基于遗传神经网络的误分类代价敏感网络入侵检测①

蒋贤特，周晓慧

(杭州电子科技大学 计算机学院，杭州 310018)

摘 要：针对传统的基于遗传神经网络的入侵检测模型未考虑误分类代价的不足，将误分类代价敏感的特征集成到基于遗传神经网络的网络入侵检测模型中，从而克服了传统模型中错误分类时可能导致代价过大的缺点。通过实验结果表明，增加了误分类代价敏感特征后的遗传神经网络能较好地控制网络入侵检测系统误报、漏报攻击时所产生的代价。

关键词：入侵检测；遗传算法；神经网络；误分类代价

Network Intrusion Detection Based on Genetic Neural Network Misclassification Cost Sensitive

JIANG Xian-Te, ZHOU Xiao-Hui

(Institute of Computer Science, Hangzhou Dianzi University, Hangzhou 310018, China)

Abstract: The paper aims at the insufficient of traditional intrusion detection based on genetic neural network not consider the misclassification cost, integrate the misclassification cost-sensitive features into the network intrusion detection model which based on genetic neural network, to overcome the defect of the traditional model's error classifying result in excessive costs. The experiment results show that after the genetic neural network increased the misclassification cost-sensitive features, it can control the cost caused by the network intrusion detection's false report、omit report attacks preferably.

Keywords: intrusion detection; genetic algorithm; neural network; misclassification cost

1 前言

互联网应用已取得长足的发展，网络安全问题显得尤为重要。传统的安全保护技术和防火墙技术有其自身的局限性，已暴露出明显的不足和弱点，如网络安全后门问题、网络内部攻击、病毒入侵等。

入侵检测系统（Intrusion Detection System，IDS）作为一种积极主动的安全防护措施，能检测未授权对象对系统的恶意攻击，并监控授权对象对系统资源的非法操作，阻止入侵行为[1]。入侵检测可以看作是一个分类问题，即在已有数据的基础上学习一个分类函数或构造一个分类模型，该函数或模型能够把数据库中的记录映射到给定类别中的某一个上，从而应用于数据预测。

入侵检测的方法有很多，其中，异常检测技术[2]主要包括基于统计方法的异常入侵检测、基于特征选择的异常检测、基于模式预测异常检测、基于神经网络的异常入侵检测技术。但这些系统都存在误报、漏报及实时性差等缺点，需要大量完备的训练数据集才能达到比较理想的检测性能。

2 遗传神经网络综述

2.1 遗传神经网络

人工神经网络相对于传统的模式识别表现出了极好的优越性，但其本身还存在着明显的缺陷，就是训练过程非常缓慢，这主要是由于反向传播过程中计算量非常大所导致的。遗传算法具有并行性，它只需搜索少数结构,利用群体的适应值信息，通过简单的复制、杂交和变异算子[6],遗传算法能以很大的概率找到

① 收稿时间:2010-09-29;收到修改稿时间:2010-11-13

全局最优解，适合于处理传统搜索方法解决不了的复杂和非线性问题。

因此，利用遗传算法[3]对神经网络结构、初始连接权、初始阈值以及学习率进行优化设计，在解空间中定位出较好的搜索空间。从而解决了人工神经网络中固有的收敛速度慢，容易陷入局部极小值等缺点。

3 代价敏感遗传神经网络的设计

3.1 误分类代价

当每类样本的误分类代价相等，且每类样本的数目大致相等时，标准分类算法是有效的[4]。但对于误分类代价不对称的数据分类问题，基于误分类率最小的标准分类器通常不能实现误分类代价最小，此时，需要把误分类代价最小化作为分类器的优化目标。

若给定把一类样本误分类为另一类样本的代价，可构造代价矩阵 C，其元素 C(i,j)表示样本的真实类标号为 j，而神经网络输出为类 i 所付出的代价。i=j 表示正确的分类，i≠j 表示错误的分类。这样，获得样本 x 所属类的代价敏感预测需要最小化如下方程：

$$R(x,j)=\sum_{j}P(j\mid x)C(i,j) \qquad (1)$$

其中，$P(j\mid x)$ 表示给定样本 x 属于类 j 的条件概率，标准的分类算法得到每个样本 x 属于各类的条件概率，然后把它分类为条件概率最大的那一类。同类样本的误分类代价可以是不同的，即代价矩阵中的元素不是常数。在现实应用中，可用“收益”替代“代价”，此时收益

$$B(i,j)=-C(i,j) \qquad (2)$$

对样本集中的每个样本 x，根据如下等式计算其属于任意一类 i 的代价

$$R(x,i)=\sum_{j}P(j\mid x)C(i,j) \qquad (3)$$

我们说样本 x 属于类 k，只要对于任意给定的类 i 如下不等式成立

$$R(x,k)\le R(x,i) \qquad (4)$$

在遗传神经网络中，利用 1/(cost + a)作为适应度函数。a 为一个取值较小的常数，代价越小，则网络的适应度越高，以此实现误分类代价最小化。

表 1　代价矩阵 C

	normal	dos	u2r	r2l	probe
normal	0	0.5	2	2	2
dos	2	0	1	1	1
u2r	0.5	0.5	0	1	1
r2l	0.5	0.5	1	0	1
probe	0.5	0.5	1	1	0

