DOI:10.19651/j.cnki.emt.2416680

基于多尺度注意力特征融合的恶意 URL 检测研究*

马栋林 陈伟杰 赵 宏 宋佳佳
(兰州理工大学计算机与通信学院 兰州 730050)

摘 要: 针对当前恶意 URL 检测模型在处理复杂结构和多样化字符组合的 URL 时,存在特征提取单一和检测精度不高的问题,提出了一种基于多尺度注意力特征融合的恶意 URL 检测模型。首先,采用 Character Embeddings 和 DistilBERT 方法分别对字符和单词进行编码,以捕获 URL 字符串中字符级和词级特征表示。其次,通过改进卷积神经网络(CNN)提取不同尺度的字符结构特征和词级语义特征,并结合双向长短期记忆网络(BiLSTM)进一步提取深层次序列特征。此外,为了实现字符级与词级多尺度特征的动态融合,创新性地引入注意力特征融合模块(AFF),有效降低信息冗余并提升对长距离序列特征的提取能力。实验结果表明,所提模型与其他基准模型相比,准确率提升了 0.32%~4.7%,F1 分数提升了 0.46%~5.5%,并在 ISCX-URL2016 等数据集上也达到了较好的测效果。

关键词: 恶意 URL 检测;多尺度特征;卷积神经网络;双向长短时记忆网络;注意力特征融合
中图分类号: TN 391 **文献标识码:** A **国家标准学科分类代码:** 510.2020

Research on malicious URL detection based on multi-scale attention feature fusion

Ma Donglin Chen Weijie Zhao Hong Song Jiajia
(School of Computer Science and Communication, Lanzhou University of Technology, Lanzhou 730050, China)

Abstract: To address the issues of single feature extraction and low detection accuracy in current malicious URL detection models when handling URLs with complex structures and diverse character combinations, this paper proposes a malicious URL detection model based on multi-scale attention feature fusion. First, Character Embeddings and DistilBERT are employed to encode characters and words separately, capturing both character-level and word-level feature representations in URL strings. Next, an improved convolutional neural network (CNN) is used to extract multi-scale character structural features and word-level semantic features, while a bidirectional long short-term memory (BiLSTM) network is employed to further extract deep sequence features. Additionally, an innovative attention feature fusion (AFF) module is introduced to dynamically fuse multi-scale features at both the character and word levels, effectively reducing information redundancy and enhancing the extraction of long-range sequence features. Experimental results show that the proposed model outperforms other baseline models, with accuracy improvements ranging from 0.32% to 4.7% and F1 score improvements from 0.46% to 5.5%, achieving excellent detection performance on datasets such as ISCX-URL2016.

Keywords: malicious URL detection; multi-scale features; convolutional neural network; bidirectional long short-term memory network; attention feature fusion

0 引 言

随着互联网技术的快速发展,网络攻击事件愈发频繁,木马、病毒、钓鱼网站等恶意行为层出不穷,对互联网的稳定运行造成了严重威胁。根据反网络钓鱼工作组(anti-phishing working group, APWG)的报告表明,2023 年钓鱼攻击事件接近五百万起。这种攻击持续上升的趋势表明,恶意 URL 在不断的更新和蔓延,给网络安全带来了极大的挑战[1]。因此,如何有效检测和防御恶意 URL 对于保障网络安全是一项亟需解决的问题。

在恶意 URL 检测任务中,传统的黑名单检测方法依赖于检测已知黑名单列表中的恶意 URL[2]。尽管该方法

收稿日期:2024-08-16
* 基金项目:国家自然科学基金(62166025)项目资助

能够保持较低的误报率，但仍需网络安全专家定期更新，以应对不断变化的新威胁。为了解决这一问题，研究者提出了机器学习的方法[3-4]，相较于黑名单检测方法有了明显的提升，但机器学习方法通常需要手动提取特征，且难以自动捕捉未知的恶意特征，攻击者可以通过设计相应特征避免被检测到，这使得基于机器学习的检测系统维护非常困难。近年来，研究者们不断提出基于深度学习的恶意 URL 检测方法[5-6]，通过自动提取和利用 URL 字符串中的复杂组合特征，更有效地识别和应对各种恶意 URL 攻击。例如，Le 等[7] 提出的 URLNet 模型利用卷积神经网络(convolutional neural network，CNN)同时处理 URL 中的字符和单词，克服了传统词袋模型特征提取的局限性。然而该模型在处理高维稀疏数据时对稀有词汇的处理过度依赖于词嵌入的泛化能力。相比之下，Zhu 等[8] 提出的 CCBLA 模型结合了 CNN 和双向长短期记忆网络(Bi-directional long short-term memory，BiLSTM)，并通过引入注意力机制，提升了对 URL 各部分特征重要性的识别。但该模型在分段处理 URL 字符串时，可能忽视部分跨段的全局特征。Nanda 等[9] 提出的 BiLSTM-GHA-CNN 模型通过集成高速公路网络和门控机制，实现了对重要特征的快速捕捉和网络的高效收敛。随着 Transformer 模型在自然语言处理任务中的广泛应用，Wang 等[10] 提出的 TCURL 模型结合了 CNN 与 Transformer，较好地平衡了局部相关性和长期依赖关系的捕捉。双分支架构有助于多种信息的融合，同时也可能会在特征融合过程中引发信息冗余或冲突问题。

