融合对抗主动学习的网络安全知识三元组抽取

李涛，郭渊博，琚安康
（信息工程大学密码工程学院，河南 郑州 450001）

摘 要：针对当前网络安全领域知识获取中所依赖的流水线模式存在实体识别错误的传播，未考虑实体识别与关系抽取任务间的联系，以及模型训练缺乏标签语料的问题，提出一种融合对抗主动学习的端到端网络安全知识三元组抽取方法。首先，将实体识别与关系抽取通过联合标注策略建模为序列标注任务；然后，设计融合动态注意力机制的 BiLSTM-LSTM 模型实现实体与关系的联合抽取，并形成三元组；最后，基于对抗网络训练一个判别器模型，增量地筛选出高质量的待标注数据进行标注，并通过迭代训练不断提升联合抽取模型的性能。通过实验表明，所提方案中实体–关系联合抽取模型优于现有的网络安全知识抽取方案，并验证了对抗主动学习方法的有效性。

关键词：知识三元组；网络安全；联合抽取；对抗网络；主动学习
中图分类号：TP391
文献标识码：A
doi: 10.11959/j.issn.1000−436x.2020174

Knowledge triple extraction in cybersecurity with adversarial active learning

LI Tao, GUO Yuanbo, JU Ankang
Department of Cryptogram Engineering, Information Engineering University, Zhengzhou 450001, China

Abstract: Aiming at the problem that using pipeline methods for extracting cybersecurity knowledge triples may cause the errors propagation of entity recognition and did not consider the correlation between entity recognition and relation extraction, and training triple extraction model lacked labeled corpora, an end-to-end cybersecurity knowledge triple extraction method with adversarial active learning was proposed. For knowledge triple extraction, the conventional entity recognition and relation extraction were modelled as sequence labeling task through joint labeling strategy firstly. And then, a BiLSTM-LSTM-based model with dynamic attention mechanism was designed to jointly extract entities and relations, forming triples. Finally, with adversarial learning framework, a discriminator was trained to incrementally select high-quality samples for labeling, and the performance of the joint extraction model was continuously enhanced by iterative retraining. Experiments show that the proposed joint extraction model outperforms the existing cybersecurity knowledge triple extraction methods, and demonstrate the effectiveness of proposed adversarial active learning scheme.

Key words: knowledge triple, cybersecurity, joint extraction, adversarial network, active learning

1 引言

对威胁信息的持续跟踪与分析，已成为增强网络安全防护的一项重要举措。以威胁情报为主的网络威胁信息通常以网络安全文本的形式披露，包括各类网络安全社区发布的博客、白皮书；软件厂商发布的安全公告等。然而，近年来层出不穷的网络安全事件导致网络威胁情报海量增长。由于非结构化形式的网络威胁情报不能被机器所理解，继续依靠人工方式从文本形式的威胁情报中识别、提取诸如新型漏洞信息、漏洞利用方式、网络攻击工具以及攻击模式等关键威胁信息，并进行关联分析已无

收稿日期：2020–06–04；**修回日期**：2020–07–23
基金项目：国家自然科学基金资助项目（No.61501515）
Foundation Item: The National Natural Science Foundation of China (No.61501515)

法满足网络安全防御的现实需求。为此，利用信息抽取技术，从网络安全文本中自动地抽取安全相关的实体、概念以及关系，将非结构化的数据转换成易于共享和集成的结构化表达，形成网络安全链接数据[1]，并构建网络安全知识图谱，赋予机器认知智能以实现网络安全文本的挖掘与智能化分析，将在网络安全主动防御体系的构建中发挥重要作用。

信息抽取作为文本挖掘的关键技术，已被广泛应用于摘要生成、自动问答、知识图谱等领域[2]。信息抽取可细分为命名实体识别、实体关系抽取和事件抽取 3 个子任务，其中通过实体识别和实体关系抽取来获取语义三元组，是构建知识图谱、理解自然语言的重要前提。由于网络安全领域所关注的实体包括攻击者、攻击方式、漏洞名、资产等特定类别，且关系抽取所针对的是此类与网络威胁相关的特定实体间的语义表达，而现有的信息抽取系统无法适用于网络安全领域实体与关系的抽取。为满足应用的需求，需开发面向网络安全领域的知识抽取系统。

相较于在通用领域以及金融、法律、生物医学等领域的成功应用，面向网络安全领域的知识抽取研究才刚起步。2018 年，第十二届国际语义评测比赛中的任务 8 旨在运用自然语言处理技术实现网络安全文本的语义信息抽取[3]，其中包含了针对恶意软件相关文本中实体、关系及其属性进行标签预测的子任务。当前面向网络安全领域的知识抽取研究是通过流水线模式进行的，即首先通过命名实体识别来获取网络安全相关的实体[4]，在此基础上再根据预定义的实体关系类别进行候选实体对间的关系预测[5]，进而得到网络安全知识三元组。尽管流水线框架具有集成不同数据源和学习算法的灵活性，但也存在一定的问题[6]：1) 关系抽取依赖实体识别的结果，而实体识别阶段产生的错误将传播到关系预测阶段，影响关系抽取效果；2) 将实体识别与关系抽取分开执行，无法充分利用 2 个任务间的语义联系；3) 先识别实体，再进行关系预测，导致流水线框架下信息抽取效率较低。

不同于流水线框架，实体与关系的联合抽取旨在对 2 个任务同时建模，当前实体与关系的联合抽取得到研究者的广泛关注[7]。早期关于联合抽取的研究依赖复杂的特征工程以及自然语言处理工具，随着深度神经网络的广泛应用，研究人员提出端到端的实体−关系联合抽取模式。现有的实体关系联合学习方法包括基于参数共享的方法和基于序列标注的方法。前者对实体识别和关系抽取任务通过共享编码层模型进行联合学习，其在训练时共享部分参数，此方法本质上仍将 2 个任务分开执行，会产生冗余信息；后者将实体与关系联合抽取任务转换成序列标注问题，基于实体−关系的联合标注策略进行建模，直接得到实体−关系三元组[2]。Zheng 等[8]首次提出基于序列标注的实体与关系联合抽取方法，并在通用领域的知识抽取中取得了较好的效果。但由于其假设一个实体只有一个关系标签，无法适用于存在重叠关系的领域文本。为解决面向生物医学文本实体与关系联合抽取中的重叠关系问题，曹明宇等[7]改进 Zheng 等[8]提出的联合标注模式，在药物−药物关系抽取中取得了较好的效果。通过类比生物医学文本发现，在网络安全文本中同一个实体也可能参与多个语义关系，因此面向网络安全领域的实体与关系联合抽取中也存在重叠关系问题。

尽管端到端的神经网络模型在诸多任务中性能突出，但其在实际应用中依赖大规模的标签数据。相较于通用领域大量可获取的标注语料，网络安全领域的标注语料极其缺乏，导致同一模型应用于网络安全领域的实体识别与关系抽取任务时效果不佳。而不需要标签数据的无监督学习方法性能通常弱于监督学习。为此，面向网络安全领域的语料标注仍然是提升实体识别与关系抽取性能的一项关键任务。然而针对网络安全文本的标注通常存在两方面的问题：1) 需要网络安全领域的专家或具备一定网络安全知识的从业人员才能完成对网络安全文本的标注；2) 相较于通用领域的文本语料，网络安全文本中含有更多的对象实例，因此需要投入更多的人工成本去标注。为减轻人工标注数据的负担，主动学习算法能够从未标注数据池中增量地采样出富有信息的样本，由专家进行标注后补充到标签数据集中，并通过迭代训练提升模型学习的性能。然而，尽管现有的主动学习算法在数据分类任务中性能良好，但此类采样策略应用于具有丰富标签空间的序列标注任务时将变得极其复杂。

