DOI:10.3969/j.issn.1671-0673.2021.04.010

一种基于 MPAN 的钓鱼 URL 检测方法

张 桥[1,2],卜佑军[2],陈 博[2],曹东伟[1],张稣荣[2]

(1.郑州大学 中原网络安全研究院,河南 郑州 450001;2.信息工程大学,河南 郑州 450001)

摘要:为解决日益严峻的网络钓鱼问题,提出一种基于消息传递注意力网络(Message Passing Attention Network, MPAN)的钓鱼 URL 检测方法。此方法相对于传统的机器学习和黑名单检测方法,无需人工提取特征且能识别新出现的钓鱼网页。首先基于敏感词分词的方法对 URL 分词,以提升利用 URL 数据信息的程度。然后通过 MPAN 获取 URL 中长距离、非连续的单词交互信息,基于自动提取的特征检测钓鱼网页。实验结果表明,基于 MPAN 的钓鱼网页检测方法能够达到较高的准确率、召回率、*F*1 值。

关键词:消息传递注意力网络;钓鱼 URL;机器学习;黑名单;URL 分词

中图分类号:TP393.08 **文献标识码**:A **文章编号**:1671-0673(2021)04-0443-07

Phishing URL Detection Method Based on MPAN

ZHANG Qiao[1,2], BU Youjun[2], CHEN Bo[2], CAO Dongwei[1], ZHANG Surong[2]

(1. Zhongyuan Network Security Research Institute, Zhenzhou University, Zhengzhou 450001, China;
2. Information Engineering University, Zhengzhou 450001, China)

Abstract: To solve the increasingly serious problem of phishing, a phishing URL detection method based on message passing attention network (MPAN) is proposed. Compared with traditional machine learning and blacklist detection methods, this method does not need to extract features manually, and can recognize new phishing web pages. Firstly, URL is segmented based on sensitive word segmentation method to improve the degree of using URL data information. Then, the long-distance and discontinuous word interaction information in the URL is obtained through MPAN, and the phishing web page is detected based on the automatic feature extraction. Experimental results show that the method based on MPAN can achieve high accuracy, recall and *F*1 value.

Key words: message passing attention network; phish URL; machine learning; blacklist; URL segmentation

0 引言

网络钓鱼是一种利用社会工程和技术手段窃取用户个人信息和金融账户数据的犯罪机制。近年来,网络钓鱼攻击迅速增长,根据中国反钓鱼联盟 APAC[1]的报告,截至 2020 年 10 月,钓鱼网站数量达到了 475 655 个,钓鱼网站数量巨大,给大众带来了严重的危害。因此,如何及时、有效地检测钓鱼网站已经成为亟待解决的问题。

收稿日期:2021-01-22;**修回日期**:2021-02-26

基金项目:国家重点研发计划资助项目(2017YFB0803201, 2017YFB0803204, 2016YFB0801200);国家自然科学基金资助项目(61572519, 61802429, 61521003);上海市科学技术委员会科研计划项目(16DZ1120503);中国博士后基金资助项目(44595)

作者简介:张 桥(1992-),男,硕士,主要研究方向为深度学习、网络安全。

针对钓鱼网站的检测问题,相关研究人员提出了各种基于黑名单和基于机器学习的方法来提高网络钓鱼检测的准确性[2]。但这些方法需要手工提取 URL、主机信息和网站内容的特征,费时费力且对新出现的钓鱼网页的检测准确率较低。这使得能够自动提取数据特征、准确检测钓鱼网站以抵御网络钓鱼攻击的技术继续成为迫切需要。

为解决上述问题,本文提出了一种基于深度学习的方法,通过从 URL 中自动提取特征来准确识别钓鱼网站。根据文献[3]的研究,网络钓鱼攻击将网络钓鱼 URL 嵌入到伪造的邮件中来传播。因此,开发有效检测网络钓鱼网址的技术,在很大程度上有助于抵御网络钓鱼攻击。本文使用消息传递注意力网络自动获取 URL 中单词之间的交互信息来识别钓鱼网站,主要贡献如下:

①本文提出了一个由消息传递神经网络(Message Passing Neural Network, MPNN)和基于注意力 Atention 的新型神经网络 MPAN,用于网络钓鱼 URL 检测。这是首次将图网络应用于钓鱼 URL 检测的一种方法。实验结果证明,通过 MPAN 识别钓鱼 URL 是有效的。

②本文提出了一个基于敏感词分词的新型分词方法,在数据预处理阶段对网址分词时不会导致敏感词丢失有效信息和无法识别新出现的单词。

③本文所提出的用于钓鱼网站检测的方法仅关注 URL 本身,相比其他考虑网站内容或主机信息的方法能够更快速地钓鱼网站。此外该方法能够应用于任何可以嵌入 URL 的地方,如电子邮件、网站等。

1 相关工作

1.1 基于黑名单的检测

黑名单方法是一种较为简单的检测方法,只需进行简单的数据库查询操作。黑名单中包含已确认钓鱼网址,当用户访问某一网址时,将其与黑名单中的数据进行匹配。若匹配成功,则将其判别为钓鱼网址。Malware Domain List[4]、Google Safe Browsing API[5]等使用的都是基于黑名单的检测方法。然而当今,网址生成算法比较成熟,每天都会出现大量的钓鱼网址,黑名单数据库无法及时包含所有的钓鱼网址。根据文献[6]的研究,约 47%~83%的钓鱼网址在钓鱼事件发生 12 小时之后才被列入黑名单中。文献[7]中指出约有 93%的钓鱼网页没有被主流的黑名单收录。基于黑名单检测钓鱼网址的局限性在于要不断收集钓鱼网站样本并及时更新黑名单数据库。

1.2 基于机器学习的检测

为了解决这些局限性,有研究人员将钓鱼 URL 的检测看作一个文本分类或聚类的问题,使用相应的机器学习技术检测钓鱼 URL。目前用于钓鱼 URL 检测的机器学习方法包括无监督和有监督方法。无监督机器学习方法又称聚类方法,该方法将 URL 划分为若干簇,使得同一簇中的数据对象之间相似度较高,不同簇之间的数据相似度低。最后对不同的簇进行标记来区分钓鱼 URL 和合法 URL。文献[8]利用域名请求的先后顺序对 URL 进行聚类操作,将伴随出现的 URL 划分为同一族,然后对其标记分类。有监督机器学习方法又称为分类方法,该方法首先从标记的 URL 数据中获得适当的特征表示,然后使用 URL 的这种表示来训练基于机器学习的预测模型,最后使用已训练的预测模型对待测 URL 进行分类。文献[9]采用给定 URL 中单词的分布式表示形式,然后使用 7 种不同的机器学习算法来预测它是否是网络钓鱼 URL。尽管上述方法表现出令人满意的性能,但它们受到以下限制:无法处理看不见的特征,因为待检测的 URL 可能包含训练集中不存在的未知单词。

1.3 基于深度学习的检测

为解决上述缺陷,已有研究者使用了深度学习技术,利用其从原始数据中自动学习特征的特点来实现对钓鱼网页的分类。文献[10-13]通过利用 URL 的潜在特征来检测网络钓鱼 URL,均表现出了良好的性能,具有较高的检测精度。本文所提方法与之前的基于深度学习的方法的主要区别是,本文使用图网络来检测钓鱼网址,不把 URL 看作一个序列,而是将其看作一组共现词,将这些词构建成图,从而将文本分类问题转化为图分类问题。本文将每个 URL 都转换为图数据,并提出了一种用于将其分类的消息传递注意力网络(MPAN)来获取 URL 中长距离、非连续的单词交互信息,有利于提高网络钓鱼检测性能。

2 检测方法

2.1 问题定义

由于钓鱼 URL 检测的目标是确定 URL 是否是钓鱼网站,因此本文将此问题描述为一个以 URL 作为输入的二进制分类问题。将$[v_1,v_2,\cdots,v_n]\in S$表示为 URL u 的字符序列。通过训练一个网络$f:S\rightarrow\{0,1\}$来将 u 分类为钓鱼 URL 或合法

