基于软提示微调和强化学习的网络安全命名实体识别方法研究

田泽庶，刘春雨，张云婷，张嘉宇，孟超，张宏莉
（哈尔滨工业大学计算学部，黑龙江 哈尔滨 150001）

摘　要：随着网络技术的迅猛发展，新型网络安全威胁不断涌现，网络安全命名实体识别重要性日益增加。针对现有基于大语言模型的命名实体识别方法在网络安全领域识别准确率差的问题，提出了一种结合软提示微调和强化学习的网络安全命名实体识别方法。通过结合软提示微调技术，针对网络安全领域的复杂性，精细调整大语言模型的识别能力，提升模型对网络安全命名实体的识别准确率，同时优化训练效率。此外，提出了基于强化学习的网络安全实体筛选器，可以有效去除训练集中的低质量标注，从而提升识别准确率。在2个开源基准网络安全实体识别数据集上评估了所提方法，实验结果表明，所提方法的F1值优于现有最佳的网络安全命名实体识别方法。

关键词：网络安全命名实体识别；软提示微调；强化学习；大规模预训练模型
中图分类号：TP389.1
文献标志码：A
DOI：10.11959/j.issn.1000–436x.2024183

Research on named entity recognition method in cybersecurity based on soft prompt tuning and reinforcement learning

TIAN Zeshu, LIU Chunyu, ZHANG Yunting, ZHANG Jiayu, MENG Chao, ZHANG Hongli
Faculty of Computing, Harbin Institute of Technology, Harbin 150001, China

Abstract: As network technology rapidly advanced, new cybersecurity threats constantly emerged, increasing the importance of cybersecurity named entity recognition. To address the problem of poor recognition accuracy in named entity recognition methods based on large language models in the cybersecurity domain, a novel cybersecurity named entity recognition method that combined soft prompt tuning and reinforcement learning was proposed. By integrating the soft prompt tuning technique, the method precisely adjusted the recognition capabilities of large language models to handle the complexity of the cybersecurity domain, improving recognition accuracy for cybersecurity named entities while optimizing training efficiency. Additionally, a reinforcement learning-based instance filter was proposed, which effectively removed low-quality annotations from the training set, further enhancing recognition accuracy. The proposed method was evaluated on two benchmark cybersecurity NER datasets, with experimental results demonstrating superior performance in F1 score compared to state-of-the-art cybersecurity NER methods.

Keywords: cybersecurity named entity recognition, soft prompt tuning, reinforcement learning, large-scale pre-trained models

收稿日期：2024–07–04；**修回日期：**2024–09–29
通信作者：张宏莉，zhanghongli@hit.edu.cn
基金项目：国家重点研发计划基金资助项目（No.2016QY03D0501, No.2017YFB0803304）；黑龙江省自然科学基金资助项目（No.LH2023F018）
Foundation Items: The National Key Research and Development Program of China (No.2016QY03D0501, No.2017YFB0803304), The Natural Science Foundation of Heilongjiang Province (No.LH2023F018)

0 引言

在网络技术迅猛发展的当下，新型网络安全威胁层出不穷，带来了海量的威胁情报数据。这些威胁情报大多通过报告形式呈现，涵盖了安全简报、公共漏洞披露报告[1]和国家漏洞数据库报告等[2]。然而，这些报告往往是非结构化的，来源多样，增加了理解网络攻击及做出风险决策的复杂性。随着深度学习技术的进步，安全分析人员越来越多地应用自然语言处理技术对网络攻击进行提取和深入分析[3]。特别是命名实体识别（NER, named entity recognition）[4]、关系提取[5]以及文本摘要技术[6]，在学术界和实践中都得到了广泛的应用和讨论。本文主要集中于网络安全命名实体识别任务的探索与应用。

命名实体识别旨在识别特定领域中具有特殊含义的名词或名词短语。在传统领域，命名实体识别的目标一般包括人名、地名和组织机构名称等专有名词。然而，在网络安全领域，命名实体识别的目标则是识别和分类与网络安全相关的实体，例如应用程序、供应商、相关术语、操作系统、版本、编程语言和硬件等。

目前，命名实体识别的方法主要基于深度学习技术，这些方法利用循环神经网络和双向长短时记忆（Bi-LSTM, bi-directional long short-term memory）网络等结构，学习文本的潜在上下文向量，从而通过高阶特征实现实体识别。此外，近年来基于预训练模型（如BERT、GPT等）的方法逐渐兴起，这些模型通过大量数据预训练，能够有效捕捉语义信息，提升通用实体识别的能力。与此同时，提示微调方法应运而生，通过将输入与提示词相结合，引导模型的输出，显著减少了计算资源的消耗，并保持了识别效果[7]。尽管当前研究在传统命名实体识别领域取得了显著进展，但这些方法在网络安全命名实体识别领域仍面临许多挑战。

首先，传统方法在网络安全命名实体识别的能力不足。目前结合大规模预训练模型的命名实体识别方法在传统命名实体识别领域表现出了强大的能力[8]，然而，网络安全实体与传统的命名实体有所不同，传统方法在识别应用程序名称、供应商、操作系统、版本号等复杂实体时往往不够准确[9]。例如，在CVE中的句子“libraries/libldap/tls_m. c in OpenLDAP, possibly 2.4.31 and earlier.”中包含了应用（OpenLDAP）、应用号（2.4.31）和文件路径（libraries/libldap/tls_m.c）等复杂实体，这些实体在大规模预训练模型的语料库中并不常见。

其次，现有网络安全命名实体识别的开源数据集中，专家标注常常存在错误，这对实体识别的准确率产生了重要影响。由于网络安全实体识别的专业性较强，实体粒度更为细化，涵盖网络攻击实体、攻击手段、攻击途径和攻击模式等多种类型，因此，开源网络安全命名实体识别数据集的标注准确度良莠不齐，影响了基于大规模预训练模型的提示微调的准确性。

为了解决上述挑战，本文提出了一种新的方法，即“基于软提示微调和强化学习的网络安全命名实体识别方法”（CNER-SPT-RL, cybersecurity named entity recognition method based on soft prompt tuning and reinforcement learning）。该方法包含2个部分，即基于软提示微调的网络安全命名实体识别方法和基于强化学习的网络安全实体筛选器。为了解决传统提示学习在网络安全领域中因缺乏专业领域知识而导致的效果不佳问题，本文首次提出基于大语言模型的软提示微调命名实体识别方法，设计了一种结合单词表和提示生成器的软提示微调策略。此外，针对现有开源网络安全实体识别数据集中存在的错误标注，本文设计了一个基于强化学习的网络安全实体筛选器，训练智能体对数据集中的样本进行筛选，以去除低质量的标注，从而提升模型的效果。本文的贡献如下。

1) 本文提出了通过微调少量软提示参数以适应网络安全实体识别任务的策略。该策略利用软提示微调技术，显著减少了计算资源的消耗，同时保持了识别准确率。

2) 本文首次引入智能体筛选网络安全实体标签，并利用强化学习技术优化智能体的决策过程，使其有效地识别和剔除低质量标注数据，从而提升网络安全命名实体识别的准确率。

3) 本文提出了一种结合软提示微调和强化学习的网络安全命名实体识别方法，整合了两者的优势，显著提升了实体识别的性能。本文在2个基准网络安全命名实体识别数据集进行了实验，

实验结果显示，相比现有最优的网络安全命名实体识别方法，本文方法在F1值上分别提升了2.42%和4.07%。

1　相关工作

本节将探讨网络安全领域的命名实体识别方法，将这些方法分为基于传统机器学习的方法和基于深度学习的方法。

1.1　基于传统机器学习的网络安全命名实体识别

在网络安全命名实体识别的早期研究中，基于规则的方法依赖于领域专家手工编写的规则，这些规则通常结合了词表和句法-词汇模式[10]。此外，基于统计的方法应用了多种机器学习算法，包括隐马尔可夫模型[11]、支持向量机[12]、感知机[13]和条件随机场（CRF, conditional random field）[14]。例如，Mulwad等[15]通过从维基百科中提取的知识库中抽取漏洞和攻击概念，生成了机器可理解的断言。Lal[16]利用基于Stanford NER的CRF模型训练，将命名实体识别视为序列标注任务。Weerawardhana等[17]则提出了基于机器学习和词性标注策略的方法来从在线漏洞数据库提取情报。

1.2　基于深度学习的网络安全命名实体识别

随着深度学习技术的快速发展，基于传统神经网络的方法在NER任务中显示出了显著的优势。Collobert等[18]提出的统一神经网络架构大大减少了对手工特征的依赖，并简化了特征学习过程。Huang等[19]通过将Bi-LSTM和CRF结合起来，首次为序列标注任务提供了一种有效的深度学习解决方案，从而使基于循环神经网络的模型在NER任务中取得了主导地位。Kim等[20]利用深度Bi-LSTM-CRF网络自动从网络威胁情报报告中提取关键信息。Qin等[21]通过卷积神经网络（CNN, convolutional neural networks）提取字符特征，并结合Bi-LSTM网络来学习全局词特征表示。

