基于提示问答数据增强的小样本网络安全事件检测方法

汤萌萌[1]，郭渊博[2]，张晗[3]，白庆春[4]，陈庆礼[1]，张博闻[5]

（1.信息工程大学密码工程学院，河南 郑州 450001；2.海南大学网络空间安全学院，海南 海口 570100；
3.郑州大学网络空间安全学院，河南 郑州 450001；4.上海开放大学上海开放远程教育工程技术研究中心，上海 200082；
5.郑州浪潮数据技术有限公司，河南 郑州 450001）

摘　要：针对网络安全领域的事件识别标注数据较为匮乏且场景和语义复杂，难以构建准确的事件识别模型的问题，提出了一种基于提示问答数据增强的小样本网络安全事件检测方法。首先利用提示信息获取事件表示知识，并结合标签词映射网络安全事件类型，从未标注的文本中生成新的数据来扩充训练数据；然后使用生成的高置信度的伪标注实例和原始数据来微调模型，以增强模型对网络安全事件的语义理解能力；最后在2个网络安全领域数据集上进行了实验验证。结果表明，与其他基线方法相比，所提方法在低资源网络安全事件检测任务上具有很强的优越性。

关键词：网络安全；事件检测；提示问答；数据增强；小样本

中图分类号：TP391

文献标志码：A

DOI：10.11959/j.issn.1000–436x.2024105

Few-shot cybersecurity event detection method by data augmentation with prompting question answering

TANG Mengmeng[1], GUO Yuanbo[2], ZHANG Han[3], BAI Qingchun[4], CHEN Qingli[1], ZHANG Bowen[5]

1. Department of Cryptogram Engineering, Information Engineering University, Zhengzhou 450001, China
2. School of Cyberspace Security, Hainan University, Haikou 570100, China
3. School of Cyberspace Security, Zhengzhou University, Zhengzhou 450001, China
4. Shanghai Open Distance Education Engineering Technology Research Center, Shanghai Open University, Shanghai 200082, China
5. Zhengzhou Inspur Data Technology Co., Ltd., Zhengzhou 450001, China

Abstract: The cybersecurity field lacks sufficient annotated data for event recognition, and the scenarios and semantics are complex, making it difficult to construct accurate event recognition models. A few-shot cybersecurity event detection method by data augmentation with prompting question answering was proposed. Firstly, event representation knowledge was obtained using prompt information and combined with label words to map cybersecurity event types. New data was generated from unlabeled text to expand the training data. Then, the generated high-confidence pseudo-annotated instances and raw data were used to fine-tune the model to enhance its semantic understanding of cybersecurity events. Experimental verification was conducted on two datasets in cybersecurity. The result showes that the proposed method's substantial superiority in low-resource network security event detection tasks compared to other baseline methods.

Keywords: cybersecurity, event detection, prompting question answering, data augmentation, few-shot

---

收稿日期：2024–02–05；**修回日期：**2024–05–14

基金项目：国家自然科学基金资助项目（No.62276091, No.62307028）；河南省重大公益专项基金资助项目（No.201300311200）；上海市自然科学基金资助项目（No.23ZR1441800）；上海市启明星项目扬帆专项基金资助项目（No.23YF1426100）

Foundation Items： The National Natural Science Foundation of China (No.62276091, No.62307028), Major Public Welfare Project of Henan Province (No.201300311200), Shanghai Natural Science Foundation (No.23ZR1441800), Shanghai Sailing Program (No.23YF1426100)

0　引言

事件检测和事件元素抽取是事件抽取任务的2个核心组成部分。其中，事件检测的目标是对给定的文本进行判断，确定是否发生了某一预定义事件类型，而这是事件抽取的先决条件。网络安全领域中的事件检测需要从新闻博客、威胁情报和日志信息等文本中检测出网络威胁等相关的安全事件信息，以对网络安全情报研究、趋势分析、技术监测等进行有效支撑[1-2]。当前大多数事件检测研究专注于通用领域，涵盖了人类日常生活的各个方面，并不依赖于特定的专业背景，但是对网络安全领域事件检测相关的研究则相对较少。

与通用领域检测的人类日常生活事件不同，网络安全领域事件检测面临以下挑战。

1) 数据的敏感性、稀缺性以及强领域性导致获取高质量的标注数据尤其困难。网络安全数据的敏感性使其难以公开和共享，同时事件的稀缺性和多样性增加了数据整理的难度。此外，领域数据需要专业人员深入分析和标注，使得高质量标注数据难以获取。因此，缓解数据匮乏、增强扩充数据变得尤其关键。

2) 网络安全领域的场景更为复杂，且实体表达形式更为多样化，涵盖了大量专业术语，使得在进行事件识别分类时，误判率相对较高。网络安全领域术语和缩写丰富多样，同一缩写可能有不同含义。比如，IDS通常解释为“入侵检测系统（intrusion detection system）”，但也可能指代“集成数据服务（integrated data service）”。因此，正确识别和分类关键词是网络安全事件检测领域的一个重要的挑战。

3) 检测事件类型具有内在复杂性、多样性、聚焦性，因而仅依赖通用领域的模型往往难以提供高质量的检测结果。网络安全事件检测专注于识别网络攻击和异常行为，需要特定的知识和技术。通用模型往往不能适应，特别是在资源受限情况下，难以充分捕捉复杂语义信息，达到模型预期性能。

因此，相比通用领域的事件检测模型，适用于网络安全领域的检测模型更需要结合安全领域的背景知识来增强对事件语义的理解。当前有监督的事件检测方法取得了良好的性能[3-5]。这些方法都依赖于大规模的标注数据集，而网络安全领域往往无法提供这样的数据条件。

近年来，小样本事件检测引起了更多的关注，许多小样本算法已经被应用于小样本目标检测中[6]。在网络安全领域，针对小样本事件检测的研究相对较少，Bayer等[7]采用了3种适用于数据量较少情况的技术用于训练出高效的分类器，分别是迁移学习、数据增强和小样本学习，受通用领域小样本事件检测方法[8-10]的启发，本文将网络安全事件检测的目标重新定义为小样本学习问题的实例化。通过将低资源场景下的事件检测任务转化为小样本学习问题，用少量训练样本微调模型以适应分类任务，并且通过提示问答提升模型对上下文的理解能力。

在低资源的网络安全场景中，标注数据非常有限，这使得准确识别和分类网络安全事件变得困难。为了有效应对这个问题，需要提高模型的泛化能力、鲁棒性和准确性，同时也需要减轻过拟合问题。这可以通过增加训练数据集的多样性来实现，数据增强[11-13]提供了有效的解决方案，为小样本学习提供了支持。数据增强的主要思路是应用不同的变换和扩充技术从现有的有限数据中生成更多样的样本，扩大训练集，为模型提供更多的学习材料。然而，同义词替换、随机插入、删除等增强方法[14]可能引入不准确的语义信息，从而改变文本的意思或导致文本意思失真。

为了保持数据的语义一致性，同时增加数据集的多样性，本文使用提示问答标注实例的方法进行数据增强。在参考文献[15]数据增强方法的基础上，本文进一步增加了标签词的使用，帮助模型更好地识别和区分重要的网络威胁事件特征，具体为采用提示模型融合标签词，设计合适的提示或问题来减少所需的标注数据量，并构建标签词加强模型输出和标签之间的映射，帮助模型更好地理解任务的上下文信息，充分利用有限的标注数据，直接从未标注的文本中生成新的训练实例。

