基于 Bert 和 BiLSTM-CRF 的 APT 攻击实体识别及对齐研究

杨秀璋[1,2]，彭国军[1,2]，李子川[1,2]，吕杨琦[1,2]，刘思德[1,2]，李晨光[1,2]

（1. 武汉大学空天信息安全与可信计算教育部重点实验室，湖北 武汉 430072；
2. 武汉大学国家网络安全学院，湖北 武汉 430072）

摘 要：针对高级可持续威胁（APT）分析报告未被有效利用，缺乏自动化方法生成结构化知识并形成黑客组织特征画像问题，提出一种融合实体识别和实体对齐的 APT 攻击知识自动抽取方法。首先，结合 APT 攻击特点设计 12 种实体类别；其次，构建融合 Bert、双向长短期记忆（BiLSTM）网络和条件随机场（CRF）的 APT 攻击实体识别模型，利用 Bert 预训练标注语料，BiLSTM 学习上下文语义信息，注意力机制突出关键特征，再由 CRF 识别实体；最后，结合实体对齐方法来生成不同 APT 组织的结构化知识。实验结果表明，所提方法能有效识别 APT 攻击实体，其精确率、召回率和 F_1 值分别为 0.929 6、0.873 3 和 0.900 6，均优于现有模型。此外，所提方法能在少量样本标注的情况下自动抽取高级可持续威胁知识，通过实体对齐能生成常见 APT 组织的结构化特征画像，从而为后续 APT 攻击知识图谱构建和攻击溯源提供支撑。

关键词：高级可持续威胁；威胁情报抽取；实体识别；实体对齐；深度学习

中图分类号：TP309

文献标志码：A

DOI: 10.11959/j.issn.1000−436x.2022116

Research on entity recognition and alignment of APT attack based on Bert and BiLSTM-CRF

YANG Xiuzhang[1,2], PENG Guojun[1,2], LI Zichuan[1,2], LYU Yangqi[1,2], LIU Side[1,2], LI Chenguang[1,2]

1. Key Laboratory of Aerospace Information Security and Trusted Computing of Ministry of Education, Wuhan University, Wuhan 430072, China
2. School of Cyber Science and Engineering, Wuhan University, Wuhan 430072, China

Abstract: Aiming at the problems that APT (advanced persistent threat) analysis reports have not been fully utilized, and there is a lack of automation methods to generate structured knowledge and construct feature portraits of the hacker organizations, an automatic knowledge extraction method of APT attacks combining entity recognition and entity alignment was proposed. Firstly, 12 entity categories were designed according to the characteristics of APT attacks. Then, an APT attack entity recognition method that combined Bert, BiLSTM (bidirectional long and short-term memory) network, and CRF (conditional random field) was proposed. The Bert model was used to pre-train the annotated corpus. The BiLSTM model was constructed to learn contextual semantic information. The attention mechanism was built to extract key features. Moreover, the CRF algorithm was proposed to identify entities. Finally, the entity alignment method was designed to generate structured knowledge of different APT organizations. Experimental results show that the proposed method can effectively identify APT attack entities, with a precision of 0.929 6, a recall of 0.873 3, and an F_1-score of 0.900 6, superior to existing models. In addition, the proposed method can automatically extract advanced persistent threat knowledge with a small number of annotated samples and generate the structured portraits of APT groups through entity alignment, thus providing support for subsequent knowledge graph construction of APT attacks and attack tracing.

Keywords: advanced persistent threat, threat intelligence extraction, entity recognition, entity alignment, deep learning

收稿日期：2022−03−23；**修回日期：**2022−05−18

通信作者：彭国军，guojpeng@whu.edu.cn

基金项目：国家自然科学基金资助项目（No.62172308, No.U1626107, No.61972297, No.62172144）

Foundation Item: The National Natural Science Foundation of China (No.62172308, No.U1626107, No.61972297, No.62172144)

0　引言

随着互联网技术的飞速发展，网络入侵、蠕虫感染、勒索病毒、分布式拒绝服务攻击等网络攻击事件越来越频繁，给社会和企业带来了巨大的安全威胁。在这些潜在威胁中，高级可持续威胁（APT, advanced persistent threat）攻击会造成更大的危害。APT 攻击是一种新型的网络攻击模式，以刺探、收集和监控情报为目的，具有极强的组织性、隐蔽性和威胁性[1]。360 公司在 2021 年发布的全球 APT 报告显示我国是 APT 攻击的主要受害者。此外，APT 攻击的恶意代码往往会利用攻击目标的相关信息有组织地躲避杀毒软件及防火墙检测，导致其攻击痕迹很难被溯源。面对当前复杂变化的网络安全环境，如何对抗 APT 攻击已成为整个安全界亟须解决的问题。安全公司生成的海量 APT 分析报告和威胁情报具有极其重要的研究价值，它们能有效地提供 APT 组织的动态，从而辅助网络攻击事件的溯源分析，但其主要以非结构化文本为主。因此，如何利用多源异构的安全文本知识自动化识别 APT 攻击实体并构建 APT 组织画像和领域知识图谱将成为新的研究热点。

近年来，深度学习技术已经被广泛应用于安全领域中，如恶意流量检测、恶意代码分析、个人隐私保护和工业控制系统安全检测等[2-3]。利用深度学习实现自动安全防护和 APT 攻击检测已经成为重要的研究方向并取得一定进展。在安全日志方面，Milajerdi 等[4]提出了一种检测 APT 攻击的 HOLMES 系统，旨在利用主机日志中的活动和警告信息来识别 APT 攻击。在网络流量方面，Marchetti 等[5]通过大量分析网络流量实现 APT 活动检测，发现与数据窃取恶意行为相关的 APT 攻击并识别出可疑主机。在恶意行为分析方面，Han 等[6]针对 APT 攻击缓慢可持续且使用 0day 漏洞特点，构建一种基于溯源图的 APT 攻击检测系统 Unicorn，其能在没有预先设定攻击特征的情况下识别隐蔽异常行为。

然而，上述方法需要获取大量 APT 样本、流量数据或日志信息，但 APT 攻击隐蔽性极强，特别是利用 0day 漏洞的样本很难被捕获，因此这些方法更偏向于模拟 APT 攻击，在实际应用中存在一定的局限性。此外，传统方法缺乏对 APT 组织画像构建，未有效利用各大安全厂商已形成的 APT 分析报告，较难从碎片化、海量化的威胁情报数据中挖掘 APT 知识，最终导致攻击检测效果不理想。因此，亟须提出一种面向 APT 攻击的知识抽取方法。