URL,将 S 输入 f 中,若 f 输出为 0,则将 u 分类为合法 URL,否则将其分类为钓鱼 URL。

2.2　检测流程

本文所提方法检测 URL 的流程如图 1 所示,将 URL 分词后通过词嵌入层转化为词嵌入向量,然后构造成图送入 MPAN 模型检测 URL 是否为钓鱼网址。

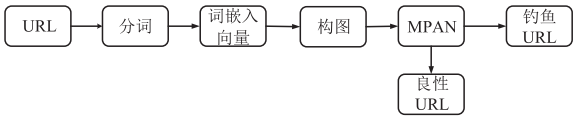

图 1　URL 检测流程

2.2.1　分词

深度学习模型只能处理经过数值化的向量,因此在对 URL 数据提取特征时需要先将其分词、编码并转化为 d 维词嵌入向量,不同词在 d 维空间的距离来表示它们之间的语义相似度。当前使用深度学习方法检测 URL 常用的分词方法有两种:基于单词划分 URL;基于字符划分 URL。基于单词划分 URL 使其转化为单词级词向量,利用“@”,“.”等特殊字符分割 URL 可能会使单词的数量相当大,造成该数据集的特征也按比例地增大,通常会大于相应训练数据集中 URL 的数量,导致在进行特征向量的转换时内存受到限制,在测试时无法获得新出现的单词的嵌入向量。相比于按单词划分 URL,基于字符划分 URL 使 URL 转化为字符级词向量能够在测试数据时获得新的 URL 的嵌入向量,从而避免了无法从不可见的单词中提取特征的问题。另外由于字符总数是固定的,在进行特征向量的转换时不会受到内存的限制且字符级分类器的大小保持固定。但是将 URL 划分为单个的字符会导致一些敏感词丢失部分有效信息,而钓鱼 URL 中确实存在一些特有的敏感词,如“login”、“password”、“registed”等。因此,根据字符划分 URL 不足以使神经网络分类器从 URL 字符串中获取全面的信息。

针对上述分词方法存在的问题,本文提出了一种基于敏感词分词的方法,如表 1 中以 www.ccd.cn.bank.com 举例。首先根据特殊字符和敏感词对 URL 进行单词级别划分,并将特殊字符看作单词处理以获得特殊字符的有效信息。然后对其中的非敏感词进行字符级别划分,而将其中的敏感词 bank 作为一个整体与其余字符进行区分,这样能够明显标记 URL 中的重点信息,有利于神经网络分类器提取更具有代表性的特征。

表 1　URL 分词

URL 分词方法	分词结果
基于单词划分	www　ccd　cn　**bank**　com
基于字符划分	w w w . c c d . c n . **b a n k** . c o m
基于敏感词划分	w w w . c c d . c n . **bank** . c o m

2.2.2　词嵌入向量

根据 URL 数据集和敏感词汇表(表 2)确定每条 URL 中字符及关键字的总长度 L 为 300。若 URL 长度超过 300,则在 URL 末尾将多余的字符截断,若 URL 的长度小于 300,则在其末尾用<pad>标记作为附加词填充。若 URL 中出现未知字符,则用未知字符标记<unk>表示。分析 URL 数据集及敏感词汇表,不同的字符、敏感词,再加一个附加词标记<pad>和一个未知字符标记<unk>总数量为 121,构建映射表为字符和敏感词赋予唯一编码,如表 3 所示。通过词嵌入矩阵将 URL 的数字编码转化为相应的词嵌入向量,该嵌入矩阵随机初始化,并与模型的其余部分联合优化。

表 2　敏感词汇

敏感词	account　admin　administrator　auth　bank　client　confirm　cmd　email host　login　password　pay　private　registed safe　secure　security　sign　service　signin　submit　user　update　validation　verification　webscr

