

基于可分离卷积的轻量级恶意域名检测模型

杨路辉¹, 白惠文¹, 刘光杰^{1,2}, 戴跃伟^{1,2}

(1. 南京理工大学自动化学院, 江苏 南京 210094;
2. 南京信息工程大学电子与信息工程学院, 江苏 南京 210044)

摘要: 考虑到基于深度学习的恶意域名检测方法计算开销大, 难以有效应用于真实网络场景域名检测实际, 设计了一种基于可分离卷积的轻量级恶意域名检测算法。该模型使用可分离卷积结构, 能够对卷积过程中的每一个输入通道进行深度卷积, 然后对所有输出通道进行逐点卷积, 在不减少卷积特征提取效果的情况下, 有效减少卷积过程的参数量, 实现更加快速的卷积过程并不降低模型的准确性。同时, 为了减轻模型训练过程中正负样本数量不平衡与样本难易程度不平衡的情况对模型分类准确率的影响, 引入了一种聚焦损失函数。所提算法在公开数据集上与 3 种典型的基于深度神经网络的检测模型进行对比, 实验结果表明, 算法能够达到与目前最优模型接近的检测准确率, 同时能够显著提升在 CPU 上的模型推理速度。

关键词: 可分离卷积; 域名生成算法; 深度学习; 网络安全

中图分类号: TP309

文献标识码: A

doi: 10.11959/j.issn.2096-109x.2020084

Lightweight malicious domain name detection model based on separable convolution

YANG Luhui¹, BAI Huiwen¹, LIU Guangjie^{1,2}, DAI Yuewei^{1,2}

1. School of Automation, Nanjing University of Science and Technology, Nanjing 210094, China
2. School of Electronic & Information Engineering, Nanjing University of Information Science and Technology, Nanjing 210044, China

Abstract: The application of artificial intelligence in the detection of malicious domain names needs to consider both accuracy and calculation speed, which can make it closer to the actual application. Based on the above considerations, a lightweight malicious domain name detection model based on separable convolution was proposed. The model uses a separable convolution structure. It first applies depthwise convolution on every input channel, and then performs pointwise convolution on all output channels. This can effectively reduce the parameters of convolution process without impacting the effectiveness of convolution feature extraction, and realize faster

收稿日期: 2020-01-03; 修回日期: 2020-05-21

通信作者: 杨路辉, yangluhui005@foxmail.com

基金项目: 国家自然科学基金 (U1836104)

Foundation Item: The National Natural Science Foundation of China (U1836104)

论文引用格式: 杨路辉, 白惠文, 刘光杰, 等. 基于可分离卷积的轻量级恶意域名检测模型[J]. 网络与信息安全学报, 2020, 6(6): 112-120.

YANG L H, BAI H W, LIU G J, et al. Lightweight malicious domain name detection model based on separable convolution[J]. Chinese Journal of Network and Information Security, 2020, 6(6): 112-120.

convolution process while keeping high accuracy. To improve the detection accuracy considering the imbalance of the number and difficulty of positive and negative samples, a focal loss function was introduced in the training process of the model. The proposed algorithm was compared with three typical deep-learning-based detection models on a public data set. Experimental results denote that the proposed algorithm achieves detection accuracy close to the state-of-the-art model, and can significantly improve model inference speed on CPU.

Key words: separable convolution, domain generation algorithm, deep learning, cyber security

1 引言

近年来,越来越多的僵尸网络(Botnet)和高级可持续威胁攻击(APT, advanced persistent threat)成为网络安全中不可忽视的安全事件类型。在这些网络攻击事件中,攻击者通常会先向目标网络中植入恶意软件,恶意软件向命令与控制(C&C, command and control)服务器传输数据并接受控制命令,实现对目标的持续控制。在通信过程中,恶意软件为了提高隐蔽性,尽可能增加存活时间,往往会使用域名生成算法(DGA, domain generation algorithm)随机生成一组域名,并注册为C&C服务器的域名,然后,恶意软件只需内置DGA的随机种子,即可计算出可用的C&C服务器域名,进而与之通信。通过DGA的方法,能够避免使用硬编码的域名和IP地址,增加了安全软件对可疑通信行为的发现和阻断难度。因此,高效准确地检测DGA生成的域名,能够有助于及时发现网络攻击事件中的可疑通信行为,对于网络攻击事件的分析和防御有十分重要的意义。

为了有效地检测DGA域名,早期的研究者主要通过提取域名的字符特征的方法进行检测。文献[1]分析了域名的相对熵、雅克比系数、编辑距离等特征,通过这些字符特征进行恶意域名检测。文献[2]在提取字符特征的基础上,引入了线性回归算法进行分类。文献[3]提取了15个维度的域名特征,并引入了机器学习分类的方法,采用J48决策树进行分类,并且在真实网络场景下的数据集中验证了分类模型的有效性。文献[4]针对基于单词的DGA域名,提出了一种基于词关联特征提取结合混合分类器的检测算法。这些使用域名特征结合机器学习分类器的方法能够在部分场景中有效使用。文献[5]提出仅从域名上提取特征不足以实现精确的检测,于是在提取域名字

