第 45 卷 第 10 期 计 算 机 工 程 2019 年 10 月
Vol. 45 No. 10 Computer Engineering October 2019

·安全技术· 文章编号：1000-3428(2019)10-0183-06 文献标志码：A 中图分类号：TP309

基于支持向量机与 Adaboost 的入侵检测系统

池亚平[1,2]，凌志婷[1,2]，王志强[1]，杨建喜[1]

（1. 北京电子科技学院 网络空间安全系，北京 100070；
2. 中国科学院信息工程研究所 中国科学院网络测评技术重点实验室，北京 100093）

摘 要： 入侵检测系统在大数据量的情况下误报率高、泛化能力弱，且单一机器学习算法不能较好地应对多种攻击类型。为此，设计一个基于支持向量机（SVM）与 Adaboost 算法的入侵检测系统。依托 Snort 系统，利用主成分分析方法对提取的特征做降维处理，并将 SVM-Adaboost 集合算法作为检测引擎。采用 NSL-KDD 数据集进行训练和测试，实验结果表明，该系统的正确率达到 97.3%，较 SVM 算法和 Adaboost 算法分别提高 4.8% 和 14.3%。

关键词： 支持向量机；Adaboost 算法；数据降维；入侵检测系统；接受者操作特征曲线

中文引用格式： 池亚平，凌志婷，王志强，等. 基于支持向量机与 Adaboost 的入侵检测系统[J]. 计算机工程，2019，45(10)：183-188，202.

英文引用格式： CHI Yaping，LING Zhiting，WANG Zhiqiang，et al. Intrusion detection system based on support vector machine and Adaboost[J]. Computer Engineering，2019，45(10)：183-188，202.

Intrusion Detection System Based on Support Vector Machine and Adaboost

CHI Yaping[1,2]，LING Zhiting[1,2]，WANG Zhiqiang[1]，YANG Jianxi[1]

（1. Department of Cyber Space Security，Beijing Electronic Science and Technology Institute，Beijing 100070，China；
2. Key Laboratory of Network Assessment Technology of CAS，Institute of Information Engineering，
Chinese Academy of Sciences，Beijing 100093，China）

【Abstract】 The Intrusion Detection System（IDS）has high false alarm rate and weak generalization ability in the case of large amount of data，and the single machine learning algorithm can not cope with multiple attack types well. To address this problem，this paper designs an IDS based on Support Vector Machine（SVM）and Adaboost algorithm. It relies on Snort system，which uses Principal Component Analysis（PCA）method to reduce the dimension of extracted features and uses the SVM-Adaboost clustering algorithm as detection engine. NSL-KDD dataset is used for training and testing. Experimental results show that the accuracy of the proposed system reaches 97.3%，which is improved by 4.8% and 14.3% respectively compared with the SVM algorithm and Adaboost algorithm.

【Key words】 Support Vector Machine（SVM）；Adaboost algorithm；data dimension reduction；Intrusion Detection System（IDS）；Receiver Operating Characteristic（ROC）curve

DOI： 10.19678/j.issn.1000-3428.0051976

0 概述

随着互联网的普及和大数据技术的应用，网络传输的信息量大幅增加，网络安全业务需求也在快速增长。据 CNCERT 监测，2017 年我国境内遭受 DDoS 攻击严重，攻击峰值流量持续攀升，攻击者的攻击手段也在不断更新换代，例如 2017 年 5 月 12 日 WannaCry 蠕虫病毒事件爆发，随后便迅速出现了多款变种[1]，网络安全形势严峻，对其防护技术的要求也越来越高。入侵检测作为网络安全技术之一被广泛应用，现有的入侵检测系统（Intrusion Detection System，IDS）主要以误用检测的模式匹配技术为主，以异常检测技术为辅。误用检测通过提取各种已知入侵行为特征并存为特征库，将数据流量与特征库进行对比从而判断是否为入侵行为。这种基于规则匹配的入侵检测系统在面对海量数据时效率较低，存在误报率高、泛化能力弱的问题。因此，如何降低误报率、提高泛化能力是网络安全领域内的重要研究课题[2]。

随着人工智能和机器学习技术的不断发展，利用机器学习来处理入侵检测系统的日志，达到从海量数据中提取关键信息的目的，对提高检测精度、降低误报率十分必要[3]。此外，机器学习能够将数据