为解决网络安全领域知识抽取中存在的上述问题，本文提出一种融合对抗主动学习的实体与关系联合抽取方案。基于联合标注策略将实体识别与关系抽取任务转化为序列标注问题，并通过对抗学

习机制训练一个判别器模型来筛选出富有信息量的样本进行人工标注，实现以较低的数据标注代价完成联合模型的训练。本文的主要贡献包含 3 个方面。

1) 不同于流水线模式的网络安全实体识别与关系抽取，本文将 2 个子任务联合起来建模为序列标注，提出一种基于端到端的网络安全实体与关系联合抽取框架。

2) 面向网络安全文本知识抽取，基于长短时记忆（LSTM, long short-term memory neural）网络和双向长短时记忆（BiLSTM, bidirectional LSTM）网络，提出一种融合动态注意力机制的 BiLSTM-LSTM 序列标注模型。

3) 针对网络安全领域标注语料缺乏的问题，基于主动学习思想，并融合对抗学习机制，提出一种对抗主动学习框架下的待标注语料采样方法。

2 相关工作

随着网络威胁的激增，详细的威胁内容以非结构化的自然语言文本形式存在，诸如安全报告、白皮书、博客、公告等。而针对此类威胁信息的分析与集成对于安全人员来说是烦琐且复杂的工作。因此，对威胁信息的自动化提取是亟待解决的问题。Liao[9]开发了一套 iACE 系统，用于自动地从威胁情报文本中提取威胁失陷指标（IoC, indicator of compromise）及其上下文关系。Panwar[10]基于 IoC 的提取框架，可以从 Cuckoo 沙箱结果中生成结构化威胁信息表达（STIX, structured threat information expression）格式的 IoC。Gasmi 等[11]将自然语言处理领域中的命名实体识别方法非结构化安全信息的抽取中，结合 LSTM 模型和条件随机场（CRF, conditional random field），提出一种基于 LSTM-CRF 的模型，对安全领域文档中相关实体，如产品、版本以及攻击名称等进行识别。Chambers 等[12]基于自然语言处理（NLP, natural language processing）的思想，通过训练前馈神经网络和文档主题生成（LDA, latent Dirichlet allocation）模型，从社交媒体数据中提取表征攻击行为的实体，进而实现分布式拒绝服务（DDoS, distributed denial of service）攻击的检测。Zhou 等[13]和 Long 等[14]运用端到端的神经网络并结合注意力机制，针对威胁情报语料建立模型，训练得到 IoC 提取器，在实际的 IoC 抽取效果上表现出较高的准确率。由于对威胁情报的利用不仅限于 IoC，威胁情报报告中提供了更多有关网络攻击的详细信息，尤其是有关攻击者、攻击技术、攻击工具等的语义信息。秦娅等[15]在对威胁情报语料分析的基础上，利用卷积神经网络（CNN, convolutional neural network）获取语料字符嵌入特征，提出一种融合特征模板的 CNN-BiLSTM-CRF 的网络安全实体识别方法，在对网络安全文本数据涉及的人名、地名、组织名、软件名、网络相关术语以及漏洞编号的识别上取得了不错的效果。张若彬等[16]针对安全漏洞领域的命名实体，提出一种基于 BLSTM-CRF 的识别模型，并结合领域词典对识别结果进行校正，实现对漏洞编号、漏洞名、漏洞类型、漏洞利用条件（软件供应商、操作系统、应用软件）、攻击方式共 7 类漏洞相关命名实体的有效识别。此外，Pingle 等[5]开发了一套基于深度学习的威胁情报语义关系抽取系统，从开源威胁情报中获取语义三元组，并与安全运营中心结合进一步提升网络安全防御能力。上述研究均属于流水线模式，而目前尚未出现面向网络安全领域的知识联合抽取研究。

主动学习算法旨在逐步选择用于标注的样本，从而以较低的标记成本实现模型较高的分类性能。当前主动学习领域的研究包括基于样本生成的主动学习和基于池的主动学习。基于样本生成的主动学习方法属于生成模型范畴，通过生成富有信息的样本，再由专家进行样本标记。Zhu 等[17]首次通过生成式对抗网络（GAN, generative adversarial network）来合成待标注样本，建立主动学习模型。但由于 GAN 模型存在训练困难以及模式崩坏的情况，生成的样本可能不满足真实样本的数据分布，当生成无意义的样本时，很难对其进行人工标注。因此，此类方法依赖于生成样本的质量和多样性。

基于池的主动学习是从数据池中筛选样本进行人工标注，当前基于池的主动学习算法是主动学习的主要研究领域，并已在图像分类、语音识别、文本分类以及信息检索等诸多实际场景中得到广泛应用。基于池的主动学习方法中具有代表性的采样策略包括基于不确实性的方法、基于集成的方法以及基于核心集的方法等。Culotta 等[18]利用最小置信度准则评估线性 CRF 模型在序列预测任务上的不确定性，实现主动学习算法在序列标注任务上的应用。Houlsby 等[19]提出了一种贝叶斯不一致主动学习算法，其中采样函数通过训练样本关于模型参

数的互信息来进行不确定性度量。Gal 等[20]通过揭露不确定性和正则化之间的关系来度量神经网络预测中的不确定性，并将其应用于主动学习。Sener 等[21]提出基于核心集的主动学习算法，该算法使采样数据点和训练模型的特征空间中未采样点间的欧几里得距离达到最小化。Kuo 等[22]提出一种基于集成的主动学习算法来表示不确定性，但该算法容易造成对样本的冗余采样。此外，Shen 等[23]将深度主动学习算法运用于命名实体识别任务中，并比较了最小置信度算法、贝叶斯非一致主动学习和最大归一化对数概率这 3 类采样策略的性能。

3　方法描述

3.1　模型架构

本节对所提模型进行详细描述，模型整体架构如图 1 所示。模型由 2 个模块组成：实体−关系联合抽取的序列标注模块和对抗主动学习模块，其中联合抽取模块包含表示层、编码层、动态注意力层、解码层。

对于三元组联合抽取模块，首先，在表示层利用 word2vec 基于所收集的网络安全文本训练得到词向量表，将输入序列映射成对应的词向量表示，此外，获取每个词所对应的字符特征向量，并将其与词向量进行拼接，组成模型的输入；然后，利用 BiLSTM 作为编码层，得到输入数据的特征编码，并结合动态注意力机制进一步捕获序列的上下文依存特征，将所得注意力向量输入 LSTM 解码层得到标签序列的向量表示；最后，根据 softmax 分类器的标签得分来输出文本的标签序列。在对抗主动学习模块，基于 BiLSTM 得到标注语句与未标注语句的特征向量，将其输入判别网络通过比较数据分布的相似性，筛选出需要标记的数据交由专家进行标记，并将标记后的数据加入标签训练集中，以此对联合抽取模型迭代进行训练。

3.2　标注策略及匹配规则

本节对所采用的标注策略进行详细阐述。Zheng 等[8]首次将实体与关系的联合抽取问题转换成序列标注任务，提出了实体与关系的联合标注策略。然而，由于其无法解决重叠关系问题，曹明宇

