

DOI: 10.3969/j.issn.1671-0673. 2026. 01. 010

基于多门共享一私有混合专家的漏洞 CVSS 度量预测系统

王晓龙^{1,2}, 杜 晔^{1,2}, 郑天帅^{1,2}, 陈奇芳³, 关昌昊⁴

(1. 北京交通大学 智能交通数据安全与隐私保护技术北京市重点实验室, 北京 100044;
2. 北京交通大学 网络空间安全学院, 北京 100044; 3. 北京交通大学 电气学院, 北京 100044;
4. 北京交通大学 计算机科学与技术学院, 北京 100044)

摘要: 针对专家人工评定的通用漏洞评分系统(CVSS)主观性强、效率低下以及现有自动化预测方法无法有效利用 CVSS 各度量特征间的关联与差异的问题, 提出一种基于多门共享一私有混合专家(MSPMoE)的漏洞 CVSS 度量预测模型。该模型可通过共享一私有混合专家机制, 协同利用各度量间的共享特征与特有特征。首先, 通过收集到的漏洞描述语料库对 DistilBERT 预训练模型进行微调, 作为模型的特征编码器; 其次, 设计动态特征提取模块, 以自适应地选择最优特征提取策略; 最后, 使用多门共享一私有混合专家作为分类器, 其中共享专家网络用于捕获度量间的共享特征, 私有专家网络则专注于提取各度量的私有特征, 并通过门控网络动态控制不同专家的权重, 使得模型能够平衡共享知识与私有知识, 从而实现 CVSS 度量的精准预测。实验结果表明, 所提模型在 7 个 CVSS 度量上的预测准确率优于现有最佳方法, 在 8 个度量上的平均准确率达到 83.32%。

关键词: CVSS 度量预测; 漏洞评估; 多任务学习; 自然语言处理; 大模型

中图分类号: TP391.1

文献标识码: A

文章编号: 1671-0673(2026)01-0072-09

A CVSS Metrics Prediction Model of Vulnerability Based on Multi-Gate Shared and Private Mixture of Experts

WANG Xiaolong^{1,2}, DU Ye^{1,2}, ZHENG Tianshuai^{1,2}, CHEN Qifang³, GUAN Changhao⁴

(1. *Beijing Key Laboratory of Security and Privacy in Intelligent Transportation, Beijing Jiaotong University, Beijing 100044, China;* 2. *School of Cyberspace Science and Technology, Beijing Jiaotong University, Beijing 100044, China;*
3. *School of Electric Power, Beijing Jiaotong University, Beijing 100044, China;*
4. *School of Computer Science and Technology, Beijing Jiaotong University, Beijing 100044, China*)

Abstract: To address the subjectivity and inefficiency of manual common vulnerability scoring system (CVSS) assessment and the limitations of existing automated methods in capturing inter-metric relationships, a CVSS metrics prediction model of vulnerability based on a multi-gate shared and private mixture of experts (MSPMoE) is proposed. In the model, a shared-private expert mechanism is used to employed enyto jointly leverage shared and unique features across different metrics. Firstly, the DistilBERT model is fine-tuned on a collected vulnerability description corpus as the feature encoder. Then, a dynamic feature extraction module is designed to adaptively select optimal feature representation strategies. Finally, the MSPMoE classifier is adopted, where shared experts are used to capture

收稿日期: 2025-10-09; 修回日期: 2025-11-10

基金项目: 国家重点研发计划(2022YFB3105105); 北京市自然科学基金(L254063)

作者简介: 王晓龙(2000—), 男, 硕士生, 主要研究方向为网络与信息安全。E-mail: 23120481@bjtu.edu.cn

通信作者: 杜 晔(1978—), 男, 教授, 博士, 主要研究方向为系统安全、网络与信息安全。E-mail: ydu@bjtu.edu.cn

common patterns across metrics, while private experts are used to extract metric-specific features. A gating network is used to dynamically weight each expert's contribution, to balance shared and private knowledge for accurate CVSS metrics prediction. Experiments show that the proposed model outperforms state-of-the-art methods on seven CVSS metrics, achieving an average accuracy of 83.32% across all eight metrics.

Key words: CVSS metrics prediction; vulnerability assessment; multi-task learning; natural language processing; large language models

近年来,全球网络安全漏洞态势日益严峻。据美国国家漏洞数据库(National Vulnerability DataBase, NVD)统计,截至2025年中,累计漏洞已超30.3万条,自2017年起漏洞披露明显加速,日均新增数量激增^[1]。这一增长趋势加剧了网络威胁,预计2025年全球因网络犯罪的年损失可能高达10.5万亿美元^[2-3]。在此背景下,快速准确评估漏洞危害性至关重要。尽管通用漏洞评分系统(Common Vulnerability Scoring System, CVSS)标准被广泛用于量化漏洞严重性^[4],但其评估仍主要依赖安全专家人工完成,易受主观认知差异影响,一致性和客观性有限。更严重的是,人工评估效率低下,造成评分严重滞后,近50%的新披露漏洞在初期缺乏CVSS评分^[5],而有限的安全专家资源^[6]难以应对漏洞的持续增长。

因此,突破人工CVSS评估在时效、一致性、覆盖范围及资源消耗方面的限制,已成为网络安全领域的核心挑战。目前,学术界正致力于运用机器学习与深度学习技术,通过挖掘结构化漏洞描述文本来构建精准的CVSS度量预测模型。早期研究侧重于采用机器学习方法预测CVSS向量。Yamamoto等^[7]采用SLDA模型挖掘漏洞描述的潜在语义结构;Khazaei等^[8]融合支持向量机、随机森林与模糊系统构建多模型预测框架;Spanos等^[9]则基于随机森林与boosting算法实现CVSS向量的量化评估。然而,此类方法依赖特征工程,需大量人工参与。为降低对人工特征的依赖,后续研究引入深度学习技术。Han等^[10]首次结合skip-gram词嵌入与CNN自动提取文本特征,实现严重性分级,但未实现CVSS度量预测。Sahin等^[11]利用Word2vec与LSTM预测CVSS最终分数,发现LSTM性能优于CNN与XGBoost。然而,仅预测综合分数难以反映风险的具体构成,限制了其对风险处置的指导价值。为降低文献[10]方法对样本平衡性的依赖,Gong等^[12]采用多任务学习框架联合预测所有CVSS度量,通过共享表征,有效利用了度量间关联,提升了参数效率,但其过度依赖共享表征导致对单