基金项目： 国家重点研发计划“云计算与大数据”重点专项（2018YFB1004101）。
作者简介： 池亚平（1969—），女，教授，主研方向为云计算安全、可信网络；凌志婷（通信作者），硕士；王志强，讲师；杨建喜，副教授。
收稿日期： 2018-07-02 **修回日期：** 2018-08-15 **E-mail：** guaiguaigo123@163.com

分析得到的信息转化为认知、挖掘数据的统计规律，输出的模型或规则对未知的网络异常数据有较好的预警能力和泛化能力。现有研究表明，应用单一的机器学习算法无法很好地抵御错综复杂的攻击形式，而对多种机器学习算法进行有效集成，使其共同发挥作用，则能大幅提高检测效果[4]。在入侵检测中引入集成学习方法，可在先验知识不足的情况下仍保证有较好的分类正确率，从而使得入侵检测系统具有较好的检测性能[5]。

本文设计一种集成多种机器学习算法的入侵检测系统，并利用 NSL-KDD 数据集[6]对其进行训练与测试。该系统以 Snort 入侵检测系统为捕包、匹配及告警工具，应用主成分分析（Principal Component Analysis，PCA）方法进行数据降维，同时集成支持向量机（Support Vector Machine，SVM）与 Adaboost 算法，将 SVM 作为基分类器进行样本预分类，利用 Adaboost 根据预分类的正确率调整权值。

1 数据降维与特征提取

入侵检测常用的数据集为日志信息，包含多种不同特征，数据量大、维数多、计算复杂度高。因此，在利用数据集进行训练之前，必须先对训练数据集进行数据降维。传统的数据降维方法有 2 种：特征选择和特征选取。

PCA 是一种通用的基于特征提取的降维工具[7]，它通过对协方差矩阵进行特征值分解，选取较大特征值所对应的特征向量组成投影矩阵，用原始数据矩阵进行投影并得到降维后的新数据矩阵，将高维的数据通过线性变换投影到低维空间上，使不同维度间的相关性尽可能小，方差尽可能大，从而达到数据降维、降噪、去冗余的目的。

本文利用 PCA 处理数据的过程如下：

1）利用 PCA 对原始数据集进行处理。

2）将降维后的数据分为训练数据和测试数据，分别作为训练集和测试集。

3）选取不同维度的 PCA 利用测试集进行测试，得到最高精度的降维维度。

4）选取最优的降维维度形成 PCA 数据降维特征提取模块。

2 系统架构与 SVM-Adaboost 模型构建

设计入侵检测系统的目的是通过收集和分析网络流量，检测出可能为异常的行为或被攻击的迹象，并与防火墙等其他网络安全防御措施共同实行防护，对入侵行为进行告警及拦截。为达到该目的，本文设计系统架构如图 1 所示。

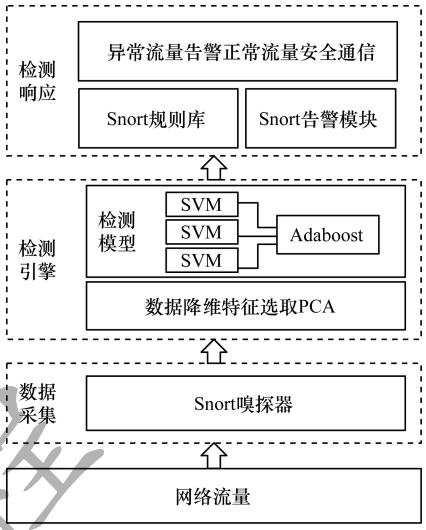

图 1 应用 SVM-Adaboost 模型的入侵检测系统架构

结合目前主流的 Snort 入侵检测系统，数据采集模块利用其嗅探功能监听、捕获数据包并进行解析，检测引擎模块利用 PCA 数据降维和 SVM-Adaboost 检测模型进行分类判断，检测响应模块利用 Snort 规则匹配功能根据配置的规则及 SVM-Adaboost 模型判断结果进行匹配，并利用 Snort 告警日志功能对异常流量产生告警并记录在日志中。

