一种基于卷积神经网络的入侵检测方法

时东阁[1] 章晓庆[1] 毛保磊[1] 李润知[1] 林予松[1,2*]

[1](郑州大学互联网医疗与健康服务河南省协同创新中心 河南 郑州 450052)

[2](郑州大学软件学院 河南 郑州 450003)

摘 要 在网络流量较大及复杂入侵环境下,传统入侵检测系统检测能力弱且精度低。针对此问题,提出一种基于卷积神经网络的检测方法 CNN-Focal。利用卷积神经网络对数据进行特征提取,使用 Softmax 回归进行多分类,并采用 Focal loss 损失函数解决 NSL-KDD 数据集不平衡的问题。实验结果表明,CNN-Focal 的精度与 F1 评分分别达到 79.25% 和 76.9%,与其他机器学习算法相比,其精度和 F1 评分有显著提高。

关键词 网络安全 入侵检测 深度学习 卷积神经网络

中图分类号 TP393.08 **文献标志码** A DOI: 10.3969/j.issn.1000-386x.2020.10.052

AN INTRUSION DETECTION METHOD BASED ON CONVOLUTIONAL NEURAL NETWORK

Shi Dongge[1] Zhang Xiaoqing[1] Mao Baolei[1] Li Runzhi[1] Lin Yusong[1,2*]

[1](*Cooperative Innovation Center for Internet Healthcare, Zhengzhou University, Zhengzhou 450052, Henan, China*)

[2](*School of Software, Zhengzhou University, Zhengzhou 450003, Henan, China*)

Abstract Under the condition of the large network traffic and complex intrusion environment, the traditional intrusion detection system has weak detection capability and low precision. To solve this problem, this paper proposes a detection method based on convolutional neural network—CNN-Focal. It used the convolutional neural network to extract the features of the data, then used Softmax regression to perform multi-classification, and used the Focal loss function to solve the imbalance problem of NSL-KDD dataset. The experimental results show that CNN-Focal has accuracy of 79.25% and F1 score of 76.9%. Compared with other machine learning algorithms, it has a significant improvement in accuracy and F1 score.

Keywords Cyber security Intrusion detection Deep learning Convolutional neural network

0 引 言

随着科学技术的发展,互联网走进了千家万户,在给我们生活带来便利的同时,也带来了网络安全问题。工业界和学术界采取各种技术和措施解决网络安全问题,其中入侵检测系统(Intrusion Detection System,IDS)是网络安全体系的重要组成部分,一直以来备受网络安全领域学者的关注。IDS 作为防火墙后第二道安全屏障,从网络中收集信息并进行分析,从中发现违反安全策略的网络行为,并和其他设备联动做出响应。虽然当前人们信息安全意识在不断提高,但网络攻击的复杂性以及攻击手段的多样性,使得信息安全事件频繁发生。海量复杂且标签不平衡的入侵数据,给入侵检测带来了重大的挑战。如何有效地从入侵数据中选择特征进行多分类,并提高入侵检测的精度,在网络安全领域具有重要的研究价值和广阔的应用前景。

1 研究现状

目前,将机器学习和深度学习应用于入侵检测领域是个重要的课题。机器学习能对特征进行学习并发

收稿日期:2019-06-17。河南省高等学校重点科研项目(17A520059)。时东阁,硕士生,主研领域:网络安全,深度学习。章晓庆,硕士生。毛保磊,实验师。李润知,副教授。林予松,教授。

现重要特征,利用机器学习的方法可以将入侵检测转化为对网络中正常和异常行为分类的问题。Balabine 等[1]提出了 SVM 算法应用于网络流量异常检测的专利;Pan 等[2]和 Tahir 等[3]将K-means和 SVM 算法结合用于入侵检测;Ambusaidi 等[4]采用网络流量的相关性检测网络的恶意行为;聚类也是常用的入侵检测方法,基于聚类的入侵检测主要包含有常规聚类方法[5]和联合聚类方法[6]。上述方法虽然有较高的检测率,但是由于数据的维度较大,特别在数据处理阶段会依据专家经验进行特征提取,这不仅需要较长的时间来选择合适的特征,还可能破坏数据之间的相关性,从而漏掉一部分有效特征。