随着预训练模型技术的大规模应用，基于预训练模型（如BERT）的方法逐渐涌现。Simran等[22]在特征生成层中应用了Bi-GRU和CNN的线性堆叠，显著提高了性能。Zhou等[23]将BERT-Bi-LSTM-CRF应用于网络安全命名实体识别任务，其中词嵌入由改进的BERT表示。Gao等[24]设计了一个数据和知识驱动的命名实体识别网络，通过与网络安全相关的外部字典生成词嵌入向量，增强了模型对网络安全模式的理解。此外，Wang等[25]提出了一个结合扰动掩码语言模型和门控注意力神经网络循环单元的特征整合和实体边界检测模型，以增强模型的表示能力。同时，Srivastava等[9]则探讨了词嵌入对网络安全命名实体识别的影响，他们的研究表明，微调的BERT词嵌入在这一任务中表现出色。

最近，大规模预训练模型（如GPT、GLM等模型）的普遍使用使得基于预训练模型的方法迎来了又一次革新。大规模预训练模型相较于传统预训练模型拥有更大的参数量，使得传统命名实体识别领域效果得到显著提升[26]。然而，在网络安全命名实体识别领域，基于大规模预训练模型和提示微调的方法的研究目前鲜有研究，这归因于网络安全实体的类型多样性和网络安全实体边界的复杂性。并且，大语言模型领域目前集中于通用领域，面对复杂的网络安全实体时表现不佳。因此，本文尝试从软提示微调方法的角度，既利用了大语言模型丰富的基础知识，同时又能学习网络安全实体的独有特征。并且提出了基于强化学习的网络安全实体筛选器，以解决训练数据集中低质量标注带来的困扰。

2　概念及问题定义

2.1　网络安全命名实体识别问题定义

在网络安全领域，命名实体识别问题可以形式化定义如下：给定网络安全领域中含有n个单词的句子$x=\{w_1,w_2,\cdots,w_n\}$，命名实体识别任务的目标是为每个单词w_i分配一个标记y_i，使得y_i表示w_i在句子中的实体类型或其是否属于一个实体。

本文采用BIOS（begin, inside, outside, single）标注方案，该方案包括以下标记：实体的开始单词B（begin）、实体的内部单词I（inside）、独立实体S（single），即仅由一个单词组成的实体和非实体的单词O（outside）。

2.2　实体类型

网络安全命名实体识别的目的是识别与网络安全相关的实体术语。这些实体术语通常包括但不限于：攻击类型（如分布式拒绝服务攻击、结构化查询语言注入攻击）、恶意软件名称（如annaCry、Stuxnet）、漏洞编号（如CVE-2021-34527）、安全

工具（如Wireshark、Metasploit）、技术术语（如缓冲区溢出、后门）。通过识别这些实体术语，可以在文本中提取有价值的信息，以辅助网络安全分析和事件响应。

2.3 软提示微调

为了提高网络安全命名实体识别的准确性和效率，本文提出了基于软提示微调的方法。软提示微调是一种参数高效的微调策略，可以在不大幅修改预训练模型参数的前提下，适应特定领域的任务需求。

软提示微调的定义如下：给定具有参数Θ的预训练语言模型和训练数据$(X,Y)=\{x_k,y_k\}_{k=1}^{N}$，常规策略是通过最大化条件概率$P(\boldsymbol{Y}|\boldsymbol{X};\Theta)$来直接微调所有参数，其中$\boldsymbol{X}$和$\boldsymbol{Y}$是$X$和$Y$的表示。然而，随着参数效率变得至关重要，软提示微调作为一种更加通用的解决方案浮现。软提示微调不是微调所有参数Θ，而是利用一组可学习的向量$\boldsymbol{P}\in\mathbf{R}^{l\times d}$与模型的输入嵌入$\boldsymbol{x}_k=[\boldsymbol{w}_1,\boldsymbol{w}_2,\cdots,\boldsymbol{w}_m]\in\mathbf{R}^{m\times d}$进行拼接，其中$\boldsymbol{w}_m$是$\boldsymbol{x}_k$的第$m$个词的嵌入表示，$l$是代表软提示长度的超参数，$d$是标签嵌入表示的维度。目标是最小化以下损失

$$\mathcal{L}_{\mathrm{PLM}}=-\sum_{k=1}^{N}\log P(\boldsymbol{y}_k|[\boldsymbol{P};\boldsymbol{x}_k];\Theta) \tag{1}$$

其中，语言模型的输入矩阵$[\boldsymbol{P};\boldsymbol{x}_k]\in\mathbf{R}^{(l+m)\times d}$是提示矩阵$\boldsymbol{P}$和输入表示$\boldsymbol{x}_k$的拼接。

2.4 基础模型

本文使用了ChatGLM-6B-V3[26]作为预训练模型。ChatGLM-6B-V3是一个开源的，支持中英双语的对话语言模型，属于自回归架构，具有62亿参数。ChatGLM作为一个自回归模型，继承了Transformer解码器的结构。它由28个相同的构件组成，每个构件称为GLM块。每个GLM块包括以下几个关键部分：输入均方根正规化层、多查询注意力（MQA, multi query attention）模块、注意力机制后的均方根正规化层和前馈神经网络（FFN, feed forward neural network）层。这些层和模块构成了ChatGLM的基础结构，使其能够处理复杂的自然语言理解和生成任务。

3 方法

本节中将介绍本文提出的网络安全命名实体识别方法，包括2个部分：基于软提示微调的网络安全实体识别方法和基于强化学习的网络安全实体筛选器。网络安全命名实体识别方法的流程分为3个主要步骤，如图1所示。首先，训练基于软提示微调的网络安全实体识别模型（步骤1）。其次，结合步骤1中的模型使用基于强化学习的网络安全实体筛选器以消除低质量的标注数据（步骤2）。最后，选择表现最佳的轮次的筛选后数据集重新训练步骤1中的基于软提示微调的网络安全实体识别模型，并用于预测和评估（步骤3）。

3.1 基于软提示微调的网络安全实体识别方法

本文提出的基于软提示微调的网络安全实体识别方法的架构如图2所示，从输入输出分为4个模块：输入层、GLM块、提示生成器、序号生成器。本节将按照从输入输出的顺序依次介绍这几个模块。

1) 输入层

考虑输入为第k条训练文本$\boldsymbol{x}_k$，如以图2中右上方的文本“The eval_js function in uzbl-core.c”为例。输入经过分词器得到对应ID，然后经过嵌入层，得到了对应位置嵌入和词嵌入，最终得到整体的文本嵌入$\boldsymbol{x}_k$。

2) GLM块

嵌入表示被送入多个GLM块组成的处理序列。在这里，以第l层GLM块为例进行说明。首先，将$\boldsymbol{x}_k$输入均方根正规化层和一个线性层，输出一个特征维度为d_{mixed}的向量$\tilde{\boldsymbol{x}}_k$，其中d_{mixed}是查询向量$\boldsymbol{Q}^l$、键向量$\boldsymbol{K}^l$以及值向量$\boldsymbol{V}^l$的特征维度之和。

$\tilde{\boldsymbol{x}}_k$之后通过3个不同的线性层分别映射为MQA模块所需的$\boldsymbol{Q}^l$、$\boldsymbol{K}^l$和$\boldsymbol{V}^l$。在MQA模块中，$\boldsymbol{K}^l$和$\boldsymbol{V}^l$的维度是一致的。最终，注意力得分通过式(2)、式(3)计算。

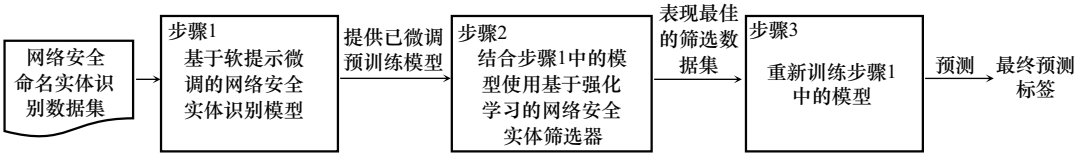


图1 网络安全命名实体识别方法的流程

$$\boldsymbol{Q}^l = \tilde{\boldsymbol{x}}_k \boldsymbol{W}_Q, \boldsymbol{K}^l = \tilde{\boldsymbol{x}}_k \boldsymbol{W}_K, \boldsymbol{V}^l = \tilde{\boldsymbol{x}}_k \boldsymbol{W}_V \tag{2}$$

$$\text{Attention}^l = \text{softmax}\left(\frac{\boldsymbol{Q}^l[\boldsymbol{K}^l]^{\mathrm{T}}}{\sqrt{d_k}}\right)\boldsymbol{V}^l \tag{3}$$

其中，$\boldsymbol{W}_Q$、$\boldsymbol{W}_K$、$\boldsymbol{W}_V$是可学习的权重矩阵，用于将输入向量$\tilde{\boldsymbol{x}}_k$转换成查询、键和值向量，d_k是键向量的维度。紧接着，通过MQA模块计算得到的注意力得分用于引导模型聚焦于输入文本的关键信息，从而增强模型的理解和预测能力。

MQA模块的输出经过注意力机制后的均方根正规化层和FFN层之后，得到输出特征$\boldsymbol{h}_k = \{\boldsymbol{h}_1, \boldsymbol{h}_2, \cdots, \boldsymbol{h}_t\}$。

3) 提示生成器

为了降低微调阶段的参数数量，本文引入了提示生成器。该生成器为原始的$\boldsymbol{Q}$、$\boldsymbol{K}$和$\boldsymbol{V}$层之上添加了可学习的软提示，以实现灵活且有效的知识转移。

