:

$$f(x) = \text{sgn} \left[\sum_{i=1}^n a_i^* y_i (x x_i) + b^* \right] \quad (5)$$

如果训练集是不可分的,或者事先并不知道其是否线性可分,希望找到一个最优的超平面,使之在训练集的分类误差的概率达到最小,通常引入松弛变量 ξ_i ,对式(1)的优化目标及约束项进行修正,可得:

$$\begin{cases} \min \frac{\|w\|^2}{2} + C \sum_{i=1}^n \xi_i \\ \text{subject to } y_i(wx_i + b) \geq 1 - \xi_i, \xi_i > 0, i=1, 2, \dots, n \end{cases} \quad (6)$$

式中, C 为惩罚因子,是对分类错误的惩罚,用于调整置信范围和经验误差的平衡。

同样可将其转化为对偶问题,最大化目标函数为:

$$\max W(a) = \sum_{i=1}^n a_i - \frac{1}{2} \sum_{i,j=1}^n a_i a_j y_i y_j K(x_i x_j) \quad (7)$$

约束条件为:

$$\begin{cases} \sum_{i=1}^n a_i y_i = 0 \\ 0 \leq a_i \leq C, i=1, 2, \dots, n \end{cases} \quad (8)$$

式中, a_i, a_j 表示拉格朗日因子, $k(x_i, x_j)$ 表示核函数,可以表示为:

$$k(x_i, x_j) = \phi(x_i) \cdot \phi(x_j) \quad (9)$$

最终最优分类函数为:

$$f(x) = \text{sgn} \left[\sum_{i=1}^n a_i^* y_i K(x_i, x) + b^* \right] \quad (10)$$

3.2 相关向量机

通过应用广义线性模型将 RVM 的训练集似然函数定义为:

$$p(t|w) = \prod_{n=1}^N y(x_n; w)^n (1 - y(x_n; w))^{1-t_n} \quad (11)$$

其中:

$$y(x_n; w) = \sigma(w' \phi(x_n)) = \frac{1}{1 + \exp(-w' \phi(x_n))} \quad (12)$$

式中, $\sigma(x) = \frac{1}{1 + \exp(-x)}$ 为 sigmoid 函数。

多核相关向量机的优化偏移 b^* 通过式(13)获得:

$$b^* = y_k - \sum_{i=1}^n a_i^* y_i \left(\sum_{s=1}^M \mu_s^* k_s(x_k, x_i) \right) \quad (13)$$

基于多核相关向量机 RVM 的分类器的决策函数可以表示为:

$$f(x) = \text{sign} \left(\sum_{i=1}^n a_i^* y_i \left(\sum_{s=1}^M \mu_s^* k_s(x, x_i) \right) + b \right) \quad (14)$$

相关向量机(RVM)通过预先选择的核函数将输入空间变换到一个高维空间,并在这个空间实现数据的线性化。设 $\psi(x_i)$ 是一个小波母函数,则可以构造平移不变形式的小波核函数为^[10]:

$$k(x, x') = \prod_{i=1}^L \psi \left(\frac{x_i - x_i'}{a} \right) \quad (15)$$

式中, L 为输入向量的维数, a 为小波核函数的伸缩因子,是需要优化选取的常量。

Mexican hat 小波母函数见式(16):

$$\psi(x) = (1 - x^2) \exp \left(-\frac{x^2}{2} \right) \quad (16)$$

相应的小波核函数为:

$$k(x, x') = \prod_{i=1}^L \left(1 - \frac{\|x_i - x_i'\|^2}{a^2} \right) \exp \left(-\frac{\|x_i - x_i'\|^2}{2a^2} \right) \quad (17)$$

