

基于神经网络的无线传感器网络异常数据检测方法

胡石 李光辉 卢文伟 冯海林

(浙江农林大学信息工程学院 临安 311300)

(浙江省林业智能监测与信息技术研究重点实验室 临安 311300)

摘要 传感器网络的异常数据检测对于环境监测具有十分重要的意义。基于BP神经网络模型和线性神经网络模型,分别提出了两种无线传感器网络异常数据检测方法。提出的方法在每个当前时刻通过最近的固定长度的历史数据集训练神经网络,来完成下一时刻的预报。通过神经网络的模型残差,确定概率为 P 的置信区间。当下一时刻数据落入置信区间内,则该数据被判为正常;反之,则为异常。为了比较和验证两种检测方法的性能,在Matlab环境下完成了仿真实验。实验结果表明,基于线性神经网络的异常数据检测方法的检测率(detection rate)达到了97.9%,误报率(false positive rate)不超过0.76%;基于BP神经网络的异常数据检测方法的检测率为96.7%,误报率不超过0.84%。

关键词 BP神经网络,线性神经网络,异常数据检测,检测率,误报率

中图分类号 TP393 **文献标识码** A

Outlier Detection Methods Based on Neural Network in Wireless Sensor Networks

HU Shi LI Guang-hui LU Wen-wei FENG Hai-lin

(School of Information Engineering, Zhejiang A & F University, Lin'an 311300, China)

(Zhejiang Provincial Key Laboratory of Intelligent Monitoring in Forestry and Information Technology, Lin'an 311300, China)

Abstract Outlier detection in wireless sensor network (WSN) is of great significance for environmental monitoring. Two outlier detection methods for WSNs were proposed based on BP neural network and linear neural network in this paper. Latest historical data with fixed length of data window was used to train a neural network model, and then these methods can predict the sensor data of the next time. A confidence interval with probability p was calculated with the help of the model residual. The new measurement will be identified as normal one if it falls inside the prediction interval. Otherwise, it will be classified as an abnormal record. In order to compare and demonstrate the performance of the proposed methods, we finished the simulation experiments in Matlab environment. The experiment results show that the detection rate of outlier detection based on linear neural network reaches 97.9%, and the false positive rate is less than 0.76%. While the detection rate of outlier detection based on BP neural network reaches 96.7%, the false positive rate is less than 0.84%.

Keywords BP neural network, Linear neural network, Outlier data detection, Detection rate, False positive rate

1 引言

随着通信技术、嵌入式计算技术和传感器技术的发展进步,具有感知能力、计算能力和通信能力的传感器网络日益显示出广阔的应用前景。无线传感器网络综合利用传感器技术、嵌入式计算技术、分布式信息处理技术和通信技术,能够实时监测、感知和采集网络分布区域内的各种环境信息,并对这些信息进行处理,并将其传送到需要这些信息的用户,以便做出科学合理的决策。

当无线传感器网络(WSNs)受外部事件(如森林火灾、地质灾害、大气污染等)的影响,或者传感器节点本身发生软件

或硬件故障(如软件缺陷、电池电量不足、电磁干扰等)时,传感器节点测量数据会出现异常^[1-3]。实时高效地检测出传感器网络中的异常数据,无论对于外部突发事件的预警防范,还是对于传感器网络本身的健康状况监测,都具有十分重要的意义。

关于无线传感器网络异常数据检测方法,国内外学术界已有诸多研究成果。针对在线数据连续积累、实时更新的特点,David J. Hill等^[5]提出了一种环境数据流明显偏离历史模式的识别方法。该方法是基于数据流及其预测区间的自回归数据驱动模型,具有执行快、数据量大且无需提前一步对异常值分类的特点。该方法适用于大规模非均匀部署的无线传感

本文受国家自然科学基金项目(61174023),浙江省自然科学基金(Y1110880, Y110791),通信网信息传输与分发技术重点实验室开放课题(ITD-U1300x/K13600xx)资助。

