基于支持向量机的机器学习IDS攻击样本的逼近代价分析

党军朋，陈运忠，李邦源

（云南电网有限责任公司玉溪供电局，云南 玉溪 653100）

摘 要：为了提高机器学习的能力，本文提出一种针基于支持向量机的机器学习IDS攻击样本的逼近代价方法，利用KKT条件把双层优化转变为单层优化的问题再对其实施求解，同时构建了单层优化的具体方法。当迭代处理λ取值为0.05与0.25的时候更易处于稳定收敛状态，λ取值为0.05时可以更快收敛；在不断的迭代过程中α值持续变小，设定取α＝1/t时逼近代价将达到一个最小收敛值。SVM攻击样本可以实现召回率的快速降低，并保持正常的流量状态，并获得比BPA方法更优的攻击效果。

关键词：机器学习；支持向量机；入侵检测；逼近代价

中图分类号：TP309.2；TP393 文献标识码：A 文章编号：1003−7241(2020)08−0032−04

Approximation Cost Analysis of Attack Samples for Machine Learning IDS Based on Support Vector Machine

DANG Jun-peng, CHEN Yun-zhong, LI Bang-yuan

(Yunnan Power Grid Limited Liability Company Yuxi Power Supply Bureau, Yuxi 653100 China)

Abstract: In order to improve the ability of machine learning, this paper proposes a method of approximate cost for IDS attack samples based on support vector machines. The KKT condition is used to transform the two-level optimization into a single-level optimization problem and then to solve the problem. At the same time, a specific method of single-level optimization is constructed. When the value of lambda is 0.05 or 0.25, it is more likely to be in a stable convergence state; when the value of lambda is 0.05, it is more likely to be in a faster convergence state. In the process of continuous iteration, the value of α keeps decreasing, and the approximation cost reaches a minimum convergence value when set a=1/t. SVM attack samples can achieve rapid reduction of recall rate, maintain normal flow state, and achieve better attack effect than BPA method.

Key words: machine learning; Support Vector Machine; intrusion detection; close cost

1 引言

当前，建立在机器学习基础上的人工智能技术获得了快速发展并被应用于许多领域，对于入侵检测系统(IDS)的安全性能提升起到了明显的促进作用[1]。在实际使用过程中，入侵检测系统都是被部署在容易受到外界攻击的网络环境中，也将其称作一种对抗环境[2]。部署IDS之后，对手将会先对IDS学习系统发起攻击[3]。在这种情况下，防止IDS学习系统受到攻陷就成为了一个关键研究内容。Ba等[3]构建得到了毒性攻击以及探测攻击共两种攻击模式。具体而言，毒性攻击是通过机器学习训练机制以及数据自适应的方式使训练数据集被污染，使IDS分类模型受到误导从而形成更易受到对手攻击的意图。

支持向量机(SVM)属于一类统计学习模型，根据前期研究资料可知，将SVM应用于入侵测试系统可以实现高度识别的性能[4−5]。虽然将SVM应用在入侵检测系统中可以实现多种优势，不过一些新的研究工作表明，SVM与

收稿日期：2019-01-29

神经网络模型都具有明显的安全突破点[6-8]。现阶段，IDS安全性获得了众多学者的分析，对网络流量实施毒性攻击的研究内容有滕翠等[9]构建的通过聚簇半径实现的异常检测模型，翟继强等[10]构建的以PCA子空间为依据的异常检测模型，本文提出了一种综合运用SVM和IDS的毒性攻击方法。同时，翟继强等[11]构建得到以SVM来实现的毒性攻击方法，具体过程是将问题通过模型转换为SVM损失最大化的优化问题，实现了对手写体识别过程的优异攻击效果。通过对毒性攻击进行建模得到一个篡改代价与目标逼近代价两者和达到最小的优化分析问题。考虑到数据篡改过程必然会受到SVM训练的限制作用，而SVM训练又则属于篡改之后得到的数据，因此互相之间形成约束，即属于双层优化的问题[12-15]。在分析SVM线性约束时，通过KKT条件把双层优化的问题转变成更易分析的单层优化过程。之后利用这些优化的方法从以上入侵检测数据集NSL中得到攻击样本，经测试发现本文构建的毒性攻击方法能够使大量的攻击数据被IDS误判成正常的流量。