为有效从 APT 分析报告中自动抽取实体，形成 APT 组织结构化知识，本文提出一种融合实体识别和实体对齐的 APT 攻击知识自动抽取方法。其贡献主要包括以下 3 个方面。

1) 设计一种混合模型来实现 APT 攻击领域的实体识别和知识融合。该模型能从 APT 报告中自动提取结构化知识，利用实体对齐和知识融合构建黑客组织画像，从而为安全研究人员网络攻击事件分析和关键特征识别提供决策支持，也为 APT 领域知识图谱构建和溯源分析提供支撑。

2) 融合主流的 ATT&CK（adversarial tactics, techniques, and common knowledge）知识框架设计 12 种命名实体类别，提出一种融合 Bert（bidirectional encoder representations from transformers）模型、双向长短期记忆（BiLSTM, bidirectional long and short-term memory）网络和条件随机场（CRF, conditional random field）的 APT 攻击实体识别方法。通过 Bert 预训练词向量，增强了模型的泛化能力，接着构建 BiLSTM-CRF 模型学习上下文语义信息并完成实体识别。该方法能有效弥补传统实体识别无法较好地抽取特定领域实体，需要大量标注信息，且对存在语义歧义、命名规则复杂的实体抽取精确率较低的不足。

3) 提出一种实体对齐方法并应用于 APT 领域的知识融合，有效地将 APT 分析报告和黑客组织指纹画像构建结合，通过实体对齐能有效提升所抽取 APT 攻击实体的质量，并形成常见 APT 组织的知识消息盒。此外，本文实验基于真实的 APT 分析报告完成，并与类似研究进行系统比较，证明了所提方法具有良好的性能和实用价值。

1　相关工作

1.1　APT 攻击研究现状

APT 是近年来形成的新型网络攻击模式，具有针对性强、组织严密、持续时间长、隐蔽性高和威胁程度大的特点，给全球政府部门、金融机构和企业带来极大的安全隐患[7]。APT 攻击给全球网络空间安全带来了严重的威胁，如何快速精准地检测 APT 攻击已成为重要的研究热点。

面对日益增多的攻击事件，工业界对 APT 攻击的防御和溯源研究越来越多。Muckin 等[8]提出网络空间安全杀伤链框架，将网络空间安全划分为 7 个

阶段，并基于攻击者视角对 APT 攻击行动进行整体分析。Mitre 公司提出 ATT&CK 知识框架，整个框架以战术、技术和过程为核心，能有效辅助自动化威胁分析。同时，FireEye、卡巴斯基、360、奇安信、安天等公司对 APT 攻击的溯源及检测都做了大量的研究。

在学术界，研究者提出了恶意代码分析[9]、主机应用保护[10]、网络入侵检测[11]、大数据分析[12]等 APT 攻击检测方法。文献[13]通过博弈论实现主动防御，在分析博弈模型的纳什均衡基础上计算使 APT 攻防双方收益最大的攻击路径和防御策略。张小松等[14]提出一种基于树形结构的 APT 攻击检测方法。Milajerdi 等[15]基于审计日志构建溯源图，结合网络威胁情报和图模式匹配设计 POIROT 系统来检测 APT 攻击。近年来，随着人工智能的火热发展，研究者将机器学习和溯源图应用到 APT 攻击检测并取得一定成果。

然而，目前主流的 APT 攻击分析框架和防御方法仍然依靠大量的专家知识，没有将安全厂商发布的 APT 分析报告有效利用，并且仍未提出一种有效的方法来自动提取并生成 APT 组织画像。此外，面对语义丰富的非结构化文本数据，传统方法抽取知识的效果较差，缺乏有效的安全知识表达，从而无法形成 APT 结构化知识。为解决上述问题，本文提出一种混合型的 APT 攻击实体识别及对齐方法。

1.2 实体识别研究现状

实体识别又称为命名实体识别（NER, named entity recognition），它在自然语言处理和知识图谱领域扮演了一个重要的角色。命名实体是指一个词或短语，用于标识一组具有相似属性的事物，命名实体识别是定位命名实体边界并提取预定义实体集合的过程[16]。目前，实体识别的方法主要有三类：基于规则的实体识别、基于统计的实体识别和基于深度学习的实体识别。

基于规则的实体识别利用词典或专家知识构造规则，为每条规则赋予权重，通过规则匹配识别命名实体。比较著名的包括 LaSIE-II[16]系统、Facile[17]系统和 DL-CoTrain[18]实体识别方法。然而，该类方法过度依赖专家知识，需要手工构造大量的规则，受领域限制严重且可移植性较差。

基于统计的实体识别是将该任务转换为多分类或序列标注问题，通过统计样本数据集的相关特征来建立识别模型。常见方法主要包括隐马尔可夫模型（HMM, hidden Markov model）[19]、最大熵（ME, maximum entropy）[20]、支持向量机（SVM, support vector machine）[21]和条件随机场[22]等。基于统计的命名实体识别方法在一定程度上对语言的依赖性更小。但是，这些方法仍然需要大量的人工参与，特征工程比较消耗时间，且严重依赖语料库和设定的特征模板，扩展性较差，缺乏对语义知识的学习。

近年来，深度神经网络被广泛应用于自然语言处理领域。基于深度学习的实体识别能够解决命名实体识别的上下文语义难以理解和数据稀疏问题，并且具有对专家知识依赖小且移植性好的优势。在实体识别任务中，常用的深度学习模型包括卷积神经网络、循环神经网络、长短期记忆（LSTM, long short-term memory）网络以及与 CRF 相结合的模型。Hammerton[23]首先将 LSTM 模型应用于实体识别。随后，ID-CNN[24]、Lattice-LSTM[25]等方法被提出。然而，这些方法缺乏预训练词向量来学习语义特征，在小样本标注场景的效果不佳，并且未融合实体对齐，从而导致其对特定领域的实体识别和知识抽取的效果不理想，如 APT 领域。

此外，针对安全领域的实体识别主要偏向于漏洞和威胁情报的实体识别[26]，其实体类别较少，场景单一，且尚无针对 APT 攻击领域的实体识别，也缺乏利用 ATT&CK 框架与结合 APT 攻击真实流程来自动提取知识的研究，从而无法为 APT 组织结构化指纹生成、攻击溯源和图谱构建提供支撑。

为了解决上述问题，本文有效地将实体识别和实体对齐任务应用于 APT 攻击领域，结合 ATT&CK 框架设计 12 种命名实体。在模型方面，本文采用 Bert 模型来预训练词向量，从而增强模型的泛化能力及适应不同的语义环境，同时构建 BiLSTM-CRF 和注意力机制模型来提取 APT 攻击领域的命令实体，并融合实体对齐提升所抽取知识的质量，最终形成常见 APT 组织知识消息盒。本文融合实体识别和实体对齐来开展 APT 攻击知识自动抽取的研究，取得了良好效果。