胡石 男,硕士生,主要研究方向为无线传感器网络, E-mail: 610479264@qq.com; 李光辉 男,博士,教授,主要研究方向为无线传感器网络、无损检测技术等, E-mail: lgh@zafu.edu.cn(通信作者)。

器网络。姜旭宝等^[6]提出了一种基于变宽直方图的异常数据检测算法,通过数据聚类的方式将网络中的动态感知数据聚类成变宽的直方图来准确检测出异常数据,同时避免了不必要的数据传输。他们对算法的性能进行了理论分析,并基于真实大规模无线传感器网络系统数据进行了实验评估,结果表明算法具有很高的准确率。肖政宏等^[7]根据 WSNs 节点的预测流量序列和实际流量序列的相关系数变化来进行异常检测,实验结果表明了该方法的有效性。潘铁彪等^[8]利用神经网络的多层感知器模型,结合滚动学习-预报机制,提出一类利用历史数据建模,估计当前数据,并根据估计值与实际测量值间的差异来检测异常值的方法。

本文基于人工神经网络(ANN)中的 BP 神经网络模型和线性神经网络模型,分别提出了两种适用于传感器网络的在线实时异常数据检测方法。提出的方法利用最近的历史数据窗口训练神经网络,预报下一时刻的传感器测量值,并给出一个预报区间。根据实际测量值是否落入预报区间来判断数据是否异常。这种方法的优势在于数据驱动,无需事先建立过程的机理模型;同时,该方法利用最近的局部数据,运算量较小,可用于在线检测。

2 BP 神经网络和线性神经网络简介

2.1 BP 神经网络结构^[9]

BP 神经网络是一种多层前馈型神经网络,由输入层、隐含层和输出层组成,其网络模型如图 1 所示。层与层之间采用全互连方式,同层节点中没有任何耦合。输入信号从输入层节点依次传过各隐含层节点,然后传到输出节点,每一层节点的输出只影响下一层节点的输出。其节点单元特性(传递函数)通常为 Sigmoid 型($f(x)=1/(1+\exp(-B_x))$, $B>0$),但在输出层中,节点的单元特性有时为线性。由图 1 可看出,网络的输入层对应传感器节点采集的温度值,网络输出层输出节点数为 1 个,对应有、无异常状态。根据所设计的网络要求,再确定隐含层以及隐含层中神经元的数目。

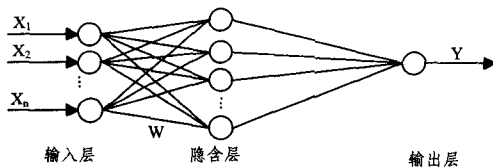


图 1 BP 神经网络模型结构

2.2 线性神经网络结构^[10]

线性神经网络是一种最简单的神经元网络,由 1 个或多个线性神经元构成。线性神经网络的模型结构如图 2 所示,这是一个具有 S 个输入单层(有 S 个神经元)的线性神经网络。其中,权值为矩阵 w ,阈值为向量 b ,这种网络也称为 Madaline 网络。线性神经元采用的传递函数为线性函数,其输入与输出之间是最简单的纯比例关系。线性神经元的输出可以取任意值,其输入与输出之间的关系为:

$$a=f(wp+b)=wp+b \quad (1)$$

由图 2 可看出,函数 f 为比例传输函数。线性神经网络是由 S 个神经元并联而成的一层网络。在进行实时数据异常检测时,输入节点对应传感器节点采集的温度值,网络的输出节点数为 1,对应有、无异常状态。

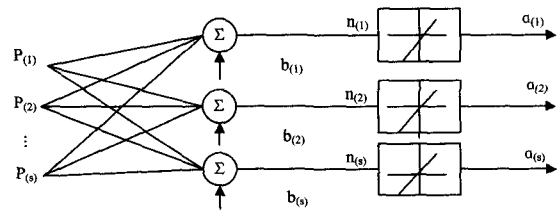


图 2 线性神经网络模型结构

3 基于神经网络的无线传感器网络异常数据检测方法

3.1 基于 BP 神经网络的无线传感器网络异常数据检测方法

设计基于 BP 神经网络的无线传感器网络异常数据检测方法时,需要考虑以下问题:①网络的拓扑结构(隐含层的层数及各层的神经元数目);②神经元变换函数的选取;③网络的初始化;④训练参数的设置。

