

基于核偏最小二乘法的物联网无线传感网络故障分析与研究

周光海¹, 宁兆龙¹, 陈志奎¹, 钟华¹, 胡月明²

(1. 大连理工大学软件学院, 辽宁 大连 116620; 2. 华南农业大学自然资源与环境学院, 广东 广州 510642)

摘 要: 随着智能化、网络化传感器技术的日益成熟, 无线传感网络在人类生活以及商业等领域有着广泛的应用, 无线传感器网络节点通常只携带有限的资源, 容易出现因资源不足而导致的故障, 对 WSN 节点进行准确、及时的故障诊断, 能够保障获得信息可靠性, 从而提高 WSN 可维护性并且延长 WSN 的使用寿命。针对该问题, 提出一种使用核偏最小二乘法来预测故障原因的方法, 该方法克服了传统线性回归方法的缺陷, 在高维的非线性空间对数据进行分析, 同时, 该方法也吸收了典型相关分析和主成分分析方法的特点, 为分析提供了更加深入、丰富的内容, 实验结果表明, 提出的方法能够有效预测到故障原因。

关键词: 无线传感网; 故障分析; 核偏最小二乘法

中图分类号: TP393

文献标识码: A

Fault analysis and research of wireless sensor network based on kernel partial least squares

ZHOU Guang-hai¹, NING Zhao-long¹, CHEN Zhi-kui¹, ZHONG Hua¹, HU Yue-ming²

(1. School of Software, Dalian University of Technology, Dalian 116620, China;

2. School of Natural Resources and Environment, South China Agricultural University, Guangzhou 510642, China)

Abstract: With the development of intelligent and networked sensor technology, wireless sensor networks were widely used in human life and commercial fields, because wireless sensor network nodes usually only carry limited resources, it is prone to failures due to insufficient resources, the accurate and timely fault diagnosis of WSN nodes can ensure the reliability of information, thus improving the maintainability of WSN and prolonging the service life of WSN. A method of using kernel partial least squares has been proposed to predict the fault reasons, the method overcomes the defects of traditional linear regression method and the nonlinear high dimensional space for data analysis. Through many experiments, the method can absorb the characteristics of canonical correlation analysis and principal component analysis method, provide a more thorough and rich content analysis, that the reason of the fault can be predicted effectively.

Key words: wireless sensor networks, fault analysis, kernel partial least squares

1 引言

随着智能化、网络化传感器技术的日益成熟, 无线传感网络已经在军事侦查、生物栖息环境监测、环境信息检测、农业生产、医疗健康监护、建筑与家居、工业生产控制以及商业等领域有着广泛

的应用, 而无线传感器节点是实现无线传感网络与物理世界交互、网络自组织、自维护等的核心, 由于无线传感器网络节点通常只携带有限的资源, 容易出现因资源不足而导致的故障^[1]。且由于无线传感器网络的部署特点, 通过人工方式对网络进行维护比较困难。加上故障节点会传输错误的传感数

收稿日期: 2017-09-23

基金项目: 国家自然科学基金资助项目 (No.61672123); 国家自然科学基金重点基金资助项目 (No.U1301253); 广东省重大科技计划基金资助项目 (No.2015B010110006); 中央高校基本科研业务基金资助项目 (No.DUT2017TB02)

Foundation Items: The National Natural Science Foundation of China (No.61672123), The State Key Program of National Natural Science Foundation of China (No.U1301253), The Key Science and Technology Planning Project of Guangdong Province (No.2015B010110006), The Fundamental Research Funds for the Central Universities (No.DUT2017TB02)

据，从而监控中心接收错误的检测信息并产生错误的决策，节点故障会导致网络提供的服务质量严重下降。因此，对无线传感器网络的故障诊断方法进行研究是非常有必要的，及时地探测并对故障节点进行修复，能有效提高网络的可用性^[2]。