2 问题描述

2.1 任务定义

APT 攻击实体识别旨在提取具有特定意义的攻击实体；实体对齐旨在精确识别不同来源的攻击组织，并将其知识融合。常见的 APT 攻击实体包括 APT 组织名称、攻击装备、攻击手法、攻击漏洞等。

该问题在本文中的定义如式(1)给出的函数 f 所示。

$$\begin{aligned} E = f(S) &= \{\text{entity}_1, \text{entity}_2, \cdots, \text{entity}_n\} \\ S &= < w_1, w_2, \cdots, w_m > \\ \text{entity}_i &= < I_{bi}, I_{ei}, t_i > \end{aligned} \tag{1}$$

其中，E 表示识别出来的 APT 攻击命名实体集，包含 n 个实体三元组；S 表示输入的 APT 攻击报告或网页单词序列，w_j 表示第 j 个位置的单词，共 m 个单词；entity_i 表示 S 中的第 i 个命名实体，$I_{bi}\in[1,m]$ 且 $I_{ei}\in[1,m]$，分别表示该命名实体在 S 中的开始和结束位置，t_i 表示该实体的类型。例如，对于输入文本序列“Lazarus usually uses phishing attacks”，模型会识别出<1,1,Lazarus>和<4,5,phishing attacks>2 个实体三元组。

图 1 详细展示了 APT 攻击实体识别及对齐任务的过程，共 7 个步骤。数据预处理和数据标注后的文本经过实体识别和实体对齐处理后，能有效提取包括组织名称、地理位置、攻击装备、攻击漏洞等结构化实体知识。实体识别其实是序列标注问题，通常采用 BIO 方法进行数据标注。其中 B 和 I 分别对应实体起始位置和实体中间位置（含结束位置），不属于任何实体的词语采用 O 表示，该方法能有效标记出实体的类型和位置。

本文针对 APT 攻击特点，结合 ATT&CK 知识框架，归纳出 12 种命名实体，包括 APT 组织（AG, APT group）、攻击装备（AEQ, attack equipment）、攻击手法（AM, attack method）、攻击漏洞（AV, attack vulnerability）、攻击事件（AE, attack event）、攻击目标（AT, attack target）、攻击行业（AI, attack industry）、恶意文件（MF, malicious files）、恶意软件家族（MFA, malware family）、区域位置（RL, regional location）、操作系统（OS, operating system）和利用软件（SI, software information），详细描述如表 1 所示。后续实验会将标记符和 BIO 方法结合，如“B-AG”表示组织的起始位置，“I-AM”表示攻击手法的中间位置。

2.2　研究动机及挑战

当前，APT 攻击非结构化文本处理存在诸多挑战，这些安全信息呈碎片化分散于互联网中，没有被有效地整合和利用。本文旨在从公开的 APT 分析报告中抽取 APT 攻击实体，构建模型自动生成 APT 结构化知识，这将为 APT 组织的恶意行为分析和攻击溯源提供线索，具体研究动机如下。

1) APT 攻击知识图谱构建。面对大规模的安全数据，传统方法主要利用系统内部产生的日志信息和入侵检测数据进行态势感知，缺乏对外部网络安全知识的有效利用和语义理解。如何描述 APT 攻击行为，自动生成攻击指纹和构建 APT 知识图谱是一个关键问题。此外，APT 分析报告存在大量嵌套、别名、缩略词及组合词，如何准确抽取知识存在挑战。基于此，本文开展 APT 领域的知识自动抽取研究。

2) 少规模数据标注和融合语义实体识别。实体识别需要丰富的训练语料来构建词语表示。然而，APT 攻击领域尚无已标注的专业语料库，传统方法需要花费大量时间去完成数据标注工作。借助专家知识，本文旨在深入分析 APT 攻击流程，实现少规模数据标注的知识抽取，设计一种融合实体识别和实体对齐的方法，有效生成黑客组织的结构化知识。