在训练相关向量机过程中,小波核函数的伸缩因子 a 是需要优化的量,其值的选取直接影响到故障诊断的性能。本文用交叉验证方法寻找最佳的参数,然后利用最佳的参数来训练模型。具体地,可将样本集分为 10 份,其中 9 份作为训练数据集,而另外的 1 份作为验证数据集。用验证集来计算均方根误差(RMSE),最小的均方根误差对应的就是最优参数^[11]。RMSE 定义为:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (t_i - \mu_i^*)^2}{n}} \quad (18)$$

式中, t_i 表示实验值, μ_i^* 为预测值, n 为样本数。

4 仿真实验及性能分析

4.1 节点故障模型及仿真

随机选取 60、120、180、240 个传感器节点,分别使用相关

向量机、支持向量机算法进行仿真诊断,并与人工神经网络算法进行诊断正确性比较。由于主要对节点的传感部件进行仿真,因此采用参考文献[8]所使用的 Simulink 仿真方法。对每次仿真的输出结果进行采样,其中故障发生之前采样 3 次,故障发生后采样 7 次。采样获取到的冲击故障、偏置故障、短路故障和漂移故障节点样本的信号输出分别如图 1—图 4 所示。

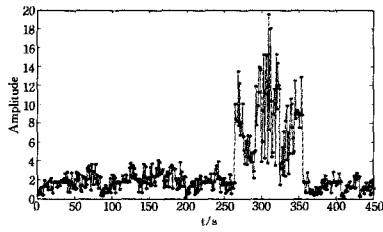


图 1 无线传感器节点冲击故障输出信号

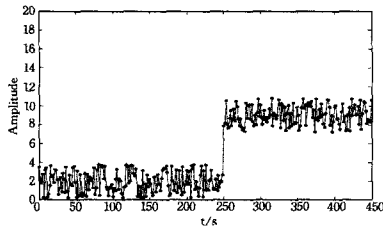


图 2 无线传感器节点偏置故障的输出信号

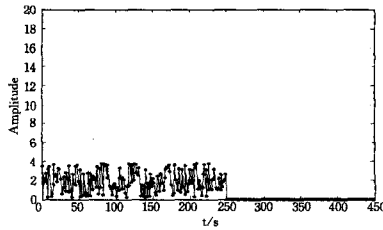


图 3 无线传感器节点短路故障的输出信号

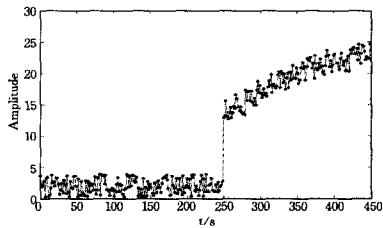


图 4 无线传感器节点漂移故障的输出信号

由于特征值的取值不在同一个数量级,输入变量差异较大,因此为了提升诊断效果,需要进行归一化处理:

$$x_i' = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}}$$

式中, x_i 是样本节点 i 归一前的数据, x_{\min} 是样本数据中的最小值, x_{\max} 是样本数据中的最大值, x_i' 是样本节点 i 归一化后的数据。

4.2 分类器选择

在进行无线传感器故障诊断过程中,使用 4 个分类器来区别正常节点和 4 种故障节点^[12],具体如表 1 所列。

CLASS1 分类器用于训练将正常节点从训练集样本中区分出来,如果一个样本节点为正常节点,则 CLASS1 的输出值设置为 +1,否则设置为 -1;在测试过程中,CLASS1 的输

出为 +1,则表示样本节点处于正常状态,如果其值输出为 -1,则表示样本节点为故障节点(即短路、偏置、漂移和冲击故障之一)。

表 1 分类器及节点状态

分类器	输出	节点状态
CLASS1	+1	正常
	-1	冲击 偏置 短路 漂移
CLASS2	+1	冲击
	-1	偏置 短路 漂移
CLASS3	+1	偏置
	-1	短路 漂移
CLASS4	+1	短路
	-1	漂移