对 WSN 节点进行准确、及时的故障诊断，能够保障获得信息可靠性，从而提高 WSN 可维护性并且延长 WSN 的使用寿命，节点通常由能量有限的电池供电，因电池耗尽而造成节点失效是非常普遍的，WSN 节点故障可以分为 2 类：硬故障和软故障。硬故障是指传感器节点在部署或长时间的工作中电源模块、CPU、传感器、通信模块发生损坏以至于产生不可自行修复的故障；软故障是指节点虽然能够继续工作并与其他节点通信，但节点所感知或发送的信息不准确，影响整个网络的监测效果^[3]。

随着无线传感网络的广泛应用，国内外有大量对 WSN 节点进行故障诊断的研究，文献[4]提出了一种用时间冗余技术来诊断传感器网络的通信和故障，但是容易在数据传输的过程中出现错误。文献[5,6]建立了一个 WSN 节点应用神经网络和粗糙集故障诊断方法，但是该算法复杂性大，不适合节点数量大、耦合密切的网络。另外一些传统的传感器节点故障诊断算法包括：粒子群算法及高斯分布法^[7]、隐马尔可夫模型^[8]、分布式故障诊断（DFD, distributed fault detection）^[9]、加权中值故障诊断（weighted median fault detection）^[10]等，但是这些传统的方法大致存在这样一个缺点，由于其对每个节点及其头节点进行一次数据融合，这样会存在大量的冗余计算，加上数据的高维性，节点自身的感知、通信或自身的数据融合都会耗费大量的能量，考虑到进行故障检测必须尽可能地减少功耗，区别于传统方案，本文提出的基于核偏最小二乘法来预测 WSN 节点的故障原因，能够有效克服数据的高维性，进而能够有效降低功耗，并且结合典型相关分析和主成分分析的特点，能够有效预测到故障原因。本文的创新点如下。

1) 本文提出一种使用核偏最小二乘法来预测 WSN 节点中的故障原因。相比于传统的预测方法，该方案能够有效克服传统线性回归方法的缺陷，在高维的非线性空间对数据进行分析。

2) 本文方案同时吸收了典型相关分析和主成分分析的特点，为 WSN 节点故障原因预测分

析提供了更加深入、丰富的内容。通过实验，能够有效预测到故障原因。

2 WSN 节点架构及故障分类

典型的 WSN 系统体系包括分布式传感器感知节点、汇聚节点、路由节点、任务感知节点和互联网。其中，高度模块化的无线传感器网络主要包括 4 个模块：传感器感知模块、处理器控制模块、无线通信模块和电源供电模块，相应的无线传感器网络故障分为传感器感知模块故障、处理器控制模块故障、无线通信模块故障和电源供电模块故障。结合已有的“故障原因工单”数据表明大致的故障具体有以下几个方面。

1) 传感器感知模块故障：基站节点故障、路由节点故障、终端节点故障。

2) 处理器控制模块故障：核心板内部节点故障、核心板外部节点故障。

3) 无线通信模块故障：数据传输接口节点故障、发射天线节点故障、网络芯片节点故障、UART 接口节点故障。

4) 电源供电模块故障：参考电压电路故障、差分运放电路故障、脉冲发射电路故障、外部接口电路故障。

3 理论模型

3.1 偏最小二乘回归

偏最小二乘回归（PLSR, partial least squares regression）^[11,12]提供一种多对多的线性回归建模的方法，特别当 2 组变量的个数很多，且都存在多重相关性，偏最小二乘回归建立的模型都有着很好的可用性。偏最小二乘回归分析在建模过程中集中了主成分分析，典型相关分析和线性回归分析方法的特点，因此，在分析结果中，除了可以提供一个更为合理的回归模型外，还可以同时完成一些类似于主成分分析和典型相关分析的研究内容，提供一些更丰富、深入的信息。