项度量私有特征的捕捉能力较弱。

随着自然语言处理技术的进步,基于预训练语言模型的方法已成为CVSS度量预测的主流方法。Rostami等^[5]借助BERT模型提取CLS语义向量,并为每个度量训练独立分类器,虽提升预测性能,却因多分类器结构导致模型复杂度显著增加。为优化效率,后续研究采用轻量化模型。Shahid等^[13]使用BERT-Small降低计算成本;Kühn等^[14]与Costa等^[15]则采用DistilBERT进一步通过词源化与术语扩展实现了网络安全领域知识迁移。

此外,多项研究持续探索多元化的预训练模型以优化预测效果,Shi等^[16]引入XLNet以利用其排列语言建模机制捕捉更丰富的语义特征;A等^[17]与Isogai等^[18]则探索了GPT系列模型在CVSS预测中的生成式潜力;Pourbehzadi等^[19]探索融合解释主成分分析和支持向量回归的漏洞自动化评分框架;Ghosh等^[20]提出CVE-LLM模型,将大语言模型应用于漏洞自动化评估;Khan等^[21]结合SMOTE与欠采样技术解决类别不平衡问题,为CVSS预测提供了一个轻量化的框架。当前CVSS度量预测模型尚未有效兼顾任务关联性与独立性。单任务模型结构清晰但忽略度量关联,难以利用度量共享特征;多任务模型虽能捕捉度量共享特征,却不能区分不同度量的语义与预测维度差异,无法充分学习度量私有特征。为此,彭真^[22]基于DistilBERT模型,引入多门混合专家(Multi-Gate Mixture of Experts, MMoE)架构^[23]预测各CVSS度量。该方法通过多个共享专家网络与度量专属门控网络,在一定程度上利用了各度量的私有特征。然而其专家网络本身为共享结构,仍难以充分捕获各度量的私有特征。

因此,提出一种基于多门共享—私有混合专家(Multi-Gate Shared and Private Mixture of Experts, MSPMoE)的漏洞CVSS度量预测模型,通过共享与私有特征的双路径学习机制,有效捕捉度量间共性并保留其私有特征。模型包含3个核心组件:经领域语料微调的DistilBERT特征编码器,增强对安全术语的语义表征;自适应特征提取器,动态捕捉

深层上下文特征;以及 MSPMoE 分类器,通过共享专家与私有专家的协同机制与门控网络,实现共享知识与私有知识的自适应融合,全面提升预测性能。

1 模型结构

基于 MSPMoE 的漏洞 CVSS 度量预测模型处理流程主要分为两个阶段:第1阶段将漏洞描述文本输入基于微调 DistilBERT 的特征编码模块,该模块由编码器和特征提取器两部分构成。编码器对文本进行特征编码,随后特征提取器从编码结果提取句子

级特征向量;第2阶段将所得句子级特征向量输入 MSPMoE 分类器,完成对各 CVSS 度量的预测,并将预测生成的分类标签转换为对应的 CVSS 向量。

1.1 基于微调 DistilBERT 的特征编码提取模块

基于微调 DistilBERT 的特征编码提取模块由两个核心阶段构成,如图1所示。领域适应微调与特征编码提取。在 DistilBERT 模型微调阶段,通过漏洞描述语料库对 DistilBERT 模型进行参数优化,旨在提升其在漏洞预测任务中的性能表现及泛化能力。特征编码提取阶段包括特征编码和特征提取两个步骤,系统性地完成从原始文本到高阶语义特征的转化。

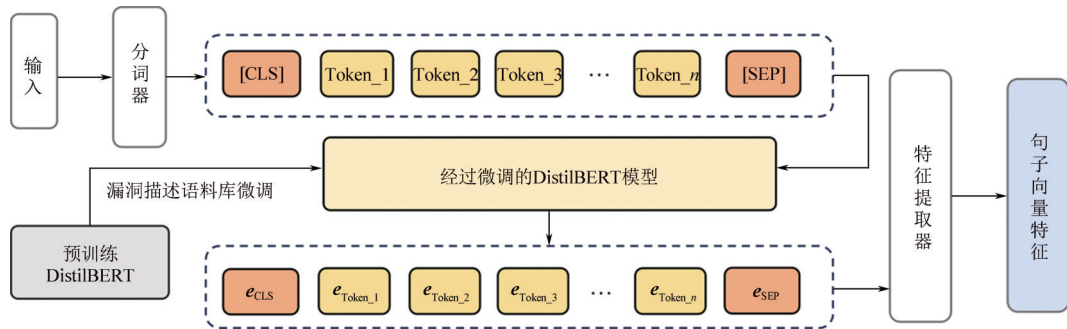


图1 基于微调 DistilBERT 的特征编码提取模块

1.1.1 DistilBERT 模型微调

对 DistilBERT 模型的微调是指利用收集到的漏洞描述语料库对 DistilBERT 预训练模型进行基于任务的微调。预训练模型本文选择在 Hugging Face 平台^[24]发布的“distilbert-base-cased”模型,该模型由6个 Transfer 层组成。然后,将1999—2021年 NVD 数据库中所有漏洞描述作为领域语料库。对 DistilBERT 的微调过程的输入输出映射关系如式(1)所示:

$$\mathbf{e} = f_{\text{DistilBERT}}(\boldsymbol{\omega}_{\text{pre}}, \mathbf{t}) \quad (1)$$