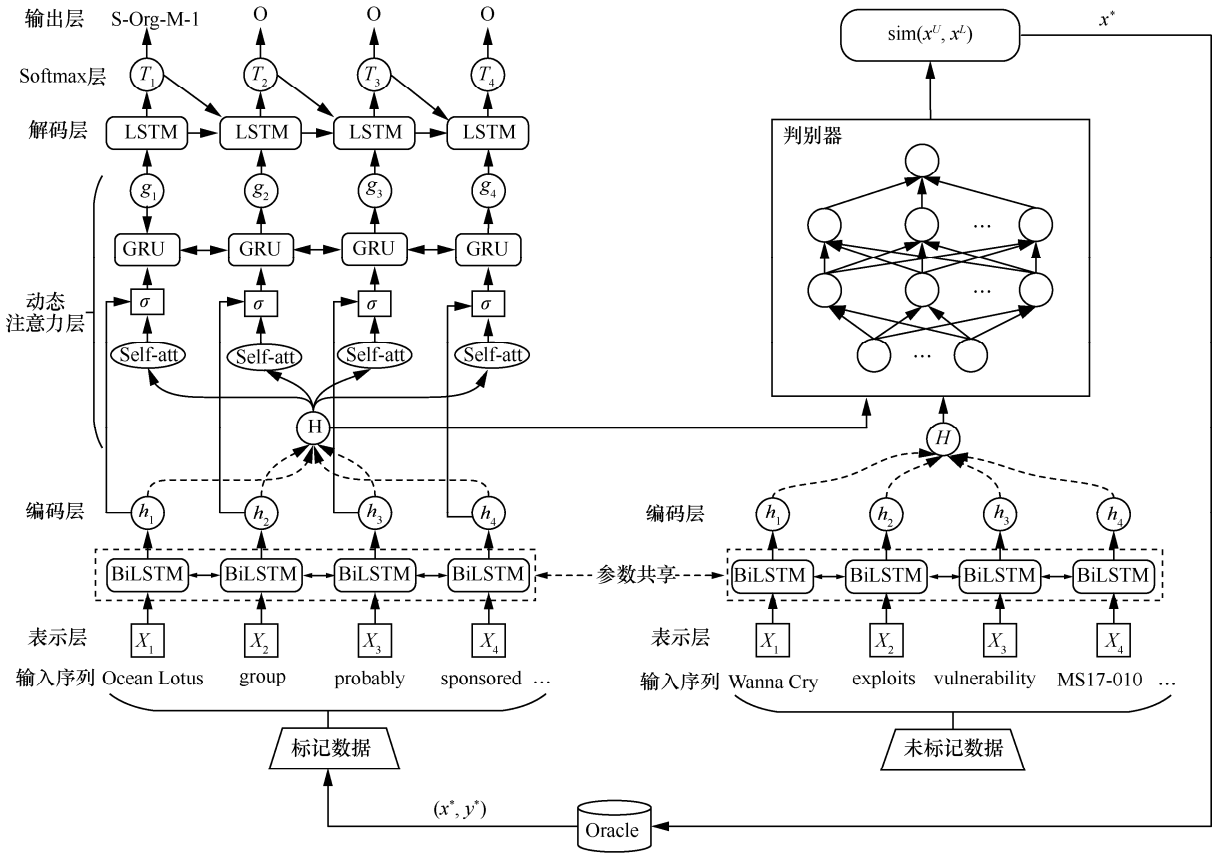


图 1　模型整体架构

等[7]在其基础上改进了联合标注策略，能够较好地解决部分情形下的重叠关系问题。在类比网络安全文本与生物医学文本的领域特性基础上，本文采用曹明宇等[7]提出的标注策略及三元组匹配规则，在实现网络安全实体−关系联合抽取的同时，解决联合抽取中部分情形下的重叠关系问题。

参照曹明宇等[7]的联合标注策略，本文在进行网络安全文本序列标注时，标签由实体边界、实体类别、关系类别、实体位置共 4 个部分组成，具体表示如下。

1) 实体边界。针对当前句子序列，采用“BIOES（B-begin，I-inside，O-outside，E-end，S-single）”模式来标识单词在实体中的位置信息，即 B 表示对应单词为实体的开始位置，I 表示对应单词为实体的内部，E 表示对应单词为实体的结束位置，S 表示对应单词为单个实体，O 表示对应单词为非实体。

2) 实体类别。分析网络安全语料并结合网络安全本体 UCO2.0[5]，给出 10 种实体类别：Organization，Location，Software，Malware，Indicator，Vulnerability，Course-of-action，Tool，Attack-pattern，Campaign。

3) 关系类别。分析网络安全语料并结合 UCO2.0，给出 9 种网络安全实体语义关系类别：comes-from，hasProduct，hasVulnerability，mitigates，uses，indicates，attributed-to，related-to，located-at。此外，增加一个 M 标签来表示当前单词所属的实体参与了多种不同类型的关系[7]。

4) 实体位置。实体位置由数字“1”“2”来标识，“1”表示该实体为三元组中的头实体，“2”表示该实体为三元组中的尾实体。

图 2 给出了该标注策略下的一个示例。输入序列中包含 2 个三元组：（OceanLotus，Comes-From，Vietnam）和（OceanLotus，Uses，watering hole attack），其中 Comes-From 和 Uses 为预定义的关系类别。根据上述标注策略，序列中每个单词都被赋予相应的标签，非实体用 *O* 来标注。如单词 OceanLotus 为单个词表示的实体，实体类别为 Organization，其同时参与了 Comes-From 和 Uses 2 种关系，此类情形属于重叠关系。此外，由于 OceanLotus 实体在两类关系中都处于头实体的位置，因此其标签为 S-Org-M-1。

在对网络安全文本输入序列完成标注的基础上，本文根据文献[7]中提出的匹配规则实现实体与关系的组合。首先根据标注结果中的实体边界和类别得到网络安全实体，进而再根据实体关系类别和实体位置形成知识三元组。对于实体关系的确定，则根据最邻近原则。当实体关系类别为非 M 时，若实体位置标识为“1”，则其向后查找与之距离最近、具有相同关系类别且实体位置标识为“2”的实体来组成三元组；若实体位置标识为“2”，则其向前查找实体位置标识为“1”的能与之匹配的实体。当实体关系类别标注为 M 时，该实体查找前后 2 个方向上能与之匹配的实体，来组成知识三元组。

3.3 表示层和 BiLSTM 特征编码层

本文利用 BiLSTM 模型实现对输入序列的特征编码，而在此之前需先在表示层将网络安全文本转化为低维、稠密的向量表示。为此，本文收集了大量网络安全文本语料，包括 AlienVault 威胁情报博客、welivesecurity 网络安全博客、思科安全威胁类博客、CVE 漏洞描述以及近年来的 APT 报告等，并利用 word2vec 训练得到 100 维的词向量表示，用于获取输入序列词级别的特征。此外，为进一步加强序列的输入特征表示，本文采用与文献[24]中相同结构的 CNN 模块来抽取所输入的网络安全句子序列的字符特征，最后将其与前述所得的词级别特征拼接后共同输入模型中。

在此基础上，利用 BiLSTM 模型实现对输入序列的特征抽取。假设当前时刻输入向量 $\boldsymbol{x}_t$，上一时刻所得隐藏状态为 $\boldsymbol{h}_{t-1}$，则在当前时刻根据上下文信息学习到的特征可简要表示为