表 3　字符和敏感词映射表

字符	编码
abcdefghijklmnopqrstuvwxyz	1-26
ABCDEFGHIJKLMNOPQRSTUVWXYZ	27-52
0123456789	53-62
./,+-=_| :&;'?"><{}( )[ ] ~! @#$%*\	63-92
account, admin, administrator, auth, bank, client, confirm, cmd, email, host, login, password, pay, private, registed, safe, secure, security, sign, service, signin, submit, user, update, validation, verification, webscr	93-119
填充字符标记<PAD>	120
未知字符标记<UNK>	121

2.2.3 构图

本文用 l 个单词表示一个 URL：$S=\{v_1,v_i,\cdots,v_l\}$，其中 v_i 表示第 i 个单词。$\boldsymbol{v}_i$ 是一个 d 维词嵌入向量，可以通过训练来更新。要为给定的 URL 构建一个图，需要将 URL 中出现的所有单词作为图的节点。每条边从 URL 中的一个单词开始，以相邻的单词结束。具体而言，URL 的图形定义为

$$V=\{v_i \mid i\in[1,l]\} \tag{1}$$

$$E=\{e_{ij} \mid i\in[1,l];j\in[i-p,i+p]\} \tag{2}$$

式中，V 和 E 是图的节点集和边集，p 表示与图中每个单词相连的相邻单词的距离。假设 p 值为 1，以 www. ccd. cn. bank. com 举例，将其构图如图 2 所示，其中节点经词嵌入矩阵转化为 d 维词向量，节点之间的边经边矩阵转为一维的向量，该矩阵是随机初始化的，且与检测模型的其余部分联合优化。

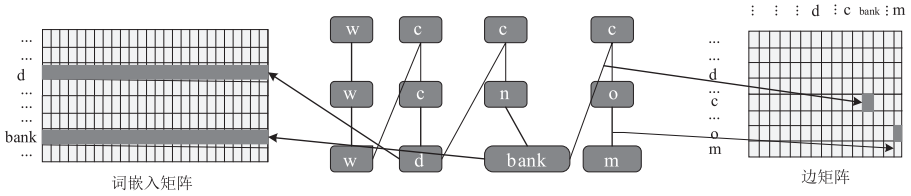

图 2 URL 构图

2.2.4 MPAN 模型

考虑到钓鱼 URL 为一种序列数据，其词与词之间存在着长距离的非连续的依赖关系，为了有效地检测网络钓鱼网址，本文提出了一个基于消息传递网络 MPNN 加注意力机制的钓鱼网页检测模型 MPAN，通过 MPNN 网络使 URL 中的单词节点充分获取与其相依赖的单词节点的信息。通过注意力机制为 URL 中对分类影响较大的单词节点赋予较大的权重。MPNN 是由文献[14]提出的一种模型框架，它包含两个核心阶段，消息传递阶段与读出阶段，在消息传递阶段，每个节点的状态信息 h_v^t 通过消息传递函数 M_t 聚合邻居节点的信息 m_v^{t+1} 然后通过更新函数 U_t 将自身信息与邻居节点信息叠加更新自己的状态，如式(3)、式(4)所示。其中，$N_{(v)}$ 表示图 G 中节点 v 的邻居，h_w^t 表示邻居节点的状态信息，e_{vw} 表示节点 v 和 w 之间的边。

$$m_v^{t+1}=\sum_{w\in N(v)} M_t(h_v^t,h_w^t,e_{vw}) \tag{3}$$

$$h_v^{t+1}=U_t(h_v^t,m_v^{t+1}) \tag{4}$$

读出阶段使用读出函数 R，根据式(5)计算 T 个时间步后整个图的特征向量 $\hat{\boldsymbol{y}}$，表示为

$$\hat{\boldsymbol{y}}=R(\{h_v^T \mid v\in G\}) \tag{5}$$

本文在 MPNN 基础上加入了注意力机制。一个网址首先被表示为一个图 G，网址中的单词或字符为图 G 的节点，在两个相邻单词之间定义了一条边。然后在图上引入消息传递注意力机制，提取图的特征，对图进行分类，模型框架如图 3 所示。