式中: $\mathbf{t} = (t_1, t_2, \dots, t_n)$ 是通过 Word2vec^[25]和 BERT 系列模型^[26-27]等 NLP 领域常用的词嵌入模型将漏洞描述 d 划分得到的一组词嵌入 token 序列,其中 n 是漏洞描述词嵌入划分的最大序列长度; $\mathbf{e} = [e_{\text{CLS}}, e_2, \dots, e_{n-1}, e_{\text{SEP}}]$ 是 DistilBERT 模型对 token 序列进行编码形成的令牌序列; $\boldsymbol{\omega}_{\text{pre}}$ 是“distilbert-base-cased”预训练模型的参数; $f_{\text{DistilBERT}}(\cdot)$ 是 DistilBERT 模型的模型结构映射函数。

通过上述微调过程,模型参数由预训练模型“distilbert-base-cased”的参数 $\boldsymbol{\omega}_{\text{pre}}$ 更新为微调参数 $\boldsymbol{\omega}_{\text{fine}}$,与从头开始训练一个 DistilBERT 模型相比,对 DistilBERT 的预训练模型进行基于任务的微调,既可以保持综合模型带来的高性能,又可以避免训

练成本极高和领域数据匮乏的问题。

1.1.2 特征编码提取

特征编码过程是将漏洞描述输入分词器,分词器会输出一组词嵌入 token 序列,微调的 DistilBERT 模型将分词器输出的词嵌入 token 序列作为输入,然后输出令牌向量,具体过程如式(2)所示:

$$\mathbf{e} = f_{\text{DistilBERT}}(\boldsymbol{\omega}_{\text{fine}}, \mathbf{t}) \quad (2)$$

式中, $\boldsymbol{\omega}_{\text{fine}}$ 为微调 DistilBERT 模型的参数,其余变量含义与式(1)相同。微调 DistilBERT 模型会输出 l 层令牌序列,一般而言,越靠近输出层,对应的令牌序列越具体,应用相关性越强。因此,本文选择最后一层令牌序列作为特征编码的结果输入特征提取器。

本文构建的特征提取模块旨在从微调后的 DistilBERT 编码器输出的令牌序列中提取得到更具信息量的句子级特征向量。尽管现有研究普遍认为令牌序列的首位 e_{CLS} 已基本融合了整个句子的特征,因此直接采用令牌序列首位 e_{CLS} 向量作为句子级特征向量,但本文发现,序列中其余令牌所蕴含的隐藏特征同样不容忽视。

这些特征既可能包含对分类任务具有判别力的关键信息,也可能引入无关噪声。为甄别并有效利用这些潜在特征,本文设计一种自适应特征提取器。

该模块以编码器输出的令牌序列作为输入,

并依据其内容动态选择特征提取策略:若令牌序列其余部分中未包含显著的任务相关特征,则采用CLS策略,直接以 e_{CLS} 向量作为句子特征向量;若令牌序列其余部分中包含有价值的任务相关特征,则采用WHOLE策略,将完整令牌序列输入BiLSTM,以融合 e_{CLS} 向量与令牌序列其余部分中包含的判别性特征,最终生成一个信息更全面的句子级特征向量。因此,将特征提取器定义如下:

$$e_{\text{sent}} = \begin{cases} e_{\text{CLS}}, & \text{CLS策略;} \\ f_{\text{BiLSTM}}(\omega_{\text{BiLSTM}}, e), & \text{WHOLE策略.} \end{cases} \quad (3)$$

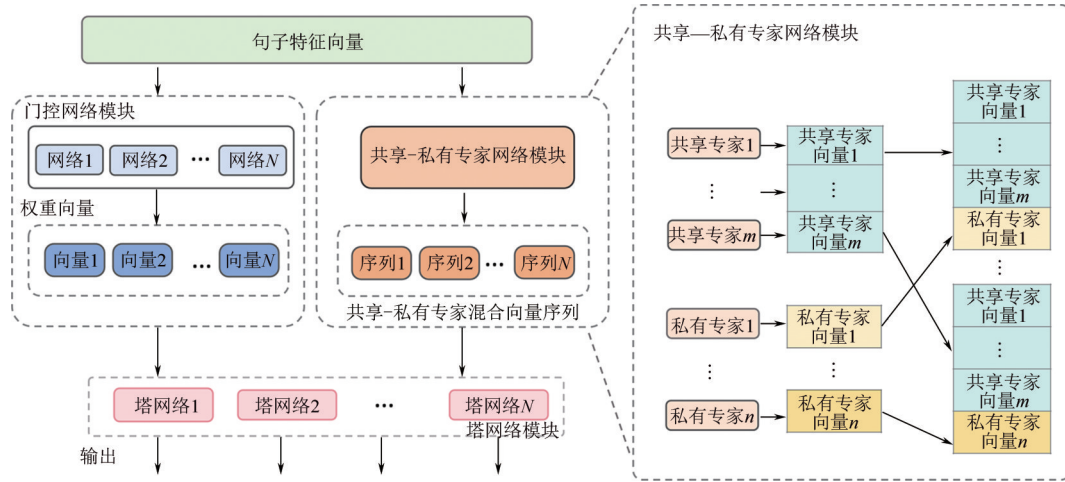


图2 多门共享-私有混合专家

共享—私有专家网络模块由共享专家网络和私有专家网络两部分构成,共享专家网络部分由多个专家网络构成,其数量与任务的数量没有直接关联;私有专家网络部分由与任务数量相同的专家网络构成,每个预测任务独享1个私有专家网络。每个专家网络是独立运算的,因此可以根据任务特点采取不同的专家网络结构。专家网络的作用是生成对漏洞描述的句子级特征向量的评价向量,具体过程如式(4)所示:

$$p = f_{\text{expert}}(\omega_{\text{expert}}, e_{\text{sent}}) \quad (4)$$

式中: e_{sent} 是漏洞描述的句子级特征向量; p 是专家网络生成的评价向量; ω_{expert} 是专家网络的权重, $f_{\text{expert}}(\cdot)$ 是专家网络的结构映射。

