

基于注意力机制与特征融合的网络钓鱼检测算法

张思睿, 延志伟, 董科军, 尉迟学彪

(中国互联网络信息中心, 北京 100190)

摘要: 网络钓鱼是攻击者实施网络诈骗犯罪的主要手段。随着国家反网络诈骗力度的不断加大, 各类网络钓鱼活动的技术对抗性也日益增加, 给网络钓鱼检测工作带来了较大压力。比如, 网络钓鱼攻击通常会利用图像来代替文本, 并对高权重的网站标志图像加以小尺度偏移或旋转等手段, 来逃避传统的基于文本或图像特征的检测算法。针对网络钓鱼技术对抗性日益增加的问题, 提出一种基于注意力机制与特征融合的网络钓鱼检测算法, 建立包含两个阶段的融合域名、网页结构、网页文本和网页图标等特征的层级分类模型, 可有效针对攻击者的各种技术对抗性策略。该算法在第一阶段利用机器学习模型的轻量特性, 融合域名、文本、网页结构特征从海量域名中预召回可疑域名子集, 在第二阶段基于候选子集, 引入注意力机制加深样本与被仿冒对象之间的全局文本关联特征的提取, 并增加样本与被仿冒对象图标间的对比特征, 建立融合文本和图像特征的深度分类模型。并在实验环节对该算法的有效性进行了验证。这种分层检测方式有效避免了对海量待检测域名的图像数据抽取, 在保证检测精度的前提下大大提升了检测效率。

关键词: 网络钓鱼检测; 层次特征融合; 注意力机制

中图分类号: TP393

文献标志码: A

DOI: 10.11959/j.issn.2096-109x.2024058

Phishing detection algorithm based on attention and feature fusion

ZHANG Sirui, YAN Zhiwei, DONG Kejun, YUCHI Xuebiao

China Internet Network Information Center, Beijing 100190, China

Abstract: Phishing has been the primary means utilized by attackers to conduct cyber fraud. As national anti-cyber fraud efforts continue to increase, the technical confrontation of various phishing activities has also escalated, bringing significant pressure to phishing detection work. For instance, current phishing attacks often employ images in place of text and apply small-scale shifts or rotations to high-weight website logo images to evade traditional detection algorithms that rely on text or image features. To address the problem of escalating adversarial phishing technologies, a phishing detection algorithm based on the attention mechanism and feature fusion was proposed, and a hierar-

收稿日期: 2024-05-21; 修回日期: 2024-07-09

通信作者: 尉迟学彪, yuchixuebiao@cnnic.cn

基金项目: 国家重点研发计划 (2022YFB3105000)

Foundation Item: The National Key Research and Development Program of China (2022YFB3105000)

引用格式: 张思睿, 延志伟, 董科军, 等. 基于注意力机制与特征融合的网络钓鱼检测算法[J]. 网络与信息安全学报, 2024, 10(4): 123-131.

Citation Format: ZHANG S R, YAN Z W, DONG K J, et al. Phishing detection algorithm based on attention and feature fusion[J]. Chinese Journal of Network and Information Security, 2024, 10(4): 123-131.

chical classification model was established. This model included two stages of fusion involving domain names, web structure, web text, and web icons, capable of effectively countering various technical adversarial strategies employed by attackers. In the first stage, the algorithm leveraged the lightweight characteristics of the machine learning model to pre-recall a subset of suspicious domain names from a multitude of domain names. This was achieved by fusing the structural features of domain names, text, and web pages. In the second stage, based on the candidate subset, the attention mechanism was introduced to enhance the extraction of global text association features between the samples and the counterfeited objects. Additionally, the contrast features between the samples and the icons of the counterfeited objects were intensified, and a deep classification model fusing text and image features was established. The effectiveness of the algorithm was ultimately verified. This hierarchical detection method effectively avoids the extraction of image data from a large number of domain names to be detected, significantly improving detection efficiency while ensuring the accuracy of detection.