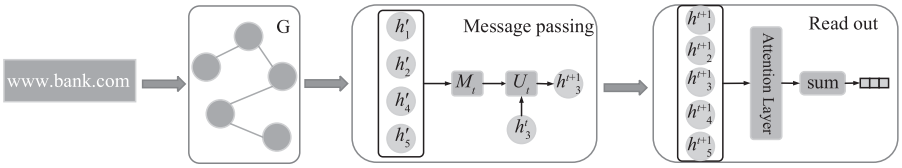

图 3 模型框架

MPAN 首先从相邻节点收集信息，并根据其原始表示和收集的信息更新其表示，定义为

$$m_v^{t+1}=\underset{w\in N_v^p}{mean}(e_{vw}h_w^t) \tag{6}$$

$$h_v^{t+1}=GRU(h_v^t,m_v^{t+1}) \tag{7}$$

式中：m_v^{t+1} 是节点 v 在 $t+1$ 时刻从其邻居接收到的消息；$mean$ 是一个均值函数，它将邻居节点的状态信息每个维度上的值相加求平均值，形成一个新的向量作为输出；N_v^p 为与节点 v 的距离为 p 的单词节点；$e_{vw}\in R^1$ 是从节点 v 到节点 w 的边缘权重，可以在训练过程中进行更新；h_v^t 为节点 v 在时刻 t 的隐藏状态表示；h_v^{t+1} 为节点 v 在 $t+1$ 时刻的更新表示。更新函数使用门控循环网络 GRU，其定义如下：

$$z^{t+1}=\sigma(W_z m_v^{t+1}+U_z h_v^t+b_z) \tag{8}$$

$$r^{t+1}=\sigma(W_r m_v^{t+1}+U_r h_v^t+b_r) \tag{9}$$

$$\tilde{h}^{t+1}=tanh(W_h m_v^{t+1}+U_h(r^{t+1}\otimes h_v^t)+b_h) \tag{10}$$

$$h_v^{t+1}=(1-z^{t+1})\otimes h_v^t+z^{t+1}\otimes\tilde{h}^{t+1} \tag{11}$$

式中：σ 是 sigmoid 函数，输出 0 至 1 之间的值，用于更新或遗忘数据，所有的 W、U 和 b 都是可训练的权重和偏差；z 和 r 分别作为更新门和重置门来

确定邻居信息对当前节点嵌入的贡献程度,更新门用于控制节点在当前时刻 t 的状态信息被带入到节点在 $t+1$ 时刻的状态信息中的程度,更新门的值越大说明节点在 t 时刻的状态信息带入到 $t+1$ 时刻的信息越多。重置门控制节点在 t 时刻的状态信息有多少被带入到 $t+1$ 时刻的候选状态信息 $\tilde{h}^{t+1}$ 上,重置门越小,节点在 T 时刻的状态信息被遗忘的就越多,那么被带入到节点在 $t+1$ 时刻的状态信息就会越少。节点在 $t+1$ 时刻的状态信息由候选状态信息 $\tilde{h}^{t+1}$、t 时刻的信息和更新门 z^{t+1} 决定。在迭代 T 个时间步充分更新单词节点之后,将自注意力机制应用于所有的词节点,并将它们聚合为图级的向量表示形式,读出函数定义为

$$u^T = \sum_{i=1}^{l} \alpha_i^T h_{v_i} \tag{12}$$

式中:α_i^T 为节点对应的注意力权值,最后,通过将图级向量送入 Softmax 层来分类,通过交叉熵函数最小化损失,如式(13)、式(14)所示;W 和 b 为权重和偏置;y_i 为样本的真实标签;$\hat{y}_i$ 为模型的预测标签。