现有的深度学习技术在提升恶意 URL 检测的性能方面取得了的进展，但在应对复杂结构和多样化字符组合的 URL 时，仍存在检测效果较差的问题。传统方法往往局限于表面字符特征的分析，忽略了 URL 的深层次语义特征，导致在处理混合字符集、特殊符号及复杂域名变换的恶意 URL 时，难以全面捕捉其规律。其中，CNN 擅长提取局部字符特征，但在理解 URL 整体语义和上下文关系上有所不足；而 LSTM 虽能处理序列数据中的长期依赖关系，但在面对大规模、高复杂度的 URL 数据时，计算效率和泛化能力受到限制[11]。为了解决这些问题，本文提出了一种基于多尺度注意力特征融合的恶意 URL 检测模型。该模型采用高效的词汇与字符编码结合技术，有效捕捉 URL 的多维尺度特征表示，并结合 DCNN 和 BiLSTM 的优势，提取精细的字符级结构特征以及深层次的词级语义特征。最后创新性的引入注意力融合模块，该模块能够动态地识别并筛选出对模型关键的特征，且能有效融合不同尺度的特征。解决了当前研究方法在面对构词复杂的 URL 时误报率高和检测效率低的问题。

1　恶意 URL 检测模型设计

1.1　模型结构

图 1 展示了提出的基于多尺度注意力特征融合的恶意 URL 检测模型，由数据预处理、特征提取、特征融合以及特征分类模块构成。在数据预处理阶段，首先对 URL 进行标准化处理，并按照字符和 WordPiece 方法进行分词，之后，使用 Character Embeddings 和 DistilBERT 技术对 URL 进行编码，分别捕捉字符尺度的细节特征和词级尺度的语义特征表示。在特征提取阶段，模型采用双路径特征提取方式，通过 DCNN 对编码后的字符级和词级向量进行多层卷积运算，以初步提取 URL 中的组合特征和语义特征。随

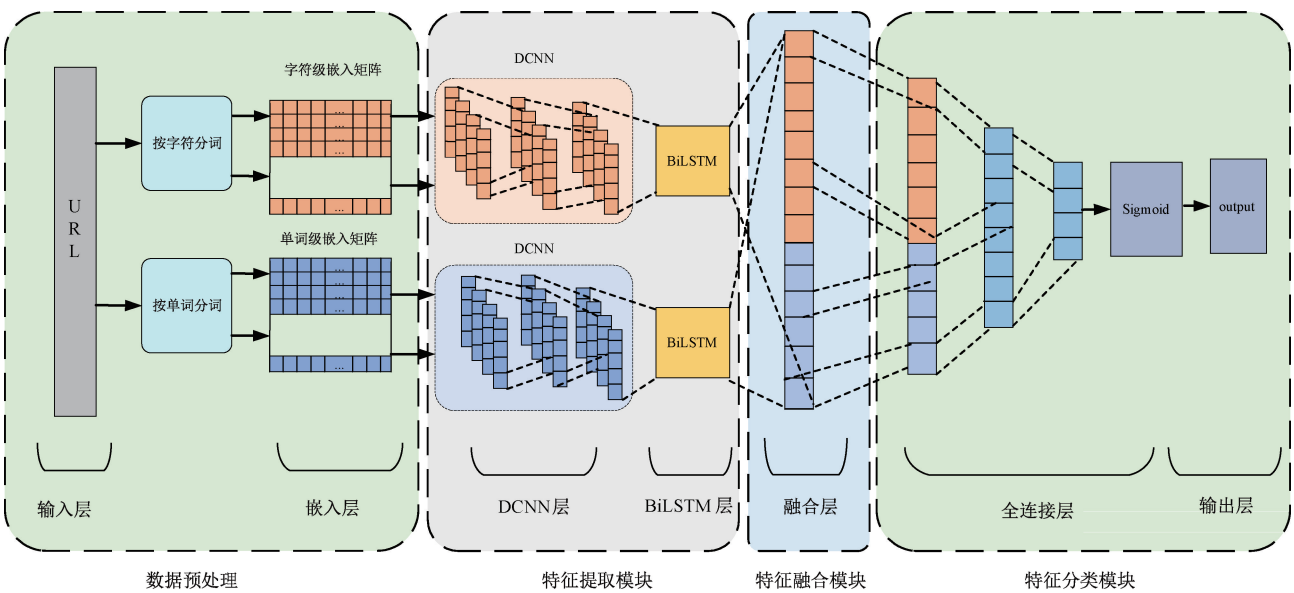

图 1　基于多尺度注意力特征融合的恶意 URL 检测模型

Fig. 1　Malicious URL detection model based on multi-scale attention feature fusion

后,结合 BiLSTM 对这些特征进行处理,以提取 URL 前后关联的深层次序列特征。为了进一步增强特征的表示能力,模型引入了多尺度注意力特征融合模块,通过关注不同尺度特征中的关键信息并进行加权,来实现多尺度特征的融合。最后,通过全连接层对融合后的特征进行整合,并输出最终的分类结果,实现恶意 URL 的检测。

1.2 URL 预处理和向量化编码

1) URL 数据预处理

给定一个去重后的数据集,根据对 URL 字符长度的数据统计分析如图 2 所示,发现其长度主要集中在 200 字符以内。因此,将每个 URL 表中的 u_i 长度 len 设置为 200 字符。如果 URL 长度 len 超过 200,则截断多余的字符;如果长度 len 不足 200,则用"0"标记进行填充[12]。