提示生成器包含2个线性层。提示生成器的输入是一个固定长度为10的向量。首先，通过第一个线性层将其映射到一个较大的特征空间，然后通过第2个线性层进一步映射到维度为$2 \times 28 \times d_v^l$的空间，对应28层GLM块，d_v^l是键/值向量的特征维度。生成的向量包括用于拼接到$\boldsymbol{K}^l$和$\boldsymbol{V}^l$的向量$\boldsymbol{P}_k^l$和$\boldsymbol{P}_v^l$。

在此过程中，2个线性层是可学习的，而GLM块中的参数保持冻结。通过这种方式，与微调整个模型相比，提示生成器中的2个线性层的参数量减少，从而有效降低了微调所需计算资源。

软提示的拼接进一步优化了模型对于特定网络安全实体的识别，使得模型在进行实体识别时能够更好地利用预训练知识和任务特定的提示信息。这一过程的数学表示为

$$\text{Attention}^l = \text{softmax}\left(\frac{\boldsymbol{Q}^l[\boldsymbol{P}_k^l; \boldsymbol{K}^l]^{\mathrm{T}}}{\sqrt{d_k}}\right)[\boldsymbol{P}_v^l; \boldsymbol{V}^l] \tag{4}$$

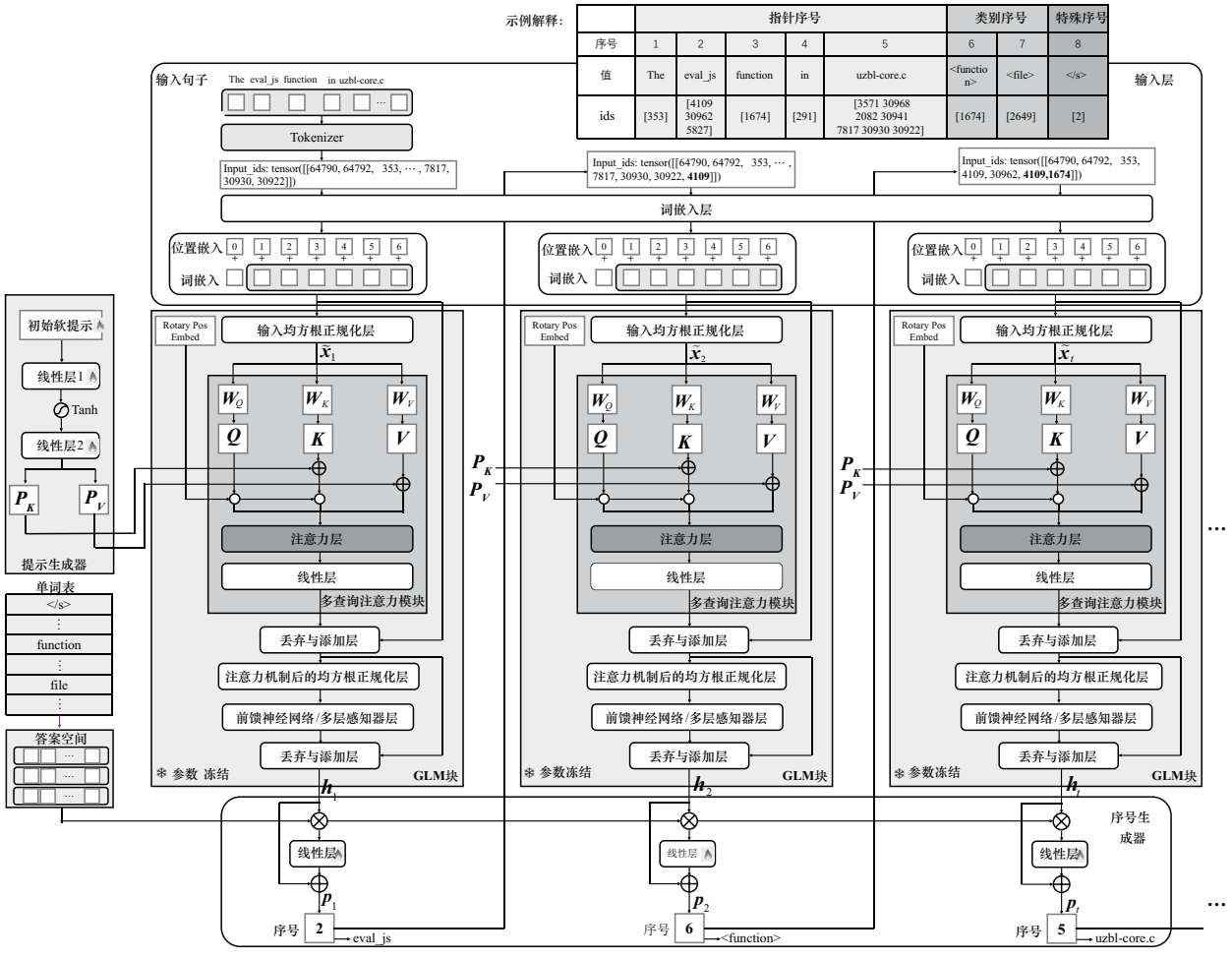


图2　基于软提示微调的网络安全实体识别方法的架构

其中，$\boldsymbol{P}_k^l$和$\boldsymbol{P}_v^l$是从提示生成器产生的，用于增强键向量$\boldsymbol{K}^l$和值向量$\boldsymbol{V}^l$的软提示向量，通过调整$\boldsymbol{P}_k^l$和$\boldsymbol{P}_v^l$的权重，模型能够更加精细地调节注意力机制的焦点，式(4)被用于替换式(3)。

4) 序号生成器

为了解决生成式语言模型在网络安全实体识别中的不准确问题，本文设计了一种新的输出特征映射方法，旨在提高实体识别的全面性和准确性，命名为序号生成器。

传统生成式语言模型在直接回答实体时，可能会出现识别不全面或跨度不准确的问题。例如，在示例“The eval_js function in uzbl-core. c”中，“uzbl-core.c”应被识别为一个文件实体，但语言模型可能会被误识别为“uzbl-core”实体。为了解决这一问题，本文设计了一种输出特征映射机制，将GLM块的输出向量直接映射为标记序号和类别序号。

具体来说，将句子中的每个标记的序号和实体类别序号进行映射。例如，在上述句子中，模型可以将预测输出表示为“26578”，其中2和5分别代表句子中的第2和第5个token，表征实体跨度，是标记序号，6和7代表实体类别，对应<function>和<file>是类别序号，8代表句子的结束符号</s>，是特殊序号。这种设计方式显著提高了实体跨度的准确度。

上述的过程可以形式化地表示如下：本文设计了一个分类器将输出特征$\boldsymbol{h}_t$映射为概率分布$\boldsymbol{p}_t \in \mathbf{R}^m$，其中$m$包含句子中的词数量和实体类别数量以及特殊符号的数量（如</s>），最后$\boldsymbol{p}_t$会映射为输出y_t，y_t表示标签，其可以通过一个字符到字符的映射函数，映射为命名实体标注，映射函数的具体形式化描述为

$$y_t = \begin{cases} X_{y_t}, y_t\text{是标记序号} \\ C_{y_t-n}, y_t\text{是类别序号} \\ S_{y_t-n-m}, y_t\text{是特殊序号} \end{cases} \tag{5}$$

其中，X是输入文本，C是网络安全实体类别的集合（如“Hardware”“File”…），S是ChatGLM中特殊符号的集合，$n=|X|$，$m=|C|$。

最终，基于软提示微调的网络安全实体识别方法的整体流程如算法1所示。

算法1 基于软提示微调的网络安全实体识别方法

输入 训练数据集$D=\{x_1,x_2,\cdots,x_k\}$，初始化的ChatGLM、提示生成器、提示向量

输出 实体识别结果$\{y_1,y_2,\cdots,y_k\}$

1) for 第k条输入文本$x_k \in D$ do
2) 将输入句子嵌入为表示$\boldsymbol{x}_k$
3) for 第l个GLM块 do:
4) 将表示$\boldsymbol{x}_k$输入均方根正则化层，$\tilde{\boldsymbol{x}}_k=$ Input_RMSNorm$(\boldsymbol{x}_k)$。
5) 根据式(2)将$\tilde{\boldsymbol{x}}_k$映射为查询、键和值向量$\boldsymbol{Q}^l,\boldsymbol{K}^l,\boldsymbol{V}^l$。
6) 由提示生成器生成软提示$\boldsymbol{P}_k^l$和$\boldsymbol{P}_v^l$。
7) 将软提示$\boldsymbol{P}_k^l$和$\boldsymbol{P}_v^l$注入MQA层，并根据式(4)计算注意力得分Attention^l。
8) 将注意力得分Attention^{l}输入注意力机制后的均方根正规化层和FFN层中得到第l个GLM块的输出表示$\boldsymbol{h}_k=\{\boldsymbol{h}_1,\boldsymbol{h}_2,\cdots,\boldsymbol{h}_t\}$。
9) end for
10) 依次将输出表示$\boldsymbol{h}_t$输入序号生成器映射为概率$\boldsymbol{p}_t$，并通过交叉熵计算损失，并参与反向传播。
11) 将输出表示$\boldsymbol{h}_k$输入序号生成器中，生成对应序号$y_k=\{y_1,y_2,\cdots,y_t\}$。
12) 将生成序号序列转换为命名实体。
13) end for

