文章编号：2096-1472(2017)-09-49-03

基于PSO-BP神经网络的入侵检测技术优化算法的研究

雷宇飞，林玉梅
(泉州信息工程学院，福建 泉州 362000)

摘　要：针对基于BP神经网络的IDS技术收敛速度较慢，易陷入局部最优值、网络瘫痪，系统稳定性差等问题，本文提出了基于PSO−BP神经网络的入侵检测技术优化算法。利用粒子群优化算法优化BP网络的权重，首先利用PSO算法优化得到一个最优初始值，然后通过BP网络算法修正误差值，从而获得最优值。

关键词：粒子群优化算法；神经网络；入侵检测
中图分类号：TP301　　**文献标识码**：A

An Intrusion Detection Technology Optimization Algorithm Based on the PSO-BP Neural Network

LEI Yufei,LIN Yumei
(*Quanzhou Institute of Information Engineering*,*Quanzhou* 362000,*China*)

Abstract: Many problems are found in the IDS technology based on the traditional BP neural network,including low convergence speed,easily falling into the local optimal value,network paralysis,poor system stability,etc.This paper presents an intrusion detection technology optimization algorithm based on the PSO-BP neural network which optimizes the BP network weight with the Particle Swarm Optimization(PSO)algorithm.Firstly,an optimal initial value is obtained by using the PSO algorithm,and then the error value is corrected with the BP network algorithm to obtain the optimal value.

Keywords: Particle Swarm Optimization(PSO)algorithm;neural network;intrusion detection

1 引言(Introduction)

网络安全是网络通信技术中的关键技术之一，也是近年来的研究热点。传统的入侵检测技术主要有误用检测(Misuse Detection)和异常检测(Anomaly Detection)技术[1]，本文针对现有入侵检测技术算法的收敛速度较慢的问题，提出基于PSO−BP神经网络入侵检测技术优化算法，利用粒子群优化算法优化BP网络的权重，首先利用PSO算法优化得到一个最优初始值，然后通过BP网络算法修正误差值，从而获得最优值。

2 基于PSO-BP神经网络入侵检测技术优化算法(Technology optimization algorithm based on PSO-BP neural network)

2.1 BP神经网络算法

BP神经网络算法[2]本质上采用梯度下降搜索方法，由于该算法的不足，如收敛速度慢，容易陷入误差函数的局部最优值的问题，且对于较大搜索空间多峰值和不可微函数等搜索全局最优值也无能为力，而粒子群算法[3]能有效地解决这些问题，因为它是一种基于群体协作的随机搜索算法，它可以被纳入多主体优化系统，是群集智能的一种，能避开局部极小值，且在进化过程中无须提供所要解决问题的梯度信息。粒子群算法和遗传算法一样，都随机初始化种群，并使用适用值来评价系统，依据适应值进行一定的随机搜索。粒子群算法依据自身速度来决定搜索。大部分的情况下，所有的粒子可能更快的收敛于最优解，且没有遗传算法遇到的问题。因此，本文提出一种基于PSO−BP神经网络入侵检测算法。

2.2 改进PSO算法描述

本文要解决的是当全局最优粒子陷入局部最优解后，变动粒子的移动方向，跳出局部最优解，对gbest参数重新更新。所以，应当粒子群中先选择参考粒子群，比如一部分最优粒子子群，每次迭代时，若全局最优粒子gbest有一定次数没有更新，则随机生成新的位置及速度的值，并更新最优粒子子群，将更新后的最优粒子子群的适应度值与当前全局最优粒子的适应度值进行比较，来决定当前全局最优粒子是否被新的粒子所取代，在后续的迭代中，粒子群将向新的全局最优粒子靠近。

在本文中，判断当前全局最优粒子是否能被取代的策略如下：

IF f(Sub)≤f(gbest) than 此粒子取成为全局最优粒子；

基金项目：福建省中青年教师教育科研项目，应用型本科院校基于大数据背景下的智慧校园建设(JAT160613).

IF f(Sub)>f(gbest) than 计算概率P，如公式(1)

$$P=\begin{cases}1 & f(Sub)\le f(gbest)\\ \exp(-S*(f(Sub)-f(gbest))) & f(Sub)>f(gbest)\end{cases} \quad (1)$$