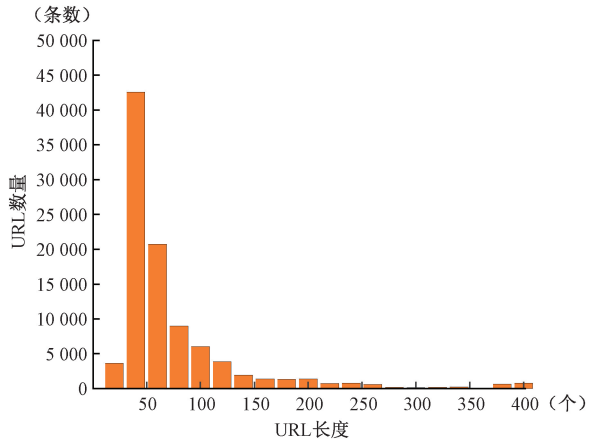

图 2 URL 长度分布图

Fig. 2 Distribution of URL lengths

2)向量化编码

不同的向量化表示方法可能会影响模型的检测效率和准确率。为了找到最佳的预处理方案,对比了多种词向量编码方式,包括传统的 One-hot 编码、Word2Vec[13]、BoW 和先进的 BERT 及其变体 DistilBERT[14] 等。发现 BERT 在词级编码上表现较好,能够捕捉丰富的语义信息;而 Character Embedding[15] 方法在字符级编码上展现出良好的鲁棒性,对词边界的结构处理更为精确。鉴于 DistilBERT 在保持 BERT 性能的同时又能减少计算资源的消耗,最终选择 DistilBERT 进行词级特征编码。

(1)字符级向量编码处理方式

采用 Character Embedding 编码方法将 URL 字符串编码成固定长度的向量表示。首先,定义一个包含 URL 数据库中所有可能字符的字典 D,并为每个字符分配唯一的索引,这些索引与字典中的字符一一对应,确保每个字符在编码过程中都能被准确识别和转换。通过这种方式,构建了一个包含 83 个字符的 URL 字符集。其次,采用字符级嵌入技术为每个字符编码生成一个稠密的向量表示,通过拼接每个字符向量来构建 URL 的字符级特征向量 $S=[e_1、e_2、\cdots、e_l]$,$S \in R^{l\times d}$,其中 $e_i \in R^d$ 为第 i 字符的矩阵向量,d 为字符嵌入的向量维度。

(2)词级向量编码处理方式

DistilBERT 是一种轻量级的预训练语言模型,由 BERT 蒸馏得到,与 BERT 相比,DistilBERT 在模型结构和参数数量上进行了精简。在使用 DistilBERT 对 URL 进行词级编码前,首先,对原始的 URL 数据进行预处理。使用内置的 *WordPiece* 算法进行分词,该过程从基本字符开始,逐渐通过统计分析字符组合的频率,将常见的字符对合并成子词。这一迭代过程持续地创建新的合并规则,直到构建出一个全面的子词库。如果 URL 中的最后几个字符无法构成一个有意义的词,这些字符将被视为特殊字符并作为单独的元素附加到字符串末尾。最终,每个单词都能准确地映射到词汇表中的对应字词,为深度学习模型提供精确和高效的输入数据。使用 DistilBERT 预训练模型将 URL 映射为一个固定维度的嵌入向量 $W=\{w_1, w_2, \cdots, w_n\}$,其中 w_i 为第 i 个单词的向量。

1.3 特征提取模块

特征提取模块采用双路径网络设计,分别对 URL 的字符级和词级编码特征 S 和 W 进行提取。两个路径均采用 DCNN 与 BiLSTM 的组合网络,但在参数设置上进行了差异化调整,以确保在处理不同尺度特征时保持结构一致性,从而增强模型对 URL 多尺度特征的捕捉能力。

1)DCNN 模块设计

DCNN 作为对 CNN 的进一步优化和深化,凭借其强大的自动特征学习和表达能力,能够更深入地挖掘 URL 中的关键信息。通过叠加多层卷积和池化操作,能够逐层提取 URL 中的高级抽象特征,从而实现对 URL 更深入的理解和更准确的分类。具体结构如图 3 所示。

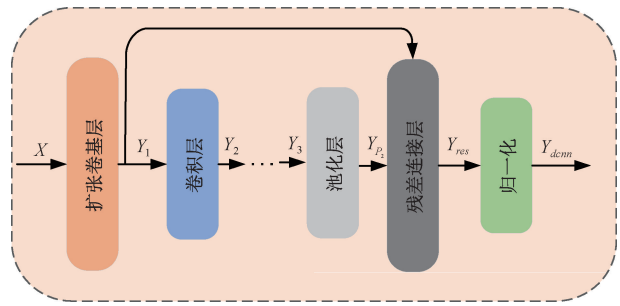

图 3 DCNN 模块结构

Fig. 3 Structure of the DCNN module

为了捕获 URL 中的广泛结构信息,第 1 层采用扩张卷积技术,通过扩大卷积核的感受野来增强对深层次特真的捕捉能力。第 2 层和第 3 层采用标准卷积层和池化层,用于进一步提取 URL 的局部关键特征。为提升网络的学习效率并防止梯度消失。第 4 层引入了残差连接,对更深层次的特征进行整合。最后采用批归一化进一步稳定了学习过程,使得模型在处理复杂 URL 结构时表现更加稳