由此，在本文中可以将问题进行形式化的描述，考虑 m 个故障因素 $\{x_1, x_2 \cdots, x_m\}$ ， n 个故障原因 $\{y_1, y_2 \cdots, y_n\}$ 。建立偏最小二乘的回归方程为

$$Y = XB + Z \quad (1)$$

其中， B 为回归系数矩阵， Z 为残差矩阵。首先在故障因素 X 中提出第一成分 t_1 ， t_1 是故障因素的线性组合，尽可能地提取 X 中的变异信息；同时在故障原因

中也提取第一成分 u_1 ，并要求 t_1 与 u_1 之间的相关程度达到最大。其中， X 与 Y 可被分解为如下形式：

$$X = TP^T + E \quad (2)$$

$$Y = UQ^T + F \quad (3)$$

其中， T 与 U 是由成分 t 与 u 组成的得分矩阵， E 与 F 是残差矩阵。 P 和 Q 分别由载入向量 p 和 q 组成， p 和 q 可以通过以下式(4)和式(5)进行计算。

$$p_i = \frac{X_i^T t_i}{t_i^T t_i} \quad (4)$$

$$q_i = \frac{Y_i^T u_i}{u_i^T u_i} \quad (5)$$

同时， t_1 与 u_1 之间的相关程度可通过式(6)进行计算。

$$\begin{aligned} [\text{cov}(t_i, u_i)]^2 &= [\text{cov}(Xw_i, Yc_i)]^2 \\ &= \max_{|r|=|s|=1} [\text{cov}(Xr, Ys)]^2 \\ &= \max_{|r|=|s|=1} \left\{ \text{var}(Xr) [\text{corr}(Xr, Ys)]^2 \text{var}(Ys) \right\} \end{aligned} \quad (6)$$

其中， cov 表示样本之间的协方差， var 表示方差， corr 表示相关系数， w 和 c 为权值向量。建立 Y 与 t_1 之间的回归，如果回归方程能达到满意精度则算法终止。否则继续提取第二成分 t_2 ，直到达到满意精度。最终对 X 提出成分 t_1, t_2, \dots, t_i ，建立 t_1, t_2, \dots, t_i 与 y_1, y_2, \dots, y_n 之间的回归式，然后再表示为 X 与 Y 之间的回归方程。最终，回归系数矩阵可以表示为

$$B = X^T U (T^T X X^T U)^{-1} T^T Y \quad (7)$$

偏最小二乘回归是一个线性的回归模型，然而在本文的故障分析中可能存在有非线性问题，所以本文使用了核偏最小二乘回归对故障因素进行建模^[13]。

3.2 核偏最小二乘回归

核偏最小二乘回归(KPLSR, kernel partial least squares regression)^[14,15]的关键思想是把输入数据映射到一个高维特征空间 H 中，将数据映射到一个核希尔伯特空间中，非线性结构在该空间中可能是一个线性形式。输入数据 X 可以通过非线性映射函数 $\phi(x)$ 映射到特征空间 H 中。进一步地，引入核函数 $K = \phi(x)\phi(x)^T$ 避免了显式的非线性映射和特征空间点积的计算，大大提升了运算效率^[16]。其中核函数的选择需要满足 Mercer 定理，目前运用较广泛的核函数包括高斯径向基核、多项式核、拉普拉斯核等^[17]。通过大量的研究观测，高斯核具有广泛的可用性，所以本文的模型中使用了高斯径向基核，它可以定义为

$$K(x_i, x_j) = e^{-\frac{\|x_i - x_j\|^2}{2\sigma^2}} \quad (8)$$

在该模型中，首先需要对 K 进行中心化，它可以被表示为 $K = \left(I - \frac{1_n 1_n^T}{n} \right) K \left(I - \frac{1_n 1_n^T}{n} \right)$ ，其中， I 是 n 维的单位矩阵， 1_n 是全为 1 的 n 维向量，紧接着，和偏最小二乘法中的过程类似，第一成分 t_1 被提取，它可以被表示为

$$w_i = \frac{\phi^T u_i}{\|\phi^T u_i\|} \quad (9)$$

$$t_i = \phi w_i = \frac{\phi \phi^T u_i}{\|\phi^T u_i\|} = \frac{K_i u_i}{\sqrt{u_i^T K_i u_i}} \quad (10)$$