本文主要的贡献如下。

1) 提出了一种基于提示问答数据增强的小样本网络安全事件检测方法，以充分捕捉语义信息来有效降低对标注数据的需求。该方法利用数据增强增加了数据集的多样性，并通过设计适当的提示问题来增强网络安全事件检测模型对复杂语义信息的理解，分析关键术语出现的上下文来判断其具体含义，减少实体多义性产生的歧义，提高事件分类的

准确性。

2) 提出了一种基于提示问答的文本事件数据增强方法，以缓解网络安全域事件检测任务中面临的数据缺乏问题。该方法将提示问答融合标签词映射作为数据增强策略，构建标签词映射加强模型输出和标签之间的关系，生成更多样化的样本实例。

3) 在2组数据集上对网络安全小样本事件检测任务进行广泛实验，结果表明，本文方法比其他基线方法性能良好。

1 相关工作

1.1 网络安全事件检测

事件检测是指在文本数据中识别、提取出具有特定意义或重要性的事件。这些事件有多种类型，包括经济事件、社会事件、自然灾害等。一些研究致力于利用社交媒体平台Twitter来分析特殊事件，例如洪水等自然灾害事件[16-18]。就研究范围而言，网络安全事件检测属于网络安全领域的信息提取，其研究的信息主要来自每日发布的新闻博客文章、威胁情报和日志信息等。Yagcioglu等[19]用卷积神经网络（CNN, convolutional neural network）和长短期记忆（LSTM, long short-term memory）网络以单词级别的元嵌入作为输入，并结合上下文嵌入来检测Twitter中的网络安全事件。Qiu等[20]设计并比较了研究我国新闻网站攻击的多个特征表示模型，这些模型从网络上的新闻资源中自动提取网络攻击事件及其参与实体。Ritter等[21]将所研究的事件分为3种类型，包括账户劫持、分布式拒绝服务（DDoS, distributed denial of service）攻击和数据泄露，并通过弱监督的方法从Twitter中提取事件来检测广泛的网络攻击。Satyapanich等[1]基于新闻博客的网络安全进行文本信息提取，构建了包含5种事件类型的数据集，并提出了一种从新闻博客文本信息中提取网络安全事件的系统。Luo等[22]提出了一种从文本中提取文档级信息的系统，以端到端的方式提取文档级网络安全事件，将事件提取任务形式转变为序列标记任务，并识别分类了4种网络安全事件类型。

虽然上述工作侧重于网络安全事件检测，但是它们的模型都采用有监督的事件检测方法。这种方法需要大量的标注数据来训练模型，并且只能覆盖有限的事件类型，扩展新的事件类型面临一定的挑战。由于特定事件的报道通常数量较少，而扩展新事件类型需要大量的标注数据来训练模型以准确理解该类型的事件。因此网络安全领域的标注数据较为缺乏，同时获取高质量的实例数据成本较高，这些方法难以应用于低资源网络安全场景。

1.2 小样本事件检测

近年来各类神经网络模型在事件检测任务上表现出良好的性能，引起了研究者的关注。Chen等[23]提出了一个动态多池卷积神经网络（DMCNN, dynamic multi-pooling convolutional neural network），用于从文本中捕获句子级特征。Nguyen等[24]采用双向递归神经网络（B-RNN, bidirectional recursive neural network）学习隐藏特征表示来捕获句子中的依赖关系。此外，还有研究者[25]引入了图卷积网络（GCN, graph convolutional network），其注意力聚合机制能更好地捕获事件检测依赖的上下文信息。然而这些方法都是基于有监督的学习模型，需要标注大量的数据集。网络安全领域的相关文本数据在不断增长，但是标注数据的缺乏依然是该研究领域的主要瓶颈。

然而在小样本场景下，由于有监督学习的事件检测模型无法充分利用少量标注数据进行训练，难以对事件的特征和上下文信息进行充分学习。通常解决该问题的方式是为该类型标注更多数据，然后用新标注的数据来重新训练模型。在实际的研究中，标注数据的成本昂贵，数据分布不均衡，某些事件类型训练实例数量较少，以上这些方法容易发生过拟合的情况，难以达到理想迁移效果。对于一些专业领域事件的检测，由于缺少标记数据，其发展和应用受到了限制。

随着对低资源场景（如小样本场景）研究的深入，如何利用少量数据训练事件检测模型也受到越来越多的关注。Peng等[26]提出一种事件检测和协同引用方法进行结构化的向量表示，尝试解决注释事件难题。Cong等[10]提出了一种基于标记的统一模型，将任务转换为具有两部分标记方案的小样本标记问题。Lai等[27]研究了基于度量的小样本学习模型以使事件检测能够扩展到新的事件类型。Deng等[28]使用了动态存储器的原型网络（DMB-PN, dynamic-memory-based prototypical network）来稳健句子编码，实现小样本事件检测。

Zhang 等[29]提出了一个面向小样本事件检测任务的混合对比学习框架，其采用了自适应任务自适应阈值（TAT, task-adaptive threshold）来消除触发词的误识别。随着大模型和提示学习在小样本学习中的成功应用，也有不少研究采用提示学习来缓解标注数据缺乏问题。Du 等[30]提出一种新的基于问答的事件抽取范式解决面临的数据稀缺问题。该方法引入了一种新的事件抽取范式，并将其形式化为一个问答（QA, question answering）任务，以端到端的方式抽取事件参数。Liu 等[31]提出了一种方法，利用机器阅读理解机制来解决文档级事件论元链接任务中的数据稀疏问题。该方法将任务明确框定为机器阅读理解（MRC, machine reading comprehension）问题，其中论元抽取被视为一个问答过程，即在文档中寻找与特定事件相关的问题的答案。

尽管上述研究为少样本事件检测模型提供了有价值的见解和潜在解决方案，促进了检测任务在低资源场景中的发展。但是它们支持的训练集样本有限，训练数据缺乏多样性，导致模型可能接受不准确的表征学习，无法很好地识别事件提及，从而影响模型的性能。

此外，GPT 3.5 的到来给自然语言发展带来质的飞跃，ChatGPT 模型利用提示问答来准备训练数据，指导模型在对话中生成符合预期的回答。但是 ChatGPT 在复杂场景中很难与特定任务模型的性能相匹配，特别是在需要专业知识背景的网络安全领域[32]。

1.3　基于提示问答的数据增强实例生成

在没有足够的训练数据场景下，数据增强是一种有效提高模型性能和鲁棒性的方法，其尝试从现有的训练实例中产生新的训练数据。Wei 等[14]通过同义词替换、随机插入、交换和删除来进行数据增强。Gao 等[33]采用两次施加标准 dropout 作为隐藏表示的最小数据增强。Xie 等[34]提出用 RandAugment 和反向翻译这些高质量的数据增强方法来替代传统的噪声注入方法。这些方法虽然能在一定程度上提高对应任务的性能，但是其数据增强方法可能会曲解生成实例的含义，同时对实例的改动具有局限性，得到的增强数据多样性不足。

