DOI：10.13718/j.cnki.xdzk.2019.10.019

一种基于 BPNN 的智能巡检异常预测模型的研究①

邓小清[1]，　王伦浪[1]，　王　刚[2]，　蒲国林[1]

1. 四川文理学院 智能制造学院，四川 达州 635000；2. 安康学院 电子与信息工程学院，陕西 安康 725000

摘要：在传统网络巡检方法中，网络异常发现主要基于单一参数进行阈值触发，误报率较高，效率低. 为了高效准确地发现网络异常，提出了一种基于 BPNN 的网络异常预测模型. 首先对采集系统采集的数据进行特征提取和初始化处理；然后，将初始化后的数据作为神经网络样本进行训练，根据误差阈值调整网络参数，确定网络结构；最后，在 Matlab 环境下进行仿真实验，将提出的 BP 神经网络模型用于网络异常预测，结果表明本文提出的方法对网络异常预测有较高的预测率.

关　键　词：BPNN；智能巡检；异常预测

中图分类号：TP183　　**文献标志码**：A　　**文章编号**：1673－9868(2019)10－0142－07

巡检是网络运维中最重要的一项工作，通过巡检能够第一时间掌握设备的运行状况，及时发现潜在的隐患，在网络运维中的地位无可替代. 传统的巡检方式主要以人工巡检为主，其他方法为辅，故障与隐患一般通过电话方式上报，巡检记录主要通过手工方式来管理，因而传统的巡检方式缺乏对巡检的实时跟踪监控，缺乏对巡检维护质量的定量考核，也不能够实现巡检结果和故障隐患的自动上报和处理，巡检管理工作效率比较低. 目前，人们已开始考虑将人工智能引入巡检系统，构造智能化巡检，并利用机器学习的方法去预测网络异常. 基于此，本文提出了一种基于 BPNN(BP 神经网络)的智能巡检异常预测模型，对网络运维系统采集的数据进行处理、学习，并自动发现网络的异常，提高了网络的安全性能.

1　智能异常预测概述

网络异常是指在网络正常运行过程中，由于系统资源逐渐耗尽或者网络攻击导致数据流量异常，引起系统崩溃或者用户无法工作的现象. 网络异常主要是观察到的行为(如 CPU 负载，内存使用率等)不正常，或者由于其它不相关的进程消耗了系统资源，数据流入、流出变化超出一定的范围等行为. 当系统识别到相应的变化时，主动发出信息，让管理员进行处理，调整网络资源的配置. 在异常预测研究领域，目前有大量的研究. Ramaki 等[1]提出一种实时片段关联算法，它通过因果关联矩阵对警报数据进行因果关联，并对警报序列进行频繁项挖掘，根据得到的频繁项集结果实现异常检测. Gu 等[2]提出使用基于流的决策树分

① 收稿日期：2019－02－26
基金项目：国家自然科学基金项目(61152003)；四川省教育厅重点项目(16ZA03532016)；四川文理学院重点项目(2016KZ002Z).
作者简介：邓小清(1982－)，女，硕士，讲师，主要从事网络技术、人工智能等方面的研究.

类模型的方法进行异常预测，将当前的状态分为normal，alert，failure 3种. Tan等[3]提出了基于流处理的异常预测模型，主要是使用马尔科夫链对多变量进行预测，再将预测值进行贝叶斯分类，对预测状态分为两类，abnormal和normal，如果是类abnormal，则报警；Ding等[4]使用异常探测标识训练数据，再用决策树的方法进行分类，将状态分为alert，anomaly和normal 3类，在实时监控时使用贝叶斯分析法来识别当前的上下文，再确定使用相应的预测模型. 江务学[5]提出了一种基于改进递归神经网络模型的非线性网络流量预测方法，通过引入遗传算法进行全局最优寻解的模型建立. 这些模型大多只是基于回归技术的模型，有一定的局限性，预测结果存在较大的误差.

在智能巡检领域，智能巡检系统主要用来实现对网元数据的实时采集，并对相应的性能指标进行异常检测[6-7]. 文献[8-9]提出了一种主成分分析方法，对提取的原始数据进行特征提取，文献[10]提出了一种基于禁忌算法和BP神经网络的网络入侵检测算法，文献[11]提出了一种改进粒子群优化的小波神经网络模型，这些方法主要是针对单一阈值实现网络预测，因而误报率比较高. 而在实际网络中识别网络异常是一个由多种因素相互影响的状态，需要对异常数据作综合分析判断.