Keywords: phishing detection, hierarchical feature fusion, attention mechanism

1 引言

网络钓鱼是一种常见的网络攻击和网络犯罪形式,它利用社会工程和技术手段来窃取消费者的个人身份数据和金融账户凭证^[1]。网络钓鱼攻击者通常会制作与合法网站相似的仿冒网站,使用户误以为是合法网站而输入个人账户信息,网络钓鱼者借此手段收集用户敏感数据。根据全球反钓鱼工作组(Anti-Phishing Working Group, APWG)的最新数据,2023年全球网络钓鱼攻击事件数量相比2022年几乎翻倍,其中金融行业是网络钓鱼攻击事件最为集中的一个行业^[1],因此网络钓鱼检测工作具有重大的社会意义。网络钓鱼检测算法根据使用特征可分为基于内容的检测和基于域名空间的检测。其中,基于域名空间的检测依赖从域名关联关系、域名查询协议(WHOIS)信息中提取的特征,而基于内容特征的网络钓鱼检测工作则聚焦于网络钓鱼与合法网站在域名字符串、网页内容等方面的相似性度量,综合分析和挖掘有效特征,从而实现对合法网站和钓鱼网站的分类。网络钓鱼检测算法的发展促使了网络钓鱼者相关规避手段的不断更新。例如,网络钓鱼攻击者会对网站文本进行特殊编码,或者将文本替换为包含同样信息的图像,以逃避基于文本内容特征的检测。网络钓鱼攻击者还会将被仿冒的合法网站的标志图像进行小尺度偏移或旋转,以减少基于图像相似度特征的检测。针对这些新型规避手段,当前基于文本特征或图像特征的分类模型已经无法有效应对。近年来,基于多模态特征融合的深度机器学习模型进行网络钓鱼检测是一个热点,将不同模态的特征连接,或者将各单模

态特征的检测结果进行合并,这些方法都需要先行解析各模态的数据特征,在实际场景中待检测的域名数量庞大,特征解析将耗费大量的时间和计算资源。

针对以上问题,本文针对内容特征展开网络钓鱼检测,设计了一种基于注意力机制和特征融合的网络钓鱼检测算法,包括两个阶段:第一阶段,结合域名字符串特征和网站文本的浅层特征,预召回可疑网站候选集;第二阶段,建立基于注意力机制^[2]的文本编码器和基于卷积神经网络的图像编码器,连接两个编码的隐藏特征后,学习分类器判断网站是否是钓鱼网站。

本文作出了以下贡献。

1) 利用机器学习模型轻量的特点,设计有效的特征从海量的域名中预召回可疑域名子集,极大地减少待检测数据量。

2) 通过注意力机制学习到待检测网站与被仿冒目标深层关联的文本特征。

3) 在图像解析的特征基础上,引入图像类别概念,学习图像与其类别的模糊特征,构成更通用的图像特征。

2 相关工作

2.1 轻量级钓鱼检测算法

轻量级钓鱼检测算法包括传统的策略式检测、基于统计特征的检测和基于机器学习的检测,这类检测算法有参数量少、可理解度高和实现复杂度低的特点。当前,传统的基于黑名单的网络钓鱼检测方法是最容易实现和最有效的网络钓鱼检测技术,并已被广泛采用^[3]。大多数浏览

器通常会维护一个网络钓鱼网站黑名单并进行定期更新, 通过简单地阻止解析黑名单上的网站来提供反钓鱼服务。基于黑名单的网络钓鱼检测方法的有效性依赖于黑名单的更新频率, 并且对于黑名单之外的未知钓鱼网站无能为力。

研究者通过对大量的网络钓鱼实例统计分析, 发现网络钓鱼者生成钓鱼域名的方式通常有替换目标统一资源定位符(uniform resource locator, URL)中的几个字符以生成与目标URL相似的钓鱼域名、在目标URL后添加数字作为钓鱼域名、以及使用目标网页名称的缩写作为钓鱼域名等。因此, 可以通过URL统计规则^[4]提取一些有效的特征。

Zouina等^[5]提出了一种基于从URL中提取的6个特征的快速网络钓鱼网站检测方法。Verma等^[6]使用钓鱼URL与标准英文字符之间的距离度量作为特征, 例如相对熵距离、欧几里得距离、字符频率和编辑距离。Kumar等^[7]从URL中提取的一系列特征, 比较了不同机器学习技术用于网络钓鱼URL分类任务的效果, 并讨论如何根据给定的URL对网络钓鱼URL进行分类。Bai等^[8]从钓鱼URL的结构特征中提取了12种特征, 训练了包括4种机器学习模型的网络钓鱼网站分类器。然而, 网络钓鱼者可以轻易地避免根据过去已知的规则生成网络钓鱼域名, 因此这也是基于URL特征的网络钓鱼检测方法的主要缺点。Gandotra等^[9]集中在使用机器学习算法来检测网络钓鱼攻击, 这些攻击集中在特征工程上。Chiew等^[10]设计了一个框架来过滤特征子集以提高分类器的准确性。但只要某些功能被验证为有效, 网络钓鱼者就可以通过避免启发式方法来设计新的攻击。

2.2 基于深度学习的钓鱼检测算法

随着人工智能技术的发展, 深度神经网络常被用于学习数据中的深层隐藏信息, 大量基于深度学习的网络钓鱼检测工作应运而生, 即通过深层的网络模型学习输入样本的隐藏层特征, 来实现未知钓鱼网站的主动检测。考虑基于单模态特征的网络钓鱼检测算法的缺点, 一些研究人员正在关注多模态特征融合的算法, 例如网页的文本特征、网页的标志图像(Logo)等^[11-13]。