2.1 Snort 入侵检测系统

在众多入侵检测系统中，Snort 占主流地位。Snort 占用的资源少，功能强大，其以误用检测为主要的检测方式对已知攻击的特征模式进行匹配，是一种开源的轻量级网络入侵检测系统（Network IDS，NIDS）。目前 Snort 使用插件模式，技术开源并通过不断的功能扩充、完善，发展已经相当成熟。

Snort 分为 3 种工作模式和 5 个模块。3 种工作模式分别为嗅探器、数据包记录器和网络入侵检测系统。在网络入侵检测系统模式中，用户可以根据自己的需求定义规则进行配置，由系统分析网络数据流并根据规则采取相应的响应。5 个模块包括数据包捕获模块、解码模块、预处理模块、规则匹配模块和输出模块，各模块之间的关系如图 2 所示。其中：数据包捕获模块实现对网络中的原始数据进行监听；解码模块主要记录各协议层次的相关信息以及一些与检测相关的数据，随协议的逐层分解对数据进行解析；规则匹配模块构造一套快速匹配的数据结构，由 detect.c 函数调用检测引擎，对已加载的规则链进行一次性检测，将其中匹配成功的数据包送入输出模块做进一步处理；输出模块根据用户的实际需求将数据包直接丢弃或者生成日志，Snort 输出模式可以将报警数据记录入 SQL 数据库中。

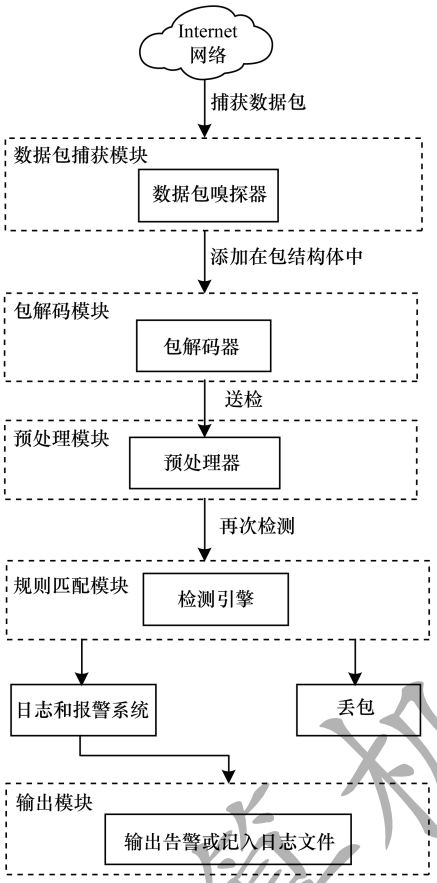

图 2　Snort 各模块关系

2.2　SVM 算法与 Adaboost 算法

SVM 算法是一种二分类算法,该算法通过找到一个超平面 $\tilde{\omega} \cdot x + b = 0$ 以及相应的分类决策函数 $f(x) = sign(\tilde{\omega} \cdot x + b)$ 将数据分隔在平面两侧,使 2 个类别的样本分开从而达到分类的目的。该算法既能有效处理非线性数据,又能限制过学习,同时具有严格的理论基础和数学基础,对样本数量的依赖性弱,又对于小样本学习应用具有很强的泛化能力[8],文献[9]通过对比分析 k 最近邻(k-Nearest Neigbour,kNN)、NaiveBayes、决策树(Decision Tree,DT)、SVM 和多层感知(Multilayer Perception,MLP)神经网络等多种机器学习算法在网络入侵检测中的应用,通过多个评价指标验证 SVM 的优越性。

Adaboost 是一种基于 Boosting 思想的机器学习算法,其将多个弱分类器进行合理结合,得到一个强分类器。该算法采用迭代思想,每次迭代只训练一个弱分类器,训练好的弱分类器将在下一次迭代中使用。Adaboost 中包括 2 种权重:数据权重和弱分类器权重。数据权重主要用于弱分类器寻找其分类误差最小的决策点,找到之后用这个最小误差计算出该弱分类器的权重;分类器权重越大说明该弱分类器在最终决策时拥有更大的发言权。Adaboost 是处理多分类问题的重要方法之一,具有高准确率和低复杂度[10],能够很好地利用弱分类器进行级联。