针对以上研究方法存在的不足，本文提出了一种基于BPNN(BP神经网络)的智能巡检异常预测模型. 首先，对网络运维系统中采集的性能数据，进行特征提取；其次，对数据进行归一化处理，再对归一化的数据进行矩阵构建，确定网络训练的样本数据，通过自主学习发现网络异常. 最后通过实验验证本文所提方法在智能巡检异常预测方面的准确率和效率.

2 基于BPNN的智能巡检异常预测模型

2.1 模型总体设计

利用神经网络进行网络异常预测的基本思想是用一系列数据单元训练神经单元，在给定一组输入数据后，预测出输出数据；反向传播(Back Propagation，BP)算法对大量数据的训练能力及其良好的鲁棒性使得它在数据分类和预测方面得到广泛应用.

先利用BP神经网络模型对网络实时监控数据进行训练，计算预测数据与实测数据的误差，通过调整后还原数据得出数据的真实区间范围，根据数据状态概率，选取概率最大的区间为预测值区间，其均值就为预测值.

2.2 BPNN异常预测模型的构建

在网络运行过程中，发现影响网络异常的性能指标很多，传统的网络异常预警主要是针对每一个性能指标设定相应的阈值进行预测告警，这就导致网络异常的误报率较多，需要人工去判断真实的情况. BP神经网络模式通过对所有网元性能参数的统计与分析，利用BP神经网络良好的非线性映射性能和学习能力，实现对网元结点性能的综合预测. BP网络模型可以调整性能参数的种类和个数，具有较强的灵活性及较强的适应能力. 还可以通过对采集到的不同时段性能数据来反复训练网络，使训练出来的网络更接近于真实网络，其预测结果更适应于真实网络.

本文使用的是中国移动某分公司2018年9月网络运维管理系统采集到的网元性能指标数据，采集的数据以小时为单位，每天采集24组数据，总共720个数据样本，每一个数据样本主要包含17个属性，如表1所示.

表1的样本数据大多数是数值类型的数据，对于非数值类型的数据可以通过转换成数值类型进行运算，对于数据值偏大的参数指标，如userSessionNum、userConcurrentNum、LinkOutFlow、LinkInFlow等，为保证网络的性能和稳定性，使用Matlab中自带的mapminmax函数对相应的数据进行归一化处理，

将数据值的范围归一到(0，1)．将数字归一化可以避免不同量级的数据之间相互影响，也可以加快网络学习的速度．将样本数据进行网络构建与训练，将训练好的网络输出进行异常预测．在样本数据中标识网络状态两类：normal 和 abnormal，因此在输出样本中只需 2 个值，设定阈值，当值小于该阈值时输出为 normal，否则为 abnormal．智能巡检异常预测模型如图 1 所示．