Yi等^[14]使用两种特征类型进行网络钓鱼网站检测, 包括URL的特征和网站之间的交互。Yang等^[15]提出了MFPD(multidimensional feature phishing detection)方法, 该方法基于深度学习驱动的URL特征、网页文本特征和网页代码特征等多维特征, 首先使用卷积神经网络(convolutional neural network, CNN)和长短期记忆(long short-term memory, LSTM)来学习给定URL的字符序列特征, 其次结合网页特征来训练钓鱼网站分类器。然而, 当钓鱼网站上没有文字而只有图片和表格时, MFPD就无法检测到此类钓鱼网站, 从而退化为简单的基于单模态特征的检测算法, 导致钓鱼网站的召回率较低。Zhang等^[16]提出了一种基于神经网络的多模态网络钓鱼检测方法(MultiPhish), 其中网站的域名和网站图标(favicon)通过深度神经网络提取特征表达, 使用变体自动编码器(variational auto encoder, VAE)来优化域名和网站图标的表示。但是, 考虑爬取网页图片需要很多时间, MultiPhish并不适合海量域名的检测任务。

基于上述工作, 本文设计了一种新的网络钓鱼检测架构, 采用分层融合的方法来减小获取数据的计算资源和时间成本, 结合文本图像和网页结构的信息来保证检测的准确性。此外, 在融合图像和文本特征时使用注意力机制来增强样本特征与目标合法网站数据集特征之间的相关性。

3 基于融合特征的钓鱼域名检测算法

网络钓鱼检测是一个从待检测域名中识别出与被仿冒目标网站相似的可疑域名的过程, 其中被仿冒目标网站可以自定义配置, 本文提到的被仿冒目标网站包括银行、证券、电商等官方网站。检测框架包括数据处理、模型结构和数据收集器3个模块, 在数据处理模块完成从域名到数据特征的构建, 提取域名特征, 域名对应网页的文本、图像和网页结构特征。在模型训练模块分阶段训练模型, 框架如图1所示, 在第一阶段中从输入的域名数据中检测出与被仿冒对象网站相似的可疑域名, 通过融合浅层的启发式特征快速召回可疑域名, 极大地减少待检测域名数据量, 基于第一阶段检测出的子数据集, 在第二阶段进一步融

合文本和图像的深层特征, 根据融合的深层特征学习分类器, 最终判定可疑域名。

3.1 数据处理

在数据处理模块中, 主要对输入的域名数据 $X: \{x_i\}_{N_x}$ 和被仿冒对象网站 $T: \{t_i\}_{N_t}$ 进行预处理。

(1) 域名数据预处理

首先, 对输入的域名做前缀扩张, 生成可访问的 URL, 并且通过访问 URL 获得网页源码文件。然后, 从网页源码中提取网页的文本内容, 最终每条域名数据都有其网页源码数据 Source_x 和网页文本数据 Text_x 。

(2) 被仿冒目标网站数据处理

与域名数据处理步骤类似, 获取被仿冒目标网站的文本内容 Text_T , 并且获取这些目标网站的 Logo 图片作为 I_T 。此外, 将被仿冒目标所属领域作为其域名类别数据 Cat_T , 例如银行、证券、电商、运营商等领域。

3.2 模型结构

3.2.1 预召回可疑域名

通过提取域名、网页文本内容和网页结构的浅层特征表达, 根据聚合后的特征建立 XGBoost 分类器^[17], 从大量待检测域名数据中召回可疑的域名数据集 X' , 缩小检测范围, 流程如图 2 所示。

其中, 针对域名和网页结构构建的启发式特征见表 1, 网页文本的浅层特征提取流程如图 3 所示。

首先对网站文本进行分词, 提取文本关键字^[18], 这些关键字可以涵盖每个网站的网站内容的主题, 表示为 Keyword_q 。在这一步中, 获取所有被仿冒对象网站文本的关键字, 并形成目标合法网站的去重关键字集合 Keyword_T 。其次, 通过 BERT 词向量模型^[19]得到关键词的词向量, 这里用 $\text{WordVec}_T \in \mathbb{R}^m \times m$ 来表示关键字