共享专家网络部分会生成一个共享专家评价向量序列 $p_{\text{share}} = (p_{\text{share}_1}, p_{\text{share}_2}, \dots, p_{\text{share}_m})$,其中 m 为共享专家数量;私有专家网络部分会生成私有专家评价向量 p_{private_l} ,其中 $l \in [1, n]$,其中 n 为任务数量。共享专家评价向量序列会分别与私有专家评价向量拼接,形成与任务对应的共享—私有专家混合向量序列 $p_{\text{share_private}_l} = (p_{\text{share}_1}, p_{\text{share}_2}, \dots, p_{\text{share}_m}, p_{\text{private}_l})$ 。

门控网络模块由多个门控网络构成,每个预测

式中: e_{CLS} 为令牌序列的首位; $e = [e_{\text{CLS}}, e_2, \dots, e_{n-1}, e_{\text{SEP}}]$ 为整个令牌序列; ω_{BiLSTM} 为BiLSTM模型的参数; $f_{\text{BiLSTM}}(\cdot)$ 为BiLSTM模型的映射结构; e_{sent} 为特征提取器输出的句子级特征向量。

1.2 多门共享—私有混合专家

为了对特征编码提取模块得到的句子级特征向量 e_{sent} 进行预测,本文提出一种兼顾CVSS度量共性特征与私有特征的多任务分类器MSPMoE,主要由门控网络模块、共享—私有专家网络模块和塔网络模块组成,结构如图2所示。

任务独享1个门控网络,因此门控网络的数量与任务数量相同。门控网络的作用是对输入的漏洞描述句子级特征向量进行处理,生成一个权重向量,这个权重向量代表了共享—私有专家模块中各个专家评判结果的权重占比,具体过程如式(5)所示:

$$g = f_{\text{gate}}(\omega_{\text{gate}}, e_{\text{sent}}) \quad (5)$$

式中: e_{sent} 是漏洞描述的句子级特征向量; $g = (g_1, g_2, \dots, g_m, g_{m+1})$ 是门控网络生成的各个专家向量的权重组成的向量; ω_{gate} 是门控网络的权重; $f_{\text{gate}}(\cdot)$ 是门控网络的结构映射。

塔网络模块由每个预测任务专属的塔网络构成,塔网络的作用是对共享—私有专家混合向量序列和权重向量加权求和,从而得到该预测任务的最终评价向量,并对这个最终评价向量进行降维处理,生成该任务的预测结果,具体过程如式(6)所示:

$$\hat{y} = f_{\text{tower}}(\omega_{\text{tower}}, p_{\text{share_private}}, g) \quad (6)$$

式中: $p_{\text{share_private}}$ 是共享—私有专家混合向量序列; $\hat{y}$ 是漏洞描述在该预测任务上的预测结果; ω_{tower} 是塔网络的权重; $f_{\text{tower}}(\cdot)$ 是塔网络的结构映射。

对于 MSPMoE 分类器每个 CVSS 度量预测任务,采取交叉熵损失函数,如式(7)所示:

$$L_i = - \sum_{j=1}^N y_j \log_2(\hat{y}_j) \quad (7)$$

式中: N 为每个任务的数据量; y_j 为真实标签; $\hat{y}_j$ 为预测标签; L_i 为第*i*个 CVSS 度量预测任务的损失。

为了利用所有度量预测任务的联合损失来训练本文模型,将这些损失函数组合在一起,得到新的分类代价损失函数,如式(8)所示:

$$L = \sum_{i=1}^n \lambda_i L_i \quad (8)$$

式中: L_i 为第*i*个 CVSS 度量预测任务的损失; λ_i 为第*i*个 CVSS 度量预测任务损失函数的可学习权重,用于平衡每个分类任务损失函数对框架整体损失函数的影响。

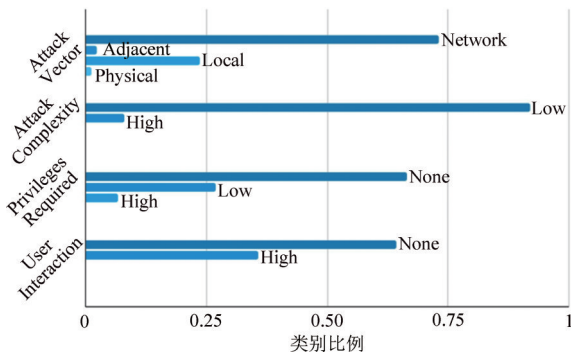
2 实验结果与分析

基于 MSPMoE 的漏洞度量预测模型开展实验,所有实验均在 Ubuntu 20.04.6 操作系统上实现,使用两块 NVIDIA GeForce RTX 4070Ti GPU,使用的深度学习框架是 pytorch。

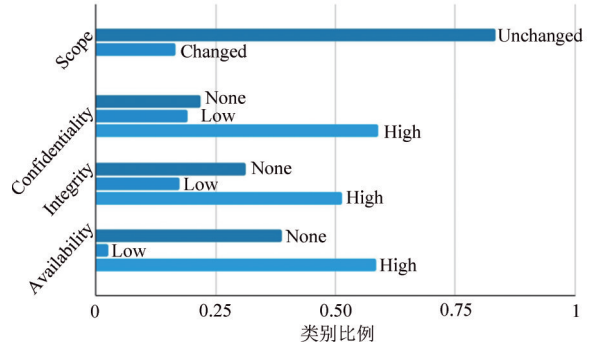
2.1 数据集与评价指标

本文所采用的数据集源自 Costa 等^[15]所构建的开源漏洞描述数据集,该数据集整合了 NVD 1999—2021 年披露的全部漏洞条目,通过数据整理得到 79 810 条数据,每条数据均包含 1 项漏洞描述信息,并附带 8 个 CVSS 3.x 版本中的基础度量指标,分别为:Attack Vector、Attack Complexity、Privileges Required、User Interaction、Scope、Confidentiality、Integrity、Availability。图 3 展示了各度量指标类别分布的视觉化结果,从中可观察到不同类别之间存在明显的样本不平衡现象。