其中， w 是权值向量， u 是 Y 的成分，在第一次获取 t_1 时， u_1 的值为随机产生的向量。进一步地，通过 t_1 来获取 Y 的成分 u ，获取过程可表示为

$$c_i = \frac{Y_i^T t_i}{t_i^T t_i} \quad (11)$$

$$u_i = \frac{Y_i c_i}{c_i^T c_i} \quad (12)$$

提取 t 和 u 的过程是一个迭代过程，提取过程直到 t 和 u 收敛为止。

算法 1 核偏最小二乘回归模型

输入 故障投诉内容 X 与故障产生原因 Y

输出 回归系数矩阵 B

参数 缩减迭代次数 d ，预置误差项 ε

1) $K = \phi(X)\phi(X)^T$

2) $K = \left(I - \frac{1_n 1_n^T}{n} \right) K \left(I - \frac{1_n 1_n^T}{n} \right)$

3) for $i=1$ to d

4) while true do

5) $w_i = \frac{\phi^T u_i}{\|\phi^T u_i\|}$

6) $t_i = \phi w_i = \frac{\phi \phi^T u_i}{\|\phi^T u_i\|} = \frac{K_i u_i}{\sqrt{u_i^T K_i u_i}}$

7) $c_i = \frac{Y_i^T t_i}{t_i^T t_i}$

8) $u_i = \frac{Y_i c_i}{c_i^T c_i}$

- 9) if $norm(t_i - t_{i-1}) < \varepsilon$ then
- 10) break
- 11) end if
- 12) end while
- 13) $K_i = \left(I - \frac{t_i t_i^T}{t_i^T t_i} \right) K_i \left(I - \frac{t_i t_i^T}{t_i^T t_i} \right)$
- 14) $Y_i = Y_i - \frac{t_i t_i^T Y_i}{t_i^T t_i}$
- 15) end for
- 16) $B = \phi^T U (T^T K U)^{-1} T^T Y$
- 17) return B

在偏最小二乘法中，通过缩减 X 与 Y 的值来判断算法的收敛性。然而这种缩减形式并不适用于核偏最小二乘法，核偏最小二乘法通过对核矩阵 K 与因变量 Y 的缩减来判断算法的收敛性。 K 与 Y 的缩减可被表示为

$$K_i = \left(I - \frac{t_i t_i^T}{t_i^T t_i} \right) K_i \left(I - \frac{t_i t_i^T}{t_i^T t_i} \right) \quad (13)$$

$$Y_i = Y_i - \frac{t_i t_i^T Y_i}{t_i^T t_i} \quad (14)$$

和偏最小二乘法模型相似，也可以得出核偏最小二乘法的模型回归系数矩阵 B ，它可以被表示如下。

$$B = \phi^T U (T^T K U)^{-1} T^T Y \quad (15)$$

并且可以得到预测模型

$$\hat{Y}_i = \phi B = K_i U (T^T K U)^{-1} T^T Y \quad (16)$$

系数矩阵 B 可以通过训练集来获取，模型的评估来通过测试集进行。具体的算法步骤被列出在算法 1 中。

4 实验

本文采用的数据集中，主要包含了 2 个数据文件，分别是“故障原因工单”和“故障节点处理方

案列表”，而本文主要考虑和分析的是“故障原因工单”，“故障节点处理方案列表”作为辅助预测表。由于 2 张表的关联程度较低，唯一值得注意的地方就是故障节点。通过模型来预测出故障原因后，进而通过故障节点来找出相应处理方案，从而进一步定位问题。

在“故障原因工单”这张表中，共有 6 103 条数据。为了进行模型评价，本文随机抽取了 500 条作为测试集，其余部分作为训练集。通过训练模型的预测，可以得到预测数据，通过度量预测数据和真实原因数据 Y 之间的相似性来预测真实原因。为了提高相似度量效率，本文对真实数据项进行了唯一化，将数据中的重复项进行剔除。