3.2 基于强化学习的网络安全实体筛选器

通过提示微调识别出的网络安全实体有时会出现实体跨度不完整等问题，而软提示微调的效果与微调所使用的数据集紧密相关。目前可供使用的开源网络安全实体识别数据集都存在着标注质量不佳的情况。例如，一些版本号和平台版本等字母数字的组合被错误地标注，这会显著影响识别结果，而这些问题在需要精准确定漏洞版本和平台等网络安全场景中是不可容忍的。因此，本文首次提出使用强化学习训练智能体，对数据集的样本标注进行筛选，其目的是不断提升数据集的质量，并找到有利于软提示微调方法的数据，使模型更好地捕获网络安全实体特征。本文使用马尔可夫决策过程来模拟环境，这在强化学习中被广泛应用。通过将之前的状态信息纳入当前状态，本文使用策略梯度算法训练了一个基于强化学习的网络安全实体筛选器。

基于强化学习的网络安全实体筛选器示意如图3所示，首先，第i−1轮待筛选的低质量标注数据集，经过训练好的预训练模型的处理，生成文本特征表示。随后，这些特征被输入包含2个线性层的策略网络中，计算出动作得分。动作得分从高到低排序，去掉一定比例的低动作得分，筛选保留下来的对应实例，形成第i轮对应的筛选数据集。最终，使用该数据集重新训练预训练模型，并评估F1值，以此与上一个周期的F1值进行比较，形成奖励。实体筛选器的各个元素将在下文中详细介绍。

1) 状态

在强化学习中，状态表示关键，它能够捕捉当前环境的全部信息，指导智能体进行决策。状态s_k需要包含文本的所有关键信息以及上一轮筛选过程中得到的反馈。因此，本文设计了一个包含文本嵌入和历史反馈的向量$\boldsymbol{S}_k$，以更全面地描述当前实例的状态。具体来说，向量$\boldsymbol{S}_k$包含2个部分：一是当前文本通过基于提示微调的ChatGLM中GLM块的输出$\boldsymbol{h}_k$，这一部分能够捕捉文本和标注的深层语义信息；二是上一个周期中被移除实例的表示的平均值$\bar{\boldsymbol{x}}_{\Omega_{i-1}}$，这一部分反映了先前筛选过程中的决策结果。通过结合这2个部分信息，状态表示能够更好地描述当前文本的特征和历史反馈，为智能体提供更全面的决策依据。

2) 奖励

假设智能体已筛选了低质量的命名实体标注，需要一个评价指标来衡量智能体的表现。如果智能体表现良好，它将获得奖励；否则，它将受到惩罚。通过在筛选后的数据集上对初始化的ChatGLM进行提示微调来进行评估。为了评估模型，本文使用验证数据集上表现最佳的周期的F1值作为评价指标。然后，根据智能体训练过程中2个相邻周期的F1值的差值来评估数据集质量，并确定第i个周期的奖励。奖励计算为

$$r_i = F_1^i - F_1^{i-1} \tag{6}$$

其中，F_1^i和F_1^{i-1}是第i和第i−1轮提示微调模型在验证集上的F1值。

3) 梯度策略网络

策略网络$\pi\left(\boldsymbol{S}_k;\theta_{\text{selector}}\right)$由2个线性层组成。第1个线性层将状态表示$\boldsymbol{S}_k$映射到与文本表示$\boldsymbol{x}_k$相同的维度。然后，映射后的表示和文本表示被拼接以获得最终的状态表示$\tilde{\boldsymbol{S}}_k=(\boldsymbol{S}_k;\boldsymbol{x}_k)$。状态表示输入第2个线性层。通过第2个线性层和Sigmoid函数后，得到动作得分ap_k。ap_k介于0（低质量）和1（高质量）之间，代表标注序列的质量。

在强化学习的早期阶段，实体筛选器获得的动作点波动剧烈。如果设置动作点阈值来过滤数据，实体筛选器可能不会收敛。因此，本文对实例的动作得分进行排序，将处于最低p百分比动作得分的实例视为低质量。p值是模型预测的数据集中的低质量标注比例。如果没有先验知识，可以使用基于模型性能的自适应调整策略，在强化学习的前若干个轮次中动态地调整p值，根据奖励动态地增加或者减少p值，如果模型在验证集上的性能提高则增加p值，反之则减少，然后在训练的后面阶段固定p值为定值p_q。训练第i轮次的p值的更新方法为

$$p_i=\begin{cases} p_{i-1}+r_i\dfrac{q-i}{q}, & i<q \\ p_q, & i\geqslant q \end{cases} \tag{7}$$

其中，q是p停止动态更新的轮数，r_i是第i轮奖励。

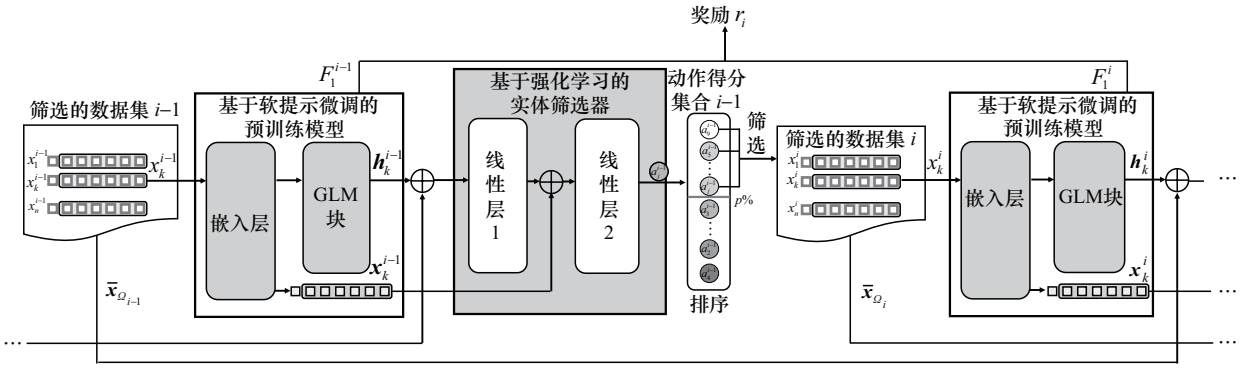


图3　基于强化学习的网络安全实体筛选器示意

根据策略梯度方法，策略网络的参数θ将进行如下更新。

$$\theta = \theta + \mu [\nabla_\theta \sum_{k \in \Omega_i} \log\pi\left(\mathrm{ap}_k|\tilde{\boldsymbol{S}}_k; \theta\right) r_i + \nabla_\theta \sum_{k \in \Omega_{i-1}} \log\pi\left(\mathrm{ap}_k|\tilde{\boldsymbol{S}}_k; \theta\right)(-r_i)] \tag{8}$$

其中，μ是学习率，$\nabla_\theta(\cdot)$表示目标函数对参数θ求梯度，Ω_i和Ω_{i-1}的计算式为

$$\Omega_i = \tilde{\Psi}_i - \left(\tilde{\Psi}_i \cap \tilde{\Psi}_{i-1}\right) \tag{9}$$

$$\Omega_{i-1} = \tilde{\Psi}_{i-1} - \left(\tilde{\Psi}_i \cap \tilde{\Psi}_{i-1}\right) \tag{10}$$

其中，$\tilde{\Psi}_i$是第i轮中移除的数据。如果第i轮的F1值增加，则只在第i轮中被移除的实例（Ω_i）将被奖励。类似地，只在第i–1轮中被移除的实例（Ω_{i-1}）将被惩罚。2轮移除数据中共享的实例不参与奖励和惩罚。通过这种方法，可以有效地识别并移除质量较低的标注数据，从而提升网络安全命名实体识别模型的准确性。

4　实验分析

4.1　数据集

由于网络安全命名实体识别任务中没有统一的数据集，本文选择了来自2个渠道的实验数据集：第一个是Bridges等[27]提供的开源数据集，该数据集从包括微软安全公告、Metasploit和美国国家计算机通用漏洞数据库在内的多个网络安全领域平台收集而来。这个数据集中包含多种实体，如应用程序、供应商、操作系统、相关术语等。第2个数据集是Alam等[28]提供的网络安全命名实体识别语料库。本文将实验数据集分为3部分：训练集（70%）、开发集（15%）和测试集（15%）。所有的语料数据都存储在JSON格式中。本文不直接使用JSON文件；相反，对数据进行预处理，并将其转换为CoNLL 2003格式。数据集的统计信息及每个实体的详细信息分别如表1~表3所示。

表1　实验数据集详情

数据集		句子/个	字符/个	实体/个
Bridges等[27]提供的数据集	训练集	11 142	584 908	110 432
	开发集	2 353	124 121	23 575
	测试集	2 283	120 716	23 577
Alam等[28]提供的数据集	训练集	2 811	68 191	2 893
	开发集	813	19 530	745
	测试集	748	19 270	892