健。第 1 层扩张卷积的计算公式如式(1)所示。

$$\boldsymbol{Y}_1 = ReLU(W_1 \times \boldsymbol{X} + b_1) \tag{1}$$

其中，$\boldsymbol{Y}_1$ 是卷积后的特征向量，W_1 是权重参数，$\boldsymbol{X}$ 是经过编码后的字符和单词的输入向量，b_1 是偏置，ReLU 是激活函数，卷积核大小为 5×5，步长为 1。第 2 层和第 3 层卷积层是对第一层提取到特征作精细处理，为了减小特征图的维度并增强特征的平移不变性，在卷积层后添加了最大池化层。计算公式如式(2)、(3)所示。

$$\boldsymbol{Y}_n = ReLU(W_n \times Y_{n-1} + b_n) \tag{2}$$

$$Y_{p_{n-1}} = MaxPool(Y_n, 2, 2) \tag{3}$$

其中，$\boldsymbol{Y}_n$ 是第 n 层卷积后的特征向量，W_n 是第 n 层权重参数，b_n 是第 n 层的偏置，$n \in [2,3]$，卷积核大小为 3×3，步长为 1。第 4 层的残差连接层和批归一化计算公式如式(4)～(5)所示。

$$\boldsymbol{Y}_{res} = Y_1 + Y_3 \tag{4}$$

$$Y_{dcnn} = BN(Y_{res}) \tag{5}$$

其中，$\boldsymbol{Y}_{res}$ 是残差链接后的特征向量，Y_{dcnn} 是 DCNN 模块最后的输出。

2)BiLSTM 模块设计

由于 URL 结构的复杂性，仅依赖 DCNN 提取特征可能会忽略 URL 字符串序列间的关联性与上下文的依赖关系。为此，本文在 DCNN 的基础上引入了 BiLSTM，以捕捉 URL 中前后关联的深层次序列特征。BiLSTM 不仅能够从前向和反向两个方向上对 URL 特征进行编码，还能更好地处理 URL 中的语义信息和复杂的上下文依赖性。BiLSTM 结构原理图如图 4 所示。

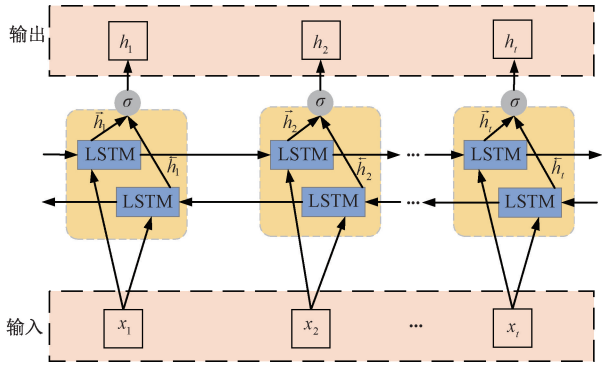

图 4　BiLSTM 模块结构图

Fig. 4　Structure of the BiLSTM module

根据双向 LSTM 的模型结构，前向和反向的输出共同构成了最终的输出，具体如式(6)～(8)所示。

$$\vec{h}_t = LSTM(X, \vec{h}_{t-1}) \tag{6}$$

$$\overleftarrow{h}_t = LSTM(X, \overleftarrow{h}_{t-1}) \tag{7}$$

$$Y_{BiLSTM} = \vec{h}_t \oplus \overleftarrow{h}_t \tag{8}$$

其中，X 是 DCNN 提取的字符级和词级特征，$\vec{h}_t$ 和 $\overleftarrow{h}_t$ 分别为前向和反向 LSTM 的隐藏层状态，Y_{Bilstm} 为 BiLSTM 网络提取到的双路径时序特征 f_{char} 和 f_{word}。

1.4　特征融合模块

通过双路径 DCNN 和 BiLSTM 网络后，获取了丰富的字符结构特征和词级语义特征，然而，这一过程极易伴随特征冗余的问题。为了优化特征表达并提升模型性能，引入了注意力特征融合模块。该模块能够动态地识别并筛选出对模型关键的特征，同时也能有效抑制不相关或冗余的信息，从而有效降低模型的计算复杂度并提升处理效率。注意力特征融合模块如图 5 所示。

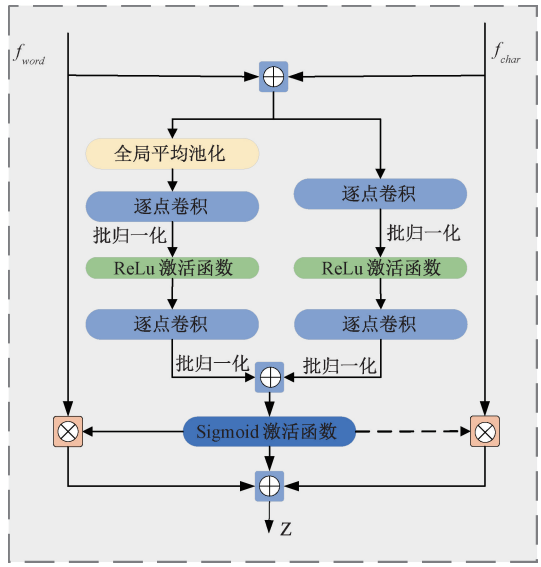

图 5　注意力特征融合模块

Fig. 5　Attention feature fusion module

在初始阶段，字符级特征首先通过逐点卷积运算和 ReLU 激活函数进行提取和优化，以增强其非线性表达能力。对于词级特征，先通过全局池化层进行降维，再利用逐点卷积和 ReLU 激活函数进一步优化。在特征融合阶段，使用 Sigmoid 函数精确计算字符级和词级特征的注意力权重，并调整其在特征表示中的重要性。最后，将调整后的特征与原始词级和字符级特征结合，生成最终的融合特征向量，从而提高模型的整体性能。注意力特征的具体计算方法如式(9)所示。

$$Z = M(f_{word} \oplus f_{char}) \otimes f_{word} + (1 - M(f_{word} \oplus f_{cher})) \otimes f_{char} \tag{9}$$