$$\hat{y} = soft\max(Wu^T + b) \tag{13}$$

$$loss = -\sum_i y_i \log(\hat{y}_i) \tag{14}$$

3　实验与分析

3.1　实验数据

本文采用的数据集包括多个平台提供的开源样本,从 PhishTank 和 MalwarePatrol 获取钓鱼 URL,从 DMOZ 和 Alexa 获取合法 URL,以此来丰富 URL 数据的来源。其中 PhishTank 是一个反钓鱼网站,用户可以在该网站提交、验证和共享网络钓鱼数据。MalwarePatrol 与 PhishTank 类似,用户可以在其中下载钓鱼 URL。DMOZ 是一个来自世界各地的志愿者共同维护与建设的最大的全球目录社区,旨在收录优秀的网站,通过它可以获取合法的 URL 数据集。Alexa 是一个专门发布网站世界排名的网站,当前拥有的 URL 数量庞大,网站排名信息详尽,收集排名靠前的网站作为合法 URL 数据集。对数据去重后,数据集中共包含 206 200 条带标签的 URL 样本,其中钓鱼样本 105 100 条,合法样本 101 100 条,二者比例大约为 1∶1。URL 部分样本示例如表 4 所示。

表 4　URL 样本示例

样本源	网址
PhishTank	https://www.faresproducts.com/wp-admin/includes/ionos-GE/#info@koalamin https://rakuten.co.jp-rktqhxtdgshsamexy.gbyhmsif.work/?signin=a&email=a
MalwarePatrol	https://demcorknitwear.com/32787231/index.php?email=aaaa@example.jp https://secure-recovery.xyz/m.checkpoint.htm
DMOZ	http://www.mjzjswh.gov.cn http://www.myzte.cn
Alexa	https://www.tmall.com http://www.babytree.com

3.2　评估标准

为验证钓鱼网页检测方法的有效性,采取了准确率(Accuracy),精确率(Precision),召回率(Recall)和 $F1$ 值作为评价指标。计算如式(14)~式(17)所示,其中 TP(True Positive)表示预测的钓鱼网页实际为钓鱼网页的数量,FP(False Positive)表示预测的钓鱼网页实际为合法网页的数量,TN 表示预测的合法网页实际为合法网页的数量,FN(False Negative)表示预测的合法网页实际为钓鱼网页的数量。Precision 表示被正确判断为钓鱼网页类别的网页占全部被判断为钓鱼网页类别的网页的比重,体现了检测方法对合法网页的区分能力,Recall 则体现了对钓鱼网页的识别能力,F1 值同时考虑到了精确率和准确率,是二者的加权平均,能综合评估检测模型的性能。

$$Accuracy = (TP+TN)/(TP+FP+TN+FN) \tag{14}$$

$$Precision = TP/(TP+FP) \tag{15}$$

$$Recall = TP/(TP+FN) \tag{16}$$

$$F1 = 2 * Precision * Recall/(Precision+Recall) \tag{17}$$

3.3　实验设置

本文将数据集按 9∶1 的比例划分为训练集和测试集,采用 10 折交叉验证法,即将样本分为 10 组,其中每组包含 10 510 条钓鱼 URL 和 10 110 条合法 URL 作为测试集,另外 9 组包含 94 590 条钓鱼 URL 和 90 990 条合法 URL 作为训练集,该过程循环 10 次,保证每组样本数据都能作为测试集预测,将得到的 10 次测试结果取平均值评测模型的检测能力。根据经验,本文使用 Adam 优化器,将学习率设为 0.01,词嵌入向量维度为 128。使用网

络中的词嵌入层初始化词向量,词嵌入层与网络其他层相结合,能够利用反向传播算法进行联合优化以学习到最符合URL数据特征的向量表示。

3.4 实验结果

本文首先进行了消融实验,根据实验结果确定超参数 T 和 P 的值。首先,在实验中将消息传递迭代次数 T 从1改为4,其结果如图4所示,随着迭代次数的增多,模型的检测准确率首先增加,但当 T 值大于2时,模型的性能反而下降。然后本文测试了模型在不同 P 值上的性能,检测结果与图4类似,如图5所示,当 P 值大于3时,模型的检测准确率停止增加。以上结果表明,虽然随着迭代次数 T 和 P 值的增加,图中的中心节点能够获得更远处邻居节点的信息,但其与中心节点不是密切相关的,反而会影响模型的检测性能。综上,本文将时间步长 T 的值设为2,数据图中的中心节点与邻居节点的距离值 P 设为3。