2.3　SVM-Adaboost 模型构建

检测引擎中的 SVM-Adaboost 检测模型是入侵检测系统的重要组成部分,集成了 2 种机器学习的算法,单纯的 SVM 算法虽然对小样本学习具有很强的泛化能力,但对于多种攻击方式的精度不高,而 Adaboost 可以解决多分类问题,将弱分类器组合为强分类器,提高准确率。基于此,本文将 SVM 作为弱分类器,每个 SVM 对应一个特征,一组 SVM 代表所有选取的特征组成一个基分类器,Adaboost 将多个基分类器组合,利用 SVM 和 Adaboost 各自的优势构建 SVM-Adaboost 模型,如图 3 所示。

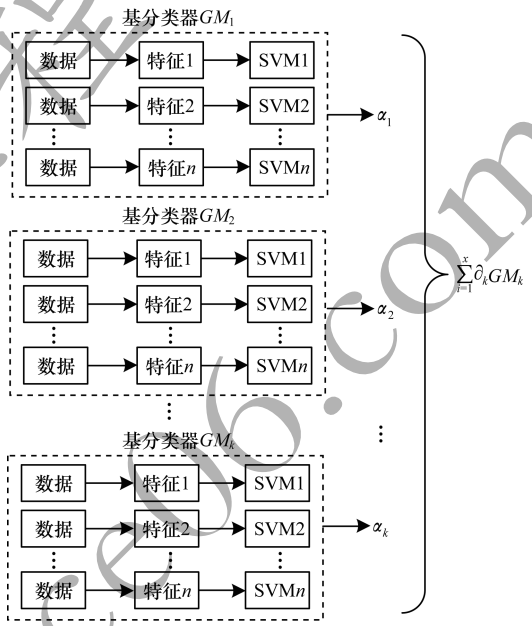

图 3　SVM-Adaboost 模型架构

在训练过程中输入训练集,每个特征用一个 SVM 弱分类器进行分类,计算每个 SVM 分类结果的错误率,所有特征对应的多个 SVM 组成一个基分类器,并由这些 SVM 中最小的错误率作为基分类器的错误率,计算基分类器的权重 ∂_k,由 ∂_k 更新数据集权重,经过多次迭代得到最终的分类器 $\sum_{i=1}^{x}\partial_k GM_k$。

在测试过程中输入测试集,遍历各基分类器,在遍历过程中由各基分类器中的弱分类器 SVM 进行预测分类,统计分类结果并对结果进行投票,选投票最高的作为基分类器的预测分类,对各分类器的分类结果进行加权累加,得到最终的分类判断。

对于 PCA 数据降维,SVM-Adaboost 模型训练过程如下:

1)输入训练数据集 $T = \{(x_1, y_1), (x_2, y_2), \cdots, (x_N, y_N)\}$,其中,$x_i \in X \subseteq \mathbb{R}^n$,$y_i \in Y = -1, 1$,预处理数据用 PCA 降维,选择 n 个特征,设迭代次数为 $k = 1, 2, \cdots, K$。

2)初始化训练样本的权重为:

$$D_1 = (\tilde{\omega}_{11}, \tilde{\omega}_{12}, \cdots, \tilde{\omega}_{1i}), \tilde{\omega}_{1i} = \frac{1}{N}(i = 1, 2, \cdots, N)$$

3)对于 $k=1,2,\cdots,K$:

每个特征用一个 SVM 进行分类,SVM 看作弱分类器 $R_S(x_i)$,共有 n 个 $R_S(x_i)$,计算每个 $R_S(x_i)$的错误率 $e_{k,s}=\sum_{i=1}^{N}\tilde{\omega}_{k,i}I(R_S(x_i)\neq y_i)(s=1,2,\cdots,n)$,得到最小的错误率 $e_{k\min}$。

将一组 n 个 $R_S(x_i)$中错误率最小的作为一个基分类器 $G_k(x_i)$,该基分类器的权重为:$\partial_k=\frac{1}{2}\lg\frac{1-e_{k\min}}{e_{k\min}}$。

更新训练数据集的权重:

$$D_{k+1}=\tilde{\omega}_{k+1,i}=\frac{\tilde{\omega}_{k,i}}{z_k}\times\exp(-\partial_k y_i G_k(x_i))$$