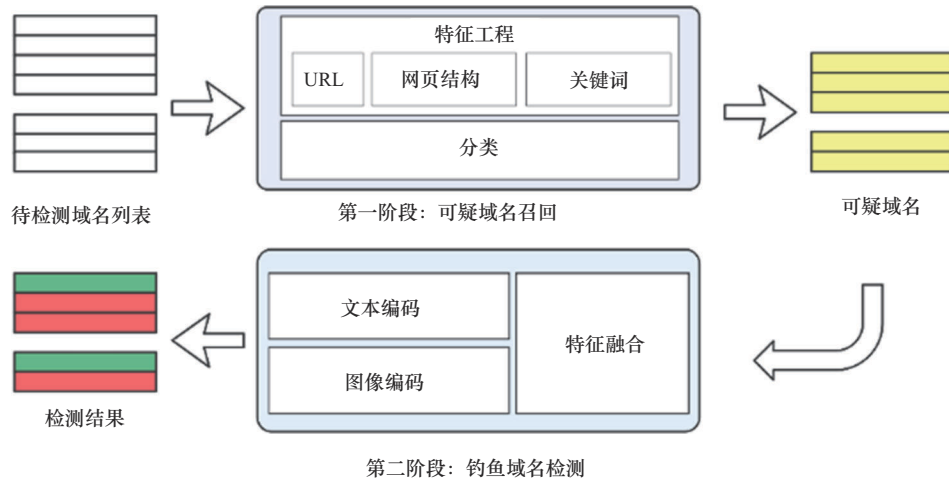


图 1 网络钓鱼域名检测算法框架

Figure 1 Framework of phishing detection algorithms

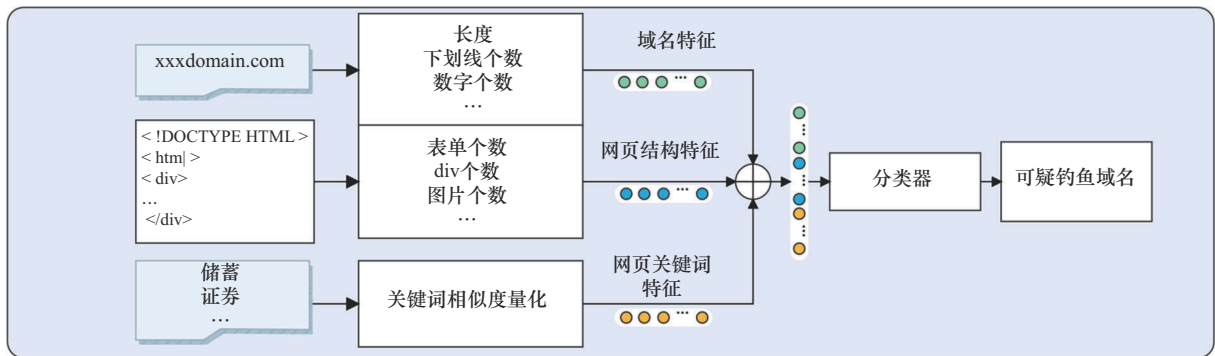


图 2 融合多模态启发式特征召回可疑域名流程

Figure 2 The process of suspicious domain names selection by fusing multimodal heuristic features

表1 阶段I中提取的浅层特征
Table 1 Shallow features extracted in stage I

数据模态	特征
网页文本	向量, 网页关键词相似度向量
	实数, 网页关键词 top-k 相似度均值
网页结构	网页是否有 logo.png/jpg
	网站是否有 favicon
	表单 form 的个数
	div 的个数
	div 的嵌套层数
	网页源码中图片个数
域名	网页内容的文本个数
	常数, 域名长度
	取值 1 或 0, 域名是否含有数字
	常数, 域名字符中的数字个数
	常数, 域名字符中“_”个数
	常数, 域名字符中“.”个数

Keyword_T的文本特征, WordVec_q ∈ ℝ^{k×m}表示关键字Keyword_q的特征, m 为词向量维度。接着, 在计算 WordVec_q 和 WordVec_T之间的余弦距离后, 得到一个相似度矩阵 M_{QT} , 矩阵中的每个值是Keyword_q和Keyword_T中关键词的相似度值。通过筛选大于经验阈值 threshold 的相似度来获得关键字对 P_{QT} 。最后, 从 P_{QT} 提取关键字特征:

$$D_{\text{keywords}} = (c_0, c_1, \dots, c_{nc}, v) \quad (1)$$

其中, c_i 是Keyword_q与Cat_T中的第 i 个类别之间的最大相似度, v 是 $M_{QT} \geq \text{threshold}$ 中的前 k 个相关词的相似度值的平均值。

3.2.2 钓鱼域名分类

在预召回可疑域名之后, 构建网站关键字编码器和网站Logo编码器来分别学习文本和图像两种模态的深层特征, 并融合两个编码器生成的特征向量, 最后基于融合的特征构建一个分类器来检测钓鱼网站, 网络结构如图4所示。