表 1　预处理数据

属　性	特征名	描　　述
CPU 利用率/%	cpuUsage	当前采集到的 CPU 的使用率
内存利用率/%	memUsage	内存的占用比率
峰值用户会话数	userSessionNum	通过巡检系统取得的 ERP 系统的同时最大用户会话数
峰值并发活动会话数	userConcurrentNum	并发的 ERP 会话数
防火墙连接数	firewallSession	网络的出口防火墙对外连接数
流入均值流速/Kbps	NeIfInOctets	给定的 t_1～t_2 时间范围内，网络流入数据的平均数
接口输入字节流量/Kbps	LinkInFlow	网络端口接收到的实时数据字节数
流出均值流速/Kbps	NeIfOutOctets	给定的 t_1～t_2 时间范围内，网络流出数据的平均数
接口输出字节流量/Kbps	LinkOutFlow	网络端口输出的字节数
带宽输入利用率/%	LinkInBandWidthUsage	输入带宽利用率＝det(ifInOctets)/ifSpeed * 100%．其中，det 代表 t_1～t_2 时段的采集值之差．
带宽输出利用率/%	LinkOutBandWidthUsage	输出带宽利用率＝det(ifOutOctets)/ifSpeed * 100%．其中，det 代表 t_1～t_2 时段的采集值之差．
链路平均时延/ms	LinkAveDelay	源端在 t 时刻向目标端发送 1 个报文，在 t_0 收到响应报文返回时延值 t_0～t．对 t_1～t_2 时段内的时延求平均值．
链路丢包率/%	LinkPkgLoseRate	给定的 t_1～t_2 时间范围内的所有报文统计丢包数 numloss，返回丢包率计算结果 numloss/n * 100%．
链路输入错包率/%	ifInErrorsRate	输入错包率＝ det(IfInErrors)/(det(ifInUcastPkts)＋det(ifInNUcastPkts)) * 100%．其中，det 代表 t_1～t_2 时段的采集值之差．
链路输出错包率/%	ifOutErrorsRate	输出错包率 ＝ det(ifOutErrors)/(det(ifOutUcastPkts)＋det(ifOutNUcastPkts)) * 100%．其中，det 代表 t_1～t_2 时段的采集值之差．
链路抖动/ms	Link Shake	网络中的延迟是指信息从发送到接收经过的延迟时间，一般由传输延迟及处理延迟组成；而抖动是指最大延迟与最小延迟的时间差，它主要标识一个网络的稳定性．
可达性(连通性)	connectivity	给定的 t_1～t_2 时间范围内，源端一次向目标端发送 n 个(n>＝3)报文，然后监听返回报文，只要收到响应报文，则“可达”；若 t_2 时刻之前没有收到任何响应报文，则“不可达”．

2.3　智能巡检异常 BPNN 结构设计与训练

BP 神经网络是一种按误差逆传播算法训练的多层前馈网络，是目前应用最广泛的神经网络模型之一[12]．BP 网络是由输入层、隐藏层和输出层构成的网络．BP 网络的结构主要包括以下几个方面：BP 网

络的层数、神经元激活函数、每层的神经元个数. BP神经网络的层数选择3层的神经网络，当训练误差不能满足阈值约束时，在每层增加新的神经元.

BP神经网络预测模型如图2所示，BP网络实现网络异常预测需要找到影响网络性能的特征数据，通过本文对网络的性能指标与网络状态的数据分析，确定网络模型的输入为已选定的17个特征向量，即 $X=(x_1, x_2, \cdots x_j, \cdots x_{17})^{\mathrm{T}}$.

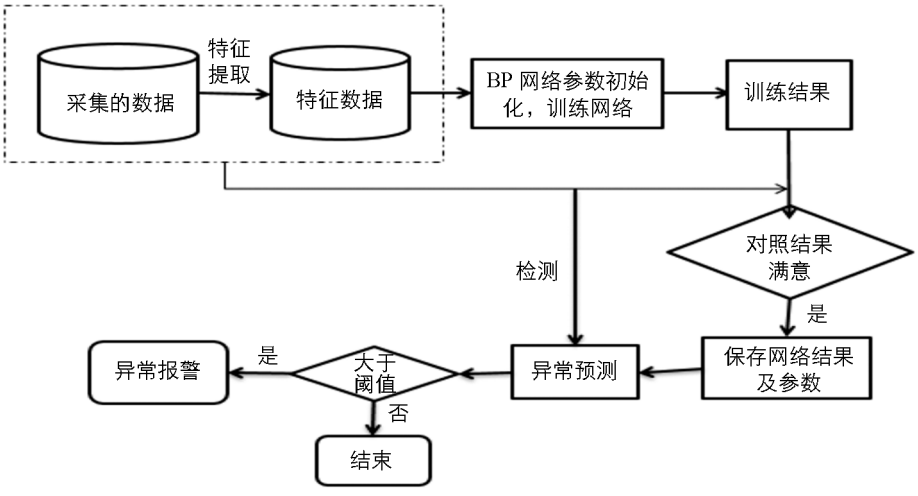

图1 基于BPNN的网络异常预测模型

输出的数据主要是综合判断网络当前的状态，网络中网络状态的标识有2类：normal和abnormal，因此在输出样本中只需2个值，设定阈值，当小于该阈值时输出判断normal，否则为abnormal.