其中,归一化因子 z_k 为 $z_k=\sum_{i=1}^{N}\tilde{\omega}_{k,i}\exp(-\partial_k y_i\ G_k(x_i))$,将样本权重带入步骤 3)求弱分类器的错误率及基分类器权重。

4)得到最终的分类器 $F(x)=\mathrm{sign}(\sum_{k=1}^{K}\partial_k G_k(x))$。

对于 PCA 数据降维,SVM-Adaboost 模型测试过程如下:

1)训练后已经得到了基分类器的权重 ∂_k,初始化得到的最终分类器。

2)输入测试数据集,遍历各基分类器 $G_k(x_i)$,遍历过程中由各基分类器中的弱分类器 $R_{S(x_i)}$进行预测分类,并统计分类结果,对分类结果进行投票。

3)将各基分类器中投票高的结果作为各基分类器的预测分类结果。

4)对各分类器的结果进行加权累加,由最终分类器得到分类判断。

5)计算准确率。

3 模型验证与对比

为验证系统模型的有效性,本文在 Windows 10、Python3.6.5 语言编程环境、2.6 GHz CPU 时钟频率、8 GB 内存环境下进行验证实验。在数据集选择方面,经过调研现有入侵检测数据集及该领域相关文献[11-14],为保证时效性和通用性,本文选取 NSL-KDD 数据集进行训练和测试;在对比分析方面,文献[12]结合 K-means 算法与朴素贝叶斯算法,利用前者进行特征选取,利用后者进行分类,文献[14]利用随机森林进行特征选取,再结合K-means ++和AdaBoost 进行分类,本文模型与文献[12,14]进行方案分析对比;在对比实验方面,本文在同一数据集上对 SVM 算法和 Adaboost 算法利用 Weka 进行正确率对比实验。

3.1 NSL-KDD 数据集

NSL-KDD 数据集是为了解决 KDD CUP99 数据集中的固有问题[15]而提出的数据集,该数据集的训练集和测试集的记录数量比较合理,与 KDD CUP99 相比,其在训练集中不包括冗余记录,因此,训练出的分类器不会更偏向于出现频繁的攻击种类,并且选中攻击并录入数据集的数量百分比与发现该攻击的难度成反比,使不同的机器学习方法在分类结果上有更大范围的不同,这对于对比不同算法的正确率评估更有效。

NSL-KDD 数据集的 41 个特征如表 1 所示,其中 4 种大类攻击类型分别是拒绝服务攻击(Denial of Service, DoS)、来自远程主机的未授权访问(R2L)、未授权的本地超级用户特权访问(U2R)和端口监视或扫描(PROBE),在 4 种攻击类型下共有 39 种不同的小攻击类型。

表 1 NSL-KDD 数据集特征

序号	类型	特征
1		duration
2		protocol_type
3		service
4		flag
5	TCP 连接基本类型	src_bytes
6		dst_bytes
7		land
8		wrong_fragment
9		urgent
10		hot
11		num_failed_logins
12		logged_in
13		num_compromised
14		root_shell
15		su_attempted
16	TCP 连接的内容特征	num_root
17		num_file_creations
18		num_shells
19		num_access_files
20		num_outbound_cmds
21		is_hot_login
22		is_guest_login
23		count
24		srv_count
25		serror_rate
26		srv_serror_rate
27	基于时间的网络流量统计特征	rerror_rate
28		srv_rerror_rate
29		same_srv_rate
30		diff_srv_rate
31		srv_diff_host_rate
32		dst_host_count
33		dst_host_srv_count
34		dst_host_same_srv_rate
35		dst_host_diff_srv_rate
36	基于主机的网络流量统计特征	dst_host_same_src_port_rate
37		dst_host_srv_diff_host_rate
38		dst_host_serror_rate
39		dst_host_srv_serror_rate
40		dst_host_rerror_rate
41		dst_host_srv_rerror_rate