2 基于SVM的IDS算法

根据SVM构建的IDS包含3部分内容，依次是预处理器、决策系统以及SVM分类器。进行预处理后可以把流量数据处理为SVM的形式。其中，SVM分类器属于IDS核心部分，需通过数据训练才能得到具有优异性能的SVM分类器。决策的作用是判断新产生的流量数据，并将其标记成“正常”、“入侵”两种不同类型。

可以将IDS所有工作过程分成二个不同阶段，具体包括训练与检测阶段。其中，在训练的时候可以利用已知与异常的数据来对SVM进行训练，由此获得支持向量与所需的参数。进行检测时，通过预处理器把未知流量数据加工为合适形式，再利用SVM分类器对其实施最终判断。

可以将双层优化的问题理解为NP问题，本文利用KKT条件把双层优化转变为单层优化的问题再对其实施求解，同时构建了单层优化的具体方法。

$$\arg \min_{X} \frac{1}{2}\left\|w-w^{*}\right\|_{2}^{2}+\frac{\lambda}{2}\left\|X-X_{0}\right\|_{F}^{2} \tag{1}$$

$$s.t. \sum_{i=1}^{N}\alpha_i y_i I(y_i(x_i^T w+b)\le 1)x_{ij}=w \tag{2}$$

上式中的I代表指示函数，可通过梯度下降的方法对其实施求解。考虑到优化变量是由数据集构成，应对其进行下述处理来完成求解过程，并得到如下的优化迭代式：

$$X_{t+1}=X_t-\alpha_t\nabla_X(\frac{1}{2}\left\|w-w^{*}\right\|_{2}^{2}+\frac{\lambda}{2}\left\|X-X_{0}\right\|_{F}^{2} \tag{3}$$

由于不能利用以上式子来直接求导X，可通过链式法将其表示成下述形式：

$$\nabla_X(\cdot)=\nabla_w(\frac{1}{2}\left\|w-w^{*}\right\|_{2}^{2}+\frac{\lambda}{2}\left\|X-X_{0}\right\|_{F}^{2}\frac{\partial w}{\partial X} \tag{4}$$

以上式子是根据SVM优化的方式获得。

3 实验

3.1 NSL数据集

通常可以利用NSL数据集对IDS系统进行测试，但在实际应用中会出现冗余、有偏的问题经改进后可以得到NSL数据集，在保留重要记录的前提下，使冗余得以去除，可以实现数据量大并保证无偏的优势。记录的特征数为42，最后一个特征属于类标签，通常将其标记成正常类型或特定攻击过程。表1给出了本文使用的数据集内含有的正常类型与攻击大类对应的记录比例。

表1 正常类型、攻击大类的数据统计

数据集	Normal	DoS	Probe	U2R	R2L
Train+	53.46%	36.46%	9.25%	0.04%	0.79%
Test+	43.08%	33.08%	10.74%	0.89%	12.22%

3.2 攻击样本逼近代价结果分析

本实验选择neptune攻击作为分析例子，由于篇幅有限，考虑到几种不同的攻击可以实现相近的测试结果。因此根据实际情况希望获得的攻击样本为：首先，正常类型的分类准确率应处于较高水平，防止被IDS发现；其次，可以明显减小neptune攻击检出率，由此实现躲避检测的效果。可以明显发现，惩罚系数λ以及搜索步长α发挥着非常重要的作用。采用不同的惩罚力度时将使算法敏感度存在明显差异，因此运用不同的λ值0.05、0.25、0.45来实施测试，通过交替优化方法来实现支持向量机w的逼近目标，从图1中可以看到进行迭代处理时逼近代价的变化规律。根据图1可知，在λ等于0.45的时候，前期的逼近代价没有得到稳定收敛状态，很容受到篡改代价的影响。当λ取值为0.05与0.25的时候更易处于稳定收敛状态，而λ取值为0.05时可以更快收敛。

通过自适应方法来设置搜索步长α。对图2进行分析可知，各步长设置方案下的目标逼近代价(Y轴)形成了不同的下降趋势。当α取0.01的很小值0.01时，前40个迭代周期内基本稳定，从第90个迭代期出现升高的情况，此时逼近代价开始显著升高。在α达到0.7的较大值时，前

40个迭代期中逼近代价发生了明显减小，处于一个收敛的变化趋势。进入第80个迭代期后，代价继续升高，导致收敛稳定性明显降低。测试不稳定收敛主要是因为步长太大时跳过了最优值，而当步长太小时则容易受到目标函数篡改代价的影响。通过实验测试可知，在不断的迭代过程中α值持续变小，设定取α＝1/t，此时逼近代价将处于一个稳定降低的过程，最后达到一个最小收敛值。