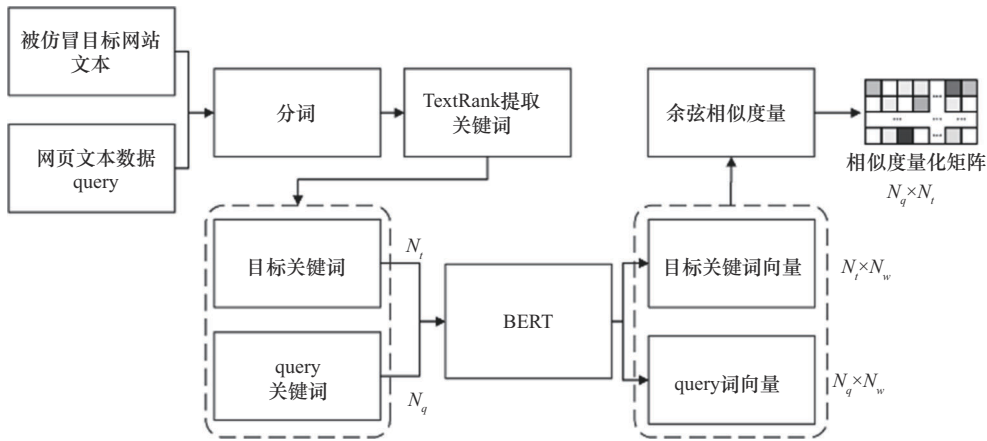


图3 浅层文本特征提取

Figure 3 Text feature extraction

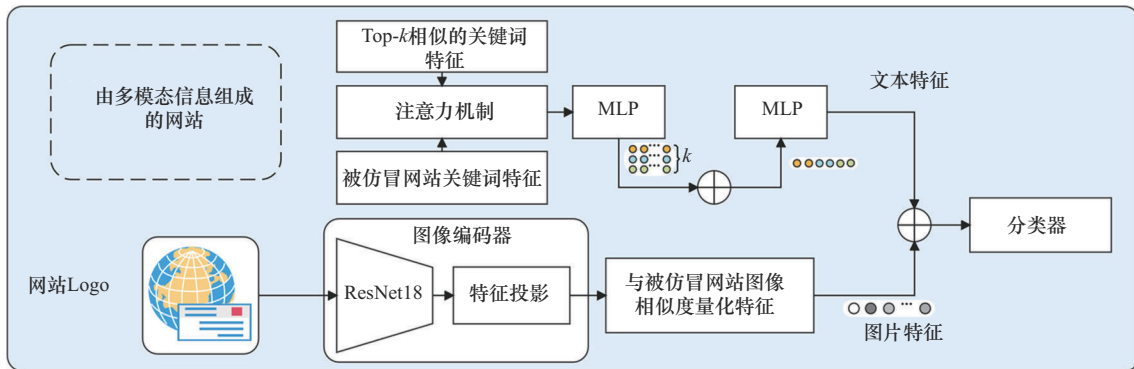


图4 钓鱼域名分类模块结构

Figure 4 The structure of the phishing domain name classification module

(1) 网站关键字编码器

对于 X' 中的每个网站, 选择与预召回阶段 top- k 相似的关键词, 并复用第一阶段 Bert 模型生成的词向量。此外, 为了获得与目标关键词集 Keyword_T 更相关的特征表示, 使用交叉注意力机制来对相似度编码。将 top- k 相似的关键词特征 $\text{WordVec}_q \in \mathbb{R}^{k \times m}$ 作为 Q , 将 Keyword_T 中的关键字的特征 $\text{WordVec}_T \in \mathbb{R}^{m \times m}$ 作为键 K 和值 V , 注意力机制可以形式化为:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{m}}\right)V \quad (2)$$

其中, m 是关键字特征的维度, $1/\sqrt{m}$ 是比例因子^[2]。

在注意力模块之后是一个多层神经网络 (multilayer perceptron, MLP), 将 k 个新的关键词特征 $\{\text{corrfeat}_i | i = 1, \dots, k\}$ 串联, 经过 MLP 层的投影得到代表输入域名的文本特征:

$$\text{textfeat}_q = \text{MLP}(\text{corrfeat}_1 \oplus \dots \oplus \text{corrfeat}_k) \quad (3)$$

(2) 网站 Logo 编码器

网站 Logo 编码器的输入是 $I_{X'}$ 中的图像, 输出是编码后的图像特征。首先根据固定大小固定通道数处理图像, 然后使用 ResNet18^[20] 和特征投影模块生成全局图像特征。特征投影模块由一个最大池化层、一个 MLP 层和一个 L2 归一化层组成。以孪生网络的形式训练使用对比损失^[21] 的编码器模型 S :

$$L_{\text{img}} = \text{label}(\sigma(\text{dist} - \text{margin}_{\text{pos}}))^2 + (1 - \text{label})(\sigma(\text{margin}_{\text{neg}} - \text{dist}))^2 \quad (4)$$