近年来,深度学习在语音识别、图像识别和自然语言处理等领域取得了不错的成果[7]。深度学习可以从原始特征提取出抽象的高层特征,不需要依据专家经验进行特征选择,因其强大的学习能力,国内外已有学者尝试将深度学习技术应用于网络安全领域中。文献[8]将 J48 决策树应用到 NSL-KDD[9]数据集上进行入侵检测;文献[10]将数据集的 70% 作为训练集,剩余部分作为测试集,利用 ANN 进行入侵预测;文献[11]将随机树应用于入侵检测中并取得了不错的成果。上述方法虽然取得了不错的效果,但是其在模型训练和测试时只使用了官方的训练集,具有一定的局限性。此外,官方测试集中攻击方法比官方训练集的攻击方法多 17 种,因此本文在模型的训练和测试中分别采用官方训练集和官方测试集,这样有利于提高模型的健壮性。

本文提出了一种基于卷积神经网络的入侵检测模型(CNN-Focal)。该模型将卷积神经网络(Convolutional Neural Network,CNN)中的门限卷积[12]和 Softmax[13]应用于入侵检测领域中进行多分类,针对不平衡数据集[14]采用 Focal Loss[15]损失函数进行优化,有效提升了入侵检测精度。

2 CNN-Focal 入侵检测模型

2.1 卷积神经网络基本原理

在深度学习领域中,卷积神经网络(Convolutional Neural Network,CNN)是一种高效的神经网络模型,已成为众多领域的研究热点。卷积神经网络的基本结构由输入层、卷积层、池化层、全连接层和输出层组成,如图 1 所示。在一个模型中卷积层、池化层和全连接层可以有一个或者多个,其中卷积层和池化层二者一般交替出现,即卷积层连接池化层,或池化层后连接卷积层,以此类推。全连接层一般紧随在池化层后面。

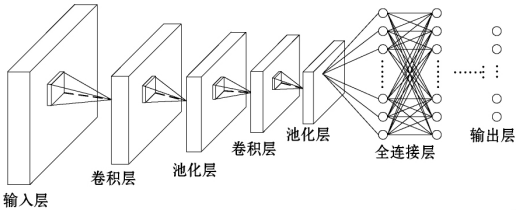

图 1 卷积神经网络基本结构

全连接层中每个神经元与前一层所有神经元进行全连接,位于 CNN 结构的末端。通常在全连接层之后是输出层,即分类层。输出层对卷积神经网络提取到的特征进行分类,分类的输出即为结果。

Softmax 回归是 Logistic 回归模型的推广,主要用于多分类。当进行 2 分类的时候,Softmax 回归会退化为 Logistic 回归。在多分类问题中,类标签 y 取两个以上的值,即类标签 y 有 k(k 为大于 2 的整数)个不同的值。给定的数据集 $\{(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)\}$,$y_i$ 表示第 i 类标签,$i \in \{1,2,\cdots,k\}$。对于给定的 x,Softmax 回归估计出 x 在 k 类标签中每一类的概率。

损失函数(Loss function)是用来评估模型的预测值 $f(x)$ 和真实值 Y 的差别程度,通常表示形式为 $L(Y, f(x))$。损失函数反映了模型的鲁棒性,即损失函数值越小,模型的鲁棒性越好。

2.2 网络结构

入侵检测问题是一个分类问题,可以通过有监督学习训练出分类模型,然后使用训练的模型对未知数据进行预测。在使用卷积神经网络时,输入层输入数据通常是二维的,入侵记录是一维数据,因此在卷积操作选择方面,本文采取一维卷积方法对入侵记录数据进行卷积操作。根据 NSL-KDD 标签不平衡和模型实际分类性能等情况,本文设计了 CNN-Focal 模型,其结构如图 2 所示。CNN-Focal 模型共有 10 层,1 个输入层、3 个卷积层、3 个 Dropout 层、1 个 Max-pooling 层、1 个全连接层和 1 个 Softmax 层。