选用平衡准确率(Balanced Accuracy)作为模型分类性能的评估指标,以更合理地反映模型在不同类别上的分类效果。各度量分类任务中平衡



(a) CVSS可利用性子度量分类的分布情况



(b) CVSS影响子度量分类的分布情况

图3 CVSS度量分类的分布情况

准确率的计算方式如下:

$$I_{\text{Recall}} = \frac{N_{\text{TP}}}{N_{\text{TP}} + N_{\text{FN}}} \quad (9)$$

$$I_{\text{BA}} = \frac{1}{n} \sum_{i=1}^n I_{\text{Recall}_i} \quad (10)$$

式中: N_{TP} 、 N_{FN} 为混淆矩阵中的元素; n 表示分类任务中类别数量; N_{TP} 表示正确预测为该分类的样本数量; N_{FN} 表示属于该分类而被预测为其他分类的样本数量; I_{Recall} 计算的是某一个分类的召回率;而 I_{BA} 对分类任务中所有分类的 I_{Recall} 取平均。

2.2 实验设置

在模型训练环节,采用基于梯度的 AdamW 优化算法结合反向传播策略实施参数更新。经超参数优化探索,学习率在 $[1 \times 10^{-5}, 1 \times 10^{-3}]$ 取值区间进行实验筛选;同时,将权重衰减系数固定为 0.01,以抑制模型过拟合风险,训练过程中设定最小批次样本量为 128,平衡计算效率与梯度估计稳定性。在文本特征表示层面,通过 WordPiece 分词机制构建词嵌入模型,并融合微调的 DistilBERT 模型进行训练,可有效捕获文本序列中的长距离依赖关系与上下文语义信息。

2.3 实验结果分析

2.3.1 模型结构实验

1)模型编码器效果分析。本文模型采用微调 DistilBERT 模型作为特征编码器。其中,微调过程基于 DistilBERT 预训练模型展开,因此预训练模型的合理选择是决定模型性能表现的核心要素。为探究不同预训练模型对最终分类任务的影响,选取并对比“distilbert-base-cased”与“cybersecurity_ner-v2”两种预训练模型。其中,“distilbert-base-cased”预训练模型属于基础型大小写敏感模型,该模型未经特定领域的针对性微调,优势在于具备良好的跨领域泛化能力,能够有效适用于多样化的分类任务场景;与之相对,“cybersecurity_ner-

v2”预训练模型则是在网络安全领域语料基础上通过NER任务微调之后的模型,可能在CVSS度量预测任务取得更好的效果。分别将上述两种预训练模型应用于所构建的模型框架中,并通过实验进行验证,实验结果见表1。

表1 编码器预训练模型平衡准确率对比 单位:%		
度量	distilbert-base-cased	cybersecurity_ner-v2
Attack Vector	76.08	74.95
Attack Complexity	77.11	74.37
Privileges Required	79.13	80.35
User Interaction	93.37	93.74
Scope	92.47	91.51
Confidentiality	86.50	86.46
Integrity	89.07	88.07
Availability	72.88	69.50

从实验结果可以看出,在6个度量上选择“distilbert-base-cased”预训练模型的分类平均准确率高于选择“cybersecurity_ner-v2”预训练模型。这表明尽管后者经过网络安全领域NER任务微调,但在CVSS分类任务中的泛化能力仍不及通用预训练模型。进一步分析可知,通过NER任务微调可能导致模型过度适应原任务,在迁移至CVSS预测时产生领域适应损耗,致使分类性能下降。因此,本文最终选用“distilbert-base-cased”预训练模型。

2)特征提取策略效果分析。本文提出的模型特征提取器可以采用不同的特征提取策略,当模型采用CLS策略,即直接使用令牌序列的第1位 e_{CLS} 作为句子特征向量,该向量经过DistilBERT模型的编码,已经融合了整个句子的特征;然而,令牌序列的其余元素可能蕴含尚未充分挖掘的隐藏信息,因此可以采用WHOLE策略对令牌序列进行处理,提取整个令牌序列的信息作为句子特征向量。分别运用上述两种策略进行特征提取,并通过实验进行验证,实验结果见表2。

从实验结果可以看出,在5个度量指标上采用CLS特征提取策略的分类准确率高出WHOLE策略,而在其余3个度量上则相反。这表明 e_{CLS} 向量确实已经融合了整个句子的特征,使用WHOLE策略可能会引入冗余或干扰特征,降低模型分类平均准确率;但是令牌序列的其余部分中的隐藏信息在部分情况下仍有助于提升分类表现。因此,实际应用中需根据任务需求选择适宜的特征提取策略。鉴于CLS策略在多数度量上表现更优,本文模型最终选用该策略。

表2 特征提取策略平衡准确率对比 单位:%		
度量	WHOLE策略	CLS策略
Attack Vector	75.25	76.08
Attack Complexity	78.41	77.11
Privileges Required	79.90	79.13
User Interaction	93.77	93.37
Scope	92.32	92.47
Confidentiality	86.45	86.50
Integrity	88.38	89.07
Availability	71.40	72.88

3)共享专家数量效果分析。本文提出模型的共享一私有专家网络模块包含多个任务私有专家网络与多个共享专家网络,任务私有专家网络的数量与任务的数量保持一致,而共享专家网络的数量作为关键可调参数,其设置对模型性能存在显著影响。本文分别将共享专家网络数量设置为2、3、4、5、6,并通过实验进行验证,实验结果见表3。