CLASS2 分类器用于训练将冲击故障节点从剩余的训练集样本(冲击、偏置、短路和漂移故障样本)中区分出来,如果一个样本节点为冲击故障节点,则 CLASS2 的输出值设置为 +1,否则设置为 -1;在测试过程中,当 CLASS2 的输出为 +1,则表示样本节点处于冲击故障状态,如果其值输出为 -1,则表示样本节点发生其他故障(即偏置、短路和漂移故障之一)。

CLASS3 分类器用于训练将偏置故障节点从剩余的训练集样本(偏置、短路和漂移故障样本)中区分出来,如果一个样本节点为偏置故障节点,则 CLASS3 的输出值设置为 +1,否则设置为 -1;在测试过程中,当 CLASS3 的输出为 +1,则表示样本节点处于偏置故障状态,如果其值输出为 -1,则表示样本节点发生其他故障(即短路和漂移故障之一)。

CLASS4 分类器用于训练将短路故障节点从剩余的训练集样本(短路和漂移故障样本)中区分出来,如果一个样本节点为短路故障节点,则 CLASS4 的输出值设置为 +1,否则设置为 -1;在测试过程中,当 CLASS4 的输出为 +1,则表示样本节点处于短路故障状态,如果其值输出为 -1,则表示样本节点发生漂移故障。

这样通过 4 个 CLASS 分类器的训练可以有效地从训练集样本中区分出正常节点、冲击故障、偏置故障、短路故障和漂移故障样本。

5 诊断结果分析

相关向量机、支持向量机及人工神经网络对无线传感器的样本故障仿真诊断研究详细结果如表 2 所列,3 种算法诊断结果比较如图 5 所示。

相关向量机的训练是在贝叶斯框架下进行的,它能产生具有较高检测精度的决策函数,同时提高了系统的实时性。本文选择 Mexican hat 小波作为其核函数,其仍具有较高的诊断结果。从表 2 可见,其平均诊断精度达到 95.6%。

支持向量机的诊断结果是基于一种统计学习理论的新的机器学习方法,其采用结构风险最小化原则在最小化样本点误差的同时最小化结构风险,提高了模型的泛化能力,因此其也有较高的精度。本文选择 RBF 高斯核函数为 SVM 的核函数,利用交叉验证方法需找其最佳的惩罚因子 c 参数和核函数宽度 σ 。虽然其具有和相关向量机相似的决策函数,但由于不能很好地实现概率预测,因此其精度与上述 RVM 模型比有差距。从表 2 可见,其平均诊断精度为 92.4%。

(下转第 343 页)

(1)分析了当前 ICS 的安全风险,并由此引出了工业控制系统信息安全审计系统。

(2)对 ICS 信息安全审计系统进行需求分析,分析了 ICS 信息安全现状及需求,并对工业控制系统用户的需求进行了详细分析。

(3)针对 ICS 信息安全现状以及系统特点,有针对性地设计了一套 ICS 信息安全审计系统,该系统能够极大地提高 ICS 的防护能力,并能提供十分丰富的审计报告。

(4)ICS 信息安全审计系统的实现。研究了系统的部署方法,并通过了相关权威机构检验。该系统在第三方权威机构的检验测试结果表明,该系统能够对工业控制系统进行全方位的安全审计,能够极大地提高工业控制系统安全防护能力。

ICS 信息安全不是一个单纯的技术问题,而是一个从意识培养开始,涉及到管理、流程、架构、技术、产品等各方面的系统工程。ICS 信息安全更是一个动态过程,需要在整个工

业基础设施生命周期的各个阶段中持续实施,不断改进。为了进一步提高工业控制系统安全防护能力,下一步我们将进一步开展 ICS 相关防御技术研究。

参考文献

[1] IEC 62443-2-1 ED. 1.0 EN;2010,Industrial communication networks - Network and system security - Part 2-1;Establishing an industrial automation and control system security program[Z]. 2010

[2] 石勇,刘巍伟,刘博.工业控制系统(ICS)的安全研究[J].网络安全技术与应用,2008(4)

[3] GB 17859-1999. Classified criteria for security protection of computer information system[S]. 1999