其中，M 代表生成的注意力权重，Z 是融合后的特征，f_{char} 和 f_{word} 分别表示字符级和词级特征，$\oplus$ 表示初始特征集成，$\otimes$ 表示各特征元素相乘。在注意力权重计算过程中，M 函数采用多尺度通道注意力模块(MS-CAM)来实现，其计算过程如式(10)～(12)所示。

$$M(X) = \mathrm{ReLU}(L(X) + g(X)) \tag{10}$$

$$L(X) = B(PWConv^{(2)}(\delta(B(PWConv^{(1)}(X))))) \tag{11}$$

$$g(X) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} X[i,j] \tag{12}$$

其中,X 表示字符与词级特征相加的特征向量,ReLU 是激活函数,$PWConv$ 表示逐点卷积操作,B 函数批量归一化,H 和 W 分别是特征 X 的高度和宽度。

1.5 特征分类模块

在特征分类阶段,采用两层全连接层处理注意力特征融合模块后的特征 Z,引入 ReLU 激活函数以提升模型捕捉复杂非线性关系的能力,并在隐藏层和第 2 层全连接层之间加入 Dropout 层以防止过拟合,从而增强模型的泛化能力。特征分类模块原理如图 6 所示。

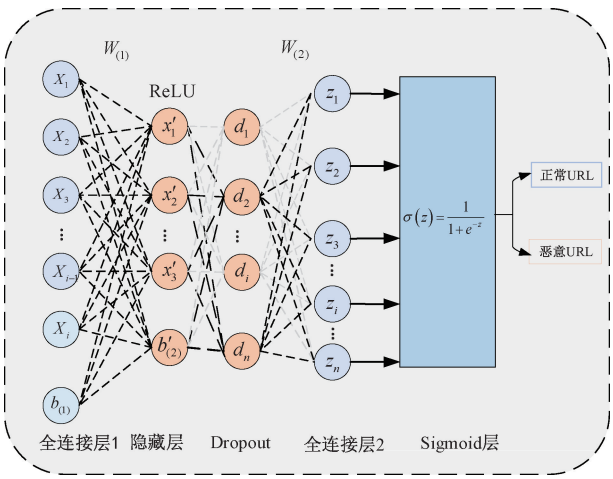

图 6 特征分类模型图

Fig. 6 Feature classification model diagram

最后,采用 Focal Loss 损失函数以提高难分类样本的预测准确性,并通过 Sigmoid 函数将输出归一化为概率值。Focal Loss 损失函数如式(13)所示。

$$L_{fl}=\begin{cases}-\alpha(1-y')^{\gamma}\log y', y=1,\\-(1-\alpha)y'\log(1-y'), y=0.\end{cases} \tag{13}$$

其中,α 是类别权重系数,y'_i 为样本真实标签,y_i 为模型预测标签。

2 实验设计与结果分析

本文实验在 Linux 操作环境下进行,硬件配置为 Tesla V100 32 G GPU。实验过程中,使用 Python 3.8.0 作为开发语言,搭建了 Tensorflow 2.6.0 和 Keras 2.6.0 深度学习框架。通过该框架,完成了模型的搭建、训练与测试,同时对模型的性能进行了评估与优化,确保实验结果的可靠性与有效性。

2.1 数据采集和预处理

为了评估所构建的模型性能,从多个公开数据集中采集 URL 样本,主要从 PhishTank、Malware Patrol、ISCX 2016 等数据集[16]中获取恶意 URL 样本,同时从 Alexa 和 Dmoz 排名最高的列表中获取合法 URL 样本,以提升数据集的多样性。在数据集做重处理后,数据集共包含 95 920 条标记明确的 URL 样本,其中包括 47 910 条恶意 URL 样本和 48 010 条合法 URL 样本,二者比例约为 1∶1。为进一步评估模型的泛化能力,按照 7∶3 的比例将恶意 URL 和合法 URL 随机划分成训练集与测试集,如表 1 所示。

表 1 训练集与测试集划分

Table 1 Training and test set division

类别	标签	训练集	测试集
合法 URL	0	33 607	14 403
恶意 URL	1	33 537	14 373
合计	—	66 144	28 776

2.2 实验设计

1)参数设计

为了全面捕获 URL 的特征,本文采用双路径网络的方式分别提取词级和字符级的多尺度特征。为了进一步提升模型的性能,通过经验值和网格搜索法确定模型超参数数值,具体结果如表 2 所示。

表 2 URL 特征提取模型的参数设置

Table 2 Parameter settings for URL feature extraction models

参数名	参数意义	字符级	词级
Length	字符(或词)的限定长度	200	200
Filter	DCNN 层 filter 个数	3	3
Kernel	DCNN 层 kernel 大小	32	6
Stride	DCNN 层步长	1	1
Dropout	Dropout 概率	0.3	0.4
Outsize	输出维度	128	64
Learning	模型权重更新的幅度	0.000 05	
Optimizer	模型权重的算法	Adam	
Epoch	训练周期	50	

2)评价指标

本文采用准确率(Accuracy)、精确率(Precision)、召回率(Recall)和 F1-score 4 项指标来评价模型性能,各项指标定义如式(14)~(17)所示。

$$Accuracy=\frac{TP+TN}{TP+TN+FP+FN} \tag{14}$$

$$Precision=\frac{TP}{TP+FP} \tag{15}$$

$$Recall=\frac{TP}{TP+FN} \tag{16}$$

$$F1-score=\frac{2\times Precision\times Recall}{Precision+Recall} \tag{17}$$

其中,TP (true positive)表示模型正确预测为恶意 URL 的样本数;FP (false positive) 表示模型错误地将合法 URL 预测为恶意 URL 的样本数;TN (true negative) 表示模型正确预测为合法 URL 的样本数;FN =(false negative)表示模型错误地将恶意 URL 预测为合法 URL