表3 共享专家数量平衡准确率对比 单位:%					
度量	2专家	3专家	4专家	5专家	6专家
Attack Vector	76.69	76.37	76.08	75.38	76.23
Attack Complexity	79.48	78.68	77.11	77.83	76.67
Privileges Required	79.95	81.16	79.13	80.21	79.16
User Interaction	93.01	93.16	93.37	92.68	93.25
Scope	92.07	91.65	92.47	90.41	91.46
Confidentiality	84.69	84.72	86.50	84.89	85.25
Integrity	87.45	87.91	89.07	86.94	87.04
Availability	70.61	72.48	72.88	69.54	67.53

从实验结果可以看出,共享专家网络数量为4的模型在5个度量分类任务上取得最佳效果,相较于其他共享专家网络数量的模型具有明显优势。进一步分析可知,共享专家网络数量过少可能导致模型无法充分提取跨度量的共性特征,而数量过多则易引发模型参数冗余、增加训练复杂度并导致过拟合风险上升,最终均致使模型性能下降。基于上述实验验证与理论分析,本文确定将共享专家网络数量设置为4。

为了更深入地了解本文模型所表现出的效果,基于上述实验确定的网络结构模型,针对8个CVSS度量,分别构建混淆矩阵,如图4所示。混淆矩阵提供了模型分类能力的综合表示,允许深入分析各种类别及其对应实际标签的正确和错误预测。通过对混淆矩阵的细致解读可知,在User Interaction、Scope、Confidentiality、Integrity等4个度量上,模型

展现出卓越的分类性能,能够精准识别并预测相应类别,体现出良好的泛化能力与稳健性。然而,在 Availability 度量的 Low 类别数据与 Attack Complexity 度量的 High 类别数据分类任务中,模型存在一定

程度的误判现象,其分类准确率有待进一步提升。这一结果表明,模型在处理该特定类别数据时,可能存在特征提取不充分或决策边界模糊等问题,这为后续模型优化与改进提供了明确的方向。

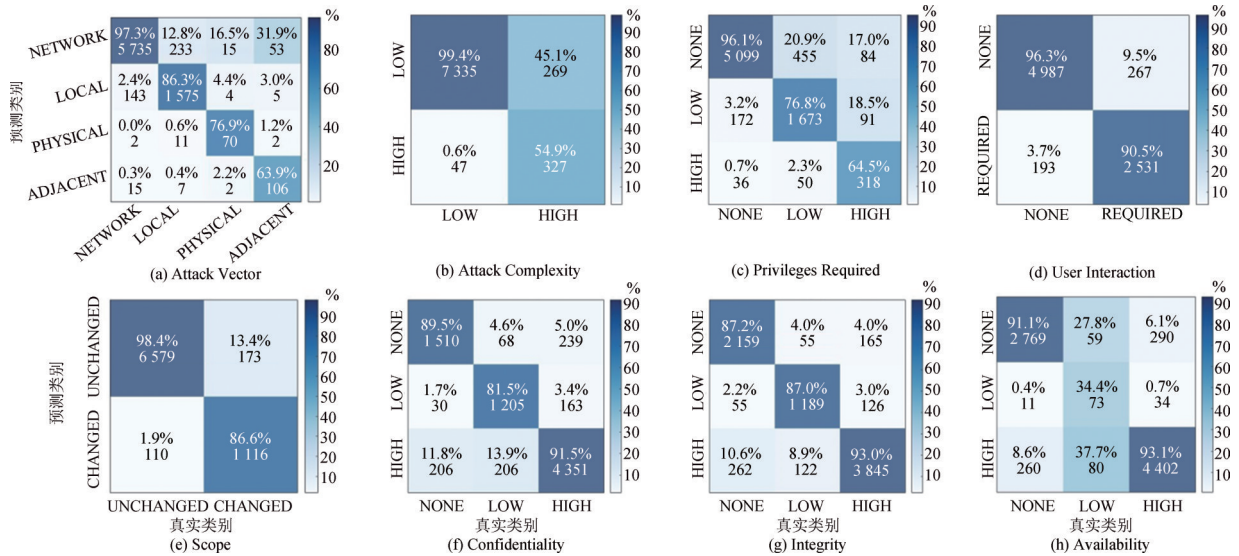


图4 模型在CVSS度量上的混淆矩阵

2.3.2 消融实验

为了验证对 DistilBERT 预训练模型进行微调以及任务私有专家网络对模型分类效果的影响,设计消融实验。移除 MSPMoE 中私有专家网络相关组件,使其退化为经典的 MMoE^[23],并将它们分别与 DistilBERT 预训练模型和微调模型进行组合,在此基础上,采用交叉组合实验设计,分别将 DistilBERT 预训练模型及微调模型与 MMoE、MSPMoE 进行组合,形成 4 组对照组:预训练模型+MMoE、微调模型+MMoE、预训练模型+MSPMoE、微调模型+MSPMoE。该实验设计通过控制变量,能够有效分离不同因素对模型性能的影响,实验结果见表 4。

从实验结果表 4 可以看出,使用 MMoE 时,微调模型在 5 个度量上的平均分类准确率高于预训

练模型;使用 MSPMoE 时,微调模型在 7 个度量上的平均分类准确率高于预训练模型。这表明预训练模型的微调过程能够有效增强其在目标领域的语义表征能力,通过参数的适应性调整,模型可更好地捕捉任务相关特征,进而提升预测任务的准确性。通过在微调模型上分别使用 MMoE 与 MSPMoE 的对比结果可以看出,使用 MSPMoE 在 6 个度量上的平均分类准确率高于使用 MMoE,这表明 MSPMoE 在有效建模 CVSS 度量间共性特征的同时,精准捕捉利用了 CVSS 度量间的私有差异特征。这种架构设计使得模型能够在共享知识与任务私有知识之间取得平衡,从而优化决策过程,显著提升预测任务的准确性。