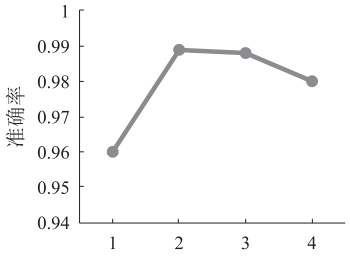

图4 SW_MPAN在不同时间步长 T 上的检测准确率

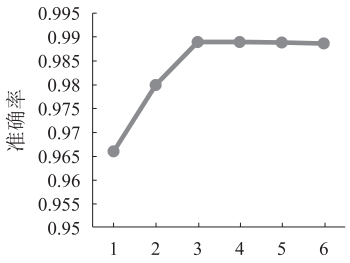

图5 SW_MPAN在不同 P 值上的检测准确率

此外,本文根据3种不同的分词方法训练了3个MPAN模型,并将它们在测试集上的检测结果进行了对比,以证明基于敏感词分词方法的有效性。具体来说,本文通过对数据采用3种分词方法来训练MPAN模型,分别为基于字符划分URL的字符级MPAN模型C_MPAN,基于单词划分URL的词级MPAN模型W_MPAN与基于敏感词划分URL的检测模型SW_MPAN。观察其在测试集上的检测效果,结果如表5所示。

W_MPAN的性能弱于C_MPAN,该结果可能源于以下3个方面:通过“.”、“\”、“?”等特殊字符对URL分词时忽略了特殊字符所具有的有效信息;为了避免内存受限,将数据集中仅出现一次的单词统一标记为<UNK>而忽略了这些单词的有效信息;无法获得新出现单词的有效信息。SW_MPAN不受上述两种分词方式的限制,在所有的评估指标中均达到最佳性能。这表明本文所提出的基于敏感词分词的方法能够有效提升对钓鱼网页的检测能力。

表5 根据不同的分词方法训练的模型在测试集上的检测效果

检测模型	准确率	精确率	召回率	$F1$ 值
C_MPAN	0.976 5	0.979 9	0.986 2	0.983 0
W_MPAN	0.953 0	0.957 6	0.958 0	0.957 8
SW_MPAN	0.988 9	0.989 9	0.993 5	0.991 7

此外,为体现本文提出的检测模型的优势,将其与深度学习模型CNN,LSTM做对比实验(这两种方法是序列数据集中应用最广泛的深度学习方法),实验结果如表6所示。

表6 所有模型在测试集上的检测效果

检测模型	准确率	精确率	召回率	$F1$ 值
SW_CNN	0.956 0	0.960 4	0.957 5	0.958 9
SW_LSTM	0.943 5	0.945 8	0.957 0	0.951 4
SW_MPAN	0.988 9	0.989 9	0.993 5	0.991 7

可看到,本文所提模型在准确率、精确率、召回率、$F1$ 值4个评估指标上均达到最佳性能,有效提升了对钓鱼网站的检测能力。

4 结束语

针对目前常用的钓鱼网站检测方法存在的需要人工提取特征,无法识别新出现的钓鱼网站的问题,本文提出了一种基于消息传递网络和注意力机制的新型网络结构MPAN,通过MPAN获取URL中长距离非连续的单词交互信息,基于自动提取的特征实现对钓鱼网页的分类。通过与深度学习中另外两种常用来处理序列数据的模型CNN、LSTM做对比实验表明,本文所提出的基于MPAN的钓鱼URL检测方法在精确率、召回率、$F1$ 值都取得了较高的结果,能够有效提升对钓鱼网站的检测能力。下一步将通过使用生成对抗网络生成钓鱼URL作为输入,对本文所提模型进行鲁棒性分析。

参考文献:

[1] 中国反钓鱼网站联盟.2020年8月钓鱼网站处理简报[EB/OL].[2021-1-20]. http://www.apac.cn/gzdt/202003/P020200320392664104846.pdf.