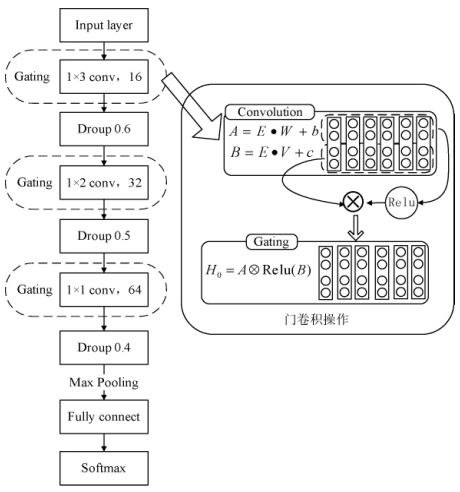

图 2 CNN-Focal 模型结构

该模型具体描述如下：

1）输入层：第一层为输入层。入侵记录是一维数据，经过数据标准化、one hot 预处理后，单条入侵记录数据由 1×41 转换为 1×122。

2）卷积层：第 2、4、6 层都是卷积层。在卷积层使用了门限卷积的概念，门限卷积中分为两部分：一部分是卷积的激活值，即 B；另一部分是直接线性得到卷积，即 A。A、B 两部分相乘得到相应的卷积值。已有不少文献证明较小的卷积核能得到更好的局部特征和分类性能[16]，因此在卷积核大小的设计方面，CNN-Focal 采用了小卷积核策略，卷积核大小分别为 1×3、1×2 和 1×1，卷积核的个数分别为 16、32、64。除此之外小卷积核可以对学习到的特征进行聚类，在一定程度上缓解卷积冗余对模型性能的影响。

3）Dropout 层：卷积神经网络模型在训练过程中容易出现过拟合现象，过拟合对模型实际性能影响很大。为了缓解此问题，CNN-Focal 模型中第 3、5 和 7 层都采用了 Dropout。Dropout 值的大小分别设置为 0.6、0.5 和 0.4，这是根据 CNN-Focal 模型实际分类效果进行设置的。

4）Max-Pooling 层：池化层能减小计算量，CNN-Focal 模型第 8 层为 Max-Pooling 层，stride 为 2，即参数数量减少为原来的一半。

5）全连接层：CNN-Focal 中全连接层神经元的个数为 200。

6）Softmax 层：在深度学习中，Softmax 回归通常作为标准的分类器用于多分类或二分类问题。CNN-Focal 采用 Softmax 回归作为多分类器。

2.3　优化策略

在模型训练中一般会遇到收敛速度慢、数据不平衡、耗费资源大等问题，因此需要引入相应的优化策略减少这些问题带来的负面影响。为了提高 CNN-Focal 模型的性能，本文采用以下三种优化策略。

1）批标准化（Batch Normalization）。在深度学习中，深度网络参数训练时内部存在协方差偏移（Internal Covariate Shift）现象。根据链式规则，随着层数的增加，偏移现象会被逐渐放大，这将影响网络参数学习。因为神经网络的本质是表示学习（Representation Learning），如果数据分布发生改变，神经网络不得不学习新的分布。批标准化对每一层都进行标准化处理，使输入样本之间不相关。通过标准化每一层的输入，使每一层的输入服从相同分布，因此克服内部协方差偏移的影响。批标准化可以看做对输入样本的一种约束，最大作用是加速收敛和提高模型分类性能。

2）损失函数优化。本文采用的入侵检测数据集的标签分布是不平衡的，数据集标签统计如图 3 所示。Focal Loss 函数适用于不平衡数据集，因此本文采用 Focal Loss 作为模型的损失函数。与其他损失函数相比，Focal Loss 损失函数修正了正负样本、难分和易分样本对损失的贡献量。

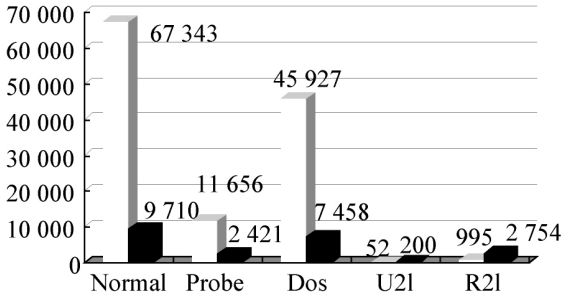

图 3　数据集标签统计

3）Adam 优化算法。Adam 算法[17]可以替代传统随机梯度下降过程的优化算法，能基于训练数据迭代地更新网络模型的权重。Adam 算法具有减少深度学习的时间以及占用计算机的资源等优势，因此被广泛应用于深度学习领域。CNN-Focal 模型采用 Adam 算法提高模型的性能。