的样本数。混淆矩阵如表 3 所示。

表 3　混淆矩阵

Table 3　Confusion matrix

混淆矩阵		真实值	
		Positive	Negative
预测值	Positive	TP	FP
	Negative	TN	TN

2.3　实验结果与分析

1)实验验证

本文采用五折交叉验证法对 URL 数据集进行评估，即将整个样本集随机分成 5 个子集，每次留出一个子集用作测试集，其中包含 12 957 条恶意 URL 和 12 536 条合法 URL。而剩余的 4 个子集则合并作为训练集，其中包含 58 306 条恶意 URL 和 56 416 条合法 URL。这一过程重复 5 次，以确保每个子集都能作为测试集使用。通过对 5 次测试结果进行平均计算，可以更全面地评估模型的整体性能。图 7 展示了本文模型在五折交叉验证条件下，训练集和测试集上准确率的平均变化曲线。

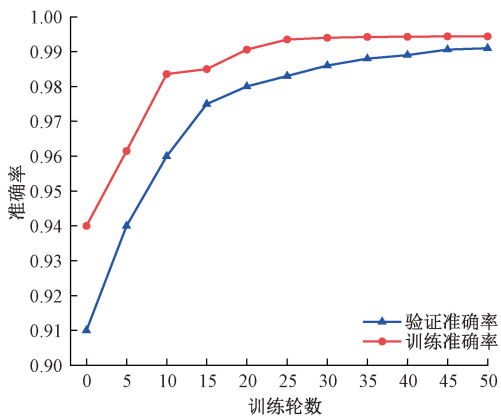

图 7　本文模型在训练集和测试集上的准确率变化曲线

Fig. 7　Accuracy change curve of the proposed model on training and test sets

根据图 4 的结果可以看出，模型在训练过程中参数收敛良好，当训练轮数达到 40 时，模型的训练准确率和测试准确率均趋于稳定。

2)消融实验

(1)不同长度 URL 的实验效果对比

为了探究不同 URL 长度对模型性能的影响，本文在相同的实验环境下，对不同长度的 URL 进行了实验对比，实验结果如图 8 所示。

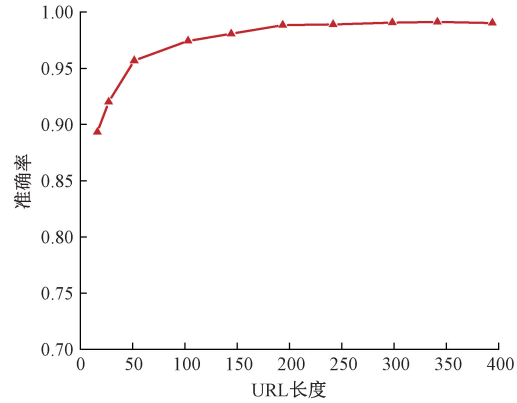

图 8　不同 URL 长度的检测准确率曲线

Fig. 8　Detection accuracy curve for different URL lengths

实验结果图 8 表明，随着 URL 长度的增加，模型的准确率呈现上升趋势，但当 URL 长度达到 200 字符时，准确率基本保持平稳。这表明，超过 200 字符的 URL 长度对模型性能的提升已不再显著。基于此，为了在模型性能和计算资源利用率之间取得最佳平衡，本文将 URL 长度设定为 200 字符。

(2)不同编码方式性能对比

为了探究不同编码方式对恶意 URL 检测模型性能的影响，本文在相同的实验环境下进行了多组对比实验，验证了多种数据特征编码技术的有效性，这些技术从传统的 One-hot 编码到更先进的词嵌入方法，如 Word2Vec 和 BERT。表 4 详细展示了采用不同编码技术的特征提取模型在恶意 URL 检测任务中的性能表现。

表 4　不同编码方式性能对比

Table 4　Performance comparison of different encoding methods　%

特征提取模型		准确率	精确率	召回率	F1 值
Char	One-hot	81.42	83.51	86.36	83.52
	Character	86.38	85.45	87.29	86.45
Word	Word2vec	76.68	76.86	76.86	75.85
	BoW	83.65	81.59	82.52	83.65
	DistilBERT	92.34	89.75	93.62	93.25
Char＋word	Character＋BERT	99.06	99.13	99.03	98.96
本文模型	Character＋DistilBERT	**99.04**	**99.12**	**99.11**	**98.97**

通过对上述对比实验结果的分析,从表 4 可知,在恶意 URL 检测任务中,Character Embeddings 在字符级特征提取上展示出了较好的性能,而 DistilBERT 在词级特征提取中则具有高效的检测效率和较低的计算资源消耗。尽管 BERT 在某些性能指标上效果较好,但其高昂的计算成本在恶意 URL 检测的实际应用场景中看来不够经济。因此,本文综合考虑字符级和词级特征,选择 DistilBERT 作为词级编码方法,有效地平衡了检测性能与计算资源之间的关系。

(3)不同分词级别性能对比与结果分析

为了验证本文模型的检测效果,本文对字符级和词级特征分别在 CNN、CNN-BiLSTM、CNN-BiLSTM-ATT 等模型上进行了消融实验。除了单独检测字符级和词级特征外,还对字符级和词级进行了融合检测。通过对各个模型在测试集上的检测效果进行分析,如表 5 所示,本文提出的模型通过充分挖掘 URL 的多尺度特征,在准确率、精确率、召回率和 F1 值等 4 个评估指标上均表现出较好的检测效果。

表 5 不同分词级别的消融实验

Table 5 Ablation study of different tokenization levels %

检测模型	准确率	精确率	召回率	F1 值
Char CNN	93.33	93.61	96.77	95.15
Word CNN	92.57	88.64	86.91	80.94
Char CNN+BiLSTM	96.42	96.78	98.25	97.93
Char CNN+ BiLSTM +ATT	96.72	97.87	95.19	95.95
Word+Char CNN	98.11	98.36	98.86	98.60
Word+Char CNN+ BiLSTM	98.39	97.85	97.75	98.05
Word+Char CNN+ BiLSTM +AFF	98.62	98.24	98.33	98.35
本文模型	**99.04**	**99.12**	**99.11**	**98.97**