其中，f(Sub)表示参考粒子子群中最优粒子的适应度值，f(gbest)则为当前全局最优粒子的适应度值。子群粒子适应度值的标准差S,计算公式如公式(2)：

$$S=\sqrt{\frac{1}{m}\sum_{i=1}^{m}(f(Sub_i)-f_{ave})^2} \quad (2)$$

其中，m为子群粒子个数，f_{ave}为子群粒子的平均适应度值，改进后算法步骤如下：

(1)根据神经网络的输入输出数量和结构确定粒子的维度。

(2)对粒子的速度和位置进行随机初始化。

(3)根据适应度函数计算得到每个粒子的适应值。

(4)将每个粒子的适应值与pbest值(当前最优粒子)作比较，如果优于pbest值,则更新pbest值及其位置,同时记录粒子群中10%的最优粒子子群。

(5)将pbest值和gbest值作比较,如果优于gbest值,则更新gbest值及其位置,同时记录gbest粒子在迭代过程中未被更新的次数Num。

(6)粒子的速度和位置改变参照公式(3)和公式(4)：

$$V_{id}=w*V_{id}+c_1*rand()*(P_{id}-X_{id})+c_2*rand()*(P_{gd}-X_{id}) \quad (3)$$

$$X_{id}=X_{id}+V_{id} \quad (4)$$

(7)若gbest粒子经过N次未更新的值达到预定数值,则取最优粒子gbest和最优粒子子群的位置进行更新,重新随机生成速度Vr和随机向量控制粒子移动方向,根据新生成的Vr和移动方向生成新的一批粒子,得到新的全局最优粒子gbest,更新完成后,gbest的未被更新次数设置为0。如果没有出现这种情况则继续进行步骤(8)。

(8)转到步骤(3)，重复执行这些步骤，直到停止。一般情况下，停止条件设定为最大所期望的适应值或者迭代次数大于设定的最大迭代数。

最终得到的最优粒子的值用来初始化BP神经网络的权值，值是否优于B。

2.3 改进后PSO算法流程设计

改进后PSO算法流程设计,如图1所示。

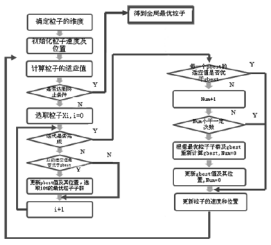

Fig.1 Improved PSO algorithm flow

3 入侵检测系统的实现(Implementation of intrusion detection system)

基于PSO-BP神经网络优化算法的入侵检测系统模型如图2所示，其中，捕获数据引擎模块的任务就是接收网络数据包，接到的网络数据包经特征提取模块，转换成对应数据包特征值的网络连接记录，这些特征值的网络连接记录经数据预处理模块归一化到一个较小的区间范围，从而减少由于记录间因字段数值差异过大而引发的网络训练不良表现。经数据库存放预处理模块处理的数据通过PSO-BP分类器分析检测；PSO-BP训练模块将从数据库模块中提取标准的数据，包括给定的样本记录向量和网络训练要达到的目标，PSO-BP分类器模块对输入中的未知的网络连接记录判别为正常或入侵，并做出相应处理，响应模块主要用以接收从分类器检测出的结果，并对入侵行为发出警报[4]。本文的重点是PSO-BP分类器的设计，其原理是利用数据及划分后的训练样本集来训练PSO-BP神经网络分类器，在PSO-BP神经网络内部建立起对训练样本的识别模型并存储攻击行为的特征模式，然后分析处理捕获到网络数据包，判断其为正常或异常网络行为，如果检测到攻击行为是未知类型时，则将其反馈到学习样本库以让PSO-BP神经网络再学习，所以可以看出算法分类引擎具有通过学习不断来扩展检测范围的能力[5]。

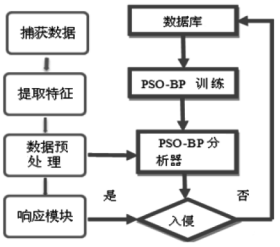

Fig.2 Intrusion detection system model

4 实验及其结果分析(Experiment and result analysis)

4.1 实验样本的选择

基于BP网络的入侵检测算法对样本的依赖性很大，并且只有公认完备、性能优异的测评数据集才能为各种入侵检测算法和技术提供比较的平台，因此，本文实验采用目前公认最优秀影响最广的基于MITLL采集整理形成的KDDCUP99入侵测试数据集，包含了目前最常见的四类攻击手段：User to Root(U2R)获取根权限攻击、Denial-Of-Service(DOS)拒绝服务攻击、R2L远程攻击、Probe探测攻击。本文选取corrected.gz数据集和10%数据集中提取相应的攻击记录，其中DOS约占78.89%、Probe约占4.13%、R2L约占6.35%、U2R约占1.54%，在此基础上，选择部分正常连接记录数据及部分DOS攻击记录，以1∶3概率加入到正常数据流量集，从而形成不均衡的数据集，从中选取150000个有效数据为训练集，并选取181500个数据为测试集，实验样本分布情况如表1所示。