为了比对偏最小二乘法与核偏最小二乘法的性能区别，在本文中做了两者的对比实验。在实验中选取了 Top 1、Top 3、Top 5 进行准确度比较，选取迭代次数 $ncomp$ 为 5。

显然从表 1 可以看出，使用 KL 距离度量的精确度结果明显好于其他距离度量。所以在后面的实验中使用 KL 距离作为相似度量。在表 1 中还发现了在 KL 距离度量下，核偏最小二乘法的结果也是要明显好于偏最小二乘法的。进一步地，接下来关注到迭代次数对预测准确度的影响。

表 2 表示了偏最小二乘法和核偏最小二乘法在迭代次数变化情况下的结果。从表 2 中可以观察到，核偏最小二乘法随着迭代次数的增加是趋于收敛的，并且它的性能是相对较好的。在实验中发现，核偏最小二乘法的在各迭代次数上的性能都是优于偏最小二乘法的。

5 结束语

随着无线传感网络在人类生活中的广泛应用，对 WSN 节点进行准确、及时的故障诊断，能够保障获得信息的可靠性，针对该问题本文提出了一种基于核偏最小二乘法的解决方案，该方法能够基于

表 1 偏最小二乘法(PLS)与核偏最小二乘法(KPLS)使用不同距离度量在 Top 1、Top 3、Top 5 中的精确度比较

距离	PLS			KPLS		
	Top 1	Top 3	Top 5	Top 1	Top 3	Top 5
NC	0.094 0	0.100 7	0.091 2	0.210 0	0.131 3	0.090 0
L1	0.060 0	0.048 7	0.074 0	0.002 0	0.018 7	0.018 4
L2	0.056 0	0.048 0	0.073 2	0.002 0	0.018 7	0.018 4
KL	0.060 0	0.100 7	0.101 6	0.220 0	0.145 3	0.103 6

表 2 偏最小二乘法(PLS)与核偏最小二乘法(KPLS)经过不同迭代次数在 Top 1、Top 3、Top 5 中的精确度比较

迭代次数	PLS			KPLS		
	Top 1	Top 3	Top 5	Top 1	Top 3	Top 5
<i>ncomp</i> =1	0.056 0	0.094 7	0.101 6	0.222 0	0.150 0	0.104 8
<i>ncomp</i> =3	0.062 0	0.096 7	0.101 6	0.224 0	0.146 7	0.103 2
<i>ncomp</i> =5	0.060 0	0.100 7	0.101 6	0.220 0	0.145 3	0.103 6
<i>ncomp</i> =10	0.060 0	0.106 7	0.101 6	0.194 0	0.135 3	0.098 0
<i>ncomp</i> =15	0.060 0	0.110 0	0.101 6	0.224 0	0.143 3	0.104 4

核学习的强大非线性映射能力，将数据映射到高维空间进行处理。它不仅克服了传统方法的缺陷，并能够更好地挖掘数据之间的非线性关系。同时吸收了典型相关分析和主成分分析的特点，实验结果表明了本文提出的方法能够有效预测到故障原因。

参考文献:

[1] 刘敏钰, 吴泳, 伍卫国. 无线传感网络 (WSN) 研究[J]. 微电子学与计算机, 2005, 22(7): 58-61.
LIU M Y, WU Y, WU W G. Wireless sensor network[J]. Microelectronics and Computer, 2005, 22(7): 58-61.

[2] ZHANG Q C, ZHU C S, YANG L T, et al. An incremental CFS algorithm for clustering large data in industrial internet of things[J]. IEEE Transactions on Industrial Information. 2017, 13(3): 1193-1201.

[3] 刘凯. 无线传感器网络节点故障诊断算法研究[J]. 江南大学, 2011: 1-2.
LIU K, The research on node fault diagnosis algorithm of wireless sensor network[J]. Jiangnan University, 2011: 1-2.