本文随机选取 10 000 条数据作为训练集、10 000 条数据作为测试集,训练集测试集各攻击数据条数占比如图 4、图 5 所示。

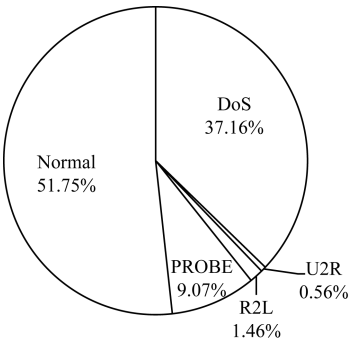

图 4　训练集数据占比

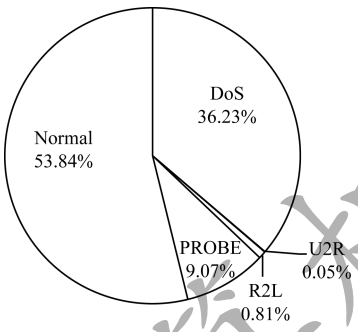

图 5　测试集数据占比

3.2　评估标准

在机器学习中有一个普遍适用的工具混淆矩阵,可以帮助了解分类过程中的错误及评估分类器的好坏,本文使用的混淆矩阵如图 6 所示。其中,真阳性率 *TP* 为将正常数据正确判断为正常类型,假阳性率 *FP* 为将正常数据错误判断为攻击类型,真阴性率 *TN* 为将攻击数据正确判断为攻击类型,假阴性率 *FN* 为将攻击数据错误判断为正常类型。

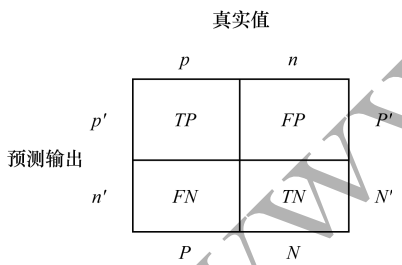

图 6　混淆矩阵

模型的正确率 A_{Accuracy} 即模型预测的精度,定义如式(1)所示,表示模型预测正确的个数占样本总个数的比例。一般情况下,模型的精度越高表明模型效果越好。

$$A_{\text{Accuracy}} = \frac{TP + TN}{TP + FN + FP + TN} \tag{1}$$

真阳率或召回率表示模型预测为正类样本的数量与总正类样本数量的比值,定义如式(2)所示。一般情况下,召回率越高说明有更多的正类样本被模型预测正确,模型的效果越好。

$$TPR = \frac{TP}{TP + FN} \tag{2}$$

假阳率或假正率 *FPR* 表示模型预测为负类样本的数量与总负类样本数量的比值,定义如式(3)所示。一般情况下,假阳率越低,说明模型效果越好。

$$FPR = \frac{FP}{FP + TN} \tag{3}$$

接受者操作特征(Receiver Operating Characteristic,ROC)曲线是一种对于灵敏度进行描述的功能图像,ROC 曲线以真阳率 *TPR* 为纵坐标,以假阳率 *FPR* 为横坐标,通过描述这两个特征的变化情况来评价分类器的好坏,在理想的情况下,最佳的分类器 ROC 应该尽可能处于左上角,这就意味着分类器在假阳率很低的同时获得了很高的真阳率。对不同的 ROC 曲线进行比较的另一个指标是曲线下的面积(Area Under the Curve,AUC),AUC 给出的是分类器的平均性能值,一个完美的分类器的 AUC 为无限趋近于 1。ROC 曲线和 AUC 常被用来评价一个分类器的优劣。

3.3　验证过程与评估结果

在利用 PCA 进行数据降维和特征选取的过程中,不同维数与正确率的关系曲线如图 7 所示,通过实验确定在将数据集由 41 维降为 10(或 9)维时正确率最高。

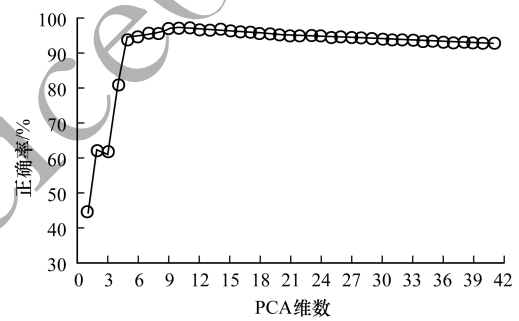

图 7　PCA 维数与正确率的关系曲线

通过实验 PCA 选取维数为 10 对数据进行预处理,经过 1 500 次迭代,模型正确率逐渐趋近平稳,约为 97.3%,利用 Python 调用 Matlab 的 API 绘制每轮迭代与正确率的关系曲线如图 8 所示。