3) APT 攻击溯源。APT 组织为抵抗恶意代码检测和防御技术，通常会使用代码混淆，从而导致

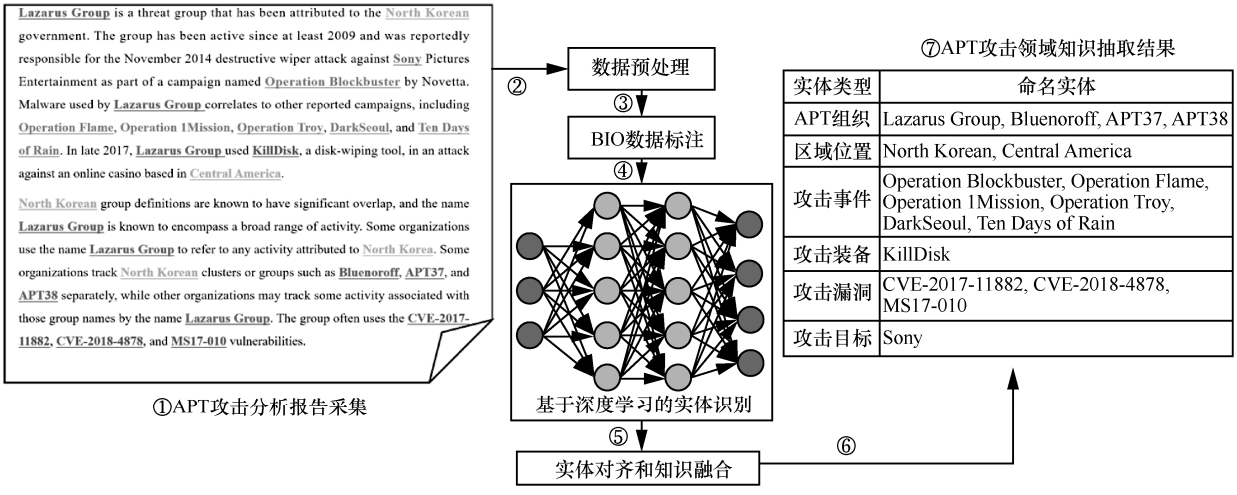


图 1　APT 攻击实体识别及对齐任务过程

表 1 APT 攻击领域的命名实体类别

标志符号	实体类别	类别定义	示例
AG	APT 组织	常见 APT 攻击的团队名称	Lazarus，APT28，OceanLotus
AEQ	攻击装备	APT 组织的装备	CobaltStrike，Metasploit，Gh0st
AM	攻击手法	APT 组织的攻击手段和技术	SQL injection，spear phishing，XSS attack
AV	攻击漏洞	APT 组织常用的漏洞，主要包括 CVE 编号标识或特定漏洞名称	CVE-2017-11882，CVE-2018-4878，EternalBlue，HeartBleed
AE	攻击事件	APT 组织近年来开展的攻击活动	Operation Blockbuster，DarkSeoul
AT	攻击目标	APT 组织攻击的公司、部门和单位	Sony，Iranian nuclear power plant
AI	攻击行业	APT 组织攻击的行业信息	financial，economic，trade policy
MF	恶意文件	APT 组织常用恶意文件、敏感目录及恶意指令，文件格式包括 exe、xls、doc	cmdl32.exe，Agent.btz，wwlib.dll
MFA	恶意软件家族	APT 组织常用的恶意软件家族	Trojan/Win32.Occamy，ZeuS
RL	区域位置	APT 组织所在区域及目标区域	North Korea，Russia，South Asia
OS	操作系统	发起 APT 攻击的操作系统环境	Windows，Mac，Linux，Android
SI	利用软件	发起 APT 攻击的软件环境	Chrome，Office，Firefox

APT 恶意代码溯源困难，仅通过异常流量和样本分析判断恶意行为的方法过于局限。如何有效利用公开的 APT 分析报告辅助恶意软件溯源并识别其所属组织具有重要意义。目前，APT 攻击的知识自动抽取研究仍处于起步阶段，安全人员的手工分析方法耗时耗力。本文对此开展研究，为后续 APT 组织画像构建和攻击溯源提供一定的帮助。

知识图谱通常包括知识抽取（实体识别、关系抽取、属性抽取）、知识表示、知识融合和知识推理等阶段。其中，APT 攻击知识图谱的关系抽取、知识表示和知识推理将在未来开展深入研究。本文将实体识别和实体对齐作为整个研究的起点，为 APT 知识图谱构建及黑客组织指纹的自动化抽取提供思路。综上，本文在这些动机的驱动下，将对 APT 攻击知识的自动抽取开展全面的研究。

3 模型设计与实现

3.1 整体框架

本文提出了一种融合实体识别和实体对齐的 APT 攻击知识自动抽取方法，其框架如图 2 所示，主要包括 5 个部分：1) 通过预处理层对语料进行数据清洗和数据标注，将预处理后的 APT 文本序列表征成向量；2) 通过 Bert 预训练，对每个词语编码并生成对应的字向量；3) 构建 BiLSTM 和 Attention 模型，利用 BiLSTM 捕获长距离和上下文语义特征，再结合注意力机制突出关键特征，将向量序列转换为标注概率矩阵；4) 通过 CRF 算法对输出预测标签间的关系进行解码，输出最优的标签序列；5) 构建语义相似度和 Birch 的实体对齐方法，通过知识匹配提升所抽取 APT 攻击知识的质量，最终融合形成各 APT 组织的知识消息盒。

整个方法的实现过程如算法 1 所示，输入为 APT 分析报告的文本词序列 S，输出为 APT 组织的命名实体及攻击知识集合 E。该算法首先经过预处理和数据标注；然后构建相关的 Bert 和 BiLSTM 模型，并经过初始化后分别对向量进行训练；最后利用 CRF 算法预测实体标签序列，再进行实体对齐。

算法 1 APT 攻击命名实体识别及对齐算法

输入 APT 分析报告的文本词序列 S，序列中的特征词 $w_{i,j}$

输出 APT 组织的命名实体集 E

1) for $w_{i,j} \in S$ do
2) 依次对特征词进行预处理和 BIO 数据标注处理，处理结果分别为 $v_{i,j}$ 和 $b_{i,j}$；
3) 构建 Bert 预训练模型，利用该模型将预处理后的特征词 $v_{i,j}$ 转换为词嵌入向量 $\boldsymbol{e}_{i,j}$；
4) 直到遍历完文本词序列结束循环。
5) end for
6) 定义深度学习模型 epoch 和 batch 参数；
7) for each epoch do
8) for each batch do
9) 构建 BiLSTM 模型，利用该模型学习经过 Bert 模型处理的向量 $\boldsymbol{e}_{i,j}$，输出结果为 $h_{i,j}$；
10) 构建注意力机制模型，通过注意力机制捕