表2　Bridges等[27]提供的数据集中每个实体的统计数据

实体类别	训练集/个	开发集/个	测试集/个
Application	14 606	3 003	2 928
Vendor	8 025	1 753	1 657
Version	21 071	4 318	4 517
Relevant term	53 899	11 753	11 504
Function	1 009	242	217
Os	2 913	647	692
Update	2 953	644	663
Edition	430	76	90
Hardware	429	71	87
File	2 208	507	507
Parameter	428	110	119
Programming	117	26	25
Language	5	1	1
Cve	2 207	398	553
Method	131	26	18

表3　Alam等[28]提供的数据集中每个实体的统计数据

实体类别	训练集/个	开发集/个	测试集/个
Malware	703	254	242
System	837	182	248
Organization	284	92	131
Indicator	1 021	208	261
Vulnerability	48	9	10

4.2　实验设置

在评估结果时，本文采用精确度（Precision）、召回率（Recall）和F1值来衡量模型在网络安全命名实体识别中的性能，具体而言，使用微F1值（micro-F1），其中部分匹配的实体没有部分得分。评价指标为

$$\mathrm{Precision} = \frac{|A|}{|T_p|}, \mathrm{Recall} = \frac{|A|}{|T_g|},$$
$$\text{F1值} = \frac{2 \times \mathrm{Precision} \times \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}} \tag{12}$$

其中，T_p是预测结果的集合，T_g是真实结果的集合，$A = T_p \cap T_g$是命中结果的集合，$|\cdot|$是表示集合中元素数量的函数。

本文采用显存为24 GB的RTX-3090 GPU来训练模型。所有实验都在一台配备12核CPU和128 GB内存的单GPU服务器上运行。对于基于软提示微

调的网络安全实体识别方法，本文使用了由Zeng等[26]提供的ChatGLM-6b-v3，使用AdamW优化器，学习率设置为1×10^{-3}来更新模型参数。以1×10^{-5}的学习率训练变换器参数，并将权重衰减设置为0.01。最大迭代轮数设置为25，耐心值设置为5。最佳轮数将根据评估结果选出。批处理大小设置为64，采用时间步骤丢弃技术对词表示进行处理，以0.1的概率避免过拟合。对于基于强化学习的网络安全实体筛选器，使用学习率为1×10^{-2}的Adam优化器来更新模型参数。设置p值初始值为0.1，到第10轮之后不再调整p值。在5个不同的随机种子上进行实验，结果取平均以增强鲁棒性。

4.3　基准模型

本节将CNER-SPT-RL与几个网络安全命名实体识别基线模型进行比较，以下是详细介绍。

1) CRF模型：Abdullah等[29]提出的这一模型是基于统计的条件概率分布模型，广泛应用于序列标记任务。

2) Bi-LSTM-CRF架构：Zhou等[23]使用此架构来执行网络安全的命名实体识别任务。具体过程中，首先通过Bi-LSTM层从输入嵌入中提取上下文特征，随后通过CRF层对序列进行解码，生成最终的标签。

3) BiLSTM-CNN模型：Wu等[30]提出的此模型在深度神经网络层中使用LSTM和CNN线性堆叠，以更有效地捕获全局和局部特征。

4) 依赖指导的LSTM-CRF架构：Jie等[31]使用这种简单而有效的架构对依赖树进行编码，将依赖树中单词间的长距离依赖关系转换为固定向量，并作为词嵌入的补充表示，随后输入Bi-LSTM层中。

5) 数据和知识驱动的NER模型：Gao等[24]为网络安全领域提出了这一模型，使用外部字典作为输入层的辅助知识数据库，以提升文本表示，并引入自注意力机制学习句子中单词内部依赖关系。

6) 特征整合与实体边界检测（FIEBD）模型：Wang等[25]开发的模型结合了新的预训练语言模型PERT和创新的GARU神经网络单元，显著提高网络安全文本中复杂实体的识别精度。

7) ChatGPT4模型：由OpenAI公司开发的大型语言模型，本文使用2024年9月的版本。

4.4　实验结果分析

2个网络安全数据集上不同模型的实验结果如表4所示。从表4可以看出，模型取得了最佳性能。例如，当在Bridges等[27]提供的数据集上测试本文方法，其精确度、召回率和F1值分别为96.35%、96.95%和96.65%，均优于所有现有的基线模型。其中一个原因是本文的提示微调方法使用了序号生成机制，使模型在面对复杂的多类型实体时，分类更为准确。此外，本文设计的序号映射机制能够将实体跨度映射为位置序号，将序列标注问题转换成分类问题，从而解决了网络安全实体边界检测难的问题。此外，本文的强化学习模块能筛除实体识别错误的低质量标注。因此，与所有基线模型相比，本文方法容易获得更高的精确度、召回率和F1值。对于Alam等[28]提供的数据集，在精确度、召回率和F1值方面也表现出色，分别达到79.14%、76.00%和77.54%，它们同样表现出色。从综合角度看，本文方法在上述2个数据集中的表现均优于所有基线模型。

在2个数据集上的每个实体类别的表现如表5和表6所示，从结果来看，本文方法在几个训练样

表4　2个网络安全数据集上不同模型的实验结果

模型	Bridges等[27]提供的数据集			Alam等[28]提供的数据集		
	精确度	召回率	F1值	精确度	召回率	F1值
CRF	87.54%	81.96%	84.66%	66.34%	63.96%	65.13%
Bi-LSTM-CRF	90.95%	91.32%	91.13%	72.25%	69.54%	70.89%
BiLSTM-CNN	91.95%	92.57%	92.26%	74.27%	71.56%	72.91%
LSTM-CRF	92.98%	92.77%	92.87%	73.89%	71.19%	72.54%
NER	93.79%	93.52%	93.65%	74.33%	71.86%	73.06%
FIEBD	94.24%	94.51%	94.37%	75.42%	73.61%	74.51%
ChatGPT4	91.05%	93.05%	92.04%	73.11%	69.31%	71.16%
CNER-SPT-RL	**96.35%**	**96.95%**	**96.65%**	**79.14%**	**76.00%**	**77.54%**

本量多的类上都有很好表现，如Bridges等[27]提供的数据集中Relevant term、Version和Application等类别和Alam等[28]提供的数据集中Indicator和Malware等类别，这说明本文方法在样本量充足的情况下有着稳定且准确的表现。

表5 Bridges等[27]提供的数据集中每个实体类别的实验结果

实体	精确度	召回率	F1值
Application	88.11%	90.78%	89.42%
Vendor	94.17%	95.78%	94.97%
Version	97.78%	98.03%	97.91%
Relevant term	99.44%	99.02%	99.23%
Function	94.88%	99.59%	97.18%
Os	92.94%	93.66%	93.30%
Update	92.99%	88.51%	90.69%
Edition	77.01%	88.16%	82.21%
Hardware	48.28%	59.15%	53.16%
File	94.02%	99.21%	96.55%
Parameter	85.16%	99.09%	91.60%
Programming	96.30%	100.00%	98.11%
Language	100.00%	100.00%	100.00%
Cve	99.75%	98.99%	99.37%
Method	74.29%	100.00%	85.25%

表6 Alam等[28]提供的数据集中每个实体类别的实验结果

实体	精确度	召回率	F1值
Malware	86.28%	80.08%	83.06%
System	74.53%	71.98%	73.23%
Organization	76.12%	68.43%	72.07%
Indicator	78.57%	79.33%	78.95%
Vulnerability	80.00%	88.89%	84.21%

4.5 基于强化学习的网络安全实体筛选器效果分析

1) 长时间测试实验

为了验证本文方法在实际应用中的稳定性，本文采用Bridges等[27]和Alam等[28]提供的2个数据集进行长时间的强化学习。实验设置为持续训练500轮，每25轮为一个阶段。在每个阶段结束时，记录提示微调模型在筛选后的数据集上的F1值，以便观察基于强化学习的网络安全实体筛选器筛选性能在长时间训练的稳定性。

基于强化学习的网络安全实体筛选器的长时间训练实验结果如图4所示，结果显示在初始阶段（前75轮），模型的F1值出现了较大的波动，表明模型对数据筛选的判断尚不稳定。然而，随着训练的深入，F1值逐渐上升，显示出筛选器的筛选能力逐步增强。在长期阶段，F1值达到一个相对稳定的高水平，验证了强化学习方法在长期训练中的有效性。

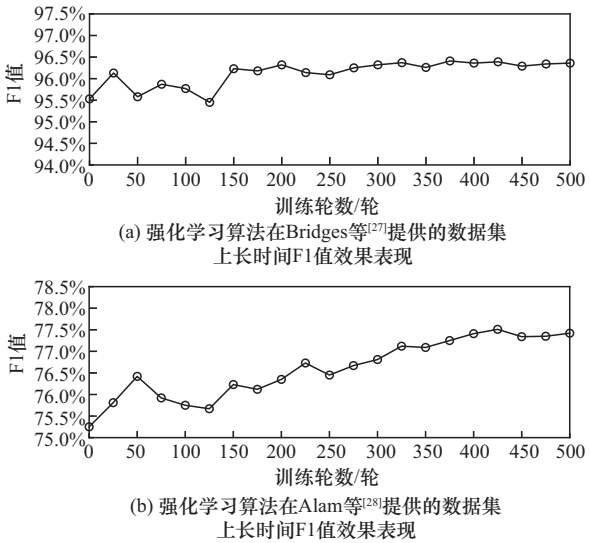


图4 基于强化学习的网络安全实体筛选器的长时间训练实验结果

通过分析实验结果，模型在各个阶段的F1值变化趋势表明，强化学习策略在实例筛选和模型训练中显著提高了数据质量和识别性能。整体而言，模型在长时间测试中保持了良好的稳定性，证实了本文方法在网络安全命名实体识别任务中的持续表现和实际应用潜力。