在最近的工作中，Liu 等[15]采用了提示机制融合自训练策略来完成 NER（named entity recognition）任务的数据增强，用 LLM（large language model）生成的数据补充人工管理的训练数据，以改进训练数据。Huang 等[35]提出了一个基于提示的小样本 NER 自训练方法。该方法使用自我培训的方式，用带标签的数据为未标记样本创建合成标签，并通过自适应标签选择和高置信度标签微调 BERT-NER （bidirectional encoder representations from transformers-NER）模型。这些方法为缓解数据稀疏问题提供了解决方案，但自训练方法存在一些缺点，如在自我培训时使用教师模型，而教师模型可能会受到标签错误或数据过拟合的影响，导致模型性能下降或训练不收敛。特别是在网络安全领域，由于需要应用专业知识来处理数据，而标注错误是一个常见的问题，这些错误的标注数据很容易导致训练模型难以收敛。因而这些方法在网络安全事件检测任务中无法达到预期的效果。

为解决以上问题，本文提出了一种新的数据增强方法用于网络安全领域的低资源事件检测，利用标签词映射来强化数据生成的多样性和语义信息，旨在加深对小样本场景的研究。特别是当获取有限标注数据时，本文方法通过 BERT 模型利用提示信息来获取事件表示的知识，不仅可以生成符合任务要求的新训练实例，同时还能通过提示问答的方式引入更广泛的变化，使训练集更加多样化，从而提高了模型的性能和泛化能力。

2　方法

为提升小样本场景中网络安全事件检测的效果，缓解检测任务面临的数据不足问题，本文设计了一种提示问答的数据增强策略来生成更多样化的样本实例。通过引入相应的事件检测提示问题，能够针对不同类型的事件有效地生成新样本数据，从而显著提高事件检测性能。本文方法基本流程如图 1 所示，可概括为 3 个关键步骤：未标注的数据使用数据增强模块生成新的训练数据；使用生成的新数据扩充原训练数据集，并通过微调事件检测模型进行训练；使用已训练的模型对文本进行分类，以预测事件类型。

图 2 概述了本文方法的框架，其中包括 2 个模块：生成训练实例模块和事件检测模块。为了最大限度地利用有限的标注数据，本文提出了一种基于提示问答的数据增强方法，用于在小样本场景中检测特定事件类型。

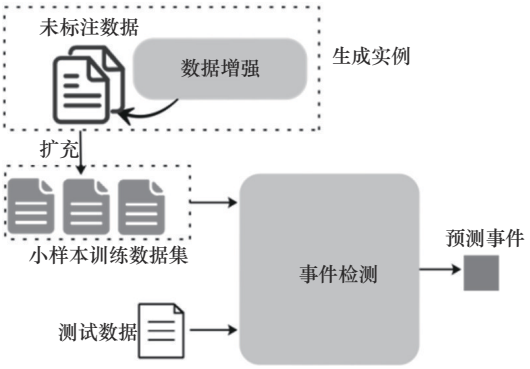


图1　本文方法基本流程

对于生成训练实例模块，首先，使用BERT预训练模型对提示问答模型中的问题和文本进行编码，将文本转换为向量表示；其次，通过融合标签词映射机制，将文本关键信息与预定义标签映射，进而精确识别事件，根据文本的向量表示和标签映射生成可能的事件类型，即伪事件类型；最后，选择置信度高于阈值的标签作为事件类型，生成新的训练实例。对于事件检测模块，利用生成的实例和原始标注的数据进行微调，以提高提示问答模型性能；对编码后的文本使用分类器进行分析，模型能够预测出具体的事件类型，实现对事件的有效识别与分类。

2.1　任务定义

为了便于描述，本节给出了如下符号以及任务定义。给定一条输入网络安全相关的文本 $X=\{x_1,x_2,x_3,\cdots,x_n\}$，其中，$x_i$ 为文本中的第 i 个单词，n 为文本总字数。事件检测任务的目标是进行文本的多分类，其结果输出为 $Y=(x_i,l)$，其中，$l\in L$ 表示事件类型的标签，L 为数据集中所有类型的集合。该方法旨在训练一个分类器，将文本映射到集合中的一个类别上，即

$$f(\cdot):X\to Y \tag{1}$$

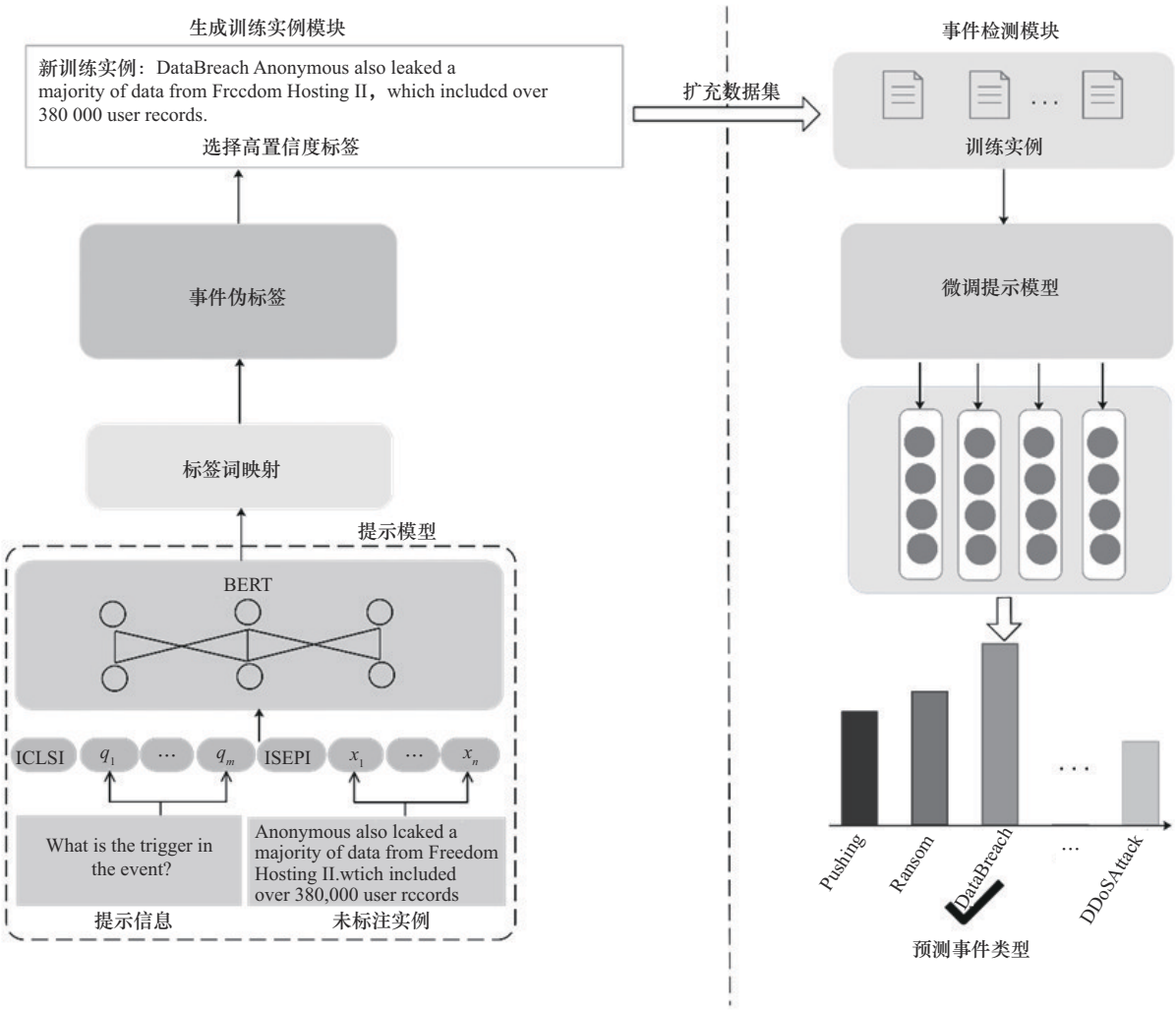


图2　本文方法的框架

2.2　基于提示问答的数据增强方法

网络安全领域的事件检测任务在训练过程中只有少量的实例可使用，这使得模型很难从有限数据中学习到足够的知识和模式，为此本节提出了提示问答的数据增强方法来解决以上问题。利用 BERT 预训练模型的知识来获取句子的语义信息，具体为使用提示问答生成针对特定事件类型的问题，然后将生成的问题与原始数据一起进行训练，在训练过程中，模型会预测事件的类别，并过滤选择具有高概率的标签，从而生成新的训练实例以扩充数据集。