2.3.3 对比实验

为了验证本文模型的有效性,将其分别与 3 种已知模型进行对比。第 1 种是 Costa 等^[15]提出的扩充词表模型,使用经过词源化预处理的漏洞描述文本中的高频专业术语,再扩充至 DistilBERT 词表,实现网络安全领域知识的有效迁移。第 2 种是彭真^[22]提出的 NLPMMoE 模型,在 DistilBERT 模型的基础上,通过 MMoE 框架实现对 CVSS 各度量的分类,为每个度量添加一个私有门控机制,以捕捉度量间的差异化特征。第 3 种是 Shi 等^[16]提出的采用 XLNet 模型替代 BERT,将其作为核心特征提取层,通过借助 XLNet 独特的排列语言建模机制,挖掘漏洞描述文本中更为丰富的语义表征。实验结果见表 5。

表4 消融实验

单位: %

度量	预训练模型	微调模型	预训练模型+MSP-	微调模型
	+MMoE	+MMoE	MoE	+MSP-MoE
Attack Vector	74.22	74.69	75.40	76.08
Attack Complexity	74.46	75.46	76.42	77.11
Privileges Required	77.62	79.36	80.88	79.13
User Interaction	92.82	92.41	92.40	93.37
Scope	92.28	92.48	91.98	92.47
Confidentiality	84.85	85.23	84.49	86.50
Integrity	87.58	87.54	87.59	89.07
Availability	69.09	68.48	70.85	72.88

表 5 所提模型与其他模型的平均准确率对比 单位:%				
度量	扩充词表 ^[15]	NLPM-MoE ^[22]	XLNet ^[18]	本文模型
Attack Vector	71.03	74.22	75.69	76.08
Attack Complexity	71.45	74.46	75.46	77.11
Privileges Required	75.14	77.62	77.87	79.13
User Interaction	92.05	92.82	93.39	93.37
Scope	91.12	92.28	92.17	92.47
Confidentiality	83.12	84.85	84.57	86.50
Integrity	85.77	87.58	87.54	89.07
Availability	67.51	69.09	69.48	72.88

从实验结果可以看出,本文模型的平均分类准确率在 7 个度量上均高于其他 3 种模型,且在 8 个量上的平均准确率达到 88.32%。这不仅验证了基于任务特性对预训练模型进行针对性微调的有效性,说明领域适配的参数优化能够显著提升语义表征与泛化能力;也揭示了在 CVSS 多度量预测任务中,模型需同时兼顾任务间的共性与差异性。具体而言,通过构建共享—私有知识交互机制,模型在提取跨任务共性特征的同时,有效捕捉各度量的私有特征,实现了知识共享与任务独立性之间的动态平衡。该机制使模型能够自适应调整决策策略,既避免过度强调共性造成的特征损失,也防止任务孤立处理导致的关联信息忽略,从而显著提升分类的准确性与稳健性。

2.3.4 模型泛化能力实验

为评估本文模型的泛化能力,构建两个新的测试数据集用于对比分析。由于 Costa 等^[15]所整理的数据集仅有 1999—2021 年 NVD 收录的漏洞信息,为验证模型在更近期及未来数据上的适应能力,分别从 NVD 中提取 2022—2023 年、2024—2025 年两个时间区间内各 500 条包含 CVSS 3.x 评分的漏洞记录,构成独立测试集,旨在系统检验模型在不同时间跨度数据上的预测稳定性与泛化表现,实验结果见表 6。从实验结果可以看出,在不同测试集上本文模型的表现存在一定差异。文献[15]数据集在 5 个评估度量上取得了最优结果;2022—2023 年测试集在 1 个度量上表现最佳;而 2024—2025 年测试集则在 2 个度量上达到最优。尽管存在上述差异,但 3 个数据集在各度量上的平均准确率较为接近,最大差距为 2 个百分点,最小差距为 0.27 个百分点,整体处于同一统计水平。这一结果反映出本文提出模型在不同时间跨度数据上均能保持稳定的预测性能,表明该模型具有良好的泛化能力与时间鲁棒性。

表 6 所提模型时间泛化能力实验结果 单位:%			
度量	文献[15]数据集	2022—2023 年数据集	2024—2025 年数据集
Attack Vector	76.08	77.84	78.08
Attack Complexity	77.11	77.99	76.81
Privileges Required	79.13	78.69	78.86
User Interaction	93.37	92.80	93.47
Scope	92.47	91.92	92.01
Confidentiality	86.50	85.83	84.97
Integrity	89.07	88.50	88.36
Availability	72.88	71.25	71.14

3 结束语

针对现有研究在 CVSS 度量预测任务中难以有效协调各度量间关联性与独立性的问题,本研究提出了一种融合共享专家网络与私有专家网络的分类器 MSPMoE,能够分别捕捉不同度量的共性特征与私有特征。基于该结构,进一步设计了配套的特征编码器与特征提取器,实现了对漏洞 CVSS 度量中共性及私有信息的协同建模与预测。为验证模型有效性,本研究与扩充词表模型、NLPMoE 模型及 XLNet 模型进行了系统对比,实验结果表明,本文模型在 7 个 CVSS 度量上的准确率均优于其他模型,在 8 个度量上的平均准确率达到 88.32%,证实漏洞 CVSS 度量间确实同时存在共性与特性,且在进行漏洞 CVSS 度量预测时需同时考虑到两方面。然而,本文模型在 8 个度量上的平均准确率存在明显的不均衡现象,最大差值可达 20.49 个百分点。未来工作中,将深入分析导致此类差异的内在原因,并在保持高准确率度量性能的前提下,重点提升低准确率度量的预测效能。

参考文献:

[1] NIST. National Vulnerability Database [EB/OL]. (2024-08-27)[2025-11-10]. <https://nvd.nist.gov>.

[2] MORGAN S. Cybercrime to cost the world \$10.5 trillion annually by 2025[EB/OL]. (2020-11-13)[2025-11-10]. <https://cybersecurityventures.com/cybercrime-damage-costs-10-trillion-by-2025>.