2) 动作得分分析

在本次实验中，本文对Bridges等[27]和Alam等[28]提供的数据集进行了数据筛选效果的分析。本文采用了基于强化学习的网络安全实体筛选器。为了直观展示筛选过程中实例的变化情况，将筛选中保留的点用浅色表示，去除的点用深色表示。实验设置的主要目的是观察动作得分在不同训练轮数下的分布变化，并通过堆积柱形图展示不同区间动作得分的实例数量比例。

图5和图6展示了在Bridges等[27]和Alam等[28]提供的数据集中，不同训练轮数下数据筛选的效果。随着训练轮数的增加，动作得分的分布逐渐出

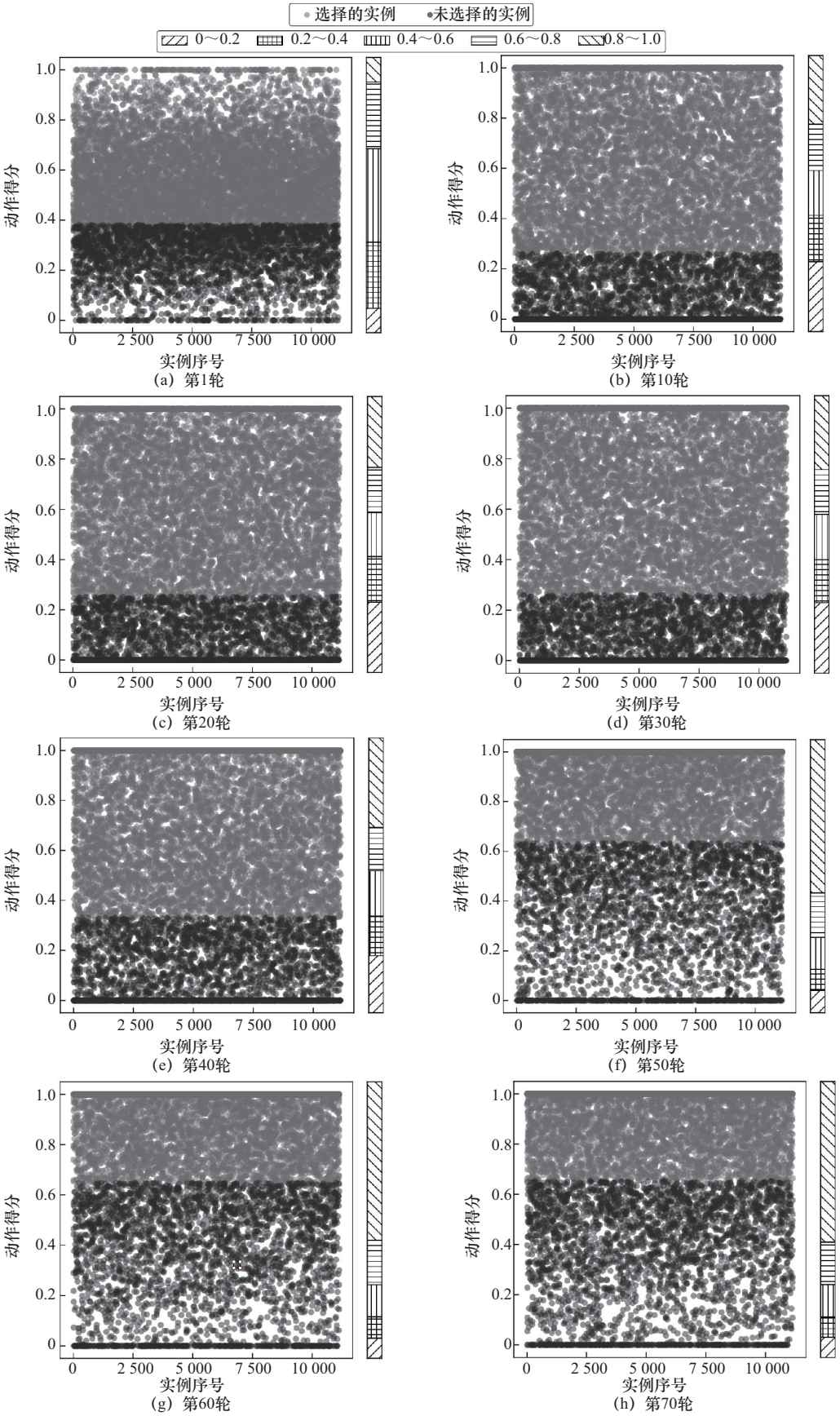


图5　Bridges等[27]提供的数据集中不同训练轮数下的动作得分分布与实例比例分布

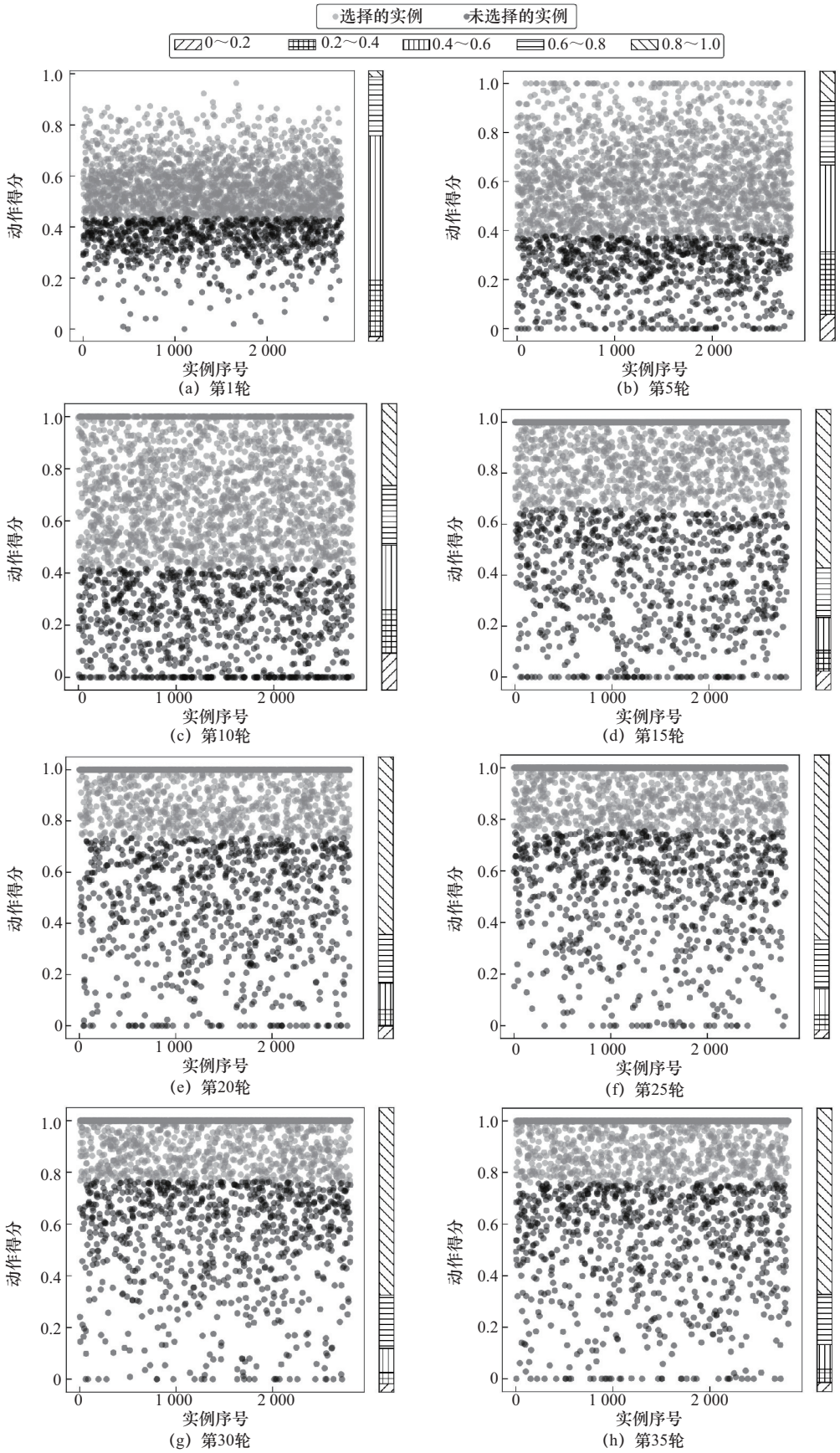


图6　Alam等[28]提供的数据集中不同训练轮数下的动作得分分布与实例比例分布

现明显的分化，筛选出的低质量数据逐渐集中在较低的动作得分范围内，而高质量数据则集中在较高的动作得分范围内。在初始阶段，模型对数据的判断不够稳定，导致部分高质量标注可能被误筛选，表现为动作得分分布的较大波动。然而，随着训练的进行，模型逐渐学习并稳定下来，低质量标注的筛选效果显著提高，数据质量得到了明显改善。堆积柱形图进一步展示了不同区间动作得分的实例数量比例。在实验的早期阶段，不同区间的实例数量分布较为均匀。而在后期，动作得分分布出现两极分化，低质量数据集中在0～0.2，高质量数据集中在0.8～1.0。实验结果证明，采用强化学习策略进行数据筛选是有效的，并且能够显著改善数据质量。