符特征的基础上,增加了流量维度的特征分析。文献[6]则进一步深入和优化了该类算法。特征工程结合机器学习分类器的方法过度依赖于人工分析和特征提取,提取的特征维度有限且容易出现检测模型失配或者过拟合问题,且分类器的选取对于分类结果影响很大。对于域名字符特征之外的通信流量特征,无法满足检测的实时性和效率要求,从域名的字符特性出发,依旧是最行之有效的检测方法。

近几年来,深度学习的发展为DGA恶意域名的检测提供了新的思路,文献[7]首次采用深度学习的方法进行DGA域名检测,其采用了长短期记忆网络(LSTM, long short-term memory)的检测模型,实现了比采用特征工程结合机器学习分类器更有效的检测算法。随后,文献[8]首次采用比LSTM计算速度更快的卷积神经网络(CNN, convolutional neural network)设计了检测模型,同样取得了比传统方法更好的检测结果。文献[9]分析和比较了多种深度网络模型对DGA域名的检测结果,发现使用了并行CNN结构的模型检测效果最好。考虑到DGA域名样本存在不平衡问题,文献[10]使用了一种结合权重调节的LSTM模型提高DGA域名的检测以及DGA家族的多分类结果,是目前检测效果最好的模型。文献[11]则使用了一种结合注意力机制的LSTM模型实现DGA域名检测与多分类。近年来的研究成果表明,使用深度学习进行DGA恶意域名检测模型检测准确率明显优于传统的特征工程结合机器学习分类器方法。然而,深度学习模型参数量大、计算复杂度较高,真实网络场景下对域名检测的时效性要求高,在保证准确率的情况下,设计更加轻量化的检测模型是恶意域名检测的一个重要研究方向。

本文基于可分离卷积结构设计了一种轻量化的恶意域名检测模型,通过对每一个输入通道进

行空间深度卷积, 然后对所有输出通道进行逐点卷积, 能够在不降低卷积特征提取效果的情况下, 有效减少卷积过程的参数量, 实现更加快速的卷积过程并实现特征提取。同时, 考虑到域名样本收集过程中, 存在正负样本数量不平衡以及不同 DGA 算法生成的域名检测难易程度不同的双重不平衡问题, 在模型的训练过程中引入了一种聚焦损失函数提高模型准确率。本文检测模型与 3 种基于深度学习的 DGA 域名检测模型进行了对比实验, 实验结果表明, 本文检测模型能够在保证较高检测率的基础上, 显著提高检测速度。本文主要贡献在于首次将可分离卷积结构引入恶意域名检测领域, 减少模型的计算参数量, 且引入了聚焦损失函数, 在不增加计算复杂度的情况下提高平均检测准确率, 同时比较和分析了多种基于深度方法的恶意域名检测模型的计算时间和检测准确率, 对于将深度学习实际应用到网络域名安全领域起到积极作用。

2 可分离卷积神经网络结构

本节介绍卷积神经网络的原理及其在恶意域名检测问题上的应用方式, 以及可分离卷积模型结构与原理, 进一步阐述了可分离卷积结构应用在恶意域名检测领域的可行性及带来的益处。

2.1 卷积神经网络

卷积神经网络最早由文献[12]提出, 后来被广泛应用于二维图像的分类和目标检测问题。其主要原理是通过卷积层提取二维图像的高维特征, 然后通过全连接层输出预测结果。经过数十年的发展, 卷积神经网络在图像分类、图像识别、图像分割等领域占据了绝对的主导地位。在自然语言处理领域, 尤其是文本分类领域, 早期的模型是基于循环神经网络(RNN, recurrent neural network)以及由 RNN 发展而来的 LSTM 网络结构进行设计, 由于文本长度不定, 且文本属于序列数据, 前后关联性较大, 早期的研究者并不认为 CNN 能够有效应用于文本分类, 直到文献[13]首次将 CNN 应用于句子的分类, 并取得了不错的效果, 此后, CNN 被广泛应用于各种文本分类任务, 近年来也被用于 DGA 恶意域名的检测。