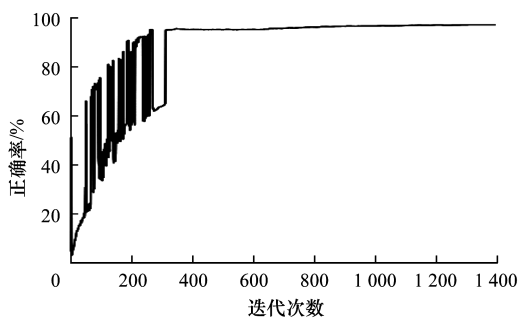

图 8　迭代次数与正确率的关系曲线

利用 Python 调用 ROC 相关函数绘制迭代 1 500 次时 ROC 曲线,并求出 AUC 的值约为 0.879,如图 9所示。

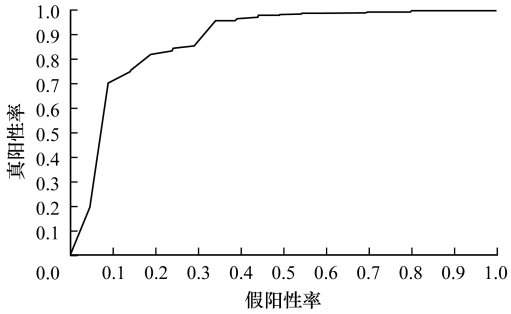

图 9 ROC 曲线

3.4 对比分析

本文改变 NSL-KDD 训练集的数据条数,利用本文提出的模型测试正确率如图 10 所示,可以看出,随着训练集数据量的增加,正确率在上升并趋于稳定至 97.3% 。

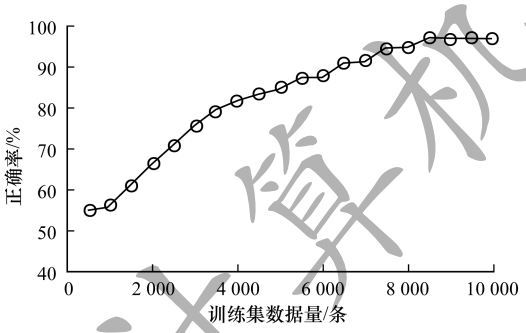

图 10 训练集数据量与正确率的关系曲线

本文利用 NSL-KDD 数据集与利用 SVM 算法和 Adaboost 算法对不同数据类型的检测正确率进行比较如表 2 所示。通过比较可知,本文算法较单一的 SVM 算法和单一的 Adaboost 算法在正确率上有明显提高。

表 2 正确率对比 %

算法	DoS	U2R	R2L	PROBE	Normal	总计
本文	92.15	99.80	95.89	93.60	99.55	97.30
SVM 算法	91.02	85.50	82.19	80.82	94.30	92.50
Adaboost 算法	91.80	73.21	75.34	69.65	91.81	83.00

文献[12]设计一种基于机器学习分类的入侵检测系统,该系统使用 NSL-KDD 数据集,在对数据进行预处理的过程中采用规范化处理,将数据特征的值缩放在一定范围内,将连续变量离散化,利用 K-means 算法进行聚类,再利用信息增益进行特征选择,并使用 Naive Bayes 算法进行分类。评估结果表明,在对连续变量离散化并进行特征选择的过程中,使用 K-means 聚类算法可以优化 Naive Bayes 对入侵攻击分类的结果。该方案只在数据预处理、特征选择及分类的过程中分别使用不同机器学习算法,并没有进行算法的集成。

文献[14]利用软件定义技术的优势和人工智能结合提出一个基于软件定义 5G 架构下的智能入侵检测系统,该系统使用 Radom forest 算法来选择基于流特征的最优子集,再将选择的特征作为输入,利用 k-means ++ 和 Adaboost 混合聚类算法把流量分成不同的攻击类,并用 KDD CUP99 和 NSL-KDD 数据集进行了验证,评估结果表明该系统有高检测精度低时间消耗,但是该模型在传统网络的应用中没有给出说明。