3　实　验

3.1　实验目的

针对本文数据集不平衡问题，本文将 Focal Loss 损失函数应用到模型中。为了验证 Focal Loss 的有效性，本文将 Focal Loss 和深度学习中常用的交叉熵损失函数进行实验对比，即将 CNN-Focal 模型的损失函数换成交叉熵损失函数，更改损失函数后的模型记为 CNN-Cross。除此之外，为了验证 CNN-Focal 模型具有更好的分类效果，本文将 CNN-Focal 与 SVM、RandomForest、DecisionTree、GaussianNB 四种方法进行了对比实验。

3.2　实验方法

在入侵检测领域，NSL-KDD 数据集被广泛应用，本文将经过数据预处理的 NSL-KDD 数据集用于 CNN-Focal 以及对比模型的训练和测试。另外本文选取了精度、准确率、召回率、F1 评分等指标对模型进行评估。

1）数据集介绍。在入侵检测的研究中，KDD CUP 99 数据集是使用最为广泛的数据集。KDD CUP 99 数据集中训练集有约 500 万条记录，测试集有约 30 万条

记录，该数据集的数据量对实验硬件环境要求较高。另外，各种统计分析显示 KDD CUP 99 数据集中存在大量冗余的记录[18]，这将使模型出现过拟合现象，且在训练过程中需要更多的计算机资源，模型收敛缓慢。

NSL-KDD 数据集解决了 KDD CUP 99 数据集中存在的问题，目前有许多研究成果都是基于 NSL-KDD 数据集的。NSL-KDD 数据集中包含 41 列特征和 1 列标签。标签列分为 5 大类：Normal，Probe，Dos，U2L 和 R2L。

NSL-KDD 数据集中训练集共有 125 973 条（约 18.2 MB），测试集共有 22 543 条（约 3.2 MB）。这个数据集的数量级对实验硬件环境的要求不高，可以在普通机器上进行实验。另外，训练集和测试集中包含的攻击方法不同，所以使用该数据集训练出的模型对于新型的攻击具有较好的检测效果。

2）数据预处理。机器学习中一般使用标称型和数值型两种数据类型，NSL-KDD 数据集中 41 列特征属性值既有标称型又有数值型。该数据集中 protcol_type、service、flag 和 label 4 列属性值类型为标称型，剩余各列为数值型。

数据的标准化（Normalization）是把数据按比例缩放到一个特定区间，使之从一个大区间落入一个小区间。标准化后的数据具有缩短模型的收敛时间、提高模型精度等优势。本文对数值型特征进行标准化操作，使之缩放到 [0,1] 之间。具体如下：

$$x_{\text{norm}} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{1}$$

式中：x、$x_{\min}$、$x_{\max}$ 分别表示原始样本数据值、样本数据的最大值、样本数据的最小值。

one-hot 编码，又称独热编码，它不仅能处理非连续型数值的特征，也会让特征之间的距离更加合理。本文中使用 sklearn 包中的 OneHotEncoder 对 protcol_type、service、flag 和 label 4 列标称型数据进行 one-hot 编码。

3）评价指标。对模型进行评估，不仅需要有效可行的实验方案，还需要有衡量模型泛化能力的评价指标，这就是性能度量。在不平衡分类任务中，最常用的性能度量有精度（Accuracy）、准确率（Precision）、召回率（Recall）以及 F1 评分四种。本文以这四种指标来评估模型。

3.3 环境构建

本文实验环境采用了 Ubuntu 16.04 64 位操作系统、Intel Core i7 处理器、64 GB 内存。实验平台采用了 TensorFlow 1.9.0、Sklearn 0.20.1 等框架，在程序实现中使用了 Python 3.5.2 进行编程。

3.4 结果分析

本文所有实验均使用 NSL-KDD 数据集。CNN-Cross 和 CNN-Focal 两组对比实验的结果如表 1 - 表 3 所示。SVM、RandomForest、DecisionTree、GaussianNB 四组实验与 CNN-Focal 的实验结果的对比结果如表 4 - 表 6所示。