观察表 5 得出,字符级特征在模型的检测效果上优于词级特征,但词级特征包含更丰富的语义信息。通过将字符级特征与词级特征进行融合,使得模型性能得到较好提升。本文模型综合考虑了字符级和词级的多尺度特征,通过引入词级特征,为字符级 URL 检测提供了额外的语义支持。这种方法不仅提高了模型对恶意 URL 的检测能力,增强了模型的泛化能力,还降低了过拟合的风险,使模型能够更好地理解和处理复杂结构的 URL。

3)主流模型对比实验

(1)评价指标对比

经过对当前主流的恶意 URL 检测模型的研究,发现大多数研究倾向于采用 CNN 与 LSTM 结合的混合神经网络模型,以应对恶意 URL 检测任务。为了验证本文提出的基于多尺度注意力特征融合的恶意 URL 检测模型的检测效果,与现有主流检测模型 URLNet[7]、CCBLA[8]、BiLSTM-GHA-CNN[9]以及 TCURL[10]模型进行了实验对比分析,实验结果如图 9 所示。

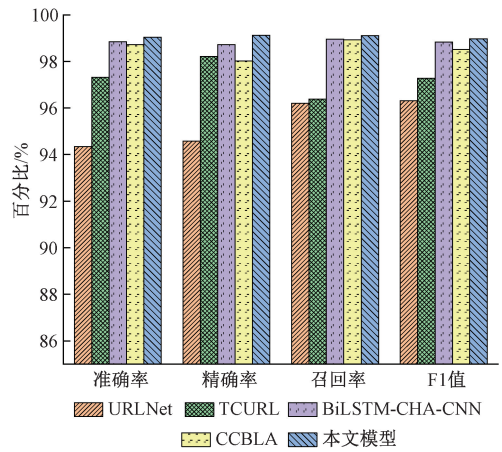

图 9 不同主流模型实验对比

Fig. 9 Experimental comparison of different mainstream models

图 9 实验结果表明,本文提出的双路径多尺度注意力特征融合的恶意 URL 检测模型在与当前主流模型对比实验中展现出较好的性能,在准确率、精确率、召回率和 F1 值上均优于现有主流模型,尤其在准确率方面,比 BiLSTM-GHA-CNN 高出约 0.22%,充分验证了所提模型在恶意 URL 识别任务中的有效性。传统模型在面对复杂结构的 URL 时,往往因特征提取能力不足或特征表示单一而难以达到理想的检测效果。相比之下,本文模型通过双路径结构,实现了对 URL 的并行且互补的特征提取,有效捕捉了 URL 中的局部细节特征和全局上下文信息。同时,结合多尺度注意力机制,模型能够自适应地聚焦于对检测任务最为关键的特征,进一步增强了特征的表征能力和模型的鲁棒性。

(2)ROC 曲线对比

为了检测模型在误报率和查全率方面的综合性能,采用 ROC 曲线进行评估。ROC 曲线中横轴表示假阳性率,纵轴表示真阳性率,曲线下面积 AUC 值越大,则模型在恶意 URL 二分类检测任务中的性能越出色。在相同的测试

集条件下，不同模型在恶意 URL 检测任务中的 ROC 曲线如图 10 所示。

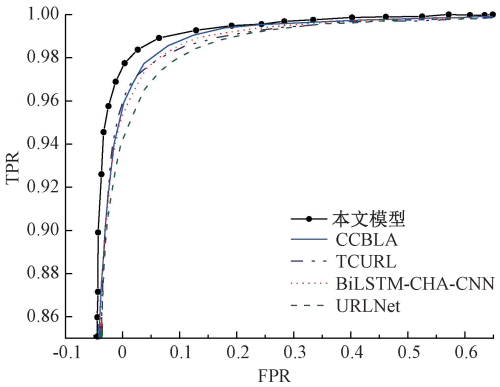

图 10　不同模型的 ROC 曲线图

Fig. 10　ROC curves of different models

观察图 10 可知，通过对 ROC 曲线的对比分析，所提模型在误报率和查全率之间取得了良好的平衡效果，其 AUC 值高于其他 4 个主流模型，这也表明本文模型在恶意 URL 二分类检测任务中具有较高的准确性。

4)模型泛化性能验证

为了验证本文提出的网络模型在不同数据集上的泛化能力，本文分别对 ISCX-URL2016、kaggle、PhishTank、Spam 数据集[17]进行二分类的实验验证。具体实验结果如表 6 所示。

表 6　不同数据集上的实验效果

Table 6　Experimental results on different datasets　%

数据集	准确率	精确率	召回率	F1-值
ISCX-URL2016	98.76	99.15	97.98	98.60
Kaggle	99.02	99.20	98.80	99.11
PhishTank	98.83	99.01	98.72	98.66
Spam	99.02	99.06	98.99	98.63
本文数据集	**99.04**	**99.12**	**99.11**	**98.97**

由表 6 显示可知，本文提出的多尺度注意力特征融合模型在恶意 URL 检测任务中表现出色。该模型通过有效的数据预处理，丰富的特征表示和多尺度特征融合等方式有效的提升了泛化能力。包括在 ISCX-URL2016、Kaggle、PhishTank 和 Spam 等数据集上准确率和 F1 值均超过 98%，精确率和召回率均达到 99%左右。这表明该模型在正确分类恶意 URL 和合法 URL 方面具有较高的准确性，并且能够在不同数据集上取得稳健的性能。