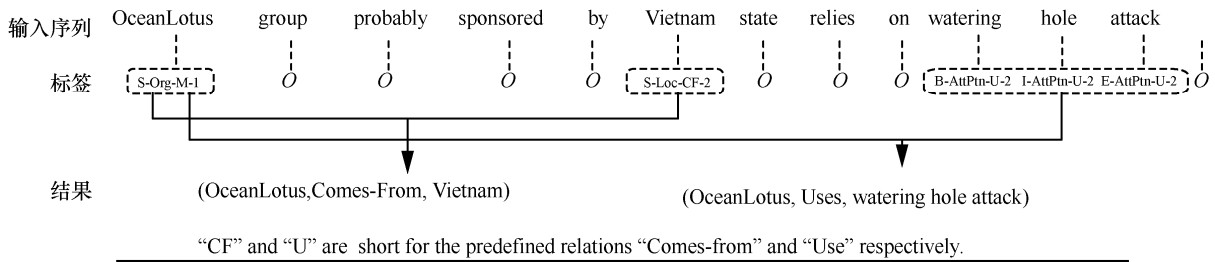


图 2 标注策略示例

$$h_t = \text{BiLSTM}(\boldsymbol{x}_t, \boldsymbol{h}_{t-1}) \tag{1}$$

进一步地，$\boldsymbol{H} = \{\boldsymbol{h}_1, \boldsymbol{h}_2, \cdots, \boldsymbol{h}_T\}$ 作为 BiLSTM 层的输出，表示对输入文本序列的特征编码。

3.4　动态注意力机制

自注意力机制能够直接学习句子中任意 2 个词之间的依存关系，并捕获句子的内部结构信息，其在机器翻译和语义角色标注中已得到成功应用。Cao 等[25]将自注意力机制用于中文命名实体识别任务中，使模型性能得到较大提升。具体地，文本序列经过 BiLSTM 层编码后，得到特征向量 $\boldsymbol{H} = \{\boldsymbol{h}_1, \boldsymbol{h}_2, \cdots, \boldsymbol{h}_T\}$，则自注意力机制计算过程为

$$\boldsymbol{Q}_i, \boldsymbol{K}_i, \boldsymbol{V}_i = \boldsymbol{H}\boldsymbol{W}_{\boldsymbol{Q}}^i, \boldsymbol{H}\boldsymbol{W}_{\boldsymbol{K}}^i, \boldsymbol{H}\boldsymbol{W}_{\boldsymbol{V}}^i \tag{2}$$

$$\boldsymbol{u}_i = \text{Attention}(\boldsymbol{Q}_i, \boldsymbol{K}_i, \boldsymbol{V}_i) = \text{softmax}\frac{\boldsymbol{Q}_i \boldsymbol{K}_i^{\mathrm{T}}}{\sqrt{d}}\boldsymbol{V}_i \tag{3}$$

其中，$\boldsymbol{Q}_i$ 对应 query 矩阵，$\boldsymbol{K}_i$ 对应 keys 矩阵，$\boldsymbol{V}_i$ 对应 value 矩阵，$\boldsymbol{W}_{\boldsymbol{Q}}^i$、$\boldsymbol{W}_{\boldsymbol{K}}^i$、$\boldsymbol{W}_{\boldsymbol{V}}^i$ 为通过训练得到的矩阵映射参数。经过 h 次的映射操作，并将计算结果进行拼接后，得到 $\boldsymbol{U} = \{\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_h\}$。再用 $\boldsymbol{W}_O$ 对矩阵 $\boldsymbol{U}$ 进一步进行映射变换

$$\boldsymbol{A} = (\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_h)\boldsymbol{W}_O \tag{4}$$

最后，得到自注意力层的输出结果为 $\boldsymbol{A} = \{\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_T\}$。

尽管上述注意力机制在序列建模过程中有助于捕获词之间的依存性，但对句子中的不同词而言，其注意力在对序列的上下文权重影响分配时保持不变，未考虑注意力分布的实际差异。本文借鉴文献[26]的思路，结合前述自注意力机制，提出一种动态注意力机制，以准确捕获词之间的相互影响。在动态注意力层，t 时刻，将 BiLSTM 层输出的特征编码 $\boldsymbol{h}_t$ 与自注意力机制对应输出 $\boldsymbol{a}_t$ 的拼接结果 $[\boldsymbol{h}_t, \boldsymbol{a}_t]$ 通过 sigmod 函数进行滤波得到 γ_t，接着进行点乘操作得到 $\boldsymbol{\varepsilon}_t$，并将其作为门控循环单元（GRU, gated recurrent unit）的输入，具体计算过程为

$$\gamma_t = \text{sigmod}(\boldsymbol{W}_s[\boldsymbol{h}_t, \boldsymbol{a}_t]) \tag{5}$$

$$\varepsilon_t = \gamma_t[\boldsymbol{h}_t, \boldsymbol{a}_t] \tag{6}$$

$$\boldsymbol{g}_t = \text{GRU}(\boldsymbol{g}_{t-1}, \boldsymbol{\varepsilon}_t, \boldsymbol{\theta}) \tag{7}$$

其中，$\boldsymbol{W}_s$ 和 $\boldsymbol{\theta}$ 为超参数矩阵。将 $\boldsymbol{G} = \{\boldsymbol{g}_1, \boldsymbol{g}_2, \cdots, \boldsymbol{g}_t\}$ 记作动态注意力层的最终输出结果，并将其输入下层进行解码。

3.5　LSTM 解码层

相比于采用 CRF 作为标签解码器，采用 LSTM 作为解码器能够显著加快模型训练速度，且能够达到与 CRF 相当的性能[23]。面对序列标注时存在的丰富的标签空间，本文参考 Zheng 等[8]使用的 LSTM 网络作为解码层以得到标签序列，解码层 LSTM 单元结构如图 3 所示。

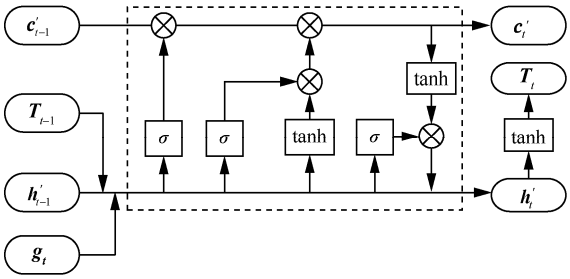


图 3　解码层 LSTM 单元结构

对于当前时刻单词 $\boldsymbol{x}_t$，使用 LSTM 解码获得其标签时，$\boldsymbol{g}_t$ 表示当前时刻从动态注意力层获得的向量表示，$\boldsymbol{T}_{t-1}$ 表示上一时刻词的预测标签向量，$\boldsymbol{h}'_{t-1}$ 表示上一时刻标签解码输出隐状态，$\boldsymbol{c}'_{t-1}$ 表示上一时刻解码所得细胞状态，则 LSTM 解码单元可形式化表示为

$$\boldsymbol{i}'_t = \sigma(\boldsymbol{W}'_{xi}\boldsymbol{g}_t + \boldsymbol{W}'_{hi}\boldsymbol{h}_{t-1} + \boldsymbol{W}_{ti}\boldsymbol{T}_{t-1} + \boldsymbol{b}'_i) \tag{8}$$

$$\boldsymbol{f}'_t = \sigma(\boldsymbol{W}'_{xf}\boldsymbol{g}_t + \boldsymbol{W}'_{hf}\boldsymbol{h}_{t-1} + \boldsymbol{W}_{tf}\boldsymbol{T}_{t-1} + \boldsymbol{b}'_f) \tag{9}$$

$$\boldsymbol{c}'_t = \boldsymbol{i}'_t \tanh(\boldsymbol{W}'_{xc}\boldsymbol{g}_t + \boldsymbol{W}'_{hc}\boldsymbol{h}_{t-1} + \boldsymbol{W}'_{tc}\boldsymbol{T}_{t-1} + \boldsymbol{b}'_c) + \boldsymbol{f}'_t\boldsymbol{c}'_{t-1} \tag{10}$$

$$\boldsymbol{o}'_t = \sigma(\boldsymbol{W}'_{xo}\boldsymbol{g}_t + \boldsymbol{W}'_{ho}\boldsymbol{h}'_{t-1} + \boldsymbol{W}'_{co}\boldsymbol{c}'_t + \boldsymbol{b}'_o) \tag{11}$$

$$\boldsymbol{h}'_t = \boldsymbol{o}'_t \tanh(\boldsymbol{c}'_t) \tag{12}$$

$$\boldsymbol{T}_t = \boldsymbol{W}_{ts}\boldsymbol{h}'_t + \boldsymbol{b}_t \tag{13}$$