3) 筛选比例p与F1值分析

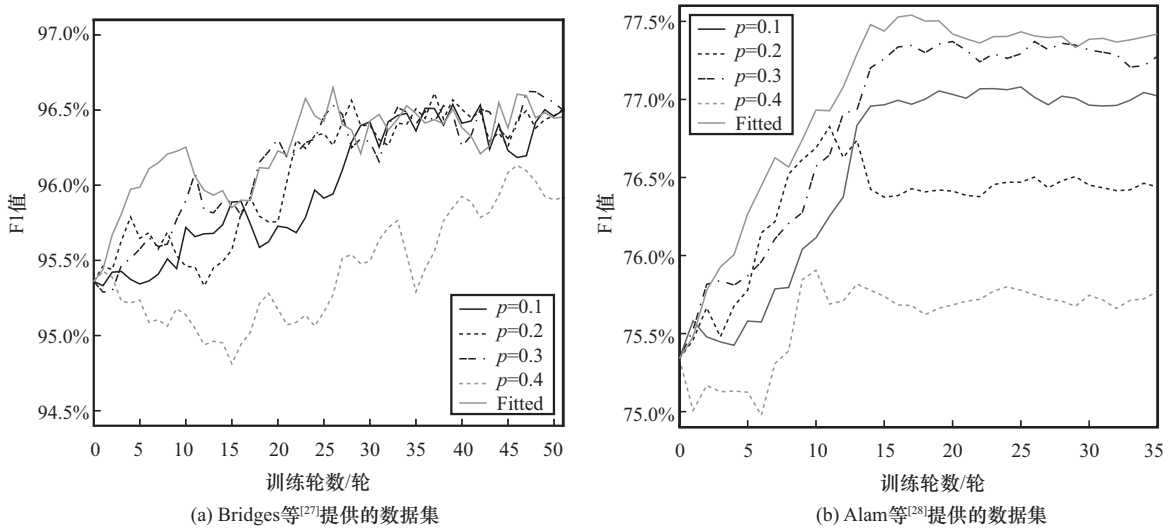
在本次实验中，针对不同的去除比例p对网络安全命名实体识别性能进行了评估。实验在2个数据集上进行，分别是Bridges等[27]提供的数据集和Alam等[28]提供的数据集。本文设置了4个不同的p值（0.1、0.2、0.3、0.4）以及根据本文方法得到的比例（Fitted），并记录了每个p值在不同训练轮数下的F1值变化情况。

实验结果如图7所示，在不同的去除比例下，F1值均表现出随着训练轮数增加而逐渐提升的趋势，最终趋于稳定。在Bridges等[27]提供的数据集中，初始F1值为95.36%，最终F1值在本文方法下达到了96.65%。不同的p值对F1值的提升效果有所差异，但整体上都表现出一定的提升。Alam等[28]提供的数据集中，初始F1值为75.34%，最终F1值在本文方法下达到了77.54%，表现出显著的提升。

从整体结果来看，较低的去除比例（p = 0.1和p = 0.2）在2个数据集上均表现出了较为稳定的提升效果，而根据本文方法得到的比例在2个数据集上的表现均优于其他比例，验证了本文方法的有效性。不同p值的去除比例通过减少低质量标注样本，有效地提升了模型的性能，且本文筛选比例p进一步优化了模型的效果。实验结果表明，适当去除低质量样本的策略可以显著提升网络安全命名实体识别模型的性能。

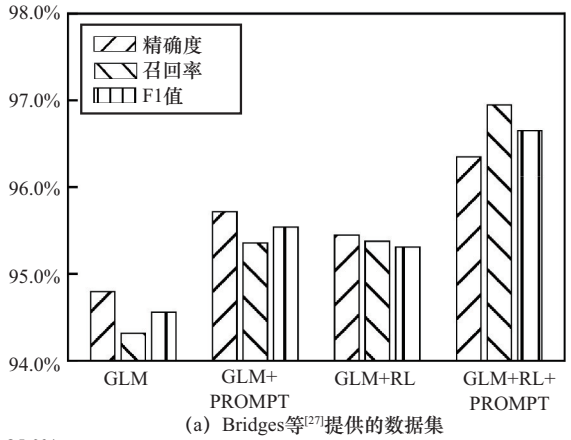
4.6　消融实验

为了验证提出模型中各组件的合理性，本节进行了几项消融实验，分别验证了基于软提示微调的网络安全实体识别方法和基于强化学习的网络安全实体筛选器。在2个数据集上的消融实验结果如图8所示。具体来说，完整模型在2个数据集上的F1值表现优异，分别为96.65%和77.54%。移除强化学习筛选数据过程后，模型性能有所下降。在第1个数据集中，F1值降低了1.11%，在第2个数据集中，F1值降低了2.2%。这一现象证明了基于强化学习的网络安全实体筛选器在筛选低质量标注数据方面的有效性。当去除基于软提示微调的网络安全实体识别方法而仅使用GLM时，3个评估指标同时下降。这表明基于软提示微调的网络安全实体识别方法能够从复杂的网络安全描述中有效学习结构化信息和依赖特征，显著提升了模型性能。

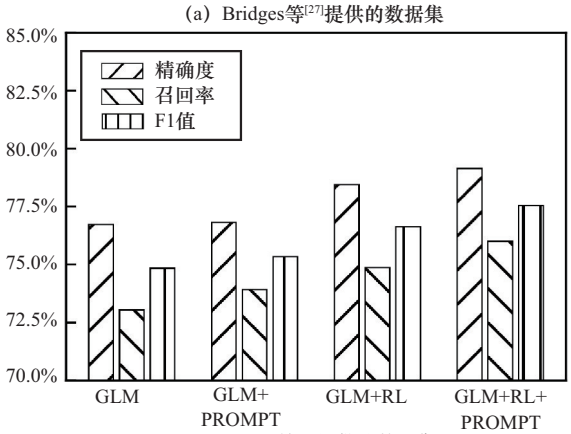


(a) Bridges等[27]提供的数据集　　(b) Alam等[28]提供的数据集

图7　不同筛选比例下模型F1值随着训练轮数的变化趋势



(a) Bridges等[27]提供的数据集



(b) Alam等[28]提供的数据集

图8　在2个数据集上的消融实验结果

4.7　案例分析

为了提供对网络安全命名实体识别任务更直观的理解，下面列出了2个数据集中的4个案例句子，以可视化实体预测结果，分别如表7~表10所示。其中，“标签”行表示原始数据集中的真实标签，“预测”行是根据给定句子由本文方法预测的标签。解码的实体标签序列使用BIOES标注方案。BIOES的含义已在第2.1节中描述。

表7　Bridge等[27]提供的数据集句子1案例分析

单词	标签	预测
The	O	O
Janrain	B-application	B-application
Capture	I-application	I-application
module	O	O
6.x-1.0	B-version	B-version
and	O	O
7.x-1.0	B-version	B-version
for	O	O
Drupal	O	O

表8　Bridge等[27]提供的数据集句子2案例分析

单词	标签	预测
Integer	B-relevant_term	B-relevant_term
overflow	I-relevant_term	I-relevant_term
in	O	O
the	O	O
i915_gem_execbuffer2	B-function	B-function
function	O	O
in	O	O
drivers/gpu/drm/i915/i915_gem_execbuffer.c	B-file	B-file

表9　Alam等[28]提供的数据集句子3案例分析

单词	标签	预测
name	O	O
Package	O	O
name	O	O
SHA	O	O
256	O	O
hash	O	O
Flash	B-System	B-System
Player	I-System	I-System
com.uxlgtsvfdc.zipvwntdy	B-Indicator	B-Indicator
728a6ea44aab94a2d0eb…	B-Indicator	B-Indicator
name	O	O

表10　Alam等[28]提供的数据集句子4案例分析

单词	标签	预测
Reznov.dll	B-Indicator	B-Indicator
—	O	O
17b8665cdbbb94482ca970a754d11d6e29c4…	B-Indicator	B-Indicator
Custom	O	O
activity	O	O
prefix	O	O
com.cact.CAct	B-Indicator	B-Indicator
Cerberus	B-Malware	B-Malware
—	O	O

在句子1中，如“Janrain Capture module”和“6.x-1.0 and 7.x-1.0”等实体术语都被本文方法正确识别。这些实体由于包含复杂的数字和版本信

息，传统模型往往难以准确识别。然而，通过本文提出的基于软提示微调的网络安全实体识别方法和基于强化学习的网络安全实体筛选器的结合，模型能够有效捕捉上下文信息和实体边界，显著提升了识别精度。

在句子 2 中，本文方法成功识别了“Integer overflow”作为相关术语实体，以及“i915_gem_execbuffer2 function”和“drivers/gpu/drm/i915/i915_gem_execbuffer.c”作为函数和文件实体。这说明本文方法不仅在处理复杂的上下文结构上表现优越，还能够准确定位技术术语和文件路径等关键信息，从而提高了实体识别的整体表现。