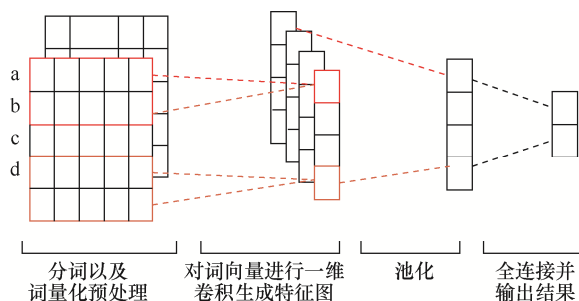


图 1 基于 CNN 的文本分类过程
Figure1 Text classification process based on CNN

图 1 展示了基于 CNN 的文本分类过程, 首先需要对文本进行分词, 然后将每个词量化成可用于卷积计算的词向量, 使用相应大小的卷积核对词向量进行卷积操作, 得到一组特征图, 对特征图进行池化, 选择其中显著性较高的特征量形成一个一维特征图, 将这个一维特征图输入全连接层中, 并输出预测结果。其中, 卷积过程可以用式(1)表示。

$$\begin{cases} c_i = f(wX + b) \\ X = [x_i, x_{i+1}, \dots, x_{i+h}] \end{cases} \quad (1)$$

其中, c_i 表示每次卷积输出的结果, f 表示非线性激活函数, w 表示每个卷积滤波器的参数, b 表示偏置项, X 表示一个窗口的词向量, h 表示卷积窗口大小即卷积核大小。

文本分类中对文本的卷积操作需要对文本中的单词进行量化, 其原理为通过某种映射方式, 将单词映射成数值向量, 目前较为流行的词向量量化方法包括 Word2vec 及 BERT 等, 然而这些量化方法依赖于大量标注过的语料库并需要训练, 量化准确率依赖于语料库的语料数量及标注的准确性。为了避免词向量的量化准确性对于文本分类结果准确性的影响, 文献[14]提出使用字符级的 CNN 进行文本分类, 只需要对每个英文单词中的单个字母进行量化, 即可进行卷积操作与全连接分类, 使用字符级的量化方法, 同样能够充分利用 CNN 进行特征提取, 并得到准确的分类结果。对于 DGA 恶意域名, 由于域名的构成元素一般为英文字母、数字以及特殊字符, 且域名生成过程大部分采用某种随机选择的方式选择一定数量的字符组合成一个域名, 所以生成的 DGA 域名不具备完整的单词构成, 使用 CNN 处

理 DGA 域名时, 必须采用字符级的 CNN 模型。同样地, 本文中的检测模型也是基于字符级的域名量化和处理的。

2.2 可分离卷积

可分离卷积是文献[15]提出的 MobileNets 模型中的高性能卷积结构, 其基本原理是将标准卷积过程分解成等效的深度卷积和逐点卷积, 能够在不损失卷积精度的情况下显著减少运算量。可分离卷积结构如图 2 所示。

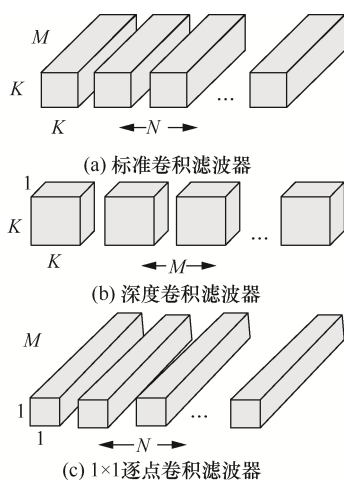


图2 标准二维卷积与深度可分离二维卷积结构
Figure2 Standard and depthwise separable 2D convolution structures

对于标准二维卷积过程, 如图 2(a)所示, 每次卷积过程使用 $K \times K$ 大小的卷积核进行卷积滤波, 输出的特征图通道数为 M 。图 2(b)表示深度卷积, 每一次卷积过程使用 M 个 $K \times K$ 大小的卷积核进行卷积, 输出特征图通道数设为 1。图 2(c)表示逐点卷积, 即使用 1×1 卷积核对图 2(b)的输出通道进行卷积操作, 图 2(b)和图 2(c)的结合使用即可取代图 2(a)所示的标准二维卷积过程, 然而计算量显著减少。图 2(a)表示的卷积过程参数量如式(2)所示。

$$K \times K \times M \times N \quad (2)$$

可分离二维卷积过程即使用图 2 (b)和图 2(c)结合, 其参数量如式(3)所示。

$$K \times K \times M + 1 \times 1 \times M \times N \quad (3)$$

因此, 可分离二维卷积与标准二维卷积参数量对比如式(4)所示。

$$\frac{K \times K \times M + 1 \times 1 \times M \times N}{K \times K \times M \times N} = \frac{1}{N} + \frac{1}{K^2} \quad (4)$$

当卷积核大小为 K 时, 使用可分离二维卷积与标准二维卷积相比, 参数量可减少约为 $\frac{1}{K^2}$ 。

因此, 使用可分离卷积能够显著提高卷积过程的计算速度。

对于恶意域名检测, 一般使用的是一维卷积, 卷积核大小为 K , 同理即可得出可分离一维卷积与标准一维卷积参数量对比为

$$\frac{K \times M + 1 \times M \times N}{K \times M \times N} = \frac{1}{N} + \frac{1}{K} \quad (5)$$