表 1 两种损失函数总体指标对比

编号	模型	Accuracy	Average precision	Average recall	F1
1	CNN-Focal	0.792 5	0.825 6	0.792 5	0.769 0
2	CNN-Cross	0.753 4	0.760 5	0.753 5	0.721 6

表 2 两种损失函数准确率对比

编号	模型	Precision				
		Normal	Probe	Dos	U2L	R2L
1	CNN-Focal	0.704 3	0.814 4	0.959 3	0.706 9	0.911 9
2	CNN-Cross	0.672 4	0.775 0	0.948 9	0.701 9	0.547 9

表 3 两种损失函数召回率对比

编号	模型	Recall				
		Normal	Probe	Dos	U2L	R2L
1	CNN-Focal	0.961 9	0.730 1	0.826 0	0.099 0	0.209 6
2	CNN-Cross	0.958 3	0.681 5	0.769 0	0.080 0	0.101 7

表 4 总体指标对比

编号	模型	Accuracy	Average precision	Average recall	F1
1	CNN-Focal	0.792 5	0.825 6	0.792 5	0.769 0
2	SVM	0.781 2	0.812 6	0.786 2	0.749 2
3	RandomForest	0.632 8	0.698 5	0.632 8	0.584 3
4	DecisionTree	0.601 9	0.695 1	0.601 9	0.548 6
5	GaussianNB	0.417 5	0.344 1	0.417 5	0.258 0

表 5 准确率对比

编号	模型	Precision				
		Normal	Probe	Dos	U2L	R2L
1	CNN-Focal	0.704 3	0.814 4	0.959 3	0.706 9	0.911 9
2	SVM	0.691 6	0.803 2	0.970 7	0	0.984 4
3	RandomForest	0.640 2	0.406 7	0.926 4	0.225 4	0.578 1
4	DecisionTree	0.636 9	0.367 0	0.897 2	0.141 7	0.681 1
5	GaussianNB	0.422 9	0	0.480 0	0	0.025 4

表 6　召回率对比

编号	模型	Recall				
		Normal	Probe	Dos	U2L	R2L
1	CNN-Focal	0.961 9	0.730 1	0.826 0	0.099 0	0.209 6
2	SVM	0.957 9	0.721 2	0.824 5	0	0.091 9
3	RandomForest	0.927 4	0.702 3	0.474 5	0.002 0	0.007 3
4	DecisionTree	0.921 5	0.742 1	0.374 3	0.003 5	0.011 7
5	GaussianNB	0.963 5	0	0.006 4	0	0.002 5

表 1 为 CNN-Focal 和 CNN-Cross 在精度、平均准确率、平均召回率和 F1 评分上的对比结果。可以看出，CNN-Focal 取得了 79.25% 的精度、82.56% 的平均准确率、79.25% 的平均召回率和 76.90% 的 F1 评分，均高于 CNN-Cross。

由表 2 可知，CNN-Focal 在 Normal、Probe、U2L、R2L 等五种类型上分别取得了 70.43%、81.44%、95.93%、70.69% 和 91.19% 的准确率，均高于 CNN-Cross。

由表 3 可知，CNN-Focal 在 Normal、Probe、U2L、R2L 等五种类型上分别取得了 96.19%、73.01%、82.60%、9.90% 和 20.96% 的召回率，均高于 CNN-Cross。

综合对比表 1 - 表 3 实验结果，Focal Loss 损失函数在本文数据集上的分类效果优于交叉熵损失函数。

表 4 列出了 CNN-Focal 模型和其他四种模型在精度、平均准确率、召回率和 F1 评分上的结果对比。可以看出，CNN-Focal 模型取得了 79.25% 的精度、82.56% 的平均准确率、79.25% 的平均召回率、76.90% 的 F1 评分，均高于其他四种模型。