其中，σ 和 tanh 表示非线性激活函数，$\boldsymbol{i}'_t$、$\boldsymbol{f}'_t$、$\boldsymbol{o}'_t$ 分别表示 LSTM 解码单元的输入门、遗忘门和输出门，$\boldsymbol{c}'_t$ 表示 t 时刻的细胞状态，$\boldsymbol{h}'_t$ 表示输出的隐状态向量，则 $\boldsymbol{T}_t$ 表示当前单词对应的预测标签的向量。

在获得预测标签向量的基础上，进一步运用 softmax 分类器计算标签概率，得到当前单词 x_t 对应第 i 个类别标签的概率为

$$p_t^i = p(i \mid x_t) = \frac{\exp(S_t^i)}{\sum_{j=1}^{N}\exp(S_t^j)} \tag{14}$$

其中，$\boldsymbol{S}_t = \boldsymbol{W}_s\boldsymbol{T}_t + \boldsymbol{b}_s$ 表示当前单词 $\boldsymbol{x}_t$ 在所有标签类别上的评分值，$\boldsymbol{W}_s$ 表示 softmax 分类器的权重矩阵，$\boldsymbol{b}_s$ 表示其偏置项。最终，得到对当前单词 $\boldsymbol{x}_t$ 的预测标签为

$$y = \arg\max p(i \mid \boldsymbol{x}_t) \tag{15}$$

模型训练时采用RMSprop算法[27]进行参数优化，最大化如式(16)所示的带偏置项的对数似然函数。

$$L_{JE} = \sum_{j=1}^{|D|}\sum_{t=1}^{L_j} \log(p_j^t(y^t \mid s_j^t;\Theta))I(O) + \alpha \log(p_j^t(y^t \mid s_j^t;\Theta))(1 - I(O))) \tag{16}$$

其中，$|D|$ 表示训练数据的规模，L_j 表示句子 s_j 的长度，p_j^t 表示句子 s_j 中第 t 个词对应标签的概率。此外，α 表示超参数，$I(O)$ 表示开关函数（如式(17)所示），当涉及非“O”的标签时，通过偏置权重 α 增强关系标签对模型的影响[8]。

$$I(O) = \begin{cases} 1, \text{tag} = 'O' \\ 0, \text{tag} \neq 'O' \end{cases} \tag{17}$$

3.6 对抗主动学习

不同于传统的基于不确定性采样的主动学习方式，本文融合对抗学习思想来评估标记数据和未标记数据间的相似性，提出对抗主动学习模式，以此增量地筛选出数据进行标记。

假设现有少量已标记的数据构成集合 S_L，未标记的数据构成集合 S_U。用 s^L 和 s^U 分别表示标记数据和未标记数据，则从标记数据集中选择样本记作 $s^L \sim S_L$，从未标记数据集中选择样本记作 $s^U \sim S_U$。在对抗主动学习模块中，标记样本 s^L 和未标记样本 s^U 首先依次经过表示层和 BiLSTM 特征编码层，得到隐层空间的特征向量表示 $\boldsymbol{H}^L$ 和 $\boldsymbol{H}^U$。通过训练一个判别器来评估隐空间中特征向量的相似性，而 BiLSTM 特征编码模型则与判别器模型构成一个对抗网络。在该对抗网络中，编码器 BiLSTM 通过特征编码，尽可能地使判别器误认为特征向量均来自标记数据集，而判别器则尽可能地对隐空间中特征向量进行区分。完成训练后，对于输入的未标记样本，判别器输出一个概率值来表示样本来自两类数据集合的概率，将其记作与标记样本的相似性得分。相似性得分高，表明未标记样本所包含的信息量已经在标记样本中有所表达；相似性得分低，表明未标记样本所包含的信息量标记样本未曾包含，该样本需交专家进行标注。将标注后的样本加入标签训练集中，并对前述的联合抽取模型重新进行训练。

在对抗主动学习模块中，BiLSTM 特征编码器类似于对抗网络中的生成器模型，其目标函数为最小化如式(18)所示的损失函数。

$$L_G = -\mathrm{E}_{s\sim S^L}[\log(D(s))] - \mathrm{E}_{s\sim S^U}[\log(D(s))] \tag{18}$$

判别器的目标函数为最小化如式(19)所示的损失函数。

$$L_D = -(\mathrm{E}_{s\sim S^L}[\log(D(s))] + \mathrm{E}_{s\sim S^U}[\log(1 - D(s))]) \tag{19}$$

综合考虑前述联合抽取模块和对抗主动学习中编码器模块，引入超参数 λ 调整两类损失函数，可得其在整个模型中的目标函数为

$$L = L_{JE} + \lambda L_G \tag{20}$$

上述对本文模型中联合抽取模块和对抗主动学习模块分别进行了详细的描述。基于对抗主动学习算法增量地筛选出需要标注的数据，将其标注后加入标签训练集，并对模型重新进行训练。本文模型实现算法如算法 1 所示。

算法 1 对抗主动学习下的网络安全知识三元组抽取算法.

输入 初始标注训练集 S_L，未标注数据集 S_U

输出 实体–关系三元组联合抽取模型

epoch 从 1 到 N 循环

1) 选择标记样本 $s^L \sim S_L$

2) 计算式(16)中的目标函数 L_{JE}

3) 选择未标记样本 $s^U \sim S_U$

4) 计算式(18)中的目标函数 L_G

5) 最小化式(20)中的目标函数 L 来更新参数 θ_{JE} 和 θ_G

6) 最小化式(19)中的目标函数 L_D 来更新参数 θ_D

7) 计算所选样本 s^U 和 s^L 间的相似性得分，并标注得分较低的样本

8) 将新标注样本添加到标注训练集 S_L

结束循环

4 实验

4.1 实验设置

本文实验所使用的数据语料采集自两部分：1) 针对 SemEval 2018 Task 8 发布的网络安全语料，从中筛选得到 500 条涉及恶意软件行为和属性的句子作为初始训练语料，另外筛选出 1 464 条句子作为测试语料，并按照前文所述联合标注策略进行标注；2) 从 AlienVault 社区、welivesecurity 社区、Amazon 安全博客以及近两年的 APT 报告中筛选得到 7 425 条威胁情报语句，标注后添加到训练集，用于对抗主动学习模型的性能测试。此外，为比较

联合抽取模型与流水线模型的性能，针对上述所采集的语料，单独构建了流水线模式中用于命名实体识别和实体关系抽取的标签数据集。

本文在评价网络安全实体−关系三元组联合抽取结果时，类比文献[7]中对药物实体−关系三元组识别结果的判断方式可得，若网络安全实体边界及其类别均被模型标记正确，则认为实体识别结果正确；若网络安全实体边界、实体类别及所属关系类别均被模型标记正确，则判定关系抽取结果正确。针对网络安全实体−关系联合抽取性能的评价，本文采用信息抽取任务中通用的评价指标，即通过准确率 P（precision）、召回率 R（recall）以及 F1 值（F1-score）3 项指标来评价联合抽取模型的性能，并将 F1 值作为评价模型性能的综合性指标。本文模型涉及的主要超参数如表 1 所示。

表 1　　本文模型涉及的主要超参数

参数名称	参数值
词向量维度	100
字符向量维度	25
卷积核个数/个	20
窗口大小	3
BiLSTM 编码层单元大小	300
LSTM 解码层单元大小	600
偏置权重	10