3　结　　论

为解决传统的检测方法在对构词复杂的 URL 检测时精度不高的问题，本文提出了一种基于多尺度注意力特征融合的恶意 URL 检测模型。主要包括数据预处理、特征提取、特征融合以及特征分类模块，通过融合 URL 的字符级结构特征和词级深层次语义特征，实现了对恶意 URL 的有效检测。实验结果表明，本文提出的多尺度注意力特征融合模型能够捕获 URL 更全面、高区分性的关键特征，从而提升恶意 URL 检测的性能。

在未来的研究中，可以通过增加对抗性样本来增强模型对对抗攻击的鲁棒性。此外，为了更全面地捕捉 URL 的特性，还可以考虑融合其他类型数据，例如 URL 的结构信息、用户行为数据等多模态信息，这些额外的数据将为模型提供更丰富、更准确的检测线索，从而能进一步提高恶意 URL 检测的性能和鲁棒性。

参考文献

[1]　陆向艳，刘峻. 网络钓鱼攻击分析和防范探讨[J]. 数字通信世界，2022(1)：179-181.
LU X Y，LIU J. Analysis and prevention of phishing attacks[J]. Digital Communication World，2022(1)：179-181.

[2]　王媛媛，吴春江，刘启和，等. 恶意域名检测研究与应用综述[J]. 计算机应用与软件，2019，36(9)：310-316.
WANG Y Y，WU CH J，LIU Q H，et al. A review of researchand application of malicious domain name detection[J]. Computer Applications and Software，2019，36(9)：310-316.

[3]　何智帆，姜和芳，刘涛，等. 基于机器学习与特征工程的恶意链接检测研究[J]. 科技风，2023(9)：63-65.
HE ZH F，JIANG H F，LIU T，et al. Research on malicious link detection based on machine learning and feature engineering[J]. Science and Technology，2023(9)：63-65.

[4]　吴森焱，罗熹，王伟平，等. 融合多种特征的恶意 URL 检测方法[J]. 软件学报，2021，32(9)：2916-2934.
WU S Y，LUO X，WANG W P，et al. Malicious URL detection method integrating multiple features[J]. Journal of Software，2021，32(9)：2916-2934.

[5]　ASIRI S，XIAO Y，ALZAHRANI S，et al. A survey of intelligent detection designs of HTML URL phishing attacks[J]. IEEE Access，2023，11：6421-6443.

[6]　ALMOUSA M，ANWAR M. A URL-based social semantic attacks detection with character-aware language model[J]. IEEE Access，2023，11：10654-10663.

[7]　LE H，PHAM Q，SAHOO D，et al. URLNet：Learning a URL representation with deep learning for malicious URL detection[J]. ArXiv preprint arXiv：1802.03162，2018.

[8]　ZHU ER ZH，YUAN Q X，CHEN ZH L，et al.

CCBLA: A lightweight phishing detection model based on CNN, BiLSTM, and attention mechanism[J]. Cognitive Computation, 2023, 15(4): 1320-1333.

[9] NANDA M, GOEL S. URL based phishing attack detection using BiLSTM-gated highway attention block convolutional neural network[J]. Multimedia Tools and Applications, 2024,83: 69345-69375.

[10] WANG CH G, CHEN Y Y. TCURL: Exploring hybrid transformer and convolutional neural network on phishing URL detection[J]. Knowledge-Based Systems, 2022, 258: 109955.

[11] 周燕.基于 GloVe 模型和注意力机制 Bi-LSTM 的文本分类方法[J].电子测量技术,2022,45(7):42-47.
ZHOU Y. Text classification method based on GloVe model and attention mechanism Bi-LSTM [J]. Electronic Measurement Technology, 2022, 45(7): 42-47.

[12] 卜佑军,张桥,陈博, 等.基于 CNN 和 BiLSTM 的钓鱼 URL 检测技术研究[J].郑州大学学报(工学版), 2021,42(6):14-20.
BU Y J, ZHANG Q, CHEN B, et al. Research on phishing URL detection technology based on CNN and BiLSTM [J]. Journal of Zhengzhou University (Engineering Edition), 2021,42(6):14-20.

[13] YUAN H P, YANG ZH G, CHEN X, et al. URL2Vec: URL modeling with character embeddings for fast and accurate phishing website detection[C]. 2018 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Ubiquitous Computing & Communications, Big Data & Cloud Computing, Social Computing & Networking, Sustainable Computing & Communications (ISPA/IUCC/BDCloud/SocialCom/SustainCom), 2018: 265-272.

[14] SANH V, DEBUT L, CHAUMOND J, et al. DistilBERT, a distilled version of BERT: Smaller, faster, cheaper and lighter[J]. ArXiv preprint arXiv: 1910.01108, 2019.

[15] ZHANG X, LECUN Y. Text understanding from scratchh [J]. ArXiv preprint arXiv: 1502.01710, 2015.

[16] AUNG E S, ZAN C T, YAMANA H. A survey of URL-based phishing detection[C]. DEIM Forum, 2019: 2-3.

[17] ALJABRI M, ALTAMIMI H S, ALBELALI S A, et al. Detecting malicious URLs using machine learning techniques: Review and research directions[J]. IEEE Access, 2022, 10: 121395-121417.

作者简介

马栋林,副教授,硕士生导师,主要研究方向为深度学习、网络信息安全等。

E-mail: 5920048690@qq.com

陈伟杰(通信作者),硕士研究生,主要研究方向为自然语言处理、网络安全。

E-mail:2900373335@qq.com

赵宏,教授,博士生导师,主要研究方向为深度学习、自然语言处理、计算机视觉。

宋佳佳,硕士研究生,主要研究方向为深度学习、说话人识别。