2.2.1　提示问答

受到文献[15]的启发，本文通过设计合适的“提示”引导模型生成特定的答案来提升模型性能。在训练过程中，这种提示问答方法会额外引入信息或约束，以指导模型生成更准确、更具针对性的输出。但与文献[15]不同的是，为提高未注释文本中生成新的事件检测训练实例的准确率，本文还引入了标签词映射的技术，将文本数据中的事件关键词与相应的标签进行关联。该方法的核心思想是在预训练的 BERT 模型上使用问句作为提示来标注文本，从而产生新的实例。假设输入序列 m 是由一个模板改编的训练文本，目标是使用预训练的 BERT 编码器对输入句子 X 和事件提示句 $Q=\{q_1,q_2,q_3,\cdots,q_m\}$ 进行编码，得到它们的上下文表示。

对于网络安全事件检测，该任务目标是输入的文本 X 与标签 $y\in Y$ 匹配。具体来说，首先用模板 $T(\cdot,\cdot)$ 将 X 和 y 包装成自然语言句子，然后判断是否匹配，开始处插入[CLS]标记和结束用[SEP]标记分隔，拼接成一个新的序列 m。例如，假设判断文本 X=“某机构网络受到攻击，其客户信息遭到泄露”是否属于 y=“数据泄露”事件类型，将句子包装为

$$m=T(X,y)=[\mathrm{CLS}]\,X?[\mathrm{MASK}],y\,[\mathrm{SEP}] \quad (2)$$

其中，[CLS]和[SEP]是特殊令牌。输入序列 m 传递给 BERT 编码器进行编码操作，可以获得每个单词在[MASK]标签的位置出现的概率。为了将单词的概率映射到匹配结果，定义了从部分单词到匹配标签的映射（第 3.2.2 节）。这种映射允许模型获得输入文本和事件类型之间的相关性。

2.2.2　标签词映射构建

根据文献[36-37]，标签词映射能在模型输出和标签之间搭建桥梁。本文创建了一个网络安全事件标签词映射，该映射将未标记的事件根据标签词分类到预先定义的类别中（如表 1 所示），以建立模型输出和标签之间的对应关系。

具体方式为：收集已定义好的事件类别，如本文采用的 2 个数据集分别定义了 9 种和 5 种事件类型，如表 1 所示。当构建网络安全领域的标签词映射时，确定需要分类的标签类别，例如“数据泄露”“恶意软件”“网络钓鱼”“漏洞检测”等。在上述示例中，“数据泄露”类别包括与数据泄露相关的单词和短语，如“data breach”“compromise”“leak”和“stolen”等为构建标签词映射的关键词。由于事件类别按照触发词进行分类，构建标签词多选用触发词相关的词语，因此可以使用基于标签词映射的方法来增强识别和分类事件。具体如式(3)所示，构建一个包含多个关键词的标签词映射。

$$f(\boldsymbol{v})=\sum_{i=1}^{n}W_i\boldsymbol{v}_i\geqslant\tau \quad (3)$$

其中，$f(\boldsymbol{v})$ 是标签词映射函数，$\boldsymbol{v}$ 是输入文本的特

表 1　　**事件标签词示例**

事件标签	事件标签词
DataBreach(Attack.Databreach)	data breach,compromised,exposed, leaked, stolen,accessed,selling,…
Phishing(Attack.Phishing)	phishing attack,tricked,impersonated, forged,masquerading,lure,…
Ransom(Attack.Ransom)	ransom, extort, demanding,ransomware attack, pay, asking, requested,…
DDoSAttack	DDoSattack, denial-of-service attack, disables , disrupt,…
Malware	virus,malware,spyware,infected,accessing,attacked,…
SupplyChain	supply chain attack,unauthorized,…
VulnerabilityExploitation	impact,caused,exploited,flaw, resulted,froze,…
VulnerabilityDiscover(Discover.Vulnerability)	discovered, found, revealed,identified,reporting,…
VulnerabilityPatch(Patch.Vulnerability)	patched, fixed, updates,…

征向量，W是每个单词的权重，τ是一个阈值。标签词映射函数可以根据特征向量$\boldsymbol{v}$和每个单词的权重W，将特征向量映射到预定义的标签集合，并通过阈值τ控制输出结果。

如果输入文本与标签词匹配，则$f(\boldsymbol{v})$将大于或等于τ，该输入文本将被归类为相应的事件类型；反之，如果$f(\boldsymbol{v})$小于τ，则该文本不属于此事件类型。例如，如果对恶意软件攻击进行分类，并选择（感染，病毒）关键词

$$f(\boldsymbol{v}) = W_1\text{Infected} + W_2\text{Virus} \tag{4}$$

则输入文本的特征向量可以表示为$\boldsymbol{v}=[v_1,v_2,v_3,v_4,v_5]$，文本包含“感染”和“病毒”，则$\boldsymbol{v}=[1,0,0,1,0]$，每个关键词的权重可以通过训练得到。标签词函数中的标签词汇将对生成的类别向量描述贡献一定的权重，从而加强模型对每个类别的语义特征理解，同时也起到了约束判定的作用。

2.2.3 实例标签选择

为了度量输入文本和每个标签的相似性，通过BERT编码后生成文本的向量和标签向量描述，采用余弦相似度函数$S(\cdot)$衡量2个向量之间的夹角余弦值。对于任意的实例表示$\boldsymbol{v}_i$和$\boldsymbol{v}_j$，余弦相似度计算式为

$$S\left(\boldsymbol{v}_i,\boldsymbol{v}_j\right) = \cos\left(\boldsymbol{v}_i,\boldsymbol{v}_j\right) = \frac{\left(\boldsymbol{v}_i,\boldsymbol{v}_j\right)}{\left(\left|\boldsymbol{v}_i\right| \times \left|\boldsymbol{v}_j\right|\right)} \tag{5}$$

其中，$\boldsymbol{v}_i$和$\boldsymbol{v}_j$分别为输入文本x_i和x_j的实例表示。使用softmax函数对相似性得分进行归一化，如式(6)所示。

$$P_j = \frac{\mathrm{e}^{S\left(x_i y_j\right)}}{\operatorname{sum}\left(\mathrm{e}^{S\left(x_i y_k\right)}\right)} \tag{6}$$