4.2　序列标注模型性能对比实验

由于本文将实体与关系的抽取联合建模为序列标注任务，本节将所提联合标注模型与 NLP 领域典型的序列标注模型进行比较，包括 CRF 模型、BiLSTM-CRF 模型、基于自注意力机制的 BiLSTM-CRF 模型、基于自注意力机制的 BiLSTM-LSTM 模型，并查看使用字符特征与否对模型性能的影响。将上述所列模型在完整的标签训练集上训练后评估模型性能，实验结果如表 2 所示。表 2 中“○”表示对应模型使用了字符级的嵌入特征，“×”表示模型未使用该特征。

在完整标签训练集上完成训练后，相比几类典型的序列标注模型，本文所提融合动态注意力机制的 BiLSTM-LSTM 模型在实体与关系联合标注任务上性能最优，取得了 64.57%的 F1 值。通过比较可以看出，相较于单一使用 CRF 模型，在增添 BiLSTM 网络进行特征获取后，模型性能得到提升，其原因可能是 BiLSTM 在一定程度上解决了序列建模过程中的长距离依赖问题，在识别过程中能够有效利用上下文信息。从表 2 中可以看出，由于字符更关注词本身的特征，在添加字符级别的嵌入特征后，BiLSTM-CRF 模型性能得到进一步的提升，其 F1 值增加了 1.6%，表明增加字符向量对于序列标注的重要性。在此基础上，再向 BiLSTM-CRF 模型添加自注意力机制 Self-att，BiLSTM-CRF 模型性能进一步得到提升。通过对比可以得出，模型性能得到提升的原因可归结为 Self-att 机制的运用，其通过捕获词之间的依存性，使所抽取的文本特征进一步得到增强，进而使模型识别性能得到提升。此外，再将自注意力机制下 BiLSTM-CRF 模型中的解码器由 CRF 更换为 LSTM 网络。可以看出，两者性能基本保持一致，表明 LSTM 解码能够达到与传统 CRF 模型相当的效果。且由于在复杂标签空间中，LSTM 解码优于 CRF 模型，模型 F1 值有 0.13%的微小提升。最后，将自注意力机制替换为本文的动态注意力机制，由于考虑了自注意力权重分布的差异性，模型性能在 F1 值上增加了 1.46%，同时也证明本文所设计的动态注意力机制是有效的。

表 2　　模型性能比较

模型	字符嵌入特征	P	R	F1
CRF	×	56.12%	55.37%	55.74%
BiLSTM-CRF	×	60.41%	58.24%	59.31%
BiLSTM-CRF	○	61.83%	60.02%	60.91%
Self-att-BiLSTM-CRF	○	63.65%	62.32%	62.98%
Self-att-BiLSTM-LSTM	○	64.06%	62.19%	63.11%
Dynamic-att-BiLSTM-LSTM	○	65.75%	63.44%	64.57%

4.3　三元组抽取方法对比

如前文所述，当前知识三元组的抽取主要分为基于先识别实体、后抽取关系的流水线框架，以及实体与关系的联合抽取方法。本节利用典型的流水线框架与联合抽取方法对网络安全知识三元组的抽取性能进行比较。

针对传统的流水线框架，首先需要识别网络安全语料中的命名实体，采用传统方法中广泛使用的 BiLSTM-CRF 模型作为序列标注工具来识别实体。在获得实体识别结果的基础上，对于网络安全语义关系抽取任务，本文分别采用端到端的 Att-PCNN_BiLSTM 模型[28]以及融合句法特征的 SDP-LSTM 模型[29]进行关系分类。针对联合抽取方法，除本文提出的联合抽取模型外，还采用传统基

于参数共享的联合抽取方法，包括性能较好的 BiLSTM_Bi-TreeLSTM 模型[30]以及 BiLSTM-CRF-Multi_head 模型[31]。实验对比结果如表 3 所示。

表 3 三元组抽取方法比较

	模型	*P*	*R*	F1
流水线抽取	SDP-LSTM	61.85%	59.03%	60.41%
	Att-PCNN_BiLSTM	63.28%	58.63%	60.87%
联合抽取	BiLSTM_Bi-TreeLSTM	62.17%	60.58%	61.36%
	BiLSTM-CRF-Multi_head	64.73%	61.65%	63.20%
	Dynamic-att-BiLSTM-LSTM	65.75%	63.44%	64.57%

通过表 3 可以看出，本文提出的融合了动态注意力机制的 BiLSTM-LSTM 模型在知识三元组抽取任务上表现出最优的性能，取得了 64.57%的 F1 值。本文方法直接实现实体与关系端到端的联合抽取，其有效利用了实体识别和关系抽取任务间的语义联系，相较于流水线方式中性能较好的关系分类模型，模型性能有较大提升，表现为 F1 值增加了 3.7%。相较于基于参数共享的联合抽取模型，本文的联合抽取方法性能也更优，评估后发现 F1 值比 BiLSTM-CRF-Multi_head 模型提升了 1.37%。虽然基于参数共享的联合抽取方法也取得了不错的效果，但由于其本质上还是先识别实体、后识别关系，仍然存在一定程度的错误传播与冗余信息。而本文提出的联合抽取模型，基于对实体与关系的联合标注策略，隐式地考虑了实体识别与关系抽取任务间的联系，表现出更优的性能。

4.4 三元组抽取实例分析

4.3 节对比了几类知识三元组抽取方法，相较于流水线模式以及传统的联合抽取方法，本文模型整体上表现更佳。本节进一步查看传统流水线的 Att-PCNN_BiLSTM 模型、BiLSTM-CRF-Multi_head 联合抽取模型以及本文联合抽取模型对网络安全知识三元组的抽取效果。示例结果如表 4 所示，“**[ ]**”加粗表示能够正确识别的实体，“[ ]”下划线表示未能被识别的实体，实体下标标识了该实体所属的关系类别。