[4] HARRISON D C, SEAH W K G, RAYUDU R. Rare event detection and propagation in wireless sensor networks[J]. ACM Computing Surveys (CSUR), 2016, 48(4): 58.

[5] LIU Y, DONG M, OTA K, et al. ActiveTrust: secure and trustable routing in wireless sensor networks[J]. IEEE Transactions on Information Forensics and Security, 2016, 11(9): 2013-2027.

[6] ZHANG Q C, CHEN Z K, YANG L T. A node scheduling model based on markov chain prediction for big data, international[J]. Journal of Communication Systems, 2015, 28(9): 1610-1619.

[7] 余成波, 李芮, 何强, 等. 基于粒子群算法及高斯分布的 WSN 节点故障诊断[J]. 振动. 测试与诊断, 2013 (1): 149-152.
YU C B, LI R, HE Q. Fault diagnosis of nodes in WSN based on particle swarm optimization and Gaussian distribution[J]. Journal of Vibration, Measurement and Diagnosis, 2013 (1): 149-152.

[8] 毛乐琦. 基于隐马尔科夫模型的无线传感网节点故障诊断算法[J]. 计算机应用与软件, 2014, 31(1): 132-135.
MAO L Q. WSN nodes fault diagnosis algorithm based on hidden markov model[J]. Computer Applications and Software, 2014, 31(1): 132-135.

[9] 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.
JIANG P. Research on an improved distributed fault detection algorithm for node failure diagnosis in wireless sensor networks[J]. Chinese Journal of Sensors and Actuators, 2008, 21(8): 1417-1421.

[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[D]. 软件学报, 2017, 18(5): 1208-1217.
GAO J L, XU Y J, LI X W. Weighted-Median based distributed fault detection for wireless sensor networks[J]. Journal of Software, 2017, 18(5): 1208-1217.

[11] ROSIPAL R. Nonlinear partial least squares: an overview[J]. Chemo in for Maniocs and Advanced Machine Learning Perspectives: Complex Amputation Methods and Collaborative Techniques, 2010: 169-189.

[12] HENSELER J, RINGLE C M, SARSTEDT M. Testing measurement invariance of composites using partial least squares[J]. International Marketing Review, 2016, 33(3): 405-431.

[13] KANEKO H, FUNATSU K. Ensemble locally weighted partial least squares as a just-in-time modeling method[J]. AIChE Journal, 2016, 62(3): 717-725.

[14] ROSIPAL R. Kernel partial least squares for nonlinear regression and discrimination[J]. Neural Network World, 2003, 13(3):291-300.

[15] WONG E, PALANDE S, WANG B, et al. Kernel partial least squares regression for relating functional brain network topology to clinical measures of behavior[C]// 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI). 2016: 1303-1306.

[16] JIA Q, ZHANG Y. Quality-related fault detection approach based on dynamic kernel partial least squares[J]. Chemical Engineering Research and Design, 2016, 106: 242-252.

[17] ZHANG Q C, ZHU C S, YANG L T, et al. An incremental CFS algorithm for clustering large data in industrial internet of things[J]. IEEE Transactions on Industrial Informatics, 2017, 13(3): 1193-1201.

作者简介:



周光海 (1995-), 男, 贵州织金人, 大连理工大学硕士生, 主要研究方向为多模态学习、图像标注。

宁兆龙 (1986-), 男, 辽宁沈阳人, 博士, 大连理工大学讲师、硕士生导师, 主要研究方向为网络优化、物联网、社交网络。

陈志奎 (1968-), 男, 辽宁大连人, 博士, 大连理工大学教授、博士生导师, 主要研究方向为大数据计算。

钟华 (1992-), 男, 山西忻州人, 大连理工大学硕士生, 主要研究方向为多模态学习、跨模态检索、图像标注。

胡月明 (1964-), 男, 广东广州人, 博士, 华南农业大学教授, 博士生导师, 主要研究方向为地理信息系统、农业物联网、土地资源。