则y_k表示所有候选标签的集合，即$\{y_1,y_2,y_3,\cdots,y_n\}$，不包括当前计算的标签$y_i$。通过余弦相似度来计算每个标签词与当前输入实例之间的相似度得分。

本文采用计算相似度来获取事件伪标签，其在注意力层分配权重，通过阈值过滤生成自适应标签。注意力层通过将Query向量与一组Key向量进行比较来计算每个伪标签的权重。然后，对相应的Value向量求和，将得到的结果用于计算加权向量，该加权向量表示注意力机制的输出。总之，该算法模型允许将每个伪标签用作输入数据，计算其与查询向量的相似度，并将相似度作为权重来获得自适应标签。

为了提高实例生成的质量和准确性，本文使用最高相似度$\text{score}_{\max}$来过滤事件伪标签，其值如式(7)所示。通过设置阈值，只有相似度得分最高的伪标签才会被选择为自适应标签。这确保了所选标签和查询向量间的最高相似度，从而提高了实例生成的质量和准确性。

$$\text{score}_{\max} = \max P_j \tag{7}$$

算法1总结了数据增强方法生成新的数据集训练的过程。其中，算法1在步骤5)~步骤6)获取生成的新实例，并将它们与原始训练样本合并，以确保生成的文本具有更高的多样性并包含更多的语料库知识。

算法1 提示问答数据增强算法

输入 标注训练数据集D_a和未标注数据集D_u

输出 训练后的事件分类模型

1) 定义预训练的BERT模型迭代次数epoch_a和事件分类的微调模型迭代次数epoch_b
2) for epoch in epoch_a do
3) 对于D_a中标注的训练样本，执行BERT模型微调训练操作
4) end for
5) 构建提示问答方法用微调过的模型，对未标注数据集D_u进行数据增强，生成增强数据集D_u^{aug}
6) 将已标注数据集D_a与增强数据集D_u^{aug}合并，形成新的训练数据集D'
7) for epoch in epoch_b do
8) 构建事件分类模型，使用新的标注训练集D'进行重新微调
9) end for
10) 返回训练完成的事件分类模型

2.3 事件检测模型的微调训练

在最后一个阶段，本文将数据扩充生成的实例与原始数据集合并，以扩充训练实例。然后，对预训练的BERT模型进行微调，重新训练网络安全事件检测模型。为了确定文本中事件的类型，本文使用提示进行辅助，具体是创建一个提示句子“This text is about <mask>”（定义的事件类型）用于检测文本中的事件类型。本文将输入文本w转换为标签序列$\left([\text{CLS}],w_1,w_2,\cdots,w_t,[\text{SEP}]\right)$。使用预先

训练的 BERT 模型对输入文本进行编码，并提取 h_{cls} 的嵌入表示[CLS]。通过训练一个额外的分类器 C，以输出标签集 Y 的概率分布，如式(8)所示。

$$P(\cdot|w)=\text{soft}\max\left(C\left(h_{\text{cls}}\right)\right) \tag{8}$$

分类器 C 和预先训练的语言模型通过最大化调整为

$$\frac{1}{N}\sum_{i}^{N}\log P\left(y_i|w_i\right),y_i\in Y \tag{9}$$

其中，y_i 是 w_i 标签。

由于存在不属于定义事件类型情况，因此增加了一个外部条件，即置信度小于阈值的类型将被排除。经过所有步骤，通过最大化条件分值来选择满足条件的最优类型作为预测文本的标签。该过程实现了使用少量标注数据检测网络安全事件类型。

3　实验

为了评估本文方法在网络安全领域的性能，本节将该方法与其他 5 种基准模型分别应用于 CASIE[1] 和 TCEDCL[38] 这 2 个网络安全数据集上，通过设置 4 种不同的训练样本（1-shot、5-shot、10-shot、15-shot），验证本文方法在网络安全领域小样本事件检测任务上的优越性。其中，对只有一个训练样本（1-shot）情况下模型的分类性能和泛化能力进行评估，观察模型在有限信息下能否快速学习新的事件类别，以验证本文方法在小样本事件检测方面的有效性。

3.1　数据集

数据集 CASIE 是由马里兰大学的 Satyapanich 等[1]研究者构建的，其标注了 1 000 篇网络安全事件的新闻报道，包含了 Attack.Databreach、Attack.Phishing、Attack.Ransom、Discover.Vulnerability 和 Patch. Vulnerability 这 5 种事件类型。数据集 TCEDCL 是本文研究团队之前工作构建的数据集[38]，收集了来自各网络安全情报的新闻、博客等的 1 000 条包含事件信息的句子，人工标注了 9 种事件类型，分别是 DataBreach（158 个）、Phishing（128 个）、Ransom（128 个）、DDoSAttack（46 个）、Malware（101 个）、SupplyChain（27 个）、VulnerabilityExploitation（83 个）、VulnerabilityDiscover（192 个）和 VulnerabilityPatch（137 个），其数量分布如图 3 所示。

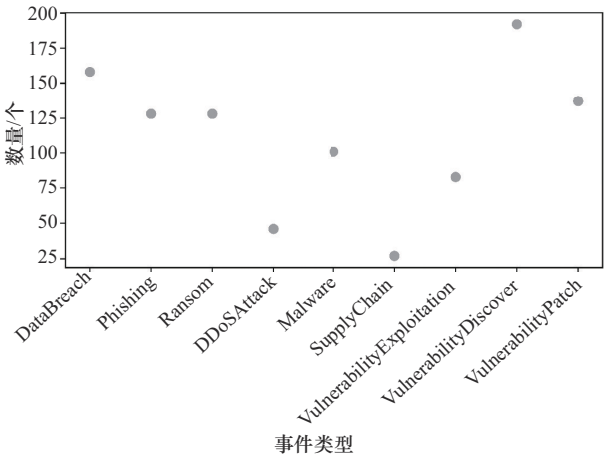


图 3　数据集 TCEDCL 包含的事件类型数量分布

对于小样本学习，本文采用了 N-way K-shot 进行 1 个、5 个、10 个和 15 个样本的实验，对每个数据集设置了 4 组，即 5-way 1-shot、5-way 5-shot、5-way 10-shot、5-way 15-shot。对于这样的 N-way K-shot 实验，本文需要从原始训练集中采样 N 类，每个类的 K 个实例组成小样本训练集。在此基础上，测试过程会从测试集中抽取部分实例参与预测，并计算预测结果和真实标签之间的准确率。

3.2　实验设置

在评估结果时，本文采用精确率 P、召回率 R 和 F_1 分数 3 个指标进行评价，其中，F_1 分数是最重要的度量指标。

$$P=\frac{\text{正确识别的事件个数}}{\text{识别出的事件总数}}\times 100\% \tag{10}$$

$$R=\frac{\text{正确识别的事件个数}}{\text{正确的事件总数}}\times 100\% \tag{11}$$

$$F_1=\frac{2PR}{P+R} \tag{12}$$

本文方法采用了 bert-base-uncased 作为基础编码器，它对最大长度为 256 的输入句子进行处理。使用 1×10^{-5} 学习率训练，在训练集上迭代训练本文方法，根据不同的训练数据集设置迭代次数，批大小为 32，并在测试集上评估其性能，得到最后的结果。训练和评估使用 PyTorch V 2.0.1，在内存为 24 GB 的 RTX 4090 GPU 上运行所有实验。同时为了在平衡性能和资源的情况下最小化数据生成所带来的噪声影响，并使事件检测结果达到最佳，需要优化增强数据实例数量的取值。因此在增强数据实例方法中，将从[5,10,20,30,…,90,100]个样本中选择，逐步调整得到最佳数值。