因此, 使用一维可分离卷积相较于标准一维卷积, 参数量可减少约 $\frac{1}{K}$ 。

3 基于可分离卷积的恶意域名检测模型设计

由于可分离卷积能够有效减少卷积过程中的参数量, 并提升计算速度, 本节详细介绍基于可分离卷积的恶意域名检测模型结构及参数设置, 同时, 针对域名进行卷积操作前, 需要将域名量化成为数值向量, 本节详细介绍域名量化过程。针对多种类型的恶意域名样本, 存在数量与难易程度不同的双重不平衡问题, 极大影响深度模型的训练准确率, 为了提高模型的准确率, 本文在模型训练过程中引入了一种聚焦损失函数。

3.1 域名量化

为了使用卷积神经网络提取域名的字符级特征, 需要对域名中的字符进行量化, 将每个字符映射成为一个数值向量。在传统自然语言处理领域中, 字符量化存在两种量化模型, 分别为独热表示模型和分布式表示模型。独热表示模型是将单词映射成为一个长向量, 向量中的所有分量中只有一个值为 1, 其他值全为 0, 数值 1 的位置代表该词在所使用的映射字典中的位置, 这种量化方法维度等于字典大小, 不能有效表达两个单词之间的联系, 且字典较大时维度过高, 不利于运算。分布式表示模型是将单词映射成一个固定维度的向量, 向量维度可控, 便于运算, 是目前更主流的量化方法。本文采用分布式表示模型对域名中的每个字符进行量化, 量化过程首先需要统

计所有域名中可能出现的字符，形成映射字典，本文基于大量域名样本统计了出现过的 87 个字符，包括大小写字母、数字、特殊符号等；然后对每个字符进行编号，将域名中的每个字符与编号进行一一映射；最后基于深度学习框架 Keras 中的 Embedding 层进行量化，即可得到每个字符的向量表达结果。

图 3 以域名“baidu”为例展示了量化过程，域名长度设定为 128 个字节，长度不够则补 0，将每个字符向量的长度设定为 128。

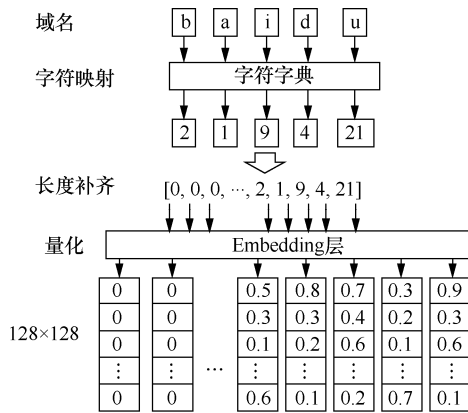


图 3 域名量化过程

Figure3 Domain name embedding process

3.2 基于可分离卷积的恶意域名检测模型与参数设计

基于可分离卷积的恶意域名检测模型基于第 3.1 节所述的字符量化方法与第 2 节所述的可分离卷积结构，辅以其他标准深度模型结构进行设计而成，其完整模型结构如图 4 所示。模型主要包含 9 层深度网络结构。第 1 层为输入层，模型输入为经过字典映射的域名向量，域名长度设为 128，长度不足则补 0。第 2 层为量化层 (Embedding)，将输入的 1×128 维域名向量量化为 128×128 维向量，具体量化过程如第 3.1 节所示。第 3 层为一维可分离卷积层 (SeparableConv1D)，卷积核大小设为 5，卷积核代表了卷积的感受野，卷积核太小，感受野不足，无法有效提取更大范围的相邻字符间的关联特征，卷积核太大，则感受野过大，容易造成忽略局部相邻字符间的关联特征，卷积核太小或太大均会对分类结果造成不利影响，本文通过实验对比了卷积核为 3~7 的分类结果，选择了表现最好的卷积核大小 5。第 4 层

是随机失活层 (Dropout)，设置随机失活概率为 0.5，可以增加模型随机性，有效防止过拟合。第 5 层是展开层 (Flatten)，用于将所有通道的卷积结果展开并合并成一个一维向量，输出维度为 1×16384 。第 6 层是全连接层 (Dense)，将 1×16384 维度的特征图进行全连接，输出 1×128 维度的特征图。第 7 层是随机失活层，同样是为了防止过拟合。第 8 层是全连接层，输入为 1×128 维度的特征图，经过全连接后输出 1×1 特征图。第 9 层是激活层 (Activation)，由于本文的模型是二分类模型，采用 sigmoid 激活函数，将输出映射为 0~1 的数值结果，输出值如果小于 0.5，则判定为合法域名，否则为恶意域名。完整的模型参数如表 1 所示。