隐藏层神经元的激励函数通常采用Sigmoid函数. 隐藏层的神经元个数可以含一个或者多个神经元，隐藏层需要在实验过程中确定，图2中省略号表示隐藏层的个数不确定. 训练样本比例为70%，检验样本量为30%. 权重更新优化采用梯度下降算法(最速下降法).

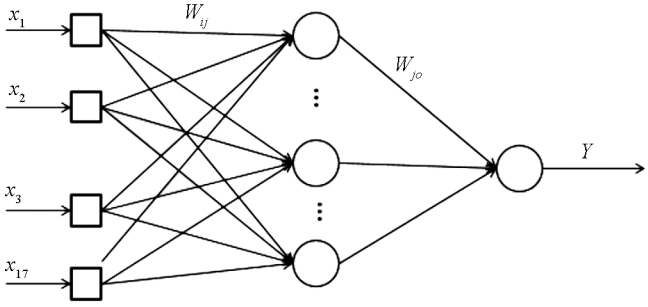

图2 BP神经网络预测模型结构图

2.4 利用训练好的网络进行异常预测

BP算法的学习过程分为正向和反向2个过程，在正向阶段，BP神经网络输入层接收到输入信息，然后从隐藏层到输出层逐层计算每个单元的输出值，如果网络的输出与期望值不一致或者还未达到终止条件，网络进入反向学习阶段. 在反向阶段，BP神经网络利用最速下降法将输出误差经过隐藏层逐层反向传递，并根据每层计算得到的误差调整对应神经元的权值和阈值. BP神经网络会在此学习算法下重复以上2个过程，直到网络输出值域与期望值域之间的误差处于误差允许范围内或者达到最大终止条件时停止学习. 根据分析，网络训练的过程如下所示：

Step1：首先对数据进行归一化处理；

Step2：网络初始化. 网络初始对权值矩阵进行随机赋值，设置网络误差 E 为0，学习率 η 为0.1，网络训练后的精度要求 $E_{\min}$ 为0.001；

Step3：输入训练样本，其中70%作为训练样本集，剩余的30%作为测试样本集，计算各层的输出；

Step4：计算网络输出的误差；

Step5：调整各层的权值；

$$\Delta w_{ij} = -\eta \frac{\delta E}{\delta W_{ij}} i = 1,2,\cdots,17;\ j = 1,2,\cdots,m$$

$$\Delta w_{jo} = -\eta \frac{\delta E}{\delta W_{jo}} j = 1,2,\cdots,m;\ o = 1,2,\cdots,p$$

Step6：检查所有样本的训练. 若 $p<5$，计数器加 1，返回 Step2，否则进行下一步；

Step7：检查网络的总体误差是否达到精度的要求，若 $E<E_{\min}$，训练结束，否则将 E 置 0，p 置为 1，返回 Step2.

最后将网络预测的序列与真实数据对照比较. 如果偏差大于 0.05，则将结果标记为异常并提交异常预警给报警系统；否则为正常. 为了确定非随机因素的影响，取相同的训练和测试样本进行多次运算，确保异常预测的准确率.

3　实验与分析

本文实验所采用的数据是中国移动某分公司 2018 年 9 月网络运维管理系统采集到的网元性能指标数据，采集的数据粒度是以小时为单位，每天 24 组数据，共采集一个月 720 个数据样本，其中每一个数据样本主要包含 17 个属性. 其中 70%作为训练样本集，30%作为测试样本集.

目前大多数网络异常预测均是基于单一指标的阈值进行检测预警，这样存在较高的误报率和漏检率. 本文将采集到的多个网元参数指标通过 BPNN 神经网络进行异常检测和预测，可以更好地保证检测系统的准确性及工作效率. 图 3 为检测样本的预测值与样本值对比图，图 4 为检测样本预测值与样本值的误差时序图，表 2 为部分检测样本预测值与样本值的数据对照表. 从图 3 可以看出，BP 算法中样本预测值曲线与真实值曲线较接近，二者贴合紧密，预测值较精确. 由图 4 和表 2 可知，曲线变化的异常点，样本误差值也大于阈值，由此可以通过图像判断这些曲线变化的异常点为检测样本中的数据异常点. 将检测的异常点输出并在数据集中查找，发现通过图像判断出的异常点和实际检测出的异常值一致，就是实际检测样本中的异常值.