本文利用 Matlab 工具来实现基于 BP 神经网络的无线传感器网络异常数据检测方法,其主要步骤如下:

- 步骤 1:网络的建立,使用 newff 函数实现;
- 步骤 2:网络的初始化,使用 init 函数实现;
- 步骤 3:网络训练,通过 train 函数实现;
- 步骤 4:网络仿真,通过 sim 函数实现。

3.2 基于 BP 神经网络的无线传感器网络异常数据检测流程

基于 BP 神经网络的无线传感器网络异常数据检测方法的实现流程如图 3 所示。

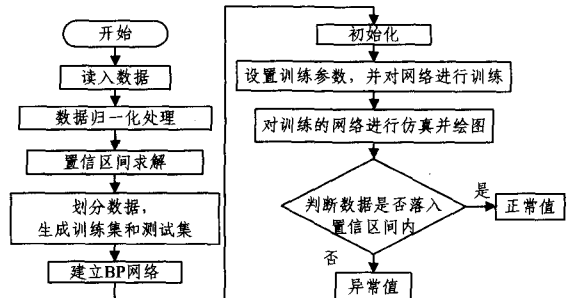


图 3 基于 BP 神经网络的无线传感器网络异常数据检测流程

该方法首先导入传感器数据流,然后对其进行归一化处理和置信区间的求解;接着将处理后的样本数据按要求进行划分,生成训练集和测试集;在建立 BP 神经网络后,对划分好的样本数据进行网络训练并利用 Matlab 仿真工具绘制图形;最后,判断数据流是否落入置信区间内,若落入该区间内,则为正常值;否则,为异常值。

3.3 基于线性神经网络的无线传感器网络异常数据检测方法

设计基于线性神经网络的无线传感器网络异常数据检测方法时,需要考虑以下问题:①网络的结构;②神经元传递函数的选取;③网络的初始化;④训练参数的设置。

类似上一节,仍然使用 Matlab 工具实现该方法,基本步骤包括:

- 步骤 1:网络的建立,使用 newlin 函数实现;
- 步骤 2:网络的初始化,通过 init 实现,该函数通过调用初始化函数 net_initFcn,并根据初始化参数 net_initParam 对网络权值和阈值进行初始化;
- 步骤 3:网络训练,通过 train 函数实现;
- 步骤 4:网络仿真,通过 sim 函数实现。

3.4 基于线性神经网络的无线传感器网络异常数据检测流程

基于线性神经网络的无线传感器网络异常数据检测方法的实现流程如图4所示。

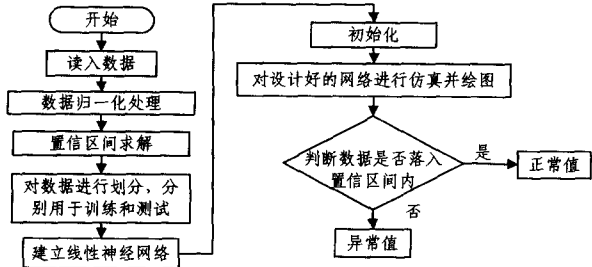


图4 基于线性神经网络的无线传感器网络异常数据检测流程

与基于BP神经网络的检测方法类似,该方法首先导入传感器数据流,然后对其进行归一化处理和置信区间的求解;接着将处理后的样本数据按要求进行划分;然后建立线性神经网络,在网络初始化后,对划分好样本的数据集进行网络训练,并利用Matlab仿真工具绘图;最后,判断数据流是否落入置信区间内,若落入该区间内,则为正常值;否则,为异常值。