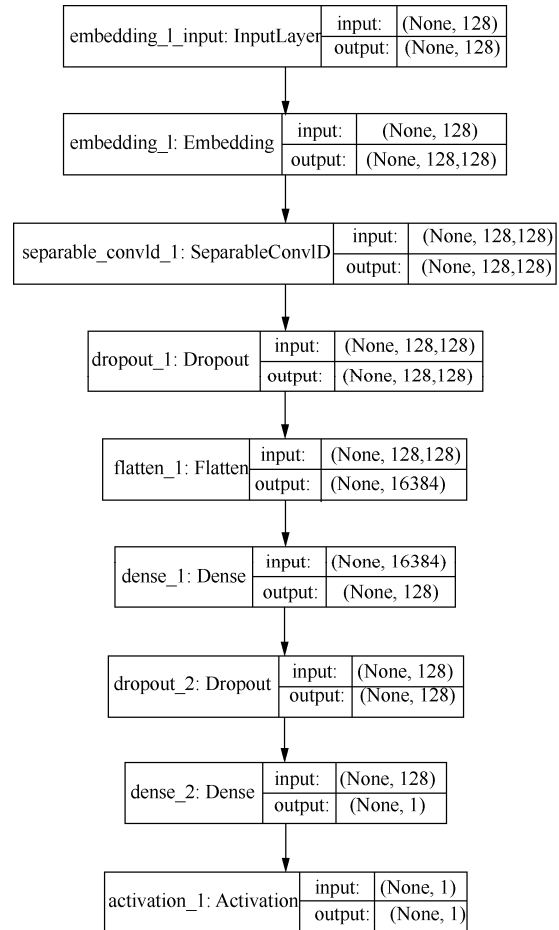


图 4 基于可分离卷积的恶意域名检测模型结构

Figure4 The structure of malicious domain name detection model based on depthwise separable convolution

3.3 聚焦损失函数

聚焦损失函数 (focal loss) 最早由文献[16]

表1 检测模型参数
Table1 The parameters of the detection model

层类型	输入大小	输出大小	参数设置
输入层 (Input)	1×128	128×128	
量化层 (Embedding)	128×128	128×128	
一维可分离卷积层 (SeparableConv1D)	128×128	128×128	Kernel_size=5, Stride=1
随机失活层 (Dropout)	128×128	128×128	Dropout_rate=0.5
展开层 (Flatten)	128×128	1×16384	
全连接层 (Dense)	1×16384	1×128	
随机失活层 (Dropout)	1×128	1×128	Dropout_rate=0.5
全连接层 (Dense)	1×128	1×1	
激活层 (Activation)	1×1	1×1	Activation= 'sigmoid'

提出, 用于缓解图像目标检测问题中的样本不均衡问题, 由传统的交叉熵损失函数改进而来, 其原理是通过调节难易样本的损失值, 增大难分类样本的损失, 减少容易分类样本的损失, 使网络模型向难分类样本聚焦, 增加对难样本的检测能力, 从而增加整体的检测准确率。传统二分类交叉熵损失函数如式(6)所示。

$$L = \begin{cases} -\log y', & y = 1 \\ -\log(1 - y'), & y = 0 \end{cases} \quad (6)$$

其中, y' 是网络模型输出的预测结果值, y 是样本真实标签, 在二分类中值为 0 或 1。当 $y=1$ 时, 预测值 y' 越接近 1, 损失值 L 越小; 当 $y=0$ 时, 预测值 y' 越接近 0, 损失值 L 越小。然而, 当训练样本中存在大量容易样本时, 即预测结果很接近真实值时, 大量容易样本带来的损失将淹没少量难样本的损失, 从而影响梯度下降的方向, 使模型无法得到最优的结果。为了减少容易样本对模型的影响, 增加难样本的影响力, 引入一个影响因子, 改进后的函数如式(7)所示。

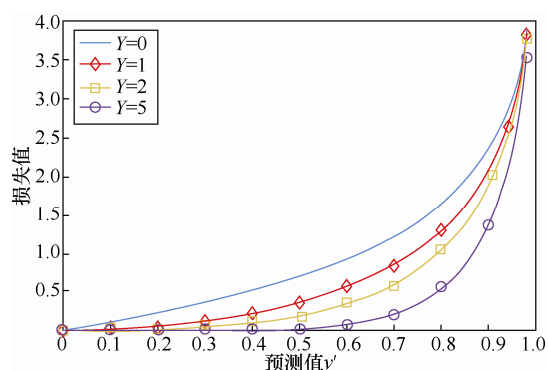
$$L_{\eta} = \begin{cases} -(1 - y')^{\eta} \log y', & y = 1 \\ -y'^{\eta} \log(1 - y'), & y = 0 \end{cases} \quad (7)$$