由表 5 可知 CNN-Focal 模型在 Normal、Probe 和 U2L 三种类型的准确率上均高于其他四种模型。CNN-Focal 模型在 Dos、R2L 两种分类上准确率略低于 SVM 模型，而高于其他三种方法。但是 SVM 在 U2L 类别上的准确率为 0，即 SVM 没有检测出 U2L 类型的攻击，而文中设计的 CNN-Focal 模型达到了 0.706 94 的准确率。总体来说，CNN-Focal 模型在准确率上优于其他四种模型。

由表 6 可知，CNN-Focal 模型在召回率上均高于其他四种模型，尤其在 U2L 攻击类型的检测效果上比其他四种更有效。

综合表 4 - 表 6 实验结果，本文提出的 CNN-Focal 模型具有更好的分类效果，与其他方法相比，本文模型在取得较高的准确率的同时，也提高了不同攻击类型的检测准确率。

4　结　语

针对传统入侵检测系统在网络流量较大及复杂入侵环境下检测能力弱且精度低的问题，本文提出了 CNN-Focal 模型。实验结果表明，与 SVM、RandomForest、DecisionTree 和 GaussianNB 等方法相比，本文模型具有较高分类精度和 F1 评分。在未来工作中，将进一步优化模型结构，减少训练时间，提高分类效果，同时还将探索二维卷积模型在入侵检测中的应用。

参　考　文　献

[1] Balabine I，Velednitsky A. Method and system for confident anomaly detection in computer network traffic：US9843488 [P]. 2017 - 12 - 12.

[2] Pan X，Luo Y，Xu Y. K-nearest neighbor based structural twin support vector machine [J]. Knowledge-Based Systems，2015，88：34 - 44.

[3] Tahir H M，Hasan W，Said A M，et al. Hybrid machine learning technique for intrusion detection system [C] //"Computer Science for Improving the Quality of Life" International Conference on Computing and Informatics，2015.

[4] Ambusaidi M A，Tan Z Y，He X J，et al. Intrusion detection method based on nonlinear correlation measure [J]. International Journal of Internet Protocol Technology，2014，8(2/3)：77 - 86.

[5] Ahmed M，Mahmood A N. A novel approach for outlier detection and clustering improvement [C] //2013 IEEE 8th Conference on Industrial Electronics and Applications (ICIEA). IEEE，2013：577 - 582.

[6] Ahmed M，Mahmood A N，Maher M J. Heart disease diagnosis using co-clustering [C] //2014 International Conference on Scalable Information Systems. Springer，2014：61 - 70.

[7] 张军阳，王慧丽，郭阳，等. 深度学习相关研究综述 [J]. 计算机应用研究，2018，35(7)：1921 - 1928，1936.

[8] Chae H S，Jo B O，Choi S H，et al. Feature selection for intrusion detection using NSL-KDD [J]. Recent Advances in Computer Science，2013，20132：184 - 187.

[9] University of New Brunswick. NSL-KDD dataset [DS/OL]. [2019-06-17]. https：//www.unb.ca/cic/datasets/nsl.html.

[10] Naoum R S，Abid N A，Al-Sultani Z N. An enhanced resilient backpropagation artificial neural network for intrusion detection system [J]. International Journal of Computer Science and Network Security，2012，12(3)：11.

[11] Thaseen S，Kumar C A. An analysis of supervised tree based classifiers for intrusion detection system [C] //2013 International Conference on Pattern Recognition，Informatics and Mobile Engineering. IEEE，2013：294 - 299.

（下转第 333 页）

结果表明灰度直方图的随机森林、灰度共生矩阵的随机森林、N-Gram 的随机森林，以及融合特征的随机森林均可以有效地进行恶意代码的分类，其中 3 种特征融合后与随机森林算法相结合其分类效果显著提升。目前更多的是静态特征融合，从代码生成的灰度图纹理和恶意代码文本两方面为切入点，取得不错的分类效果。下一步将融合一些动态特征，比如恶意软件的行为特征，观察其是否可以进一步提高恶意代码分类的准确率，从而进一步优化分类器。

参 考 文 献

[1] 国家互联网应急中心(CNCERT). 2018 年我国互联网网络安全态势报告[OL]. [2019-04-18]. https://www.freebuf.com/articles/network/201280.html.

[2] 奇虎 360. 中国互联网安全报告[EB/OL]. 2018-08-02. https://www.freebuf.com/articles/paper/179295.html.

