分层神经网络在入侵检测系统中的应用

毕 靖[1],张 琨[2]
BI Jing[1],ZHANG Kun[2]

1.北京建筑工程学院 理学院,北京 100044
2.河北工业大学 信息工程学院,天津 300130
1.Science School,Beijing University of Civil Engineering and Architecture,Beijing 100044,China
2.School of Information Engineering,Hebei University of Technology,Tianjin 300130,China
E-mail:bijing@bucea.edu.cn

BI Jing,ZHANG Kun.Intrusion detection based on hierarchical neural network.Computer Engineering and Applications,2009,45(20):106-107.

Abstract: Intrusion detection is one of important ways to improve network security,and is a main research topic in computer field.The benchmark dataset commonly used in the research of intrusion detection is adopted.Firstly,data are changed into the appropriate type for simulations;Secondly RBF network is used for raw detection;Thirdly Elman BP network is used for advanced detection;Lastly a lot of simulation results are gained from MATLAB platform,and display that hierarchical network is fit for intrusion detection.

Key words: hierarchical network;intrusion detection;network security

摘 要:入侵检测技术是提高网络安全的重要手段之一,旨在利用分层神经网络解决入侵检测问题。针对入侵检测研究的通用审计数据集,首先将数据进行预处理以便运算;其次利用 RBF 网络实现粗检测;再次利用 Elman BP 网络进行细检测,从而实现分层神经网络的入侵检测;最后在 MATLAB 平台下进行仿真实验,仿真结果表明,分层神经网络结构在入侵检测中体现出良好的特性。

关键词:分层神经网络;入侵检测;网络安全
DOI:10.3778/j.issn.1002-8331.2009.20.032 **文章编号:**1002-8331(2009)20-0106-02 **文献标识码:**A **中图分类号:**TP393.08

1 前言

随着各类网络系统的迅猛发展和广泛应用,网络安全已经成为目前计算机系统刻不容缓需要解决的问题。入侵检测技术要求在不影响网络性能的前提下,对计算机系统进行实时监控,发现并阻止已知入侵行为,同时还能够学习和适应未知入侵行为。

入侵检测技术通常可以分为特征检测和异常检测两类[1]。特征检测是基于入侵行为都具有某种特定模式或特征的假定,常用的方法包括专家系统、Petri 网等。当明确了解待检测入侵行为特定模式或特征时,该类技术最有效,反之对于未知入侵行为则很难处理。异常检测是基于入侵行为必与正常行为存在差异的假定,常用的方法包括基于统计、神经网络、数据挖掘等技术的方法。它可以通过比较差异检测到未知入侵行为,但是对已知入侵行为的误报率较高。

神经网络作为数据分析的工具广泛应用于各个领域,在入侵检测中也得到了大量使用,已有的工作给出了网络结构、参数学习方法等方面有所差异的神经网络,并对应用效果进行比较。同时,也有文献提出多个神经网络组合使用的方案,通过"举手表决"的方式判断是否存在入侵行为。

利用分层神经网络对入侵检测通用数据集进行处理,其本质也是一种多神经网络的组合结构,包含粗选和有针对性细选两个步骤,从而提高检测精度。

2 入侵检测数据预处理

本文采用的入侵检测数据为通用数据集[2],其中每组数据(含 42 个元素)的格式如下面两组数据所示,其中前 41 个元素为审计数据的输入,第 42 个元素为对应的输出。

第一组:

0,tcp,http,SF,163,20554,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,12,15,0.00,0.00,0.00,0.00,1.00,0.00,0.13,90,255,1.00,0.00,0.01,0.03,0.00,0.00,0.00,0.00,normal.

第二组:

0,udp,private,SF,105,146,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,2,2,0.00,0.00,0.00,0.00,1.00,0.00,0.00,255,254,1.00,0.01,0.01,0.00,0.00,0.00,0.00,0.00,snmpgetattack.

数据处理的目的在于:一是可以满足数值运算的要求;二

基金项目:北京市教委科技计划面上项目(the Projects of Beijing Municipal Commission of Education);河北工业大学博士启动基金(the Doctor Start-up Fund Research of Hebei University of Technology)。
作者简介:毕靖(1974-),女,讲师,主要研究方向为入侵检测;张琨(1976-),女,硕士,主要研究方向为智能理论在信息科学中的应用。
收稿日期:2009-03-05 **修回日期:**2009-04-16

是提高检测效果。采取如下 3 个步骤对数据进行预处理。

(1)将字符数据替换为数值数据,如表 1 列出第 2 输入列字符串与数值间的替换关系,以及表 2 列出输出列字符串与数值间的替换关系。其中替换结果为对应字符串第一次出现的次序,下同。输出包含 5 种可能,其中 normal 表示正常行为,其他表示入侵行为。