损失值随预测值变化如图 5 所示, 聚焦损失函数增大了难易样本产生的损失值的区分度, 使难样本产生相对于容易样本更加显著的损失值。

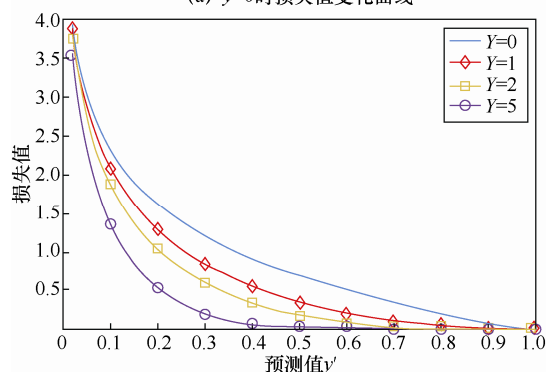
考虑到在域名检测问题中, 合法域名的样本比恶意域名样本更容易获取, 合法域名的数量明显多于恶意域名, 需要增加一个权重 α , 完整的

聚焦损失函数如式(8)所示。

$$L_{\eta} = \begin{cases} -\alpha(1 - y')^{\eta} \log y', & y = 1 \\ -(1 - \alpha)y'^{\eta} \log(1 - y'), & y = 0 \end{cases} \quad (8)$$



(a) $y=0$ 时损失值变化曲线



(b) $y=1$ 时损失值变化曲线

图5 聚焦损失函数损失值随预测值变化曲线
Figure5 The loss curve of focalloss

恶意域名检测模型的训练样本中, 合法域名数量远多于恶意域名样本数量, 且恶意域名种类较多, 不同 DGA 算法原理不同, 其生成的域名

异常程度也不相同,因此针对恶意域名进行分类时,对于部分异常程度较高的 DGA 种类检测效果较好,而针对部分异常程度低的 DGA 种类,检测效果一般不好。为了使检测模型聚焦在更难检测的 DGA 种类上,本文采用聚焦损失函数作为训练过程中的损失函数,用来缓解训练过程中样本数量和难易程度不均衡带来的问题,提高综合检测率。

4 实验与结果分析

4.1 数据集

本文从公开数据源中收集了 400 000 个合法域名和 100 000 个恶意 DGA 域名,其中,合法域名来自思科收集的 DNS 白名单数据。恶意域名来自 360 网络安全实验室收集的 DGA 域名样本,共包含 20 种恶意软件或攻击事件中的 DGA 域名样本,每种类型取 5 000 个样本。样本详情如表 2 所示,训练集与测试集按照 9:1 的比例划分。

4.2 实验平台与模型评价标准

本文实验使用的硬件平台处理器为英特尔酷睿 i7-7700HQ@2.8 GHz,训练过程使用英伟达显卡 GTX2080Ti 进行加速,深度神经网络框架采用 Keras 作为前端,Tensorflow 作为后端。在参数设置方面,本文采用 Adam 优化器,学习率设为 0.001, β_1 设为 0.9, β_2 设为 0.999,参数衰减率设为 0.000 05。

实验结果评价标准采用平均准确率、召回率和曲线下面积(AUC, area under curve)和 CPU 推理时间。平均准确率和曲线下面积用于衡量模型整体的分类效果,召回率用于比较对算法生成域名样本的分类效果,CPU 推理时间是指模型对单个域名进行预测所使用的 CPU 计算时间。准确率、召回率与 AUC 计算式如(9)所示。

$$\begin{cases} A_{\text{avg}} = \frac{A_{\text{pos}} + A_{\text{neg}}}{2}, \\ R = \frac{N_{\text{TP}}}{N_{\text{TP}} + N_{\text{FN}}}, \\ \text{AUC} = \int \left(\frac{N_{\text{TP}}}{N_{\text{TP}} + N_{\text{FN}}} \right) d \left(\frac{N_{\text{FP}}}{N_{\text{FP}} + N_{\text{TN}}} \right) \end{cases} \quad (9)$$

其中, A_{avg} 为模型分类平均准确率, A_{pos} 表示恶意域名样本的分类准确率, A_{neg} 表示合法域名样本的分类准确率, R 表示召回率, N_{TP} 为恶意域名样本预测正确的数量, N_{TN} 为合法域名样本预测正确的数量, N_{FN} 表示恶意域名样本预测错误的数量, N_{FP} 表示合法域名样本预测错误的数量, AUC 表示曲线下面积的值。

4.3 损失函数参数设置

本文在训练过程中使用了聚焦损失函数,其中有两个可调节的超参数 α 和 γ , α 由合法域名与恶意域名样本数量决定,由于使用的合法域名与恶意域名数量为 4:1,合法域名标签为 0,恶意域名标签为 1,因此取 $\alpha=0.8$ 。 γ 值与样本的难易程度有关,无法通过计算得出最优值,因此在实验过程中设置不同 γ 值,通过实验选取最佳的 γ 值,在实验过程中,分别取 $\gamma=\{0, 1, 1.5, 2, 2.5, 3, 3.5, 4, 5\}$,模型平均准确率如图 6 所示。通过调节 γ 参数后的平均准确率和 AUC 结果发现,当 $\gamma=2$ 时,模型的检测平均准确率和 AUC 最高,且实验结果表明,使用聚焦损失函数能够提高模型的检测准确率。