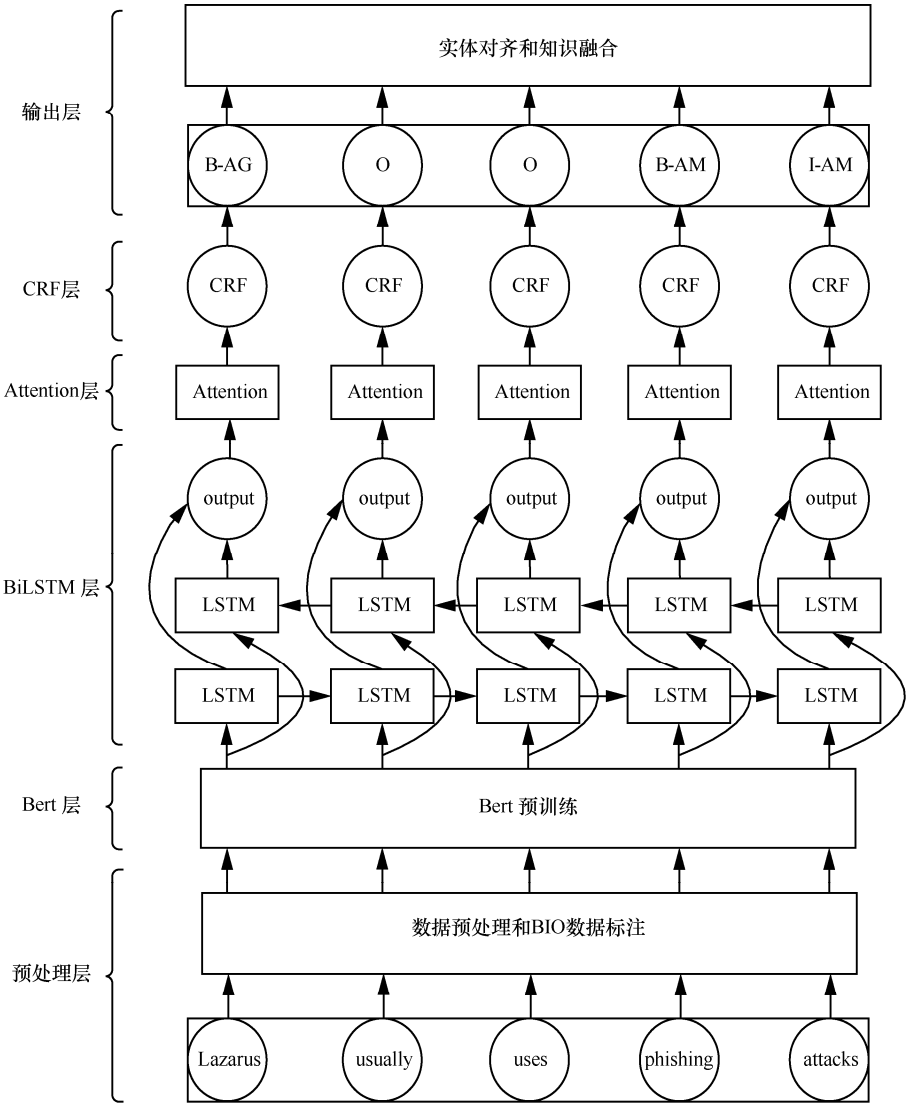


图 2　本文模型总体框架

获关键特征，并记为 $a_{i,j}$；

11) 迭代更新神经网络参数和权重；

12) 构建 CRF 模型计算关键特征所属类别，其输出结果为 $o_{i,j}$；

13) 比较命名实体识别预测的结果 $o_{i,j}$ 与真实类别 $b_{i,j}$，评估模型性能，直至模型收敛。

14) end for

15) end for

16) 构建实体对齐和知识融合算法，实现特征词 $w_{i,j}$ 和实体类别 $o_{i,j}$ 的知识融合；

17) 输出 APT 攻击结构化实体知识 E。

3.2　Bert 模型

Bert[27]是谷歌 2018 年提出的预训练语言模型，通过双向 Transformer 更好地捕捉语句中的双向关系。Bert 模型充分考虑词嵌入、句嵌入和位置嵌入的关系特征，增强了字向量的语义表示，从而获取高质量的词向量。本文通过该模型预训练 APT 领域知识，使用多个 Transformer 双向编码器对字符进行编码，其会将输入句子中的每个词都和句中所有词做注意力计算，从而获取词间的相互关系，捕获句子内部结构。这能在一定程度上反映不同词语之间存在的关系和重要程度，有效解决 NLP 中的长依赖问题，计算式为

$$\text{Attention}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V})=\text{softmax}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d_k}}\right)\boldsymbol{V} \tag{2}$$

其中，$\boldsymbol{Q}$、$\boldsymbol{K}$、$\boldsymbol{V}$ 分别表示 Query、Key 和 Value 向量，它们是编码器的输入字向量矩阵；d_k 表示输入

向量的维度。由于 Bert 模型能够将学习到的语义知识通过迁移学习应用到数据标注较少的命名实体任务上，因此本文选用该模型进行 APT 攻击命名实体识别的上游预处理任务，在一定程度上减少数据标记工作，从而更好地挖掘 APT 文本中的特征信息。

3.3 BiLSTM 模型

LSTM 是一种典型的循环神经网络，能解决训练时产生的梯度爆炸或梯度消失问题。LSTM 核心结构包括遗忘门、输入门、输出门和记忆单元。整个记忆单元由细胞状态 C_t 来调节，输入门和遗忘门共同保留重要信息，遗忘无用信息。LSTM 结构的计算式为

$$\begin{cases} i_t = \sigma(\boldsymbol{W}_{xi}x_t + \boldsymbol{W}_{hi}h_{t-1} + \boldsymbol{W}_{ci}c_{t-1} + b_i) \\ z_t = \tanh(\boldsymbol{W}_{xc}x_t + \boldsymbol{W}_{hc}h_{t-1} + b_c) \\ f_t = \sigma(\boldsymbol{W}_{xf}x_t + \boldsymbol{W}_{hc}h_{t-1} + \boldsymbol{W}_{cf}c_{t-1} + b_f) \\ c_t = f_t c_{t-1} + i_t z_t \\ o_t = \tanh(\boldsymbol{W}_{xo}x_t + \boldsymbol{W}_{ho}h_{t-1} + \boldsymbol{W}_{co}c_t + b_o) \\ h_t = o_t \tanh(c_t) \end{cases} \tag{3}$$

其中，i_t、f_t 和 o_t 分别表示 t 时刻输入门、遗忘门和输出门的结果，c_t 表示 t 时刻细胞状态，σ 和 tanh 表示 2 种不同的神经元激活函数，$\boldsymbol{W}$ 表示连接两层神经元的权重矩阵（如 $\boldsymbol{W}_{xi}$ 表示输入层到隐藏层的输入门权重），b 表示偏置项，x_t、h_t 和 z_t 分别表示 t 时刻的输入变量、隐藏层变量和增量。

本文使用 BiLSTM 模型更好地捕获 APT 数据集的语义特征和长距离依赖信息。该模型由前向 LSTM 和后向 LSTM 组成，从前后 2 个方向对 APT 攻击文本的实体进行识别，从而提高具有前后关联实体识别的性能，如“SQL Injection”“OceanLotus Group”“XSS Attack”等。

3.4 CRF 算法

CRF 是一种判别式概率无向图模型，在给定输入随机变量的情况下，能计算输出随机变量的条件概率分布。在命名实体识别中，BiLSTM 模型能够捕获长距离的文本信息，但无法感知实体及相邻标签间的依赖关系，并且 APT 攻击领域的实体依赖关系更加复杂。CRF 算法能有效解决该问题，它考虑标签之间的转移关系并计算整体标签序列的概率，从而获取全局最优的标记序列。因此，本文在 BiLSTM 模型后连接一个 CRF 模型，用以提升 APT 攻击命名实体的识别效果。

本文使用线性链条件随机场，对于任一输入序列 $\boldsymbol{X}=(\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n)$，其中 $\boldsymbol{x}_i$ 为第 i 个单词的输入向量，假定 $\boldsymbol{s}$ 是 BiLSTM 模型的输出得分矩阵，$\boldsymbol{s}$ 由 n 个单词和 k 个标签组成，s_{ij} 表示第 i 个单词的第 j 个标签的分数。对于预测标签序列 $\boldsymbol{Y}=(\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_n)$，其得分函数计算规则为分数越大则标签的可能性越高。

$$\text{Score}(\boldsymbol{X}, \boldsymbol{Y}) = \sum_{i=0}^{n} \boldsymbol{A}_{y_i, y_{i+1}} + \sum_{i=1}^{n} s_{i, y_i} \tag{4}$$