3.3 基准模型

为了验证本文方法的有效性，使用以下5个具有代表性的模型作为基准。

1) Proto[39]是一种原型网络的小样本学习方法，该方法通过神经网络学习到的表示空间中的示例来表示每个类。

2) Proto+Dot[27]是一种使用点积法计算相似度的Proto法，该方法通过计算向量间的点积来量化相互间的相似程度。

3) PA-CRF[10]利用一个原型化的摊销条件随机场（CRF）来生成转换分数，以实现基于标签原型的新事件类型的适应能力，统一解决小样本事件检测。

4) HCL-TAT[29]是一种基于任务自适应阈值的混合对比学习方法，该方法设计了一个易于自适应的阈值来缓解触发词的误识别。

5) CyberLSTM[1]使用带有CRF损失的双向LSTM网络对网络安全事件类型进行分类，其采用注意力的深度神经网络方法，以融合丰富的语言特征和词嵌入。

3.4 实验结果分析

为了探究本文方法在小样本场景下的表现，本节进行了如下的小样本模拟实验。对于每种事件类型，分别从训练集中随机抽取{1, 5, 10, 15}个样本组成小样本训练集来训练模型，再统计模型在测试集上的精确率P、召回率R和F_1分数。表2给出了2个数据集在4种不同设置上不同模型的实验结果对比，其中，粗体表示最佳结果。从表2中可以看出，在2个数据集的4种不同设置中，本文方法相较于其他基线模型实现了最佳性能，在不同的数据集和任务上都是有效的。通过观察实验结果得出如下结论。

1) 本文方法相较基线模型在精确率P指标上最优，并获得了75.31%的F_1分数值。

2) Proto原型网络的模型结构均将输入样本映射到度量空间，使用传统方式度量查询样本与支持集之间的距离，从而进行分类。而本文方法使用深度神经网络来学习度量空间中样本的相关性，处理文本分类任务时能够更好地理解和表示文本，其精确率P高达79.16%，相较于原型网络得到进一步提升，因此验证了本文方法更有利于事件检测分类任务。

3) 研究结果显示，本文方法在所有设置下性能均有不同程度的提升。具体而言，相较于HCL-TAT方法，在{1, 5, 10, 15}个样本下，本文方法的F_1分数实现了34.09%、4.71%、4.35%、4.36%的显著提升。这些结果证明了本文方法在学习更好的特征表示和实现更准确的分类结果方面的有效性。

4)本文方法在1-shot设置的情况下取得了最优

表2 2个数据集在4种不同设置上不同模型的实验结果对比

模型	TCEDCL数据集											
	5-way1-shot			5-way5-shot			5-way10-shot			5-way15-shot		
	P	R	F_1	P	R	F_1	P	R	F_1	P	R	F_1
Proto	16.32%	42.83%	23.63%	18.12%	**80.30%**	29.56%	24.12%	83.57%	37.43%	34.03%	86.80%	48.89%
Proto+Dot	5.98%	42.43%	10.48%	23.32%	79.78%	36.09%	34.03%	**84.76%**	48.56%	38.07%	84.18%	52.43%
PA-CRF	1.26%	35.27%	2.43%	38.44%	74.14%	50.63%	44.32%	75.80%	55.94%	70.20%	77.84%	73.82%
HCL-TAT	2.08%	**51.06**%	3.99%	41.90%	76.58%	54.16%	53.88%	81.87%	64.99%	58.86%	**87.20%**	70.28%
CyberLSTM	22.22%	8.00%	11.76%	50.00%	8.00%	13.79%	66.67%	16.00%	25.81%	62.50%	20.00%	30.30%
本文	**41.85**%	34.94%	**38.08**%	**74.48%**	47.78%	**58.20%**	**78.27%**	69.72%	**73.75%**	**79.16%**	71.81%	**75.31%**

模型	CASIE数据集											
	5-way1-shot			5-way5-shot			5-way10-shot			5-way15-shot		
	P	R	F_1	P	R	F_1	P	R	F_1	P	R	F_1
Proto	7.19%	13.55%	9.39%	7.49%	**47.76%**	12.95%	9.70%	**58.40%**	16.63%	11.95%	**71.38%**	20.47%
Proto+Dot	3.00%	13.43%	4.91%	20.36%	24.30%	22.16%	22.10%	28.72%	24.97%	21.35%	32.97%	25.92%
PA-CRF	2.62%	13.85%	4.40%	20.24%	26.80%	23.07%	20.28%	35.54%	25.83%	27.93%	25.47%	26.64%
HCL-TAT	11.69%	18.76%	14.40%	32.58%	33.27%	32.93%	38.69%	45.67%	41.90%	36.18%	65.41%	46.59%
CyberLSTM	**28.57%**	4.00%	7.02%	33.33%	6.00%	10.17%	42.85%	12.00%	18.75%	50.00%	12.00%	19.35%
本文	24.50%	**23.61%**	**24.05%**	**37.77%**	42.00%	**39.77%**	**42.86%**	46.01%	**44.38%**	**54.40%**	43.45%	**48.32%**

的性能，其F_1分数为38.08%。该方法证明了相较于其他基线方法，本文方法能够更充分地利用有限的数据资源。

5) 本文方法在TCEDCL数据集的改进比在CASIE数据集更大。TCEDCL数据集包含的类型和样本选择更全面，而CASIE数据集选择目标句及其前后句进行标注，包含的事件论元信息更全面，但导致事件无关的信息增多，降低了模型的准确性。

3.5　消融实验

为了验证本文方法中各个模块的有效性，本文在TCEDCL数据集上进行了消融研究，该数据集采用了小样本事件（5-way）的设置。首先探索了数据增强模块在事件检测任务的有效性；其次分析了标签词对事件检测方法的影响；最后探究了不同提示模板和增强实例数量对本文方法性能的影响。

1) 数据增强模块的影响

为了研究数据增强模块是否有助于性能提高，本文将该模块去除，对比结果如表3所示。当网络安全事件检测方法无数据增强模块时，整体F_1分数下降8.10%~18.91%。这一结果充分说明了从未标注的数据中生成新的训练数据，虽然可能会引入不正确的数据产生噪声，但是在一定程度上能丰富学习语料，更容易使事件检测方法充分利用有限的标注数据，捕获更加全面的特征表示，从而提高整个模型的性能。

表3　数据增强模块和标签词模块的影响

方法	1-shot	5-shot	10-shot	15-shot
移除数据增强模块	33.75%	48.96%	54.84%	59.31%
移除标签词模块	39.24%	50.16%	58.52%	62.70%

2) 标签词模块的影响

为了证明本文方法中数据增强模块标签词的有效性，本文探究了标签词模块的特征表达性能，实验结果如表3所示。在无标签词的情况下，未标注的事件无法借助标签词映射到已定义好的类别中，这可能会对本文方法的数据增强模块的性能产生一定影响，生成更多带有噪声的新训练数据。因而其方法可能会学习到这些错误实例信息，增加预测出错误结果的概率。

3) 不同提示模板的影响

表4中给出本文方法3种提示模板设计的F_1分数，并从中进行选择。从结果观察到，在4种不同设置下，提示设计“This text is about<mask>”的F_1分数优于其他2种。这说明设计合适的提示会改善网络安全小样本事件检测的结果。