表 1　第 2 输入列字符串替换结果

字符串	替换结果	字符串	替换结果
tcp	1	icmp	3
udp	2		

表 2　输出列字符串替换结果

字符串	替换结果	字符串	替换结果
normal	1	xlock	4
snmpgetattack	2	smurf	5
named	3		

(2)去除数值完全相同的输入列,这样可以大大简化神经网络的规模,提高运算速度。

(3)将每列所有可能取值从小到大排序,并用它对应的序号替换该数值。这样可以大大加快神经网络的收敛速度,并提高运算速度。

经过进一步处理后的数据,将能够适应神经网络的训练和学习的要求。

3　分层神经网络结构

输出列经过数据处理后已经成为整数形式,神经网络的输出结果通常为小数形式,所以结果只能通过四舍五入才能与已知输出数据进行比较。

假设输出结果为 1.00、1.12、1.20 和 1.48,它们四舍五入后的结果都为 1,即表示正常行为。显然对于 1.00 和 1.12 判断为正常行为基本无疑问;对于 1.20 则有些疑问;而对于 1.48 则有很大的疑问。

针对上述问题,提出一种分层神经网络的结构进行入侵检测,对一些有疑问的数据进行二次检测,而对其他数据仅进行一次检测。

该结构如图 1 所示,其中参数 alpha 可以由用户根据具体情况进行设置。

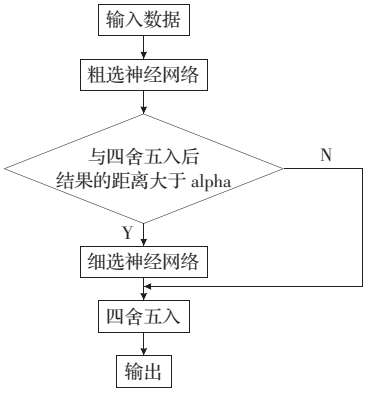

图 1　分层神经网络结构

MATLAB 提供了神经网络工具箱,使用它能够快速实现对实际问题的建模和求解。选择该工具箱中的网络作为粗选、细选神经网络,并利用相关函数建立、训练和运算网络。

下面的仿真中粗选神经网络为 RBF 网络,细选神经网络为 Elman BP 网络,这两种网络都具有收敛快和运算时间短的优点。选择不同类型的网络结构可以有效避免结果的重复。

4　仿真结果

4.1　RBF 网络仿真结果

RBF 网络具有结构简单、参数易调和收敛速度快的特点,同时相对基本 BP 网络等结构精度较高,因此选择该网络结构作为比较对象。

将 1 000 组数据的奇数组作为训练数据,将偶数组作为测试数据。表 3 为隐层结点数为 62 时的仿真结果。

表 3　RBF 网络的仿真结果

测试样本输出值	对应个数	测试输出正确个数	测试输出错误个数	正确率/(%)	错误率/(%)
1	349	331	18	94.8	5.2
2	52	27	25	51.9	48.1
3	1	0	1	0	100
4	4	0	4	0	100
5	94	78	16	83.0	17.0
总计	500	436	64	87.2	12.8

当参数 alpha 设置为 0.25 时,在判断正确的 436 组中神经网络与四舍五入后结果的距离大于 alpha 的有 33 组;而在判断错误的 64 组中大于 alpha 的有 38 组。

这里需要说明如下 3 点:

(1)在判断错误的 64 组中,小于 alpha 的有 26 组。这说明这 26 个错误输出明显偏离正确值,很难通过简单的措施进行修正。本文提供的方案未对这部分数据进行处理,当然也可以通过调整 alpha 参数值改变这部分数据的个数。

(2)将训练数据作为测试数据时,在判断正确的 472 组中神经网络与四舍五入后结果的距离大于 alpha 的有 46 组;而在判断错误的 28 组中大于 alpha 的达到 27 组。

(3)Elman BP 网络的结果与 RBF 网络相比稍有差异,其判断正确率为 86.5%。

4.2　分层神经网络仿真结果

同样将 1 000 组数据的奇数组作为训练数据,将偶数组作为测试数据,参数 alpha 选为 0.25。Elman BP 网络包含两个隐层,分别包含 62 和 310 个结点,表 4 为分层神经网络的仿真结果。

表 4　分层神经网络的仿真结果

测试样本输出值	对应个数	测试输出正确个数	测试输出错误个数	正确率/(%)	错误率/(%)
1	349	339	10	97.1	2.9
2	52	33	19	63.5	36.5
3	1	0	1	0	100
4	4	1	3	25.0	75.0
5	94	87	7	92.6	7.4
总计	500	460	40	92.0	8.0

在进入细选神经网络处理的 71 组数据中,经粗选神经网络判断正确的为 33 组,经细选神经网络判断正确的为 57 组,精度提高 33.8%。

(下转 178 页)