4 实验及结果分析

4.1 实验环境与评价指标设计

本文利用Matlab(R2010b)软件中的BP神经网络工具箱和线性神经网络工具箱,对提出的传感器网络异常数据检测方法进行仿真分析。采用的真实实验数据来源于本校的无线传感器网络野外实验系统,该系统采样频率为每隔10分钟采样1次。选择编号为1391的节点在2012年3月份测得的750个温度数据,作为实验的数据样本。

为了评价和比较两种无线传感器网络异常数据检测方法的性能,本文使用检测率和误报率两个性能指标。检测率是指算法检测到的异常数据样本数与实际的异常数据样本总数之比;误报率是指被算法误判为异常的正常数据样本数与总的正常数据样本数之比^[5]。

4.2 样本处理

首先对温度传感器所采集到的750个数据进行划分,其中500个数据用于神经网络的建立及训练,250个数据(含有部分异常数据)用来对建立的神经网络进行验证。

在进行数据分析之前,需要对样本数据进行归一化处理。神经网络是以样本在事件中的统计分布机率来进行网络的训练和预测的,归一化是统一在区间 $[-1, 1]$ 之间的统计坐标分布。此外,归一化也是为了加快神经网络的训练速度和网络的收敛性。

4.3 BP神经网络异常数据检测方法的仿真结果

BP神经网络的无线传感器网络异常检测的训练参数如表1所列。其中,show表示神经网络训练过程中的显示步数,即每隔多少步,显示一次运行结果;lr为神经网络的学习速率;epochs表示神经网络所要进行的最大训练步数;goal则为神经网络的目标精度。

表1 基于BP神经网络的无线传感器网络异常检测方法的训练参数

show	lr	epochs	goal
50	0.05	5000	0.001

利用上述设置的训练参数,对该BP神经网络进行学习训练、验证以及预测。仿真过程的均方误差(mse)如图5所示,横坐标表示网络的训练步数(epoch),纵坐标表示网络的

mse。点划线代表网络训练的mse,实线代表网络验证的mse,星号号为网络预测的mse,两条虚线的交点为该神经网络的最佳验证点。由图5可知,网络在起始阶段的mse较大,随着神经网络训练的不断进行,网络逐渐收敛,mse逐渐降低,直至达到某一稳定状态。在所显示的157步神经网络训练过程中,当网络训练步数到达151步时,均方误差达到了0.091647,正处于两条虚线的交叉点,故此点为网络的最佳验证点。

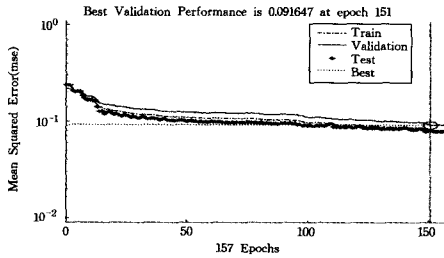


图5 基于BP神经网络的无线传感器网络异常检测方法的均方误差

其次,通过神经网络的仿真,绘制出了95%置信区间内的基于BP神经网络的无线传感器网络数据预测情况,其结果如图6所示。图6(a)表示温度传感器节点收集到的500个温度数据(经归一化后)的实时显示图,图6(b)是相应的传感器数据经BP神经网络学习、训练后,得出的实时预测情况。

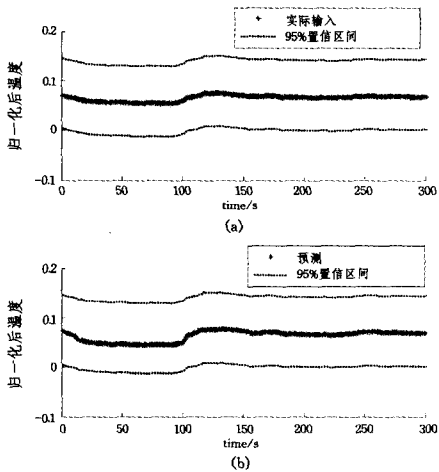


图6 基于BP神经网络的无线传感器网络数据预测结果

观察图6可以发现,在一定置信区间(95%)内,基于BP神经网络的无线传感器网络数据预测效果比较理想,整体上符合原始数据的大体变化趋势。图7描述了基于BP神经网络的无线传感器网络数据预测的误差曲线。从图7不难发现,基于BP神经网络的无线传感器网络数据预测的误差曲线基本上是在0附近波动,平均误差为0.0103。