针对示例 1 中的网络安全句子序列，相较于标

表 4 三元组抽取结果示例

模型	抽取结果
示例 1	Since the revelation of an [**Adobe Flash Player**] $_{e1,\ hasVulnerability}$ zero day exploit exposed as part of the leaked Hacking Team arsenal in 2015 designated [**CVE-2015-5119**] $_{e2,\ hasVulnerability}$.
Att-PCNN_BiLSTM	Since the revelation of an [**Adobe Flash Player**] $_{e1\ uses}$ zero day exploit exposed as part of the leaked Hacking Team arsenal in 2015 designated [**CVE-2015-5119**] $_{e2\ uses}$.
BiLSTM-CRF-Multi_head	Since the revelation of an [**Adobe Flash Player**] $_{e1\ hasVulnerability}$ zero day exploit exposed as part of the leaked Hacking Team arsenal in 2015 designated [**CVE-2015-5119**] $_{e2\ hasVulnerability}$.
Dynamic-att-BiLSTM-LSTM	Since the revelation of an [**Adobe Flash Player**] $_{e1\ hasVulnerability}$ zero day exploit exposed as part of the leaked Hacking Team arsenal in 2015 designated [**CVE-2015-5119**] $_{e2\ hasVulnerability}$.
示例 2	[**Apt 28**] $_{e1,\ M}$ which we suspect is sponsored by [**Russian**] $_{e2,\ comes\text{-}from}$ government, uses [**spear phishing emails**] $_{e2,\ uses}$ to target its victims by specific topics.
Att-PCNN_BiLSTM	[**Apt 28**] $_{e1,\ comes\text{-}from}$ which we suspect is sponsored by [**Russian**] $_{e2,\ comes\text{-}from}$ government, uses [spear phishing emails] to target its victims by specific topics.
BiLSTM-CRF-Multi_head	[**Apt 28**] $_{e1,\ comes\text{-}from}$ which we suspect is sponsored by [**Russian**] $_{e2,\ comes\text{-}from}$ government, uses [**spear phishing**] emails to target its victims by specific topics.
Dynamic-att-BiLSTM-LSTM	[**Apt 28**] $_{e1,\ M}$ which we suspect is sponsored by [**Russian**] $_{e2,\ comes\text{-}from}$ government, uses [**spear phishing emails**] $_{e2,\ uses}$ to target its victims by specific topics.
示例 3	One identified malware sample ([**75193fc10145931ec0788d7c88fc8832**] $_{e1,\ indicates}$, compiled in March 2014) uses a password-protected [**.7z**] $_{e1,\ located\text{-}at}$ to deliver the [**Etumbot installer**] $_{e2,\ M}$, which is most likely contained within [**spear phishing email**] $_{e2,\ located\text{-}at}$.
Att-PCNN_BiLSTM	One identified malware sample ([**75193fc10145931ec0788d7c88fc8832**] $_{e1,\ indicates}$, compiled in March 2014) uses a password-protected [.7z] to deliver the [**Etumbot installer**] $_{e2,\ indicates}$, which is most likely contained within [spear phishing email].
BiLSTM-CRF-Multi_head	One identified malware sample ([**75193fc10145931ec0788d7c88fc8832**] $_{e1,\ indicates}$, compiled in March 2014) uses a password-protected [.7z] to deliver the [**Etumbot installer**] $_{e2,\ indicates}$, which is most likely contained within [**spear phishing**] email.
Dynamic-att-BiLSTM-LSTM	One identified malware sample ([**75193fc10145931ec0788d7c88fc8832**] $_{e1,\ indicates}$, compiled in March 2014) uses a password-protected [.7z] to deliver the [**Etumbot installer**] $_{e2,\ M}$, which is most likely contained within [**spear phishing email**] $_{e2,\ located\text{-}at}$.

准的抽取结果，所用的三类模型均能够正确识别出相应的实体及其类别。然而，对于流水线框架，基于此实体识别结果进行关系预测时，其将实体对间原本的语义关系“hasVulnerability”错误地分类为关系“uses”，此结果可能由于未考虑实体识别与关系抽取任务间的联系所致。而两类联合抽取模型则准确地表达了其语义关系。

针对示例 2，标准抽取结果应该含有 3 个实体，且实体“APT 28”参与了多重关系，对于此标注结果，本文模型在准确识别实体的同时还能准确表达所有实体的关系类别。在流水线框架中，其未能识别出实体“spear phishing emails”。而基于参数共享的联合抽取模型，对于实体“spear phishing emails”的识别出现边界错误，且其无法处理重叠关系问题。

针对示例 3，标准抽取结果理应包含 4 个相关实体，而三类模型均未能准确识别出这些实体。对于流水线框架，其在实体识别时就产生较大的误差，相应地也只表达了恶意软件及其指示器之间的“indicates”关系。对于基于参数共享的联合抽取模型，其未能识别实体“.7z”，且实体“spear phishing email”边界识别有误，其三元组抽取结果也只表达了“indicates”关系。对于本文模型，虽然只遗漏了实体“.7z”，但其对实体“spear phishing email”的实体位置标注错误，导致其在组成三元组时未能被匹配，由此说明本文方法在实体与关系的联合识别中还有待改进与优化。

4.5　对抗主动学习算法性能评估

为验证本文所提对抗主动学习模块的有效性，分别以完整的标签训练集和通过对抗主动学习筛选获得的训练集来评估本文实体与关系联合抽取模型的性能，评估结果如表 5 所示。可以看出，随着标注数据量的增加，模型整体性能不断提升。当获得全部标注数据的 45%，并对模型重新进行训练时，模型性能与使用完整标签训练集训练后的模型性能已经非常接近，证明了本文所提对抗主动学习算法的有效性。

此外，将本文提出的对抗主动学习，与常规的主动学习算法进行比较，包括最小置信度（LC, least confidence）算法、贝叶斯非一致主动学习（BALD, Bayesian active learning by disagreement）以及最大归一化对数概率（MNLP, maximum normalized log-probability）算法。运用各模型逐次筛选得到不同规模的标签数据，并评估模型性能，结果如图 4 所示。相比之下，本文所提对抗主动学习模型表现出最优的性能。而基于 LC 的主动学习算法性能最差，其原因是 LC 算法在采样时通过序列标注模型的输出来计算不确定性，而复杂的标签空间导致其采样准确性较差。MNLP 算法和 BALD 算法虽然在主动采样时对模型不确定性的计算进行了一定的优化，但其性能仍受到标签空间的影响。不同于复杂的不确定性计算，本文所提对抗主动学习算法通过直接比较未标记数据和标记数据的相似性进行采样，在降低计算复杂度的同时提高了模型采样的准确率，其可实现以相对较低的数据标注代价来逐步提升三元组抽取效果。

表 5　不同规模标注数据下的性能比较

标注数据规模	*P*	*R*	F1
10%	33.28%	32.75%	33.01%
20%	52.72%	50.64%	51.66%
30%	62.15%	59.97%	61.04%
40%	64.95%	63.02%	63.97%
45%	65.62%	63.25%	64.41%
100%	65.75%	63.44%	64.57%

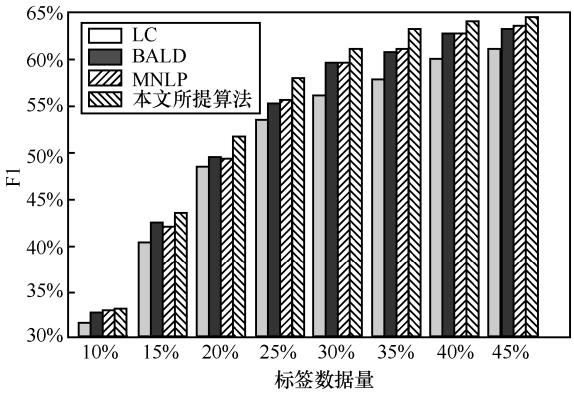


图 4　主动学习算法性能对比

5　结束语

为实现面向网络安全领域知识三元组的抽取，本文提出了一种融合对抗主动学习的网络安全实体与关系联合抽取模型。对于当前网络安全领域知识抽取流水线模式存在的问题，提出基于联合标注策略将实体识别与关系抽取同时建模为一个序列标注任务，通过 BiLSTM 网络对输入序列进行特征编码，并基于动态注意力机制准确捕

获词之间的影响权重，在此基础上利用 LSTM 对标签进行解码预测。此外，针对网络安全领域标签数据缺乏的问题，提出基于对抗主动学习框架，评估标记样本与未标记样本间的相似性得分，筛选出高质量的样本进行标注，实现以较低的标注代价来提升模型性能。实验验证了本文所提对抗主动学习框架的有效性，并对比已有网络安全实体与关系抽取模型，表明本文所提序列标注模型的性能更优。

参考文献：

[1] JOSHI A, LAL R, FININ T, et al. Extracting cybersecurity related linked data from text[C]//2013 IEEE Seventh International Conference on Semantic Computing. Piscataway: IEEE Press, 2013: 252-259.

[2] 鄂海红, 张文静, 肖思琪, 等. 深度学习实体关系抽取研究综述[J]. 软件学报, 2019, 30(6): 1793-1818.
E H H, ZHANG W J, XIAO S Q, et al. Survey of entity relationship extraction based on deep learning[J]. Journal of Software, 2019, 30(6): 1793-1818.