[4] 张帅.工业控制系统安全现状与风险分析——ICS 工业控制系统安全风险分析之一[J].计算机安全,2012(1)

[5] 王文宇,刘玉红.工控系统安全威胁分析及防护研究[J].信息安全与通信保密,2012(2)

(上接第 329 页)

表 2 诊断结果

诊断算法	节点数	诊断错误数	诊断正确率%	平均正确率%
RVM- Mexican	60	3	95.0	95.6
	120	5	95.8	
	180	8	95.6	
	240	10	95.8	
SVM-RBF	60	5	91.7	92.4
	120	8	93.3	
	180	14	92.2	
	240	18	92.5	
ANN-RBF	60	10	83.3	84.2
	120	19	84.1	
	180	28	84.4	
	240	36	85.0	

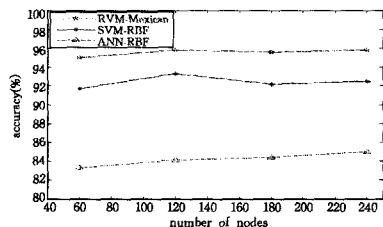


图 5 RVM,SVM 及 ANN 诊断结果比较

神经网络是基于传统统计学的基础上的,而传统统计学主要研究样本数据趋于无穷多时的统计性质,因此较上述算法有较弱的对策泛化能力。从表 2 可见,其平均诊断精度为 84.2%。

从图 5 可见,在无线传感器网络不同样本数下,诊断算法的诊断精度一致表现为:RVM-Mexican>SVM-RBF>ANN-RBF。结果说明,对无线传感器节点的诊断使用 RVM-Mexican 算法会明显好于 SVM-RBF 和 ANN-RBF。

结束语 本文将相关向量机及支持向量机算法应用于无线传感器网络的节点故障诊断研究,在分析传感器节点故障类别的基础上以 4 个分类器区别正常节点及 4 类故障节点,并与人工神经网络等算法进行了比较。仿真实验表明,相关向量

机算法比支持向量机和人工神经网络有更高的诊断精度。

参考文献

[1] Pottie G J, Kaiser W J. Embedding the Internet; Wireless integrated network sensors[J]. Communications of the ACM, 2000, 43(5):51-58

[2] Tipping M E. Sparse Bayesian learning and the relevance vector machine[J]. The Journal of Machine Learning Research, 2001, 3:211

[3] Vapnik V N. The Nature of Statistical Learning Theory(2nd ed) [M]. New York, USA; Springer-Verlag, 1995

[4] Scholkopf B, Alexander J S. Learning with Kernels; Support Vector Machines, Regularization, Optimization, and Beyond [M]. London; The MIT Press, 2001

[5] Vladimir N V. The Nature of Statistical Learning Theory[M]. New York; Springer, 1995

[6] Kropotov D, Ptashko N, Vasiliev O, et al. On kernel selection in relevance vector machines using stability principle [C] // The 18th International Conference on Pattern Recognition (ICPR 06). 2006

[7] Zhang Q, Benveniste A. Wavelet networks[J]. IEEE Trans on Neural Networks, 1992, 3(6):889-898

[8] 闫丹.基于人工免疫的无线传感器网络节点故障诊断[D].成都:电子科技大学,2009

[9] Vapnik V N. Estimation of dependencies based on empirical data [M]. Berlin; Springer-Verlag, 1982

[10] Zhang L, Zhou W D, Jiao L C. Wavelet support vector machine [J]. IEEE Trans Syst Man Cybern Part B, 2004, 34(1):34-37

[11] 何飞,黎敏,等.基于小波相关向量机的产品质量模型[J].北京科技大学学报,2009,31(7):934-938

[12] Zhao Cheng-lin, Sun Xue-bin, et al. Fault diagnosis of sensor by chaos particle swarm optimization algorithm and support vector machine[J]. Expert Systems with Applications, 2011(38):9908-9912