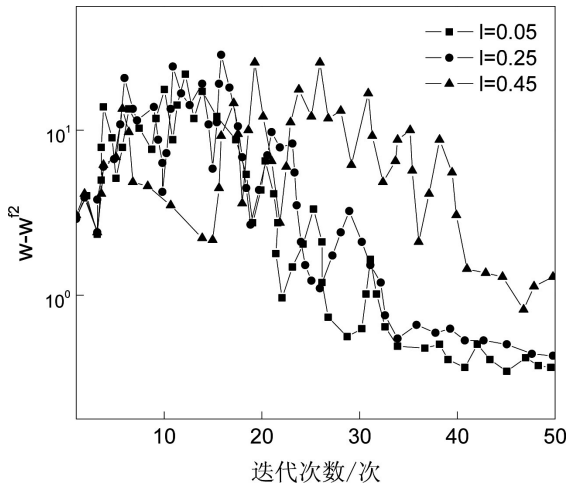

图1 不同l的逼近代价趋势

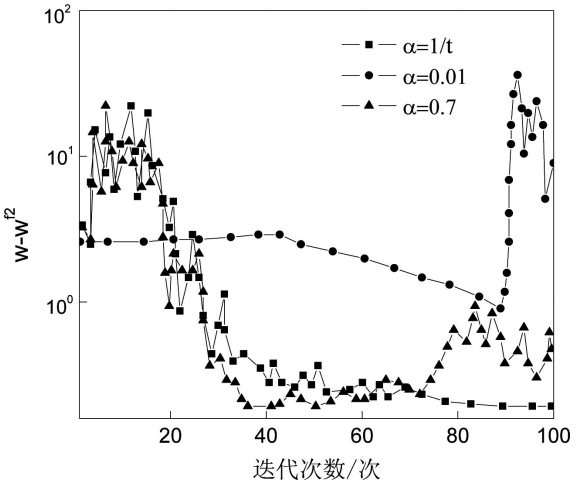

图2 不同α的逼近代价趋势

从表2中可以看到分别在攻击前与攻击后对应的SVM召回率和精度。首先，无法通过BPA方法来实现稳定的分类精度，其中neptune分类精度由85.3%增大至97.82%，但nmAP攻击由92.82%减小为54.02%。BOPA方法则实现所有攻击精度的提高，由此表明经过攻击后除了能够达到良好攻击效果，还可以保持正常流量，具有比BPA方法更优的稳定性。同时，两者都出现了召回率的减小，可见存在一定数量的攻击流量无法被IDS准确识别，而BOPA方式相对BPA发生了召回率的明显降低，特别是对于smurf攻击过程，BOPA方式由98.23%减小为57.06%，但BPA只减小到98.25%。根据以上分析可知，本文采用的毒性攻击方法存在很多攻击没有被IDS检测出来，具有比BPA方法更优的攻击效果。此外，各攻击方式训练原始样本SVM方面出现了不同的精度，neptune与satan依次等于85.63%和92.65%，差异得到了7%以上，攻击样本只有0.62%，对于BPA方式只有2.01%，这说明只有很少的正常流量发生误判，对于召回率则表现出了较大的差异。总体来看，SVM攻击样本可以实现召回率的快速降低，并保持正常的流量状态，并获得比BPA方法更优的攻击效果。

4 结束语

1）根据SVM构建的IDS包含3部分内容，依次是预处理器、决策系统以及SVM分类器。将IDS所有工作过程分成二个不同阶段，具体包括训练与检测阶段，本文利用KKT条件把双层优化转变为单层优化的问题再对其实施求解，同时构建了单层优化的具体方法。

2）迭代处理时逼近代价的变化规律可知，当λ取值为0.05与0.25的时候更易处于稳定收敛状态，而λ取值为0.05时可以更快收敛。在不断的迭代过程中α值持续变小，设定取α＝1/t，此时逼近代价将处于一个稳定降低的过程，最后达到一个最小收敛值。

3）SVM攻击样本可以实现召回率的快速降低，并保持正常的流量状态，并获得比BPA方法更优的攻击效果。

表2 原始样与攻击样本训练的SVM在测试集上的分类性能比较

攻击类型	原始样本		BOPA攻击M		BPA攻击	
	召回率	精度	召回率	精度	召回率	精度
neptune	99.06%	93.78%	74.47%	93.78%	74.47%	98.44%
smurf	98.10%	93.85%	93.86%	93.85%	93.86%	82.89%
satan	93.78%	94.01%	87.05%	94.01%	87.05%	98.06%
nmap	90.25%	94.37%	34.93%	94.37%	34.93%	55.19%

参考文献：

[1] 张彬.计算机网络安全技术分析[J].计算机产品与流通,2019(1):66.