[3] PHANDI P, SILVA A, LU W. Semeval-2018 task 8: semantic extraction from cybersecurity reports using natural language processing (SecureNLP)[C]//Proceedings of the 12th International Workshop on Semantic Evaluation. [S.n.:s.l.], 2018: 697-706.

[4] SIMRAN K, SRIRAM R, VINAYAKUMAR R, et al. Deep learning approach for intelligent named entity recognition of cyber security[J]. arXiv Preprint, arXiv: 2004.00502, 2020.

[5] PINGLE A, PIPLAI A, MITTAL S, et al. RelExt: relation extraction using deep learning approaches for cybersecurity knowledge graph improvement[C]//Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. Piscataway: IEEE Press, 2019: 879-886.

[6] HUANG W, CHENG X, WANG T, et al. BERT-based multi-head selection for joint entity-relation extraction[C]//CCF International Conference on Natural Language Processing and Chinese Computing. Berlin: Springer, 2019: 713-723.

[7] 曹明宇, 杨志豪, 罗凌, 等. 基于神经网络的药物实体与关系联合抽取[J]. 计算机研究与发展, 2019, 56(7): 1432-1440.
CAO M Y, YANG Z H, LUO L, et al. Joint drug entities and relations extraction based on neural networks[J]. Journal of Computer Research and Development, 2019, 56(7): 1432-1440.

[8] ZHENG S, WANG F, BAO H, et al. Joint extraction of entities and relations based on a novel tagging scheme[C]// Proceedings of the 55th Association for Computational Linguistics. [S.n.:s.l.], 2017: 1227-1236.

[9] LIAO X. Towards automatically evaluating security risks and providing cyber intelligence[D]. Atlanta: Georgia Institute of Technology, 2017.

[10] PANWAR A. Toward automatic generation and analysis of indicators of compromise (IoCS) using convolutional neural network[D]. Arizona: Arizona State University, 2017.

[11] GASMI H, LAVAL J, BOURAS A. Information extraction of cybersecurity concepts: an LSTM approach[J]. Applied Science, 2019, 9(19): 1-15.

[12] CHAMBERS N, FRY B, MCMASTERS J. Detecting denial-of-service attacks from social media text: applying nlp to computer security[C]//Proceedings of the North American Chapter of the Association for Computational Linguistics. [S.n.:s.l.], 2018: 1626-1635.

[13] ZHOU S, LONG Z, TAN L, et al. Automatic identification of indicators of compromise using neural-based sequence labelling[J]. arXiv Preprint, arXiv:1810.10156, 2018.

[14] LONG Z, TAN L, ZHOU S, et al. Collecting indicators of compromise from unstructured text of cybersecurity articles using neural-based sequence labelling[C]//2019 International Joint Conference on Neural Networks (IJCNN). Piscataway: IEEE Press, 2019: 1-8.

[15] 秦娅, 申国伟, 赵文波, 等. 基于深度神经网络的网络安全实体识别方法[J]. 南京大学学报(自然科学), 2019, 55(1): 29-40.
QIN Y, SHEN G W, ZHAO W B, et al. Research on the method of network security entity recognition based on deep neural network[J]. Journal of Nanjing University(Natural Science), 2019, 55(1): 29-40.

[16] 张若彬, 刘嘉勇, 何祥. 基于 BLSTM-CRF 模型的安全漏洞领域命名实体识别[J]. 四川大学学报(自然科学版), 2019, 56(3): 469-475.
ZHANG R B, LIU J Y, HE X. Named entity recognition for vulnerabilities based on BLSTM-CRF model[J]. Journal of Sichuan University(Natural Science Edition), 2019, 56(3): 469-475.

[17] ZHU J J, BENTO J. Generative adversarial active learning[J]. arXiv Preprint, arXiv: 1702.07956v5, 2017.

[18] CULOTTA A, MCCALLUM A. Reducing labeling effort for structured prediction tasks[C]//International Conference on Artificial Intelligence. Piscataway: IEEE Press, 2005: 746-751.

[19] HOULSBY N, HUSZAR F, GHAHRAMANI Z, et al. Bayesian active learning for classification and preference learning[J]. arXiv Preprint, arXiv:1112.5745, 2011.

[20] GAL Y, GHAHRAMANI Z. Dropout as a Bayesian approximation: representing model uncertainty in deep learning[C]//International Conference on Machine Learning. Piscataway: IEEE Press, 2016: 1050-1059.

[21] SENER O, SAVARESE S. Active Learning for convolutional neural networks: a core-set approach[J]. arXiv Preprint, arXiv: 1708.00489, 2017.

[22] KUO W, HANE C, YUH E L, et al. Cost-sensitive active learning for intracranial hemorrhage detection[C]//Medical Image Computing and Computer Assisted Intervention. Piscataway: IEEE Press, 2018: 715-723.

[23] SHEN Y, YUN H, LIPTON Z C, et al. Deep active learning for named entity recognition[C]//International Conference on Learning Representations. Piscataway: IEEE Press, 2018: 1-15.

[24] CHIU J P C, NICHOLS E. Named entity recognition with bidirectional LSTM-CNNs[J]. Transactions of the Association for Computational Linguistics, 2016, 4: 357-370.

[25] CAO P, CHEN Y, LIU K, et al. Adversarial transfer learning for chinese named entity recognition with self-attention mechanism[C]//The 2018 Conference on Empirical Methods in Natural Language Processing. Piscataway: IEEE Press, 2018: 182-192.

[26] 程梦, 洪宇, 唐建, 等. 面向属性抽取的门控动态注意力机制[J]. 模式识别与人工智能, 2019, 32(2): 184-192.
CHENG M, HONG Y, TANG J, et al. Gated dynamic attention mechanism towards aspect extraction[J]. Pattern Recognition and Artificial Intelligence, 2019, 32(2): 184-192.

[27] TIELEMAN T, HINTON G. Lecture 6.5-rmsprop, coursera: neural networks for machine learning[R]. University of Toronto, Technical Report, 2012.

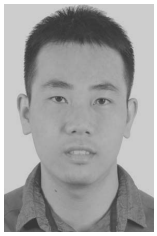
[28] 张晓斌, 陈福才, 黄瑞阳. 基于 CNN 和双向 LSTM 融合的实体关系抽取[J]. 网络与信息安全学报, 2018, 4(9): 44-51.
ZHANG X B, CHEN F C, HUANG R Y. Relation extraction based on CNN and BiLSTM[J]. Chinese Journal of Network and Information Security, 2018, 4(9): 44-51.

[29] XU Y, MOU L, LI G, et al. Classifying Relations via long short term memory networks along shortest dependency paths[C]//The 2015 Conference on Empirical Methods in Natural Language Processing. Piscataway: IEEE Press, 2015: 1785-1794.

[30] MIWA M, BANSAL M. End-to-end relation extraction using LSTMs on sequences and tree structures[C]//The 54th Annual Meeting of the Association for Computational Linguistics. Piscataway: IEEE Press, 2016: 1105-1116.

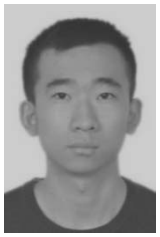
[31] BEKOULIS G, DELEU J, DEMEESTER T, et al. Joint entity recognition and relation extraction as a multi-head selection problem[J]. arXiv Preprint, arXiv: 1804.07847, 2018.

[作者简介]



李涛（1992− ），男，甘肃甘谷人，信息工程大学博士生，主要研究方向为网络威胁语义建模。

郭渊博（1975− ），男，陕西周至人，博士，信息工程大学教授、博士生导师，主要研究方向为大数据安全、态势感知。



琚安康（1995− ），男，河南辉县人，信息工程大学博士生，主要研究方向为多步攻击检测、异构安全数据融合。