表4　不同提示模板的影响

模板	1-shot	5-shot	10-shot	15-shot
A<mask>event	35.56%	53.23%	69.92%	73.28%
A <mask> event happens in this text	37.34%	56.96%	71.68%	74.79%
This text is about <mask>	38.08%	58.20%	73.75%	75.31%

4) 增强实例数量的影响

图4显示了在训练集中添加不同数量的增强文本对F_1分数的影响。随着增强实例数量的增加，不同规模的训练集表现出不同程度的F_1分数提升。然而，在增强的性能达到最高点之后，继续增加实例数量反而会导致模型整体性能下降。通过结果分析发现，生成实例的过程中存在错误预测结果，这引入了一定量的噪声。若生成实例中的噪声过强，则会改变选择实例的结果，干扰训练数据的信息，导致模型学习到错误的模式，降低事件检测的质量，从而影响整个模型的性能。

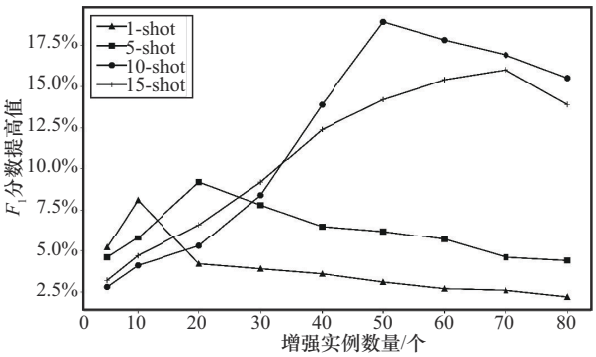


图4　在训练集中添加不同数量的增强文本对F_1分数的影响

3.6　实例分析

为了进一步阐明本文方法性能的优越性，本节给出具体案例的实验结果，比较不同模型对同一个句子的分类结果，如表5所示，其中，加粗字体表示该事件的触发词，√表示模型预测的结果与真实结果一致，×表示模型预测的结果与真实结果不一致。

举出4个实例进行说明，其中加粗单词表示事件触发词。

表5 不同模型检测实例的结果

序号	实例事件类型	模型预测结果				
		Proto+Dot	PA-CRF	HCL-TAT	Cyber-LSTM	本文方法
S_1	DataBreach	×	×	√	×	√
S_2	Ransom	√	√	√	×	√
S_3	Phishing	×	×	×	×	√
S_4	Malware	×	×	×	×	×

实例 S_1: Customer information from more than 133 000 users was **compromised** in the incident.

实例 S_2: The hackers, who identified themselves as ‘Turkish Crime Family’, **demanded** $75 000 in Bitcoin or Ethereum, another increasingly popular crypto-currency, or $100 000 worth of iTunes gift cards in exchange for deleting the alleged cache of data.

实例 S_3: Other messages posed as a rejected financial transaction, failed deposit/refund or canceled order alerts in order to **trick** the users into opening them.

实例 S_4: Last time, in March 2016, the district had to cancel school for a day to allow technology staff time to recover from the malware, which **infected** some of the district’s servers and many of its more than 600 computers.

针对实例 S_1，Proto+Dot、PA-CRF 和 CyberLSTM 模型未能很好地训练样本事件与标签之间的相互关系，错误地将“DataBreach”识别为其他事件类型；而另外 2 种模型则准确识别出相应实体。

针对实例 S_2，只有 CyberLSTM 模型未能识别出事件类型“Ransom”，此结果可能由于训练集样本较少，模型难以捕捉句义信息。CyberLSTM 模型需要大量的训练数据支持，在样本数量较少的情况下，其性能会存在不足。

针对实例 S_3，4 种基线模型均给出了错误的对应事件类型，而本文模型采用提示问答方法充分利用有限标注数据，并借助标签词的丰富语义信息，生成更多样化的训练样本，因此可正确识别出该事件并分类。

针对实例 S_4，5 种模型均未能准确识别出对应的事件类型，由此也说明本文模型在小样本事件检测中还有待改进和优化。

4 结束语

本文提出了一种利用提示问答方法进行低资源网络安全事件检测任务的方法，同时借助标签词的丰富语义信息，从未标注的文本中生成新的训练数据，为网络安全事件检测产生更多实例的表示；然后利用增强的训练数据，设计合适的提示问题，充分利用有限的标注数据，减少模型对人工标注数据的需求；最后将本文方法在 TCEDCL 和 CASIE 这 2 个数据集上进行了大量实验。实验结果表明，与其他基准模型相比，本文方法在小样本场景下性能达到了先进的水平。该方法在当前网络威胁日益严峻的背景下能对事件进行有效识别分类，同时不需要大规模高质量的标注数据，适用于小样本应用场景。在未来的工作中，笔者将尝试将该模型应用于篇章级文本的事件检测，以增强句子间的关联，减少句子级事件检测中信息的重复和冗余，同时避免句子间因缺少关联导致的事件漏检。

参考文献：

[1] SATYAPANICH T, FERRARO F, FININ T. CASIE: extracting cybersecurity event information from text[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(5): 8749-8757.

[2] 李涛, 郭渊博, 琚安康. 融合对抗主动学习的网络安全知识三元组抽取[J]. 通信学报, 2020, 41(10): 80-91.
LI T, GUO Y B, JU A K. Knowledge triple extraction in cybersecurity with adversarial active learning[J]. Journal on Communications, 2020, 41(10): 80-91.

[3] NGUYEN T H, GRISHMAN R. Event detection and domain adaptation with convolutional neural networks[C]//Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers). Stroudsburg: Association for Computational Linguistics, 2015: 365-371.

[4] LIN H Y, LU Y J, HAN X P, et al. Nugget proposal networks for Chinese event detection[C]//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Stroudsburg: Association for Computational Linguistics, 2018: 1565-1574.

[5] LU Y J, LIN H Y, XU J, et al. Text2Event: controllable sequence-to-structure generation for end-to-end event extraction[C]//Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). Stroudsburg: Association for Computational Linguistics, 2021: 2795-2806.

[6] MA Y B, WANG Z H, CAO Y X, et al. Few-shot event detection: an empirical study and a unified view[C]//Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1:

Long Papers). Stroudsburg: Association for Computational Linguistics, 2023: 11211-11236.

[7] BAYER M, FREY T, REUTER C. Multi-level fine-tuning, data augmentation, and few-shot learning for specialized cyber threat intelligence[J]. Computers & Security, 2023, 134: 103430.

[8] CHEN J W, LIN H Y, HAN X P, et al. Honey or poison? solving the trigger curse in few-shot event detection via causal intervention[C]//Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. Stroudsburg: Association for Computational Linguistics, 2021: 8078-8088.

[9] DENG S M, ZHANG N Y, LI L Q, et al. OntoED: low-resource event detection with ontology embedding[C]//Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). Stroudsburg: Association for Computational Linguistics, 2021: 2828-2839.

[10] CONG X, CUI S Y, YU B W, et al. Few-shot event detection with prototypical amortized conditional random field[C]//Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing(Voluml: Long Papers). Stroudsburg: Association for Computational Linguistics, 2021: 28-40.

[11] XIA M Z, KONG X, ANASTASOPOULOS A, et al. Generalized data augmentation for low-resource translation[C]//Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Stroudsburg: Association for Computational Linguistics, 2019: 5786-5796.