其中, dist 是两个输入图像之间的距离, $\sigma(\cdot)$ 是 ReLU 函数, margin 是距离边界, 通过缓解过拟合问题。在训练过程中, 正样本是来自同一类的 $\text{label} = 1$ 的图像, 而负样本是来自不同类的 $\text{label} = 0$ 的图像。

因此, 给定一个网站 Logo, 图像编码器 S 将输出其特征表示, 并使用欧几里得距离计算 $I_{X'}$ 中的每幅图像与 I_T 中的所有图像之间的相似度, 即对于 $I_{X'}$ 中的图像, 可得其图像相似度特征:

$$\text{imgfeat}_q = (s_0, s_1, \dots, s_m) \quad (5)$$

其中, $\text{imgfeat}_q \in \mathbb{R}^m$, 且 s_i 是 $I_{X'}$ 中的图像与 I_T 中的第 i 个图像的相似度。

(3) 网络钓鱼检测

将上述两个编码器输出的文本特征和图像特征串联在一起作为融合的特征, 并使用由 MLP 和 sigmoid 函数组成的二元分类器来判断是否为钓鱼网站:

$$y_q = \text{sigmoid}(\text{MLP}(\text{textfeat}_q \oplus \text{imgfeat}_q)) \quad (6)$$

4 实验及分析

本文在 .COM 和 .CN 域名数据上验证了所提出的算法, 在下文中提供了测试结果和消融实验的结果。由于本文提出的方法通过注意力机制和神经网络融合了文本特征和图片特征, 因此在实验部分与传统的特征融合方法和结果融合方法进行了对比。在实际钓鱼检测场景中, 最终需要依靠法务人员判定可疑网站是否是真正的钓鱼网站, 这也决定了钓鱼检测算法需要有尽可能高的召回率和相对较高的精确率, 由此减少人力资源的消耗。

4.1 实验参数设置

(1) 数据集

由于面向全量域名的实验资源消耗过大, 因此本文实验随机选取了 3 000 条钓鱼域名和 8 000 条非钓鱼域名。在数据处理模块对域名进行特征挖掘, 包括域名生成的网址、访问获取的网页源码、网页文本和网页图标数据。被仿冒目标对象数据包括 85 个关键词向量和 115 个合法网站标志图片, 通过抓取常见的银行、金融和电子支付类网站的官网获取。训练和测试数据集的比例为 5:1。

(2) 参数设置

本文从 ResNet18 中得到 512 维的图像特征。除非另有说明, 否则所有 MLP 都由两个具有层范数和 ReLU 激活函数的线性层组成。本文将 top- k 关键词的 k 设置为 7, 对没有足够关键词的数据采用零填充。使用学习率为 0.000 5 的 AdamW 优化器来训练整个网络。所有实验均在 NVIDIA GTX1080 GPU 上进行。

4.2 对比实验

本实验分别对域名、网页结构和网页文本 3 种不同的特征进行了测试, 结果如表 2 所示。从表 2 可以得出以下结论: ①与针对域名特征的研究^[22]做对

比, 当仅使用域名作为输入时, 由于缺乏足够的信息, 算法表现最差; ②网页结构在之前的反钓鱼工作中被广泛使用, 并获得了较好的效果; ③网页文本是一个有效的特征, 但一些钓鱼网站为了逃避检测, 用图片代替文字, 从而缺乏足够的关键字, 这限制了该特征的特异性, 例如, 一些钓鱼网站仅由“登录”等词组成和注册, 不能提取任何有效信息; ④当引入3个输入时, 检测性能显著提高, 这表明引入更多信息有利于网络钓鱼检测。

4.3 消融实验

在本实验部分, 测试了阶段二的性能, 阶段二的输入是阶段一的输出, 在此输入子集上评估了模型的效果。为了研究本文所提模型每个组成部分的影响, 对数据集进行了消融实验。具体来说, 为验证所提出的方法的重要性, 比较了使用MLP的文本嵌入和使用CNN的原始输出和MLP的图像嵌入。这两个对比实验的目的是分别从网络中去除目标数据集信息(目标关键字和目标网页标志图片), 为了公平比较, 其他训练或测试设置保持不变。

实验结果见表3。在方法(a)中, 不涉及目标词信息, 网络中去除了注意力机制, 实验发现这导致网络的泛化能力较差。在方法(b)中, 目标网站图标信息被去除, 因此不能使用对比学习方式训练图像嵌入模块, 相比之下, 分类器中只有CNN特征可用于特征融合。从实验结果发现, 这种方法在测试过程中也存在过拟合问题。表3中可以看出本文所提方法获得了最佳性能。该实验表明, 文本嵌入中的注意力机制和图像嵌入中的图像相似度估计有助于在网络钓鱼检测过程中获得有效的“目标相关”的特征表示。