文献[12]与文献[14]在不同网络环境下进行设计实现,前者解决传统网络环境中的入侵检测问题,而后者针对软件定义网络进行设计,在算法的选择上,前者在过程中利用 2 种算法分别进行聚类与分类,后者利用 2 种算法进行混合聚类,两者在数据集的选择上与均与本文相同但是数据数目不同。

4 结束语

入侵攻击行为具有隐蔽性高和更新快的特点,使传统的入侵检测系统难以有效防范,并且单一的机器学习方法并不能很好地应对多种攻击类型。为此,本文设计基于 SVM 和 Adaboost 的入侵检测系统。该系统以 Snort 为依托,通过 PCA 进行数据降维和特征选取,并利用 SVM 对小样本数据检测精度高的优势及 Adaboost 能够将弱分类器组合成强分类器的功能,集合成 SVM-Adaboost 模型并组成检测引擎应用于入侵检测系统。验证评估与对比实验结果表明,本文系统可有效提高整体检测精度,降低误报率,但其对部分数据类型检测的精度较低,而且本文尚未考虑除传统网络以外的其他网络应用场景,下一步将对此进行研究。

参考文献

[1] 国家计算机网络应急技术处理协调中心. 2017 年互联网网络安全态势综述[EB/OL]. [2018-04-10]. http://www.cert.org.cn/publish/main/upload/File/situation.pdf.

[2] 王有金. 数据挖掘在入侵检测系统中的应用研究[D]. 长春:吉林大学,2017.

[3] 刘华春,候向宁,杨忠. 基于改进 K 均值算法的入侵检测系统设计[J]. 计算机技术与发展,2016,26(1):101-105.

[4] ALMSEIDIN M,ALZUBI M,KOVACS S,et al. Evaluation of machine learning algorithms for intrusion detection system[C]//Proceedings of IEEE International Symposium on Intelligent Systems and Informatics. Washington D. C.,USA:IEEE Press,2017:277-282.

[5] 谭爱平,陈浩,吴伯桥. 基于 SVM 的网络入侵检测集成学习算法[J]. 计算机科学,2014,41(2):197-200.

[6] University of New Brunswick. NSL-KDD dataset[EB/OL]. [2018-04-10]. http://www.unb.ca/cic/datasets/nsl.html.

[7] 张义宏. 基于 PCA 的 BP 神经网络优化的研究与应用[D]. 沈阳:东北大学,2014.

(下转第 202 页)

(上接第 188 页)

[8] 童智靖. 不均衡数据下基于 SVM 的分类算法研究与应用[D]. 哈尔滨:哈尔滨工程大学,2011.

[9] SHAH S A R, ISSAC B. Performance comparison of intrusion detection systems and application of machine learning to Snort system[J]. Future Generation Computer Systems,2017,80:157-170.

[10] MAZINI M, SHIRAZI B, MAHDAVI I. Anomaly network-based intrusion detection system using a reliable hybrid artificial bee colony and AdaBoost algorithms[J]. Research Journal of Applied Sciences,2018,12(3):304-310.

[11] MEENA G,CHOUDHARY R R. A review paper on IDS classification using KDD99 and NSL KDD dataset in WEKA[C]//Proceedings of 2017 International Conference on Computer,Communications and Electronics(Comptelix). Washington D.C.,USA:IEEE Press,2017.

[12] EFFENDY D A, KUSRINI K, SUDARMAWAN S. Classification of intrusion detection system(IDS) based on computer network[C]//Proceedings of International Conferences on Information Technology, Information Systems and Electrical Engineering. Washington D.C., USA:IEEE Press,2018.

[13] PATIDAR R, SHARMA T, MORIWAL R. Investigation for effective anomaly-based incursion disclosure on NSL-KDD [J]. International Journal of Computer Science Information and Engineering, Technologies, 2017,3(3):1-5.

[14] LI Jiaqi, ZHAO Zhifeng, LI Rongpeng. Machine learning-based IDS for software-defined 5G network [J]. IET Networks,2018,7(2):53-60.

[15] TAVALLAEE M,BAGHERI E,LU W. A detailed analysis of the KDD CUP99 data set[C]//Proceedings of the 2nd IEEE Symposium on Computational Intelligence for Security and Defense Applications. Washington D.C.,USA:IEEE Press,2009.

编辑　金胡考