3.2 代价敏感遗传神经网络的设计

遗传神经网络的基本原理是用遗传算法对神经网络的结构和连接权值进行优化学习,利用遗传算法的寻优能力来获取最佳的网络。

本文采用的神经网络结构为三层神经网络结构，即输入层、隐含层和输出层。输入层的神经元节点数目(以下用 I 表示)由输入数据决定，隐含层节点数目[7](以下用 H 表示)则由经验公式：H＝2×I＋1 来确定，而输出层节点数目(以下用 O 表示)则由代价矩阵的维数确定，本文中代价矩阵为 5×5 的矩阵，所以输出层节点数目为 5。输入层－隐含层的权值数目(以下用 W 表示)为 W＝I×H，隐含层－输出层的权值数目(以下用 W’表示) W’＝H×O。

3.2.1 染色体编码和初始种群的产生

待优化的参数是一个三层前向神经网络所有的权重和阈值,由神经网络 I 个输入层节点，H 个隐含层节点和 O 个输出层节点可知，将有（I×H＋H×O）个待优化参数。以网络的各个权值和阀值做为基因，每个网络的各个基因组成染色体向量 V=[v1,…,vk,…vn],vk 为染色体中第 k 个基因，对应的是神经网络的第 k 个权值，采用实数形式，即用权值或阀值的实际值。对于给定结构的前馈型网络，记其权值（或阀值）为 v1，…，vk，…，vn,组成染色体 V,其中 vk 为第 k 个权值，然后随机产生初始种群 V1，V2，…， Vpop size，pop size 为初始种群中染色体个数。

3.2.2 代价嵌入与染色体遗传算子操作

通过计算适应度函数之后，对种群进行选择、交叉和变异[6]等操作。本文中所考虑的入侵检测中的误分类代价问题，因此，适应度函数设置为 f = 1 / (cost + a)，这是为了克服传统遗传算法中适应度函数的不确定性，利用欧氏范数所得出的一个表达式。cost 代表误分类之后的代价，由于样本的正确分类之后代价为 0，即 cost 等于 0，因此加入一个常量 a，本研究中设

为 0.1。0.1 是基于大量实验得出的一个比较理想稳定的值。

1) 选择：利用适应度大小，对种群进行选择，以选择较优的个体进入下一代。

2) 交叉：在父代种群中随机选取两个个体进行配对，然后随机选取交叉位，交换码串得到新个体。本文采用算术交叉，算术交叉定义为两个染色体的如下组合方式：

$$V1' = \lambda V1 + (1-\lambda)V2 \quad (5)$$

$$V2' = \lambda V2 + (1-\lambda)V1 \quad (6)$$

算术交叉可以保证产生的后代位于两个父代染色体之间。

3) 变异：变异是采用以给定的变异概率 p。选择变异位对个体进行变异。目的是增加种群的多样性，避免陷入局部最优。

4 实验数据分析

4.1 入侵检测系统和算法的评估标准

到目前为止，评价入侵检测系统和算法性能的主要指标有：1)准确性 2)处理性能 3)完备性以及 4)及时性。由此定义了入侵检测系统的两个问题：

① 误报率：是指入侵检测系统错误地将系统的正常活动定义为入侵行为。

② 漏报率：是指入侵检测系统将入侵行为定义为系统的正常行为[5]。

为了入侵检测系统更为有效，显然误报率和漏报率应越小越好。考虑到误报和漏报的代价是不一样的，所以前面引入了代价敏感的概念，以使系统的损失减至最小。

4.2 实验环境及数据分析

本文采用现有的 KDD99 数据来进行实验。整个的数据收集过程是在一个模拟的网络环境中进行。

我们首先对测试数据进行多次测试。首先随机生成神经网络种群，经过遗传算法对神经网络的优化，保留适应度大的网络作为下一代的网络种群，并在此基础上通过对优秀父代个体的交叉产生新的子代个体。