[12] HOU Y, LIU Y, CHE W, et al. Sequence-to-sequence data augmentation for dialogue language understanding[C]//Proceedings of the 27th International Conference on Computational Linguistics. Stroudsburg: Association for Computational Linguistics. 2018: 1234-1245.

[13] WU X, LV S W, ZANG L J, et al. Conditional BERT contextual augmentation[C]//International Conference on Computational Science. Berlin: Springer, 2019: 84-95.

[14] WEI J, ZOU K. EDA: easy data augmentation techniques for boosting performance on text classification tasks[C]//Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing. Stroudsburg: Association for Computational Linguistics, 2019: 6382-6388.

[15] LIU J, CHEN Y F, XU J N. Low-resource NER by data augmentation with prompting[C]//Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence. California: International Joint Conferences on Artificial Intelligence Organization, 2022: 4252-4258.

[16] BRUIJN J A D, MOEL H D, WEERTS A H, et al. Improving the classification of flood tweets with contextual hydrological information in a multimodal neural network[J]. Computers & Geosciences, 2020, 140: 104485.

[17] HOSSEINI P, HOSSEINI P, BRONIATOWSKI D. Content analysis of Persian/Farsi Tweets during COVID-19 pandemic in Iran using NLP [C]//Proceedings of the 1st Workshop on NLP for COVID-19 (Part 2) at EMNLP 2020. Stroudsburg: Association for Computational Linguistics, 2020: 1-16.

[18] LAMB A, PAUL M J, DREDZE M. Separating fact from fear: tracking flu infections on twitter[C]//Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Stroudsburg: Association for Computational Linguistics,2013: 789-795.

[19] YAGCIOGLU S, SEYFIOGLU M S, CITAMAK B, et al. Detecting cybersecurity events from noisy short text[C]//Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Lingusistics. Stroudsburg: Association for Computational Linguistics, 2019: 1366-1372.

[20] QIU X Y, LIN X X, QIU L K. Feature representation models for cyber attack event extraction[C]//Proceedings of the 2016 IEEE/WIC/ACM International Conference on Web Intelligence Workshops. Piscataway: IEEE Press, 2016: 29-32.

[21] RITTER A, WRIGHT E, CASEY W, et al. Weakly supervised extraction of computer security events from twitter[C]//Proceedings of the 24th International Conference on World Wide Web. Piscataway: IEEE Press, 2015: 896-905.

[22] LUO N, DU X Y, HE Y T, et al. A framework for document-level cybersecurity event extraction from open source data[C]//Proceedings of the 24th International Conference on Computer Supported Cooperative Work in Design. Piscataway: IEEE Press, 2021: 422-427.

[23] CHEN Y B, XU L H, LIU K, et al. Event extraction via dynamic multi-pooling convolutional neural networks[C]//Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). Stroudsburg: Association for Computational Linguistics, 2015: 167-176.

[24] NGUYEN T H, CHO K, GRISHMAN R. Joint event extraction via recurrent neural networks[C]//Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Stroudsburg: Association for Computational Linguistics, 2016: 300-309.

[25] NGUYEN T, GRISHMAN R. Graph convolutional networks with argument-aware pooling for event detection[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2018, 32(1): 5900-5907.

[26] PENG H R, SONG Y Q, ROTH D. Event detection and co-reference with minimal supervision[C]//Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. Stroudsburg: Association for Computational Linguistics, 2016: 392-402.

[27] LAI V D, DERNONCOURT F, NGUYEN T H. Extensively matching for few-shot learning event detection[J]. arXiv Preprint, arXiv: 2006.10093, 2020.

[28] DENG S M, ZHANG N Y, KANG J J, et al. Meta-learning with dynamic-memory-based prototypical network for few-shot event detection[C]//Proceedings of the 13th International Conference on Web Search and Data Mining. New York: ACM Press, 2020: 151-159.

[29] ZHANG R H, WEI W, MAO X L, et al. HCL-TAT: a hybrid contrastive learning method for few-shot event detection with task-adaptive threshold[C]//Proceedings of the Findings of the Association for Computational Linguistics. Stroudsburg: Association for Computational Linguistics, 2022: 1808-1819.

[30] DU X Y, CARDIE C. Event extraction by answering (almost) natural questions[C]//Proceedings of the 2020 Conference on Empirical Meth-

ods in Natural Language Processing. Stroudsburg: Association for Computational Linguistics, 2020: 671-683.
[31] LIU J, CHEN Y F, XU J N. Document-level event argument linking as machine reading comprehension[J]. Neurocomputing, 2022, 488: 414-423.
[32] GAO J, ZHAO H, YU C, et al. Exploring the feasibility of ChatGPT for event extraction[J]. arXiv Preprint, arXiv: 2303.03836, 2023.
[33] GAO T Y, YAO X C, CHEN D Q. SimCSE: simple contrastive learning of sentence embeddings[C]//Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. Stroudsburg: Association for Computational Linguistics, 2021: 6894-6910.
[34] XIE Q Z, DAI Z H, HOVY E, et al. Unsupervised data augmentation for consistency training[J]. Advances in Neural Information Processing Systems, 2020, 33: 6256-6268.
[35] HUANG G H, ZHONG J, WANG C, et al. Prompt-based self-training framework for few-shot named entity recognition[C]//International Conference on Knowledge Science, Engineering and Management. Berlin: Springer, 2022: 91-103.
[36] CUI G Q, HU S D, DING N, et al. Prototypical verbalizer for prompt-based few-shot tuning[C]//Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Stroudsburg: Association for Computational Linguistics, 2022: 7014-7024.
[37] HU S D, DING N, WANG H D, et al. Knowledgeable prompt-tuning: incorporating knowledge into prompt verbalizer for text classification [C]//Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Stroudsburg: Association for Computational Linguistics, 2022: 2225-2240.
[38] TANG M M, GUO Y B, BAI Q C, et al. Trigger-free cybersecurity event detection based on contrastive learning[J]. The Journal of Supercomputing, 2023, 79(18): 20984-21007.
[39] SNELL J, SWERSKY K, ZEMEL R S. Prototypical networks for few-shot learning[J]. arXiv Preprint, arXiv: 1703.05175, 2017.

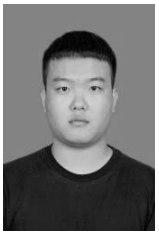
[作者简介]

汤萌萌（1989– ），女，河南信阳人，信息工程大学博士生，主要研究方向为信息安全、网络安全事件抽取。

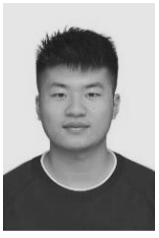
郭渊博（1975– ），男，陕西周至人，博士，海南大学教授、博士生导师，主要研究方向为网络防御、数据挖掘、机器学习和人工智能安全等。

张晗（1985– ），女，河南项城人，郑州大学讲师，主要研究方向为自然语言处理、信息安全。

白庆春（1990– ），女，上海人，上海开放大学助理研究员，主要研究方向为自然语言处理、机器学习、教育大数据分析。



陈庆礼（1998– ），男，河南新乡人，信息工程大学硕士生，主要研究方向为人工智能安全。



张博闻（1998– ），男，河南郑州人，郑州浪潮数据技术有限公司工程师，主要研究方向为计算机通信、网络安全等。