在句子 3 中，系统实体和指示符实体如“Flash Player”和“com. uxlgtsvfdc. zipvwntdy”被准确识别，与标签完全一致。这进一步证明了本文方法在不同上下文和领域中的通用性和鲁棒性。

在句子 4 中，指示符实体如“Reznov. dll”和“com.cact.CAct”以及恶意软件实体“Cerberus”也被正确识别。

这些成功的识别案例展示了本文方法在处理复杂网络安全描述和多样实体类型时的优越性，减少了漏检和误检的情况。

通过上述案例分析可以看出，本文方法在序号生成机制和基于强化学习的网络安全实体筛选器的协同作用下，显著提升了网络安全命名实体识别的精度和召回率，尤其在处理复杂多样的实体时表现出色。相比之下，传统方法难以应对复杂结构和多样实体，易导致识别不准确。

5　结束语

在这项研究中，本文提出了一种创新的网络安全实体识别方法，旨在显著提高实体识别的性能与效率。首先，结合 GLM 的软提示微调方法，不仅能够在几乎不损失精度的前提下减少显存占用，还有效地解决了网络安全领域特有的实体跨度识别不精准的问题。具体而言，提示微调过程中设计的序号映射机制，通过精准地捕捉实体的上下文信息，提升模型对复杂实体的识别能力。其次，本文首次引入强化学习用于筛选数据集中的低质量标注，通过智能体的动态学习和决策优化，显著提升了方法在网络安全实体识别数据集上的表现。实验显示，在 2 个基准网络安全数据集上，所提方法优于现有的最优方法。此外，针对现实世界场景中可用的统一网络安全数据集稀缺的挑战，未来的研究将探索小样本学习或无监督学习的应用，以期在减少人工干预的同时，进一步提高实体识别的效果和适应性。

参考文献：

[1] MOURA G C M, HEIDEMANN J. Vulnerability disclosure considered stressful[J]. ACM SIGCOMM Computer Communication Review, 2023, 53(2): 2-10.

[2] GLYDER J, THREATT A K, FRANKS R, et al. Some analysis of common vulnerabilities and exposures (CVE) data from the national vulnerability database (NVD)[C]//Proceedings of the Conference on Information Systems Applied Research. Piscataway: IEEE Press, 2021: 1-7.

[3] GIANNAKOPOULOS T, MALIATSOS K. On the usage of NLP on CVE descriptions for calculating risk[C]//European Symposium on Research in Computer Security. Berlin: Springer, 2024: 104-123.

[4] KUEHN P, BAYER M, WENDELBORN M, et al. OVANA: an approach to analyze and improve the information quality of vulnerability databases[C]//Proceedings of the 16th International Conference on Availability, Reliability and Security. New York: ACM Press, 2021: 1-11.

[5] OKUTAN A, MELL P, MIRAKHORLI M, et al. Empirical validation of automated vulnerability curation and characterization[J]. IEEE Transactions on Software Engineering, 2023, 49(5): 3241-3260.

[6] PRASAD S G, SHARMILA V C, BADRINARAYANAN M K. Role of artificial intelligence based chat generative pre-trained transformer (ChatGPT) in cyber security[C]//Proceedings of the 2023 2nd International Conference on Applied Artificial Intelligence and Computing (ICAAIC). Piscataway: IEEE Press, 2023: 107-114.

[7] DING N, QIN Y, YANG G, et al. Parameter-efficient fine-tuning of large-scale pre-trained language models[J]. Nature Machine Intelligence, 2023, 5(3): 220-235.

[8] XU Y J, TAN X B, TONG X, et al. A robust Chinese named entity recognition method based on integrating dual-layer features and CSBERT[J]. Applied Sciences, 2024, 14(3): 1060.

[9] SRIVASTAVA S, PAUL B, GUPTA D. Study of word embeddings for enhanced cyber security named entity recognition[J]. Procedia Computer Science, 2023, 218: 449-460.

[10] DOWNEY D, BROADHEAD M, ETZIONI O. Locating complex named entities in web text[C]//Proceedings of the 20th International Joint Conference on Artifical Intelligence. New York: ACM Press, 2007: 2733-2739

[11] MORWAL S. Named entity recognition using hidden Markov model (HMM) [J]. International Journal on Natural Language Computing, 2012, 1(4): 15-23.

[12] MANSOURI A, AFFENDY L S, MAMAT A. A new fuzzy support vector machine method for named entity recognition[C]//Proceedings of the 2008 International Conference on Computer Science and Information Technology. Piscataway: IEEE Press, 2008: 24-28.

[13] MCDONALD R, HALL K, MANN G. Distributed training strategies for the structured perceptron[C]//Human Language Technologies: The

2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics. New York: ACM Press, 2010: 456-464.

[14] PATIL N, PATIL A, PAWAR B V. Named entity recognition using conditional random fields[J]. Procedia Computer Science, 2020, 167: 1181-1188.

[15] MULWAD V, LI W J, JOSHI A, et al. Extracting information about security vulnerabilities from web text[C]//Proceedings of the 2011 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology. Piscataway: IEEE Press, 2011: 257-260.

[16] LAL R. Information extraction of security related entities and concepts from unstructured text[D]. Baltimore: University of Maryland Baltimore County, 2013

[17] WEERAWARDHANA S, MUKHERJEE S, RAY I, et al. Automated extraction of vulnerability information for home computer security[C]//International Symposium on Foundations and Practice of Security. Berlin: Springer, 2015: 356-366.

[18] COLLOBERT R, WESTON J, BOTTOU L, et al. Natural language processing (almost) from scratch[J]. Journal of Machine Learning Research, 2011, 12: 2493-2537.

[19] HUANG Z, XU W, YU K. Bidirectional LSTM-CRF models for sequence tagging[J]. arXiv Preprint, arXiv: 1508.01991, 2015.

[20] KIM G, LEE C, JO J, et al. Automatic extraction of named entities of cyber threats using a deep Bi-LSTM-CRF network[J]. International Journal of Machine Learning and Cybernetics, 2020, 11(10): 2341-2355.

[21] QIN Y, SHEN G W, ZHAO W B, et al. A network security entity recognition method based on feature template and CNN-BiLSTM-CRF[J]. Frontiers of Information Technology & Electronic Engineering, 2019, 20(6): 872-884.

[22] SIMRAN K, SRIRAM S, VINAYAKUMAR R, et al. Deep learning approach for intelligent named entity recognition of cyber security[C]//International Symposium on Signal Processing and Intelligent Recognition Systems. Berlin: Springer, 2020: 163-172.

[23] ZHOU S, LIU J J, ZHONG X F, et al. Named entity recognition using BERT with whole world masking in cybersecurity domain[C]//Proceedings of the 2021 IEEE 6th International Conference on Big Data Analytics (ICBDA). Piscataway: IEEE Press, 2021: 316-320.

[24] GAO C, ZHANG X, LIU H. Data and knowledge-driven named entity recognition for cyber security[J]. Cybersecurity, 2021, 4(1): 9.

[25] WANG X D, LIU J Y. A novel feature integration and entity boundary detection for named entity recognition in cybersecurity[J]. Knowledge-Based Systems, 2023, 260: 110114.

[26] ZENG A, LIU X, DU Z, et al. GLM-130B: an open bilingual pre-trained model[J]. arXiv Preprint, arXiv: 2210.02414, 2022.

[27] BRIDGES R A, JONES C L, IANNACONE M D, et al. Automatic labeling for entity extraction in cyber security[J]. arXiv Preprint , arXiv: 1308.4941, 2013.

[28] ALAM M T, BHUSAL D, PARK Y, et al. CyNER: apython library for cybersecurity named entity recognition[J]. arXiv Preprint, arXiv: 2204.05754, 2022.

[29] ABDULLAH M S, ZAINAL A, MAAROF M A, et al. Cyber-attack features for detecting cyber threat incidents from online news[C]//Proceedings of the 2018 Cyber Resilience Conference (CRC). Piscataway: IEEE Press, 2018: 1-4.

[30] WU X J, ZHANG T Q, YUAN S, et al. One improved model of named entity recognition by combining BERT and BiLSTM-CNN for domain of Chinese railway construction[C]//Proceedings of the 2022 7th International Conference on Intelligent Computing and Signal Processing (ICSP). Piscataway: IEEE Press, 2022: 728-732.

[31] JIE Z, LU W. Dependency-guided LSTM-CRF for named entity recognition[J]. arXiv Preprint, arXiv: 1909.10148, 2019.

[作者简介]

田泽庶（1997– ），男，黑龙江哈尔滨人，哈尔滨工业大学博士生，主要研究方向为信息抽取、知识图谱构建等。

刘春雨（1994– ），女，黑龙江哈尔滨人，哈尔滨工业大学博士生，主要研究方向为城市计算、服务计算等。

张云婷（1997– ），女，黑龙江哈尔滨人，哈尔滨工业大学博士生，主要研究方向为网络与信息安全、对抗文本生成等。

张嘉宇（1997– ），男，山西太原人，哈尔滨工业大学博士生，主要研究方向为知识图谱构建、社交立场分析等。

孟超（1991– ），男，河南鹿邑人，哈尔滨工业大学博士生，主要研究方向为社交网络分析、社交立场分析等。

张宏莉（1973– ），女，吉林榆树人，博士，哈尔滨工业大学教授、博士生导师，主要研究方向为社交网络分析、网络与信息安全等。