表3 阶段II实验结果和消融实验结果

Table 3 Stage II results and ablation study

方法	准确率	钓鱼域名分类效果评估		
		召回率	精确率	F1值
(a)文本嵌入+MLP	0.915 3	0.994 5	0.900 1	0.944 9
(b)图像嵌入+MLP	0.931 5	0.976 0	0.931 6	0.953 3
本文	0.947 6	0.977 9	0.951 6	0.964 6

在图5中提供了相应的ROC(receiver operating characteristic)曲线和P-R(precision-recall)曲线。从图中可以看出, 本文所提方法获得了最高的AUC分数, 作为对比, 当目标文本特征从网络中移除时, 性能显著下降, 这再次验证了本文提出的基于注意力机制的特征表达的重要性。

5 结束语

本文提出的钓鱼检测框架, 第一阶段利用机器学习模型轻量的特点, 根据对数据集的挖掘分析结果, 设计有效的特征从海量的域名中预召回可疑域名子集。第二阶段基于可疑域名子集, 将交叉注意力机制和对比表示学习引入深度神经网络, 创新地学习待检测网站和被仿冒目标网站之间的深度全局关联特征。这种分层的检测方式, 避免了对海量待检测域名的多模态数据拉取, 在保证检测精度的前提下, 提升了检测效率。最后, 将本文提出的检测框架部署至真实环境中, 开展稳定高效的网络钓鱼检测任务。此外, 本文算法具有配置可扩展的优点, 通过合理配置被仿冒网站的数据集, 可灵活应用本文所提模型检测赌博、恶意仿冒政企等其他有害网站。

表2 不同特征的实验结果

Table 2 Experimental results with different modal features

特征	准确率	钓鱼域名分类效果评估		
		召回率	精确率	F1值
仅域名	0.600 3	0.813 5	0.607 4	0.695 5
仅网页结构	0.851 3	0.869 3	0.873 2	0.871 2
仅网页文本	0.734 3	0.826 1	0.781 2	0.803 0
域名+网页结构+网页文本(本文)	0.701 4	0.988 6	0.735 0	0.843 1

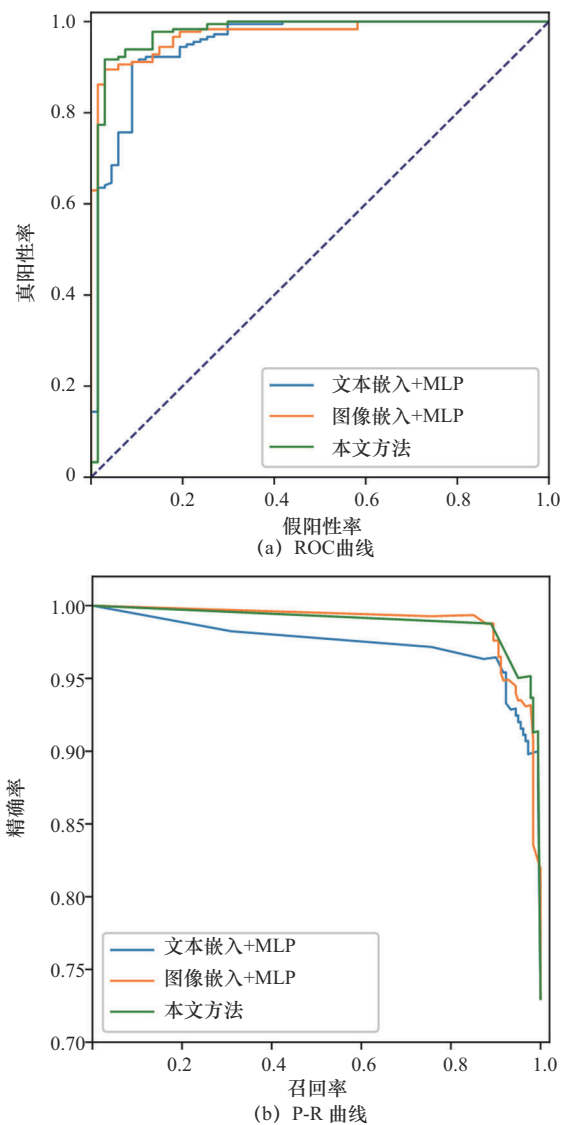


图5 ROC曲线和P-R曲线

Figure 5 ROC curve and P-R curve

参考文献:

- [1] Anti-Phishing Working Group. Phishing attack trends report[R]. 2022.
- [2] 曲强,于洪涛,黄瑞阳. 基于注意力机制的社交垃圾文本检测方法[J]. 网络与信息安全学报, 2020, 6(1): 54-61.
QU Q, YU H T, HUANG R Y. Attention-based approach of detecting spam in social networks[J]. Chinese Journal of Network and Information Security, 2020, 6(1): 54-61.
- [3] JAIN A K, GUPTA B B. A novel approach to protect against phishing attacks at client side using auto-updated white-list[J]. EURASIP Journal on Information Security, 2016, 2016(1): 9.
- [4] JEEVA S C, RAJSINGH E B. Phishing URL detection-based feature selection to classifiers[J]. International Journal of Electronic Security and Digital Forensics, 2017, 9(2): 116.
- [5] ZOUINA M, OUTTAJ B. A novel lightweight URL phishing detection system using SVM and similarity index[J]. Human-Centric Computing and Information Sciences, 2017, 7(1): 17.
- [6] VERMA R, DYER K. On the character of phishing URLs: accurate and robust statistical learning classifiers[C]//Proceedings of the 5th ACM Conference on Data and Application Security and Privacy. New York: ACM, 2015: 111 - 122.
- [7] KUMAR J, SANTHANAVIJAYAN A, JANET B, et al. Phishing website classification and detection using machine learning[C]//Proceedings of the 2020 International Conference on Computer Communication and Informatics (ICCCI). Piscataway: IEEE Press, 2020: 1-6.
- [8] BAI W H. Phishing website detection based on machine learning algorithm[C]//Proceedings of the 2020 International Conference on Computing and Data Science (CDS). Piscataway: IEEE Press, 2020: 293-298.
- [9] GANDOTRA E, GUPTA D. An efficient approach for phishing detection using machine learning[M]. Multimedia Security. Singapore: Springer, 2021: 239-253.
- [10] CHIEW K L, TAN C L, WONG K, et al. A new hybrid ensemble feature selection framework for machine learning-based phishing detection system[J]. Information Sciences, 2019, 484: 153-166.
- [11] VIGNESHWARAN P, ROY A S, SATHVIK B S, et al. Multidimensional features driven phishing detection based on deep learning [M]//International Conference on Intelligent Emerging Methods of Artificial Intelligence & Cloud Computing. Cham: Springer International Publishing, 2022: 362-369.
- [12] LI Y C, CHU S N, XIAO R. A pharming attack hybrid detection model based on IP addresses and web content[J]. Optik, 2015, 126 (2): 234-239.
- [13] CHIEW K L, CHANG E H, SZE S N, et al. Utilisation of website logo for phishing detection[J]. Computers & Security, 2015, 54: 16-26.
- [14] YI P, GUAN Y X, ZOU F T, et al. Web phishing detection using a deep learning framework[J]. Wireless Communications and Mobile Computing, 2018, 2018(1): 4678746.
- [15] YANG P, ZHAO G Z, ZENG P. Phishing website detection based on multidimensional features driven by deep learning[J]. IEEE Access, 2019, 7: 15196-15209.
- [16] ZHANG L, ZHANG P, LIU L C, et al. Multiphish: multi-modal fea-

tures fusion networks for phishing detection[C]//Proceedings of the ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Piscataway: IEEE Press, 2021: 3520-3524.

- [17] CHEN T Q, GUESTRIN C. XGBoost: a scalable tree boosting system[C]//Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2016: 785-794.

- [18] MIHALCEA R, TARAU P. TextRNN: bringing order into text[C]//Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing. Barcelona: Spain, 2004: 404-411.

- [19] DEVLIN J, CHANG M, LEE K, et al. Bert: pre-training of deep bidirectional transformers for language understanding[C]. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics. 2019: 4171-4186.

- [20] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2016: 770-778.

- [21] HADSSELL R, CHOPRA S, LECUN Y. Dimensionality reduction by learning an invariant mapping[C]//Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06). Piscataway: IEEE Press, 2006: 1735-1742.

- [22] SINGH S, SINGH M P, PANDEY R. Phishing detection from URLs using deep learning approach[C]//Proceedings of the 2020 5th International Conference on Computing, Communication and Security (ICCCS). Piscataway: IEEE Press, 2020: 1-4.

[作者简介]



张思睿（1994-），女，陕西咸阳人，中国互联网络信息中心工程师，主要研究方向为网络安全、机器学习。



延志伟（1985-），男，山西兴县人，中国互联网络信息中心研究员，主要研究方向为域名服务安全、信息中心网络。



董科军（1977-），男，浙江舟山人，中国互联网络信息中心正高级工程师，主要研究方向为互联网基础资源管理、分布式系统。



尉迟学彪（1984-），男，山东高密人，中国互联网络信息中心研究员，主要研究方向为域名服务安全、网络空间治理。