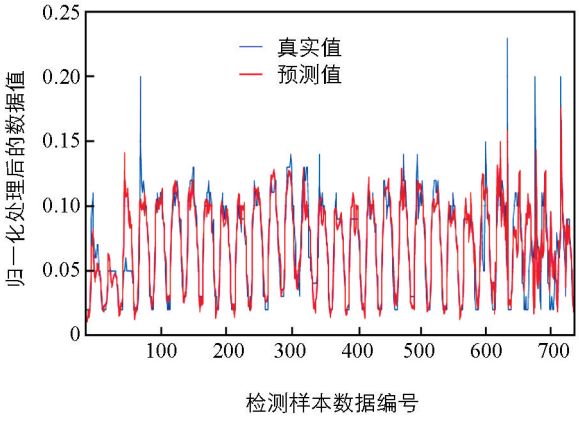

图 3　真实值与预测值

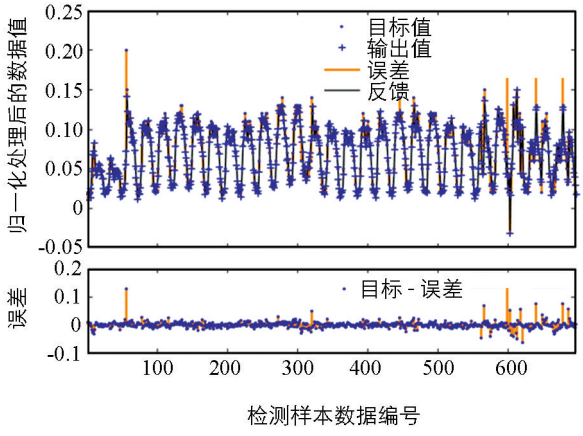

图 4　9 月预测误差时序图

在传统的巡检中，通过单一阈值触发来发现网络异常现象，工作效率极其低下. 在智能巡检系统中，能实现网络状态的间隔自动巡检，并主动将异常状态发送至告警系统，但是需要人为进行识别. 而 BP 神经网络可以通过网络提前预测，判断异常并主动将信息发送至告警系统，大大提高了预测速度，保障了网络的安全，异常预测性能对比如表 3 所示.

表2 神经网络训练部分结果

数据编号	真 实	预 测	误 差
601	0.021 846	0.086 254 556	−0.064 408 556
602	0.019 759	0.004 633 379	0.015 125 621
603	0.020 584	0.024 743 123	−0.004 159 123
604	0.020 256	0.041 526 238	−0.021 270 238
605	0.059 843	0.048 578 047	0.011 264 953
606	0.081 256	0.066 624 627	0.014 631 373
607	0.089 675	0.083 195 791	0.006 479 209
608	0.113 524	0.109 211 461	0.004 312 539
609	0.108 756	0.113 908 514	−0.005 152 514
610	0.121 543	0.103 876 012	0.017 666 988
611	0.109 527	0.112 919 602	−0.003 392 602
612	0.102 568	0.138 843 93	−0.036 275 93
613	0.112 158	0.091 307 648	0.020 850 352
614	0.118 956	0.103 335 349	0.015 620 651
615	0.121 052	0.101 546 666	0.019 505 334
616	0.110 214	0.110 192 692	2.130 76E-05
617	0.090 515	0.110 249 033	−0.019 734 033
618	0.082 053	0.100 207 143	−0.018 154 143
619	0.098 572	0.074 006 017	0.024 565 983
620	0.061 058	0.058 926 24	0.002 131 76
621	0.049 875	0.041 877 314	0.007 997 686
622	0.232 109	0.029 921 4	0.202 187 6
623	0.020 548	0.027 115 96	−0.006 567 96
624	0.018 977	0.026 360 403	−0.007 383 403

表3 异常预测性能对比

	人工巡检	智能巡检	BP网络
异常发现方式	人工巡检(间隔时间较长)	实时检测(根据巡检粒度)	提前预测
时间	异常发生后10 min	异常发生后1 min	异常发生前

通过实验和数据分析，在神经网络学习训练时，应事先对数据进行预处理才能够让神经网络有较好的识别能力；且神经网络训练周期较长，参数较多，网络训练有一定的难度，训练的时间较长. 但是一旦网络结构确定，它的检测效率相比其它巡检方法会提高，且模型存储简单、易于更新.