其中，$\boldsymbol{A}$ 表示转移矩阵，旨在完成标签之间的分数转移，A_{ij} 表示标签 i 转移为标签 j 的概率，即转移分数。同时，在 k 个标签的基础上增加“开始”和“结束”2 个标签，它们对应的 $\boldsymbol{s}$ 分数为 0。CRF 模型会采用对数最大似然估计计算损失函数使正确的序列的概率最大，再解码得到最大分数的输出序列，作为最终 APT 攻击实体识别的标注结果。

3.5 实体对齐算法

实体对齐旨在确定 2 个待消解的实体是否指向同一个目标实体，又称为实体消解。本文通过上述步骤抽取不同来源的 APT 攻击知识，包含 APT 组织、攻击装备等在内的 12 种命名实体，接着构建基于语义相似度和 Birch 聚类的实体对齐算法，将每篇 APT 分析报告的实体映射成融合上下文语义的词向量，再对 APT 组织名称执行实体对齐与知识融合，从而自动化生成 APT 组织的知识消息盒（InfoBox）。

实体对齐算法流程如图 3 所示。通过融合 Bert 和 BiLSTM-CRF 模型完成命名实体识别，抽取 APT 文本的结构化知识，再利用 Word2Vec 转换成融合语义的词向量，构建实体相似度计算模型，通过语义相似度距离来实现基于 Birch 算法的聚类，从而将相似的命名实体进行对齐，最后将不同 APT 组织共现的实体进行知识融合，形成最终的 APT 组织指纹特征库以及知识消息盒。

4 实验结果与分析

4.1 实验数据集及预处理

本文采集来自 Mitre ATT&CK、FireEye、McAfee、Kaspersky 等安全公司的 APT 分析报告，经过数据清洗及文本拆分后共形成包含 466 篇文本的数据集，覆盖了全球各地区的 APT 组织，典型的 APT 组织包括 APT28、APT29、APT32、APT33、APT37、Lazarus、Turla 等，其 12 种 APT 领域实体

类别数量分布如表 2 所示。整个实验数据集按照一定比例随机划分为训练集、测试集和验证集，并且测试集中存在未知的 APT 命名实体。

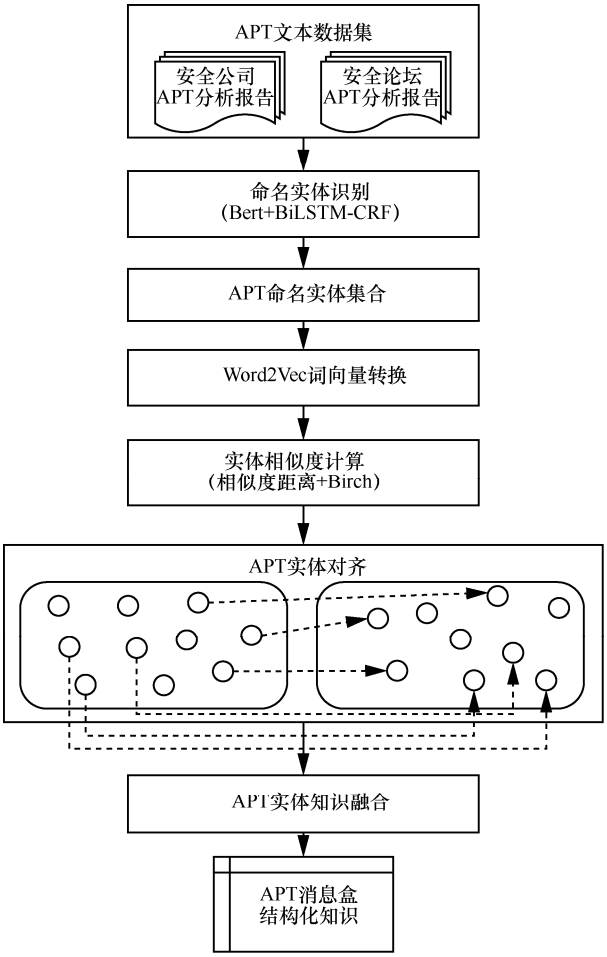


图 3　实体对齐算法流程

接着，利用 BIO 标记法分别对不同类别的实体进行数据标注。单词起始位置使用“B-”字符加表 1 定义的标志符号，如“B-AM”表示攻击手法的起始位置；单词中间位置使用“I-”字符加表 1 定义的标志符号；其他不属于任何实体的单词使用 O 表示，通过该方法能有效给出实体对应的类型及位置。图 4 是第 2 节介绍示例对应的实体标注结果。从图 4 可以看到，APT 组织、地理位置、攻击事件、攻击装备、攻击漏洞等均被标注。

表 2　APT 领域实体类别数量分布

实体类别	实体数量/个	实体类别	实体数量/个
APT 组织	128	攻击行业	34
攻击装备	651	恶意文件	116
攻击手法	65	恶意软件家族	38
攻击漏洞	60	区域位置	72
攻击事件	16	操作系统	5
攻击目标	31	利用软件	48

4.2　评价指标

实验采用 4 个常用于评价实体识别的指标来衡量算法的有效性，分别是精确率（Precision）、召回率（Recall）、F_1 值（F_1-score）和准确率（Accuracy），计算式分别为

$$\text{Precision}=\frac{\text{TP}}{\text{TP}+\text{FP}} \tag{5}$$

$$\text{Recall}=\frac{\text{TP}}{\text{TP}+\text{FN}} \tag{6}$$

$$F_1\text{-score}=\frac{2\times\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}} \tag{7}$$

$$\text{Accuracy}=\frac{\text{TP}+\text{TN}}{\text{TP}+\text{TN}+\text{FP}+\text{FN}} \tag{8}$$

其中，精确率表示正确识别的实体数量占识别出的

Lazarus(B-AG) Group(I-AG) is(o) a(o) threat(o) group(o) that(o) has(o) been(o) attributed(o) to(o) the(o) North(B-RL) Korean(I-RL) government(o). The(o) group(o) has(o) been(o) active(o) since(o) at(o) least(o) 2009(o) and(o) was(o) reportedly(o) responsible(o) for(o) the(o) November(o) 2014(o) destructive(o) wiper(o) attack(o) against(o) Sony(B-AT) Pictures(o) Entertainment(o) as(o) part(o) of(o) a(o) campaign(o) named(o) Operation(B-AE) Blockbuster(I-AE) by(o) Novetta(o). Malware(o) used(o) by(o) Lazarus(B-AG) Group(I-AG) correlates(o) to(o) other(o) reported(o) campaigns(o), including(o) Operation(B-AE) Flame(I-AE), Operation(B-AE) 1Mission(I-AE), Operation(B-AE) Troy(I-AE), DarkSeoul(B-AE), and Ten(B-AE) Days(I-AE) of(I-AE) Rain(I-AE). In(o) late(o) 2017(o), Lazarus(B-AG) Group(I-AG) used(o) KillDisk(B-AEQ), a(o) disk-wiping(o) tool(o), in(o) an(o) attack(o) against(o) an(o) online(o) casino(o) based(o) in(o) Central(B-RL) America(I-RL).