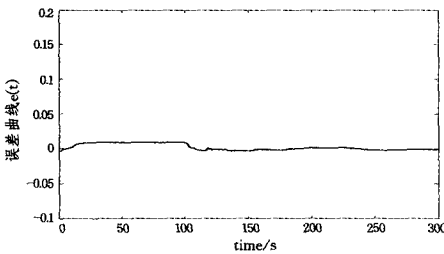


图7 基于BP神经网络的无线传感器网络数据预测误差曲线

图8给出了利用BP神经网络对余下的250个数据样本进行异常检测的结果。若检测到的数据点落于两条虚线(即

95%置信区间)内,则数据点为正常点;反之,则为异常点。由图 8 可知,样本中的个别异常值能被实时准确地检测出来。

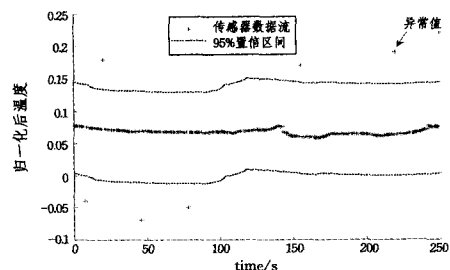


图 8 基于 BP 神经网络的无线传感器网络异常检测实验结果

为了评价该神经网络性能的好坏,此处引用 ROC(receiver operating characteristic)曲线来表示网络的检测率和误报率之间的关系^[4]。该网络的 ROC 曲线如图 9 所示,横坐标表示网络的误报率,纵坐标表示网络的检测率,曲线为基于 BP 神经网络的无线传感器网络异常检测的 ROC 曲线。即使网络的检测率最先达到 90% 以上,相应的误报率也不超过 1%。

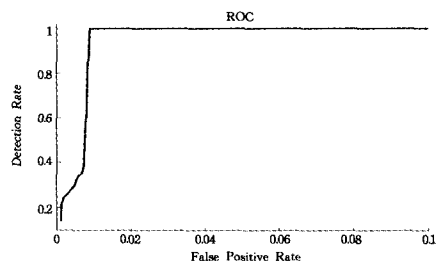


图 9 基于 BP 神经网络的无线传感器网络异常检测方法的 ROC 曲线

4.4 线性神经网络异常数据检测方法的仿真结果

线性神经网络的无线传感器网络异常检测的训练参数设置如下:学习速率 lr,利用 maxlinr 函数取样本数据的最大学习速率;训练步数设为 500;目标精度设为 0.0001。经仿真后,得到的结果如图 10 所示。图中的 PI 表示置信区间。

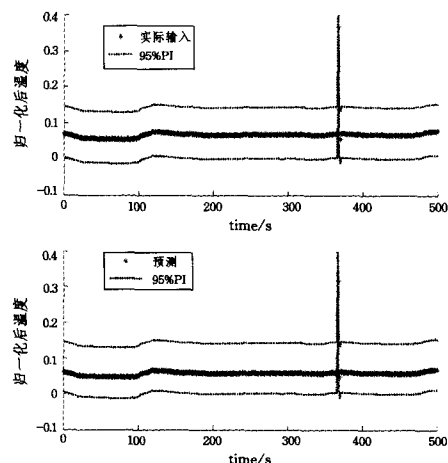


图 10 基于线性神经网络的无线传感器网络数据预测结果

通过比较图 6 与图 10,可以发现基于 BP 神经网络的无线传感器网络异常检测方法,其传感数据预测曲线与原始数据有略微的差别,但基于线性神经网络的传感数据预测曲线与原始数据几乎无差别。图 11 描述了基于线性神经网络的无线传感器网络数据预测误差曲线,其平均误差为 0.0096。对比图 7 与图 11,也能得出相应的结论。由此得出,线性神经网络的学习能力更强一些,且其网络的泛化能力明显强于