4 结 论

本文针对传统网络异常发现主要基于单一的性能参数进行阈值触发，存在误报率较高，效率低的问题，提出了一种基于BPNN的智能巡检异常预测模型. 首先，对网络性能参数进行特征提取，数据初始化处理，使其作为神经网络样本进行训练学习；然后，根据误差阈值调整网络参数，直到误差低于阈值. 通过实验证明BP神经网络异常预测模型可有效实现网络异常的发现，具有较高的准确率，大大提高了巡检的质量和效率. 未来可进行的研究方向有：(1) 进一步对网络采集的数据通过工作日与周末、闲时与忙时的细化来提高预测的精准度；(2) 对后期训练得到的数据进一步优化处理；(3) 通过网络数据与高斯分布结合对网络异常进行预测.

参考文献：

[1] RAMAKI A A, AMINI M, EBRAHIMI A R. RTECA: Real Time Episode Correlation Algorithm for Multi-Step Attack Scenarios Detection [J]. Computers & Security, 2015, 49: 206-219.

[2] GU X H, PAPADIMITRIOU S, YU P S, et al. Toward Predictive Failure Management for Distributed Stream Processing Systems [C]//The 28th International Conference on Distributed Computing Systems. Beijing, China: IEEE, 2008: 825-832.

[3] TAN Y M, GU X H, WANG H X. Adaptive System Anomaly Prediction for Large-Scale Hosting Infrastructures [C]// Acm Symposium on Principles of Distributed Computing. New York, NY, USA: ACM, 2010: 173-182.

[4] DING S F, SU C Y, YU J Z. An Optimizing BP Neural Network Algorithm Based on Genetic Algorithm [J]. Artificial Intelligence Review, 2011, 36(2): 153-162.

[5] 江务学. 基于结构优化递归神经网络的网络流量预测 [J]. 西南大学学报(自然科学版), 2016, 38(2): 149-154.

[6] 饶小毛，郭　鑫，周锦伟. 以物联网技术为核心的运维智慧巡检研究 [J]. 电信技术，2014，1(6)：85-87.

[7] 陈新慧. 交换机智能巡检系统设计与实现 [D]. 成都：电子科技大学，2012.

[8] 黄思慧，陈万忠，李　晶. 基于 PCA 和 ELM 的网络入侵检测技术 [J]. 吉林大学学报(信息科学版)，2017(5)：106-113.

[9] 高　妮，高　岭，贺毅岳，等. 基于自编码网络特征降维的轻量级入侵检测模型 [J]. 电子学报，2017，45(3)：730-739.

[10] 周丽娟. 一种改进的基于 TS-BPNN 的网络入侵检测方法 [J]. 陕西理工大学学报(自然科学版)，2018，34(5)：45-49.

[11] 张　兰. 改进 PSO 的 WNN 模型在短期负荷预测中的应用 [J]. 西南师范大学学报(自然科学版)，2017，42(6)：100-104.

[12] 阎平凡，张长水. 人工神经网络与模拟进化计算 [M]. 北京：清华大学出版社，2005：26-30.

A Study on a BPNN-Based Anomaly Prediction Model for Intelligent Inspection

DENG Xiao-qing[1], WANG Lun-lang[1], WANG Gang[2], PU Guo-lin[1]

1. *School of Intelligent Manufacturing, Sichuan University of Arts and Science, Dazhou Sichuan 635000, China;*
2. *School of Electronic and Information Engineering, Ankang University, Ankang Shaanxi 725000, China*

Abstract: In traditional network inspection methods, the network anomaly discovery in intelligent patrol inspection is triggered by a single threshold based on performance parameters, with high false positive rate and low efficiency. In order to predict network anomalies effectively and accurately, we propose a network anomaly prediction model based on BPNN (BP neural network)in this paper. The collected data are extracted, preprocessed and trained as neural network samples, and then the network parameters are adjusted according to the error threshold to determine the network structure. A simulation experiment is carried out in Matlab environment, and the proposed BP neural network model is used for network anomaly prediction. The results show that the method proposed in this paper has satisfactory prediction rate for network anomaly prediction.

Key words: BPNN(Back Propagation Neural Network); intelligent inspection; anomaly prediction

责任编辑　崔玉洁