4.4 对比实验结果

本文对比了提出的基于可分离卷积的检测算法、文献[8]中基于标准卷积的检测算法、文献[7]中基于 LSTM 的检测算法以及文献[10]中的检测算法,对比结果如表 3 所示,样本中每种域名的分类准确率如图 7 所示。

与文献[8]中基于标准卷积的检测算法和文

表 2 数据集描述
Table2 Dataset description

样本标签	样本描述	数量/个
合法域名	样本来自思科收集的 DNS 请求白名单	400 000
恶意域名	20 种恶意软件生成的 DGA 样本, 具体类型如下: Gameover、Murofet、Dircrypt、Tinba、Necurs、Ramdo、Ranbyus、Cryptolocker、Emotet、Corebot、Banjori、Qakbot、Rovnix、Kraken、Ramnit、Locky、Pykspa、Simda、Symmi、Virut	100 000

献[7]中基于 LSTM 的算法相比,本文提出的算法平均准确率和召回率更高。与文献[10]中的检测算法相比,本文算法的平均准确率与其相差不大。然而,本文提出的算法在 CPU 上的计算速度与 3 种对比算法相比均有明显优势,推理时间仅为 0.6 ms,用时仅为文献[8]算法的 57%,是文献[7]和文献[10]的 29%。本文在保证整体检测效果与目前最优算法接近的基础上,显著提升了模型计算速度。具体到针对 20 类不同种类的 DGA 域名的检测率,详细结果如图 7 所示。由于每种 DGA 域名的生成原理不同,其异常程度和检测难度也不相同,文献[8]中的算法对大部分 DGA 域名能够实现超过 95%的准确率,本文算法针对这些 DGA 种类保持了相近的准确率。然而,对于部分较难检测的 DGA 域名种类,文献[8]中的算法检测率较低,对 Symmi 的检测率仅为 67.87%,对 Virut 的检测率仅为 66.25%,对 Simda 的检测率为 84.82%,由于本文使用了聚焦损失函数,使模型能够聚焦到这几类难样本,对 Symmi 的检测率提升至 80.09%,对 Virut 的检测率提升至 81.27%,对 Simda 的检测率提升至 89.03%,提升效果明显。

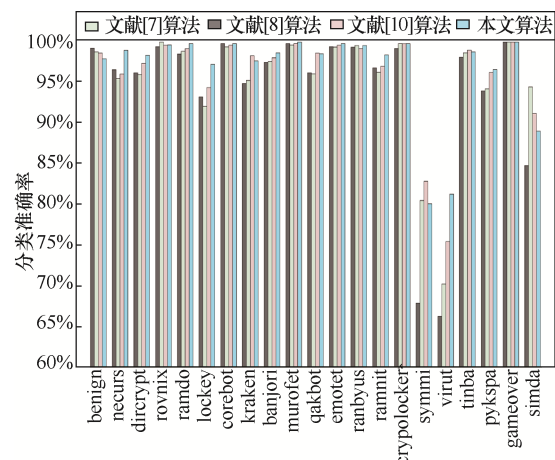


图7 域名分类结果对比详情

Figure7 Details of the classification result

实验结果表明,本文提出的基于可分离卷积的轻量级恶意域名检测模型能够在保证较高检测率的情况下,与对比算法相比明显提升了在 CPU 上的模型推理速度,在使用聚焦损失函数后,能够有效提升部分较难检测的 DGA 类型的检测准确率,与对比算法相比,本文提出的检测模型更适合在真实网络场景中应用。

5 结束语

考虑到在真实网络环境中应用深度网络模型进行恶意域名检测,需要兼顾准确率和计算速度,本文提出了一种基于可分离卷积的轻量级恶意域名检测模型,该模型使用可分离卷积结构取代标准卷积结构,能够使用深度卷积核逐点卷积相结合的方式代替标准卷积过程,显著提升卷积计算速度,同时保证不降低准确率,为了减轻样本数量和难易程度对模型训练结果的影响,引入了一种聚焦损失函数。本文从公开数据源上采集了大量合法域名与恶意域名样本,并且将提出的检测模型与基于标准卷积的检测模型、基于 LSTM 的检测模型以及目前检测率最高的检测模型进行对

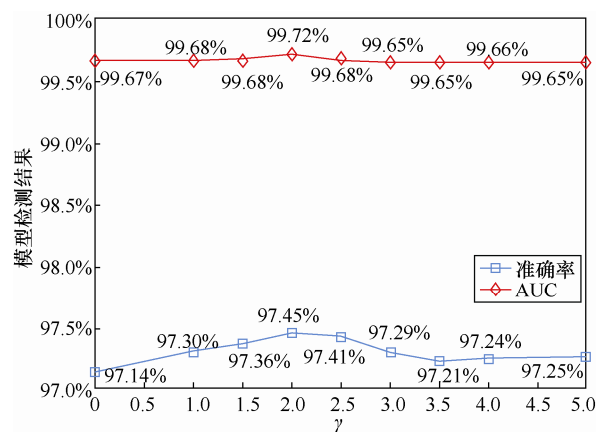
图6 模型检测结果随参数 γ 值变化曲线Figure6 Detection result with different value of γ