表1 实验样本分布

Tab.1 Experimental sample distribution

数据集名称	训练样本	测试样本
DOS	82346	95677
U2R	254	345
R2L	13453	16774
PROB	4563	4255
NORMAL	42616	52908
UN	13455	16457
SUM	144579	186316

4.2 样本特征的选取

KDDCUP99数据集存储格式是文本格式，每条记录格式都由字符数据与数值构成。其中有的还会附加一个表示攻击类型的特征。这些特征分成四类：一是网络连接的基本特征，包括连接时间、服务、基于什么样的协议、连接时间等；二是网络连接的内容特征，如ROOT用户登录次数、访问敏感数据的频率等；三是基于时间的网络流量特征，如相同主机的连接和相同服务的连接等；四是基于主机的网络流量特征。但由于每条记录都包含41种特征，使得系统的计算量过大，所以，需要对记录进行特征选择，本文选择了14个能够体现用户行为的特征来作为研究对象，分别是：duration连接持续时间(秒)、protocol_type协议类型：TCP、UDP、flag连接状态、service服务类型、src_bytes源到目的站的数据字节数等。最后将样本特征按类型以数值形式进行统一的编码处理后，进行归一化，即将输入或输出映射到[0，1]区间。

4.3 实验测试与结果

本文在基于Matlab 7.0对本文提出的改进PSO算法和标准BP算法进行了检测验证仿真实验。计算出各网络训练所用时间、检测率、漏报率及误报率。通过对结果的分析来证明该改进算法是否具有较高的入侵检测效率、较低的误报率和漏报率。

(1)确定网络结构及参数，本文采用三层的BP神经网络结构，输入层采用14个节点(因为所采用14个特征值的样本)，输出层对应5种分类结果，设置三个节点。根据一些学者提出的公式，本文分别采用6、7、8、9作为隐含层节点数。网络的权、阈值初始化[-1,1]的随机数，各学习率赋初值为0.1，误差精度为.001，最大迭代次数为3000次。PSO的参数设置为：粒子总数500个,全局最优粒子gbest的最多未更新次数设置为10,超过此次数则对gbest和最优粒子子群进行更新；PSO的迭代次数设置为3000次。

(2)分别用训练样本对标准BP，改进PSO-BP算法进行了训练网络，选用不同的参数进行多次训练，结果发现，本文改进后的BP算法在误差进行平坦区后能较快跳出，检测精度也提高了，收敛更快。

(3)测试网络的性能，对测试结果进行了归类，并与期望值进行了比较，得出已知行为检测率、未知行为检测率、误报率等。入侵检测效果比较如表2所示。

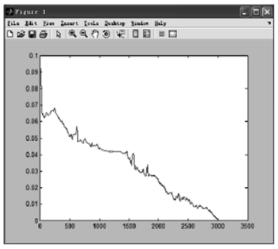

Fig.3 The error performance of Improved PSO-BP algorithm

表2 入侵检测效果比较

Tab.2 Comparison of intrusion detection results

	BP神经网络算法	改进PSO-BP算法
已知行为检测率	0.9238	0.9875
未知行为检测率	0.8273	0.9076
误报率	0.0348	0.0073

从表2可以算出，基于改进PSO-BP的入侵检测算法，对未知行为也有较好的检测率，与传统的入侵检测方法相比，在入侵检测性能上取得比较好的检测效果。而且使用改进PSO优化的BP神经网络，在这四个入侵类型的检测准确率上都有提高。而随着检测准确率的提高，自然而然就降低误报率和漏报率。

5 结论(Conclusion)

为提高IDS的检测效率，本文使用改进的PSO-BP算法设计了一种新的入侵检测模型，通过与传统的BP神经网络入侵检测算法的检测性能对比实验，表明基于改进的PSO-BP神经网络的入侵检测算法对入侵检测系统的检测性能有较大的提升，并实验验证该模型在入侵检测方面的性能，准确率较高，收敛较快，达到了本文的预期要求。

参考文献(References)

[1] 华辉有,等.一种融合Kmeans和KNN的网络入侵检测算法[J].计算机科学,2016,43(03):158-162.

[2] 谢康.基于神经网络的入侵检测相关技术研究[D].山东大学,2016.

[3] 李民政,蓝剑平.时空域信息融合的智能家居入侵检测算法[J].北京邮电大学学报,2017,40(3):76-84.

[4] 徐振华.基于BP神经网络的分布式入侵检测模型改进算法研究[J].网络安全技术与应用,2016(2):77-78.

[5] 杨云峰,唐凤仙.改进遗传算法优化神经网络的入侵检测研究[J].河池学院学报,2017,37(2):77-83.

作者简介：

雷宇飞(1981-)，男，硕士，讲师.研究领域：计算机网络，数据库技术.

林玉梅(1982-)，女，硕士，讲师.研究领域：计算机网络.