[2] 卢明星.一种基于迁移学习的入侵检测技术的探讨[J].电子技术与软件工程,2018(24):179-180.

[3] 陈颖聪,陈智明,丘美景.对入侵检测系统恶意流量的分类研究[J].机电工程技术,2018(12):88-90.

[4] 马苗苗,何诺.一种入侵检测系统中告警信息量化评估方法[J].数据通信,2018(6):15-18.

[5] 王利.基于数据挖掘的网络数据库入侵检测系统[J].电脑知识与技术,2018(36):7-8.

[6] 徐海明,胡永亮,施雄健.基于气象部门网络安全体系构建的探究[J].信息技术与信息化,2018(12):116-118.

[7] 王禹程.抵抗Web攻击的异常入侵检测算法[J].电子设计工程,2018,26(24):126-130.

[8] 李挺,洪镇南,刘智勇,肖体正.基于增量单类支持向量机的工业控制系统入侵检测[J].信息与控制,2018,47(6):756-761.

[9] 滕翠,梁川.联动式网络安全系统的防御体系设计分析[J].现代信息科技,2018,2(12):174-175,181.

[10] 翟继强,马文亭,肖亚军.Apriori-KNN算法的警报过滤机制的入侵检测系统[J].小型微型计算机系统,2018,39(12):2632-2635.

[11] 翟继强,肖亚军,杨海陆,王健.改进的人工蜂群结合优化的随机森林的U2R攻击检测研究[J].信息网络安全,2018(12):38-45.

[12] 王美荣.数据平面和控制平面相分离的网络入侵检测系统[J].安庆师范大学学报(自然科学版),2018,24(4):40-43.

[13] 王博.优化遗传神经网络在入侵检测系统的研究和应用[J].电脑知识与技术,2018(34):168-169.

[14] 乔加强.计算机网络的安全威胁及防御技术[J].电子技术与软件工程,2018(22):188.

[15] 于家杰.基于预防策略的计算机网络防御技术分析[J].齐鲁师范学院学报,2018,33(6):71-77.

作者简介：党军朋(1988-),男,本科,技师,研究方向:变电站综合自动化系统运维。

(上接第5页)

参考文献：

[1] 崔宝珍.基于小波分析的滚动轴承故障诊断方法研究与应用[D].太原:中北大学硕士学位论文,2005.

[2] 林水泉,张清华,孙国玺,等.离心泵转子的故障诊断与修复[J].广东石油学工学院学报,2016,26(4):47-50.

[3] 高芮.基于神经网络的机械故障诊断技术的研究[D].青岛:青岛科技大学硕士学位论文,2016.

[4] 刘杰.基于小波分析和神经网络的模拟电路故障诊断方法研究[D].贵州:贵州大学2017届硕士研究生学位论文,2017.

[5] 周国宪,伍星,刘韬.基于多传感器的神经网络和D-S证据理论在故障诊断中的应用[J].测试技术学报,2017,31(4):290-297.

作者简介：林水泉(1988-),男,实验师,从事旋转机械故障诊断监测与诊断、转子动平衡等研究。

(上接第10页)

参考文献：

[1] 田景文,高美娟.人工神经网络算法研究及应用[M].北京:北京理工出版社,2008:119-132.

[2] 宋清昆,刘一.免疫遗传算法小波神经网络控制器设计[J].哈尔滨:哈尔滨理工大学学报,2015,20(4):55-59.

[3] 河东中,贡丽霞,白艳萍.小波变换和神经网络的电路故障诊断[J].现代电子技术,2020,43(10):30-35.

[4] PENGFEI LIANG,CHAO DENG,JUN WU,ZHIXIN YANG.Intelligent fault diagnosis of rotating machinery via wavelet transform,generative adversarial nets and convolutional neural network[J].Measurement,2020(5):15-17.

[5] 王月,刘亚秋,郭继峰,等.基于离散粒子群优化的云计算Qo S调度算法[J].计算机工程,2017,43(6):111-117.

作者简介：宋昌江(1983-),男,本科,副研究员,研究方向:计算机视觉与模式识别的技术研究与技术应用。