[3] 崔鸿雁，徐帅，张利锋，等. 机器学习中的特征选择方法研究及展望[J]. 北京邮电大学学报，2018，41(1)：1-9.

[4] 高程程，惠晓威. 基于灰度共生矩阵的纹理特征提取[J]. 计算机系统应用，2010，19(6)：195-198.

[5] 周绮凤，洪文财，杨帆，等. 基于随机森林相似度矩阵差异性的特征选择[J]. 华中科技大学学报(自然科学版)，2010，38(4)：58-61.

[6] 王卫红，朱雨辰. 基于 N-Gram 与加权分类器集成的恶意代码检测[J]. 浙江工业大学学报，2017，45(6)：604-632.

[7] Breiman L. Random forest[J]. Machine Learning，2001，45：5-32.

[8] Nataraj L，Karthikeyan S，Jacob G，et al. Malware images：visualization and automatic classification[C]//8th International Symposium on Visualization for Cyber Security. ACM，2011.

[9] 戴逸辉，殷旭东. 基于随机森林的恶意代码检测[J]. 网络空间安全，2018，9(2)：70-75.

[10] Kaggle[OL]. https://kaggle.com/c/maleware classification/.

(上接第 322 页)

[9] Rndic N，Laskov P. Practical evasion of a learning-based classifier：A case study[C]//2014 IEEE Symposium on Security and Privacy. IEEE，2014：197-211.

[10] 张思思，左信，刘建伟. 深度学习中的对抗样本问题[J]. 计算机学报，2019，42(8)：1886-1904.

[11] Papernot N，McDaniel P，Wu X，et al. Distillation as a defense to adversarial perturbations against deep neural networks[C]//2016 IEEE Symposium on Security and Privacy(SP). IEEE，2016：582-597.

[12] Carlini N，Wagner D. Defensive distillation is not robust to adversarial examples[EB]. arXiv：1607.04311，2016.

[13] Graese A，Rozsa A，Boult T E. Assessing threat of adversarial examples on deep neural networks[C]//2016 15th IEEE International Conference on Machine Learning and Applications(ICMLA). IEEE，2016：69-74.

[14] Shaham U，Yamada Y，Negahban S. Understanding adversarial training：Increasing local stability of supervised models through robust optimization[J]. Neurocomputing，2018，307：195-204.

[15] Zhang F，Chan P P，Biggio B，et al. Adversarial feature selection against evasion attacks[J]. IEEE Transactions on Cybernetics，2016，46(3)：766-777.

[16] Bhagoji A N，Cullina D，Mittal P. Dimensionality reduction as a defense against evasion attacks on machine learning classifiers[EB]. arXiv：1704.02654，2017.

[17] Biggio B，Corona I，Fumera G，et al. Bagging classifiers for fighting poisoning attacks in adversarial classification tasks[C]//International workshop on multiple classifier systems. Springer，2015：350-359.

[18] Xu W L，Qi Y J，Evans D. Automatically evading classifiers[C]//Proceedings of the 2016 Network and Distributed Systems Symposium，2016：21-24.

(上接第 327 页)

[12] Dauphin Y N，Fan A，Auli M，et al. Language modeling with gated convolutional networks[EB]. arXiv：1612.08083，2016.

[13] 冉鹏，王灵，李昕，等. 改进 Softmax 分类器的深度卷积神经网络及其在人脸识别中的应用[J]. 上海大学学报(自然科学版)，2018，24(3)：352-366.

[14] Krawczyk B. Learning from imbalanced data：open challenges and future directions[J]. Progress in Artificial Intelligence，2016，5(4)：221-232.

[15] Lin T Y，Goyal P，Girshick R，et al. Focal loss for dense object detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence，2018，42(2)：318-327.

[16] Simonyan K，Zisserman A. Very deep convolutional networks for large-scale image recognition[EB]. arXiv：1409.1556，2014.

[17] Kinga D P，Adam J B. Adam：A method for stochastic optimization[C]//International Conference on Learning Representations(ICLR)，2015.

[18] Tavallaee M，Bagheri E，Lu W，et al. A detailed analysis of the KDD CUP 99 data set[C]//2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications. IEEE，2009：1-6.