[3] 汤萌萌,罗鹏,方晨,等.网络安全领域的事件抽取研究综述[J].信息工程大学学报,2025,26(2):224-230.

[4] MELL P, SCARFONE K, ROMANOSKY S. Common vulnerability scoring system [J]. IEEE Security & Privacy, 2006,4(6):85-89.

[5] ROSTAMI A, MOHAMMADI M, KARIMIPOUR H. Reliability assessment of cyber-physical power systems consid-

- ering the impact of predicted cyber vulnerabilities[J]. International Journal of Electrical Power & Energy Systems, 2023, 147: No.108892.
- [6] AKSU M U, DILEK M H, TATLI E İ, et al. A quantitative CVSS-based cyber security risk assessment methodology for IT systems[C]//Proceedings of the 2017 International Carnahan Conference on Security Technology. Piscataway, USA: IEEE, 2017. DOI: 10.1109/CCST.2017.8167819.
- [7] YAMAMOTO Y, MIYAMOTO D, NAKAYAMA M. Text-mining approach for estimating vulnerability score[C]//Proceedings of the 2015 4th International Workshop on Building Analysis Datasets and Gathering Experience Returns for Security. Piscataway, USA: IEEE, 2015: 67-73.
- [8] KHAZAEI A, GHASEMZADEH M, DERHAMI V. An automatic method for CVSS score prediction using vulnerabilities description[J]. Journal of Intelligent & Fuzzy Systems, 2015, 30(1): 89-96.
- [9] SPANOS G, ANGELIS L, TOLOUDIS D. Assessment of vulnerability severity using text mining[C]//Proceedings of the 21st Pan-Hellenic Conference on Informatics. New York, USA: ACM, 2017: No.49.
- [10] HAN Z B, LI X H, XING Z C, et al. Learning to predict severity of software vulnerability using only vulnerability description[C]//Proceedings of the 2017 IEEE International Conference on Software Maintenance and Evolution. Piscataway, USA: IEEE, 2017: 125-136.
- [11] SAHIN S E, TOSUN A. A conceptual replication on predicting the severity of software vulnerabilities[C]//Proceedings of the Evaluation and Assessment on Software Engineering. New York, USA: ACM, 2019: 244-250.
- [12] GONG X, XING Z C, LI X H, et al. Joint prediction of multiple vulnerability characteristics through multi-task learning[C]//Proceedings of the 2019 24th International Conference on Engineering of Complex Computer Systems. Piscataway, USA: IEEE, 2019: 31-40.
- [13] SHAHID M R, DEBAR H. CVSS-BERT: explainable natural language processing to determine the severity of a computer security vulnerability from its description[C]//Proceedings of the 2021 20th IEEE International Conference on Machine Learning and Applications. Piscataway, USA: IEEE, 2021: 1600-1607.
- [14] KÜHN P, RELKE D N, REUTER C. Common vulnerability scoring system prediction based on open source intelligence information sources[J]. Computers & Security, 2023, 131: No.103286.
- [15] COSTA J C, ROXO T, SEQUEIROS J B F, et al. Predicting CVSS metric via description interpretation[J]. IEEE Access, 2022, 10: 59125-59134.
- [16] SHI F, KAI S F, ZHENG J H, et al. XLNet-based prediction model for CVSS metric values[J]. Applied Sciences, 2022, 12(18): No. 8983.
- [17] A M, KOTA K, BABU A S, et al. CVE severity prediction from vulnerability description-a deep learning approach[J]. Procedia Computer Science, 2024, 235: 3105-3117.
- [18] ISOGAI S, OGATA S, KASHIWA Y, et al. Toward extracting learning pattern: a comparative study of GPT-4o-mini and BERT models in predicting CVSS base vectors[C]//Proceedings of the 2024 IEEE 35th International Symposium on Software Reliability Engineering Workshops. Piscataway, USA: IEEE, 2024: 127-134.
- [19] POURBEHZADI M, JAVIDI G, HOWELL C J, et al. Enhanced (cyber) situational awareness: using interpretable principal component analysis (iPCA) to automate vulnerability severity scoring[J]. Decision Support Systems, 2024, 186: No.114308.
- [20] GHOSH R, FARRI O, VON STOCKHAUSEN H M, et al. CVE-LLM: automatic vulnerability evaluation in medical device industry using large language models[J]. ArXiv e-Prints, 2024: arXiv:2407.14640.
- [21] KHAN T, MALIK M, AZIZ Z, et al. Threat and vulnerability classification using NLP and machine learning techniques on text-based security data[J]. Journal of Computers and Intelligent Systems, 2025, 3(2): 139-152.
- [22] 彭真. 配电网信息物理系统网络安全风险评估技术研究[D]. 北京: 北京交通大学, 2024: 1-81.
- [23] SHAZEER N, MIRHOSEINI A, MAZIARZ K, et al. Outrageously large neural networks: the sparsely-gated mixture-of-experts layer [DB/OL]. (2017-01-23) [2025-11-10]. <https://arxiv.org/abs/1701.06538>.
- [24] Hugging Face. The AI community building the future [DB/OL]. (2023-12-23) [2025-11-10]. <https://huggingface.co>.
- [25] MIKOLOV T, CHEN K, CORRADO G, et al. Efficient estimation of word representations in vector space [DB/OL]. (2013-01-16) [2025-11-10]. <https://arxiv.org/abs/1301.3781>.
- [26] DEVLIN J, CHANG M 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. Stroudsburg, USA: ACL, 2019: 4171-4186.
- [27] CHI P H, CHUNG P H, WU T H, et al. Audio bert: a lite bert for self-supervised learning of audio representation[C]//Proceedings of the 2021 IEEE Spoken Language Technology Workshop. Piscataway, USA: IEEE, 2021: 344-350.

(编辑: 冯 春)