North(B-RL) Korean(I-RL) group(o) definitions(o) are(o) known(o) to(o) have(o) significant(o) overlap(o), and(o) the(o) name(o) Lazarus(B-AG) Group(I-AG) is(o) known(o) to(o) encompass(o) a(o) broad(o) range(o) of(o) activity(o). Some(o) organizations(o) use(o) the(o) name(o) Lazarus(B-AG) Group(I-AG) to(o) refer(o) to(o) any(o) activity(o) attributed(o) to(o) North(B-RL) Korea(I-RL). Some(o) organizations(o) track(o) North(B-RL) Korean(I-RL) clusters(o) or(o) groups(o) such(o) as(o) Bluenoroff(B-AG), APT37(B-AG), and(o) APT38(B-AG) separately(o), while(o) other(o) organizations(o) may(o) track(o) some(o) activity(o) associated(o) with(o) those(o) group(o) names(o) by(o) the(o) name(o) Lazarus(B-AG) Group(B-AG). The(o) group(o) often(o) uses(o) the(o) CVE-2017-11882(B-AV), CVE-2018-4878(B-AV), and(o) MS17-010(B-AV) vulnerabilities(o).

图 4　APT 文本经过 BIO 序列标注示例

所有实体数量的百分比，从查准的角度评估模型；召回率表示正确识别的实体数量占所有该类标注实体数量的百分比，从查全的角度评估模型；F_1值表示精确率和召回率的调和平均数，从查准和查全2个角度综合反映模型的效果，本文用来对实体识别实验进行整体评估，F_1值越大，表明模型正确识别的实体数量越多且越全；准确率表示分类预测正确实体数量占该类别实体总数量的比值。为验证本文模型的有效性和真实性，最终结果为10次实验结果的平均值，从而避免噪声影响。

4.3 实体识别实验

实验数据集按照6:3:1的比例随机划分为训练集、验证集和测试集，实验环境为Windows10 64位操作系统，GPU为GTX 1080Ti，内存为16 GB，CPU处理器为Inter(R) Core i7-8700K，编程语言为Python3.7。模型参数设计方面，文本序列长度设为500，BiLSTM模型2个方向的神经元数设置为256，采用Adam优化器，Epoch设置为15，初始学习率设置为0.001，Transformer层数设置为12，并且增加Dropout防止过拟合，其参数设置为0.4。

本文提出一种融合Bert和BiLSTM-CRF模型的APT攻击实体识别方法，与现有常见的实体识别方法（CRF、LSTM-CRF、GRU-CRF、BiLSTM-CRF、CNN-CRF和Bert-CRF）的对比结果如表3所示。

表3 各模型实体识别结果对比

模型	Precision	Recall	F_1-score	Accuracy
CRF	0.800 2	0.719 0	0.757 4	0.741 3
LSTM-CRF	0.877 7	0.762 8	0.816 3	0.816 9
GRU-CRF	0.889 6	0.748 9	0.813 2	0.805 2
BiLSTM-CRF	0.951 4	0.764 3	0.847 6	0.846 6
CNN-CRF	0.853 6	0.810 4	0.831 4	0.826 4
Bert-CRF	0.923 6	0.754 2	0.830 3	0.814 5
本文模型	0.929 6	0.873 3	0.900 6	0.900 4

由表3可知，本文模型在APT攻击领域的实体识别任务中能够取得较好效果，其精确率、召回率和F_1值分别为0.929 6、0.873 3和0.900 6，比现有6种模型均有一定程度的提升。相比于CRF，本文模型的F_1值提升了0.143 2；相比于融合卷积神经网络的CNN-CRF，本文模型的F_1值提升了0.069 2；相比于LSTM-CRF和BiLSTM-CRF，本文模型的F_1值分别提升了0.084 3和0.053 0；相比于GRU-CRF，本文模型的F_1值提升了0.087 4；相比于Bert-CRF，本文模型的F_1值提升了0.070 3。同时，本文模型的准确率为0.900 4，比其他6种模型的平均值高0.098 5。通过实验发现，融合Bert和BiLSTM-CRF及注意力机制的实体识别模型具有最佳的效果，其主要原因是Bert预训练能更好地表示APT攻击领域知识，并且BiLSTM网络能学习上下文语义信息，注意力机制能有效突出关键特征。

为更形象地表现本文模型的良好性能，使用训练和验证数据进一步评估模型的学习过程，得出如图5所示的准确率变化曲线和如图6所示的误差变化曲线。由图5可知，与其他模型相比，本文模型训练过程更加稳定，整个曲线收敛速度更快，能在较少训练周期下取得较高的准确率。图6展现了各模型的误差随训练周期变化的曲线，本文模型的误差随训练周期收敛速度更快，曲线更平缓，这进一步体现了本文模型应用到APT实体识别是可行的。

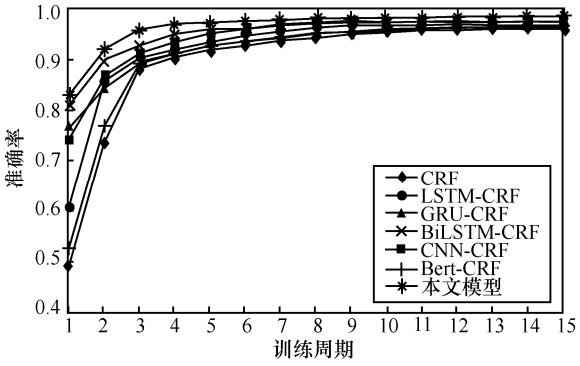


图5 各模型准确率变化曲线

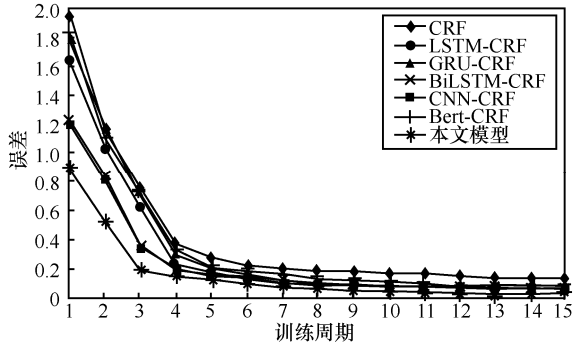


图6 各模型误差变化曲线

为进一步衡量本文模型对APT攻击领域不同类别实体的识别效果，本节进行了详细的对比实验，得出如表4所示的12种APT攻击实体类别的识别结果。

表 4　APT 攻击领域不同类别的实体识别结果

实体类别	Precision	Recall	F_1-score
APT 组织	0.937 0	0.877 3	0.906 2
攻击装备	0.951 7	0.892 2	0.921 0
攻击手法	0.947 3	0.908 5	0.927 5
攻击漏洞	0.907 4	0.859 6	0.882 9
攻击事件	0.930 4	0.856 0	0.891 7
攻击目标	0.931 8	0.872 3	0.901 1
攻击行业	0.925 2	0.900 4	0.912 6
恶意文件	0.910 2	0.832 9	0.869 8
恶意软件家族	0.890 9	0.830 5	0.859 6
区域位置	0.947 5	0.874 1	0.909 3
操作系统	0.942 1	0.890 6	0.915 7
利用软件	0.933 8	0.885 3	0.908 9
平均结果	0.929 6	0.873 3	0.900 6