[2] DOU Z C,KHALIL I,KHREISHAH A,et al. Systematiza-

tion of knowledge (SoK): a systematic review of software-based web phishing detection[J]. IEEE Communications Surveys & Tutorials,2017,19(4):2797-2819.

[3] GUPTA B B, TEWARI A, JAIN A K, et al. Fighting against phishing attacks: state of the art and future challenges[J]. Neural Computing and Applications, 2017, 28(12):3629-3654.

[4] CANALI D, COVA M, KRUEGEL C, et al. A fast filter for the large-scale detection of malicious web pages[C]// Proceedings of the 20th international conference on World wide web, 2011:197-206.

[5] PRIYA M, SANDHYA L, THOMAS C. A static approach to detect drive-by-download attacks on webpages[C]// 2013 International Conference on Control Communication and Computing (ICCC), 2013:298-303.

[6] SHENG S, WARDMAN B, WARNER G, et al. An empirical analysis of phishing blaclists[C]//The 6th Conference on Email and Anti-Spam, 2009:59-78.

[7] ALEROUD A, ZHOU L N. Phishing environments, techniques and countermeasures: a survey[J]. Computers & Security, 2017, 68:160-196.

[8] 彭成维,云晓春,张永铮,等. 一种基于域名请求伴随关系的恶意域名检测方法[J]. 计算机研究与发展, 2019,56(6):1263-1274.

[9] SAHINGOZ O K, BUBER E, DEMIR O, et al. Machine learning based phishing detection from URLs[J]. Expert Systems With Applications, 2019, 117:345-357.

[10] ZHANG M, XU B Y, BAI S, et al. A deep learning method to detect web attacks using a specially designed cnn[C]//International Conference on Neural Information Processing, 2017:828-836.

[11] 崔艳鹏,刘咪,胡建伟. 基于 CNN 的恶意 Web 请求检测技术[J]. 计算机科学,2020,47(2):281-286..

[12] BAHNSEN A C, BOHORQUEZ E C, VILLEGAS S, et al. Classifying phishing URLs using recurrent neural networks[C]//2017 APWG Symposium on Electronic Crime Research (eCrime), 2017:1-8.

[13] LE H, PHAM Q, SAHOO D, et al. URLNet: learning a URL representation with deep learning for malicious URL detection[C]//Research Collection School Of Computing and Information Systems, 2018:1-13.

[14] GILMER J, SCHOENHOLZ S S, RILEY P F, et al. Neural message passing for quantum chemistry[C]//Proceedings of the 34th International Conference on Machine Learning, 2017:1263-1272.

(编辑:高明霞)

(上接第 437 页)

[10] SUTSKEVER I, VINYALS O, LE Q V. Sequence to sequence learning with neural networks[C]//Proceedings of the 27th International Conference on Neural Information Processing Systems-Volume 2, 2014:3104-3112.

[11] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems, 2017:6000-6010.

[12] LIU Y H, OTT M, GOYAL N, et al. RoBERTa: A robustly optimized BERT pretraining approach[EB/OL]. (2019-12-24)[2021-4-1]. https://openreview.net/forumid=SyxS0T4tvS.

[13] DEVLINJ, CHANGM W, 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: Human Language Technologies, Volume 1, 2019:4171-4186.

[14] WU H C, LUK R W P, WONG K F, et al. Interpreting TF-IDF term weights as making relevance decisions[J]. ACM Transactions on Information Systems, 2008, 26(3):1-37.

[15] ROBERTSON S, ZARAGOZA H. The probabilistic relevance framework: BM25 and beyond[J]. Foundations and Trends in Information Retrieval, 2009, 3(4): 333-389.

(编辑:刘彦茹)