BP 神经网络。

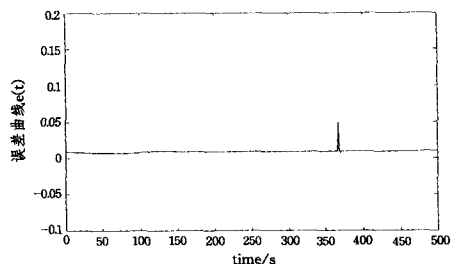


图 11 基于线性神经网络的无线传感器网络数据预测误差曲线图

结束语 异常数据检测的关键是如何形成一个用户或系统正常活动的特征轮廓。神经网络的自学习性和自适应性吸引了越来越多的学者研究如何将其运用到异常检测中,它主要通过大量样本训练,不断地学习,调整主体的特征模式,从而构建一个具有自适应性的特征轮廓。该轮廓可以获得正常和异常数据的分类知识,从而能够对异常数据进行识别。另外,神经网络在学习阶段需要花费较长的时间和较多的系统资源,但是在测试和检测阶段它的速度都较快,而且资源的占用率也较低,完全可以用于实时检测。

本文提出了基于 BP 神经网络和线性神经网络两种无线传感器网络异常数据检测方法。通过利用神经网络的模型残差,确定概率为 P 的置信区间。当下一时刻数据落入置信区间内,则该数据被判为正常;反之,则为异常。实验结果证实两种方法的检测率和误报率都达到了令人满意的效果。基于 BP 神经网络的方法检测率达到了 96.7%,误报率不超过 0.84%;而基于线性神经网络的方法检测率达到了 97.9%,误报率不超过 0.76%。

参考文献

- [1] McDonald D, Sanchez S, Madria S, et al. A Survey of Methods for Finding Outliers in Wireless Sensor Networks[J]. Journal of Network and Systems Management, 2013: 1-20
- [2] Shahid N, Naqvi I H, Qaisar S B. Characteristics and classification of outlier detection techniques for wireless sensor networks in harsh environments; a survey[J]. Artificial Intelligence Review, 2012: 1-36
- [3] Sheng B, Li Q, Mao W, et al. Outlier detection in sensor networks[C]//Proceedings of the 8th ACM international symposium on Mobile ad hoc networking and computing. ACM, 2007: 219-228
- [4] Zhang Y, Meratnia N, Havinga P. Outlier detection techniques for wireless sensor networks; A survey[J]. Communications Surveys & Tutorials, IEEE, 2010, 12(2): 159-170
- [5] Hill D J, Minsker B S. Anomaly detection in streaming environmental sensor data; A data-driven modeling approach[J]. Environmental Modeling & Software, 2010, 25(9): 1014-1022
- [6] 姜旭宝, 李光耀, 连朔. 基于变宽直方图的无线传感器网络异常数据检测算法[J]. 计算机应用, 2011, 31(03): 694-697
- [7] 肖政宏, 谢赞福, 陈志刚. 无线传感器网络中一种基于流量预测和相关系数的异常检测方法[J]. 微电子学与计算机, 2009(7): 214-216
- [8] 潘轶彪, 袁景洪, 朱凯, 等. 基于多层感知器的异常数据实时检测方法[J]. 上海交通大学学报, 2011, 45(8): 1226-1229
- [9] 郭晶, 杨章玉. MATLAB 6.5 辅助神经网络分析与设计[M]. 北京: 电子工业出版社, 2003
- [10] 张德丰. MATLAB 神经网络应用设计[M]. 北京: 机械工业出版社, 2012