本文通过对比实验，即加入误分类代价与不加入误分类代价之间的结果进行对比。

实验结果如下：

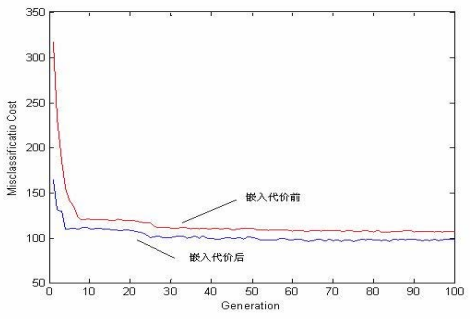


图 3(a) 误分类代价前后对比

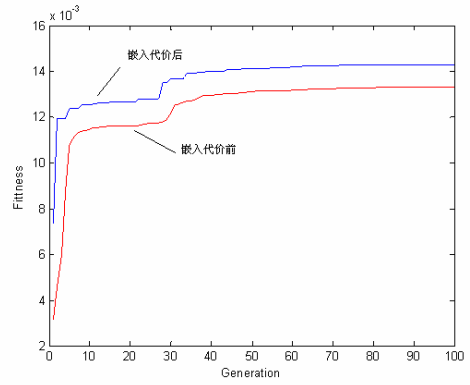


图 3(b) 分类精度前后对比

从上面的实验结果，我们可以看出，遗传神经网络解决了神经网络收敛速度慢的问题。随着遗传代数的增加，分类精度逐渐提高，误分类代价不断降低 。并且，到达一定的遗传代数之后，误分类精度维持在一定的数值保持不变，恰好证明了遗传算法已经达到或近似达到了最优解。

5 结束语

当样本的误分类代价不相等时，传统的基于精度的分类器不能实现代价最小值。本文通过把样本的不同误分类代价集成到遗传神经网络中，引入代价矩阵 C，从而实现了样本不同误分类代价的最小值。

实验结果同时也表明，引入了代价敏感的遗传神经网络的误分类代价比传统的分类器有更加小的代价值。

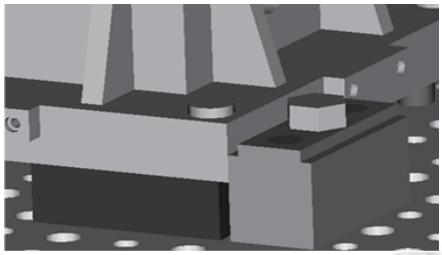
参考文献

1 Sherif JS. Intrusion Detection System and Models Computer Society. Proc. of the Eleventh IEEE International Workshops

(下转第 48 页)

图 6(b)所示，然后通过相邻功能元之间的耦合进行回溯。

最终得出原始的需求原型，在此基础上进行生长型设计得出另两种夹具结构方案，如图 7 所示。同样根据功能原型可得到更多的设计方案。然后可结合系统的力学综合等模块以及制造工艺的知识原理进行评价选出最佳方案。



(a)　反求夹具结构一

(b)　反求夹具结构二

图 7　反求后新型组合夹具

5　总结

Pro/Engineer 作为当今三维建模软件的优秀代表，其参数化造型及统一数据库的建模思想为机械结构设计带来了巨大的方便。同时利用其支持自定向下的概念设计，可以进一步提高设计的效率。其自带 Pro/TOOLKIT 开发接口为设计者提供了开发平台，在此基础上结合功能表面、分解重构原理、广义定位原理，开发了产品结构生长型设计理论的产品反求系统，从而为产品的创新设计提供一条实用的技术途径。

参考文献

1 刘影,杭九全,万耀青.反求工程与现代设计.机械设计,1998,15(12):1−4.

2 杨铁牛.面向逆工程的原始设计参数还原的研究与实践[博士学位论文].西安:西安交通大学,2000.

3 曹树坤.机械产品概念结构生长型设计力学综合技术研究[博士学位论文].济南:山东大学,2002.

4 杨志宏.产品公差与概念结构同步综合进化设计理论与技术研究[博士学位论文].济南:山东大学,2003.

5 黄克正,艾兴,张承瑞.复杂曲面的分解重构原理及其应用.中国科学 E 辑,1997,27(1):89−96.

6 陈洪武,黄克正,杨波.基于功能表面的产品结构设计自动化研究与实现.机械设计与研究,2004,20(3):24-27.

7 吴立军,陈波.Pro/Engineer 二次开发技术基础.北京:电子工业出版社,2005.

8 张继春.Pro/Engineer 二次开发实用教程.北京:北京大学出版社,2003.

9 李剑峰,陈建,李方义.Pro/TOOLKIT 技术及其在 Pro/Engineer 二次开发中的应用.网络与信息化,2003,(5):39−43.

(上接第 51 页)

on Enabling Technologies Infrastructure for Collaborative Enterprises (WETICE' 02). 2002.

2 李昆仑,黄厚宽,田盛丰,等.入侵检测的 1 类支持向量机模型.中国安全科学学报,2003,13(6):72−75.

3 Goldbergde. Genetic Algorithms in Search, Optimization and Machine Learning. New York: Addison-Wesley, 1989.

4 Han J, Kamber M. Data Minig: Concepts and Techniques. San Francisco CA: Morgan Kaufmann, 2001.

5 闻新,周露,王丹力.神经网络应用设计.北京:科技出版社,2001.172−158.

6 杨平,郑金华.遗传算子的比较和研究.计算机工程与应用,2007,43(15):59−65.

7 Dudaro, Hart PE, Storkdg. Pattern Classification. 2nd Ed. New York:Wiley, 2001. 311.