表3 对比实验结果
Table3 Comparative experimental results

算法	召回率	平均准确率	AUC	CPU 推理时间/ms
文献[8]算法	94.06%	96.61%	0.9968	1.05
文献[7]算法	95.36%	96.97%	0.9960	2.07
文献[10]算法	96.26%	97.51%	0.9971	2.09
本文算法	96.75%	97.46%	0.9971	0.60

比实验, 实验结果表明, 本文提出的模型能够达到与目前最优算法接近的检测准确率, 同时能够显著提升在 CPU 上的模型推理速度。

本文设计的恶意域名检测模型是一种通用的检测模型, 针对部分较难检测的 DGA 域名种类准确率不够高, 未来的研究工作可以针对一些特定种类 DGA 域名的特征设计专用的检测模型, 同时考虑到近年来人工智能芯片的快速发展, 可以结合人工智能芯片的特性, 设计更加快速而准确的检测算法。

参考文献:

- [1] YADAV S, REDDY A K K, REDDY A L N, et al. Detecting algorithmically generated malicious domain names[C]//Proceedings of the 10th ACM SIGCOMM Conference on Internet Measurement 2010. 2010: 48-61.
- [2] YADAV S, REDDY A K K, REDDY A L N, et al. Detecting algorithmically generated domain-flux attacks with DNS traffic analysis[J]. IEEE/ACM Transactions on Networking, 2012, 20(5): 1663-1677.
- [3] BILGE L, SEN S, BALZAROTTI D, et al. EXPOSURE: a passive DNS analysis service to detect and report malicious domains[J]. ACM Transactions on Information and System Security (TISSEC), 2014, 16(4): 1-28.
- [4] YANG L, ZHAI J, LIU W, et al. Detecting word-based algorithmically generated domains using semantic analysis[J]. Symmetry, 2019, 11(2): 176.
- [5] SCHIAVONI S, MAGGI F, CAVALLARO L, et al. Tracking and characterizing botnets using automatically generated domains[J]. Computer Science, 2013(2): 217-248.
- [6] SCHIAVONI S, MAGGI F, CAVALLARO L, et al. Phoenix: DGA-based botnet tracking and intelligence[C]//International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment. 2014: 192-211.
- [7] WOODBRIDGE J, ANDERSON H S, AHUJA A, et al. Predicting domain generation algorithms with long short-term memory networks[J]. arXiv preprint arXiv:1611.00791, 2016.
- [8] YU B, GRAY D L, PAN J, et al. Inline DGA detection with deep networks[C]//2017 IEEE International Conference on Data Mining Workshops (ICDMW). 2017.
- [9] YU B, PAN J, HU J, et al. Character level based detection of DGA domain names[C]//2018 International Joint Conference on Neural Networks (IJCNN). 2018: 1-8.
- [10] TRAN D, MAC H, TONG V, et al. A LSTM based framework for handling multiclass imbalance in DGA botnet detection[J]. Neurocomputing, 2018, 275: 2401-2413.
- [11] QIAO Y, ZHANG B, ZHANG W, et al. DGA domain name classification method based on long short term memory with attention

mechanism[J]. Applied Sciences, 2019, 9:4205.

- [12] LECUN Y, BOSER B, DENKER J S, et al. Backpropagation applied to handwritten zip code recognition[J]. Neural Computation, 1989, 1(4): 541-551.
- [13] KIM Y. Convolutional neural networks for sentence classification[J]. arXiv preprint arXiv:1408.5882, 2014.
- [14] ZHANG X, ZHAO J, LECUN Y. Character-level convolutional networks for text classification[C]//Advances in Neural Information Processing Systems. 2015: 649-657.
- [15] HOWARD A G, ZHU M, CHEN B, et al. Mobilenets: efficient convolutional neural networks for mobile vision applications[J]. arXiv:1704.04861, 2017.
- [16] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, (99): 2999-3007.

[作者简介]



杨路辉 (1992-), 男, 江西黎川人, 南京理工大学博士生, 主要研究方向为网络与信息安全。



白惠文 (1992-), 男, 吉林白山人, 南京理工大学博士生, 主要研究方向为网络流量分析。



刘光杰 (1980-), 男, 江苏徐州人, 博士, 南京信息工程大学教授、博士生导师, 主要研究方向为信息安全、多媒体系统、深度学习。



戴跃伟 (1962-), 男, 江苏镇江人, 博士, 南京信息工程大学教授、博士生导师, 主要研究方向为网络与多媒体信息安全。