由表 4 可知，本文模型在“攻击手法”实体类别上的预测效果最佳，其 F_1 值为 0.927 5，这一方面是由于该类别的实体数量较多，另一方面是该类实体广泛存在于富含语义的 APT 攻击事件中，并且带有攻击行为的动作特征，从而导致其识别效果更好。接下来，识别效果较好的实体类别包括“攻击装备”“攻击行业”“APT 组织”等，这些实体也都具有上下文语义突出和广泛存在于 APT 分析报告中的特点，比如“APT-C-36”“Lazarus Group”“APT28”属于 APT 组织实体，“CobaltStrike”“PowerShell”“Mimikatz”属于攻击装备实体。然而，本文实验识别效果最差的类别是“恶意软件家族”，其 F_1 值为 0.859 6，这是由于该类实体数量较少，其构造规则缺乏规律，常混合出现于上下文段落中，较难与普通特征词区别，从而识别困难。

此外，本文结合 ATT&CK 框架，选取了 4 种类别数量较多且常出现于 APT 攻击事件中的实体类别进行 F_1 值比较，包括攻击装备、APT 组织、攻击手法和利用软件，其详细的对比实验结果（F_1 值）如图 7 所示。由图 7 可知，本文模型在 4 种类别的实体识别中 F_1 值均最高。其中，本文模型攻击装备的 F_1 值为 0.921 0，比 6 种对比模型的平均 F_1 值提升了 0.059 6；本文模型 APT 组织的 F_1 值为 0.906 2，比 6 种对比模型的平均 F_1 值提升了 0.046 0；本文模型攻击手法和利用软件的 F_1 值分别为 0.927 5 和 0.908 9，比 6 种对比模型的平均 F_1 值分别提升了 0.043 0 和 0.124 9。综上所述，本文模型能有效对 APT 攻击领域的命名实体进行识别，其效果优于传统的实体识别模型，且能在各实体类别上取得良好性能。

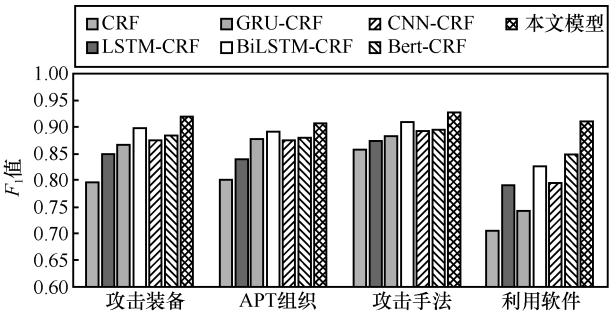


图 7　常见 4 种实体类别的 F_1 值对比结果

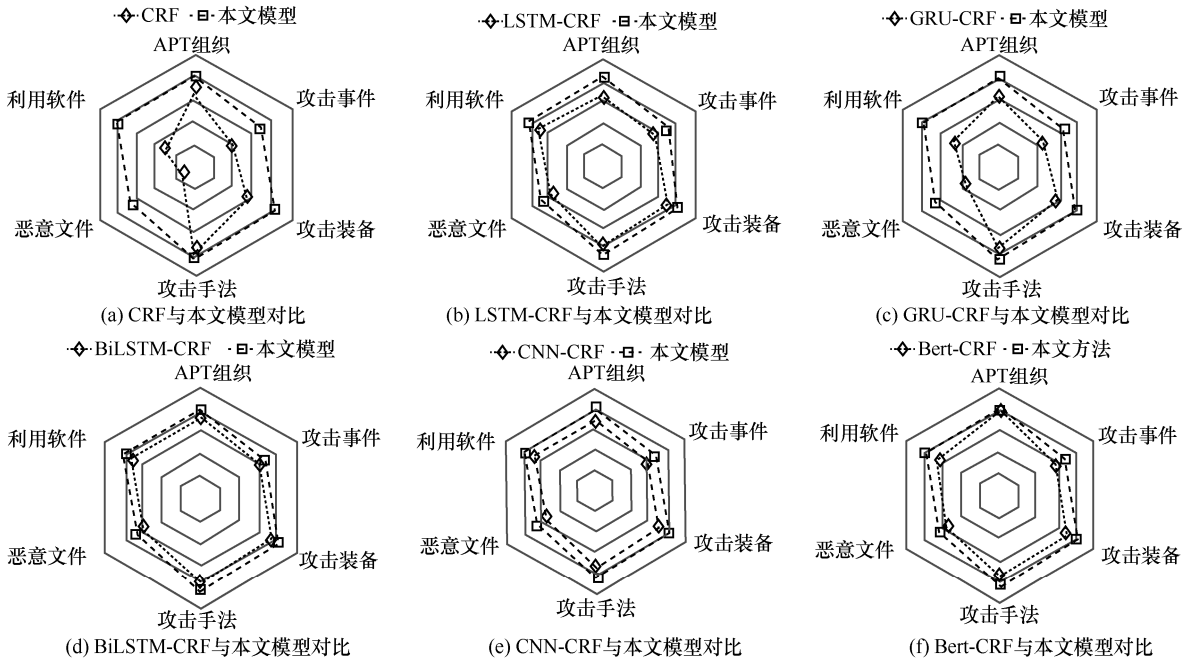
4.4　小样本标注的实体识别实验

为有效评估各模型对小样本标注的实体识别效果，本文按照 2:7:1 的比例随机划分训练集、测试集和验证集，从而降低训练集标注知识，并进行详细的对比实验，实验结果如表 5 所示。由表 5 可知，本文模型在小样本标注情况下，实体识别的精确率、召回率和 F_1 值分别为 0.780 0、0.589 4 和 0.671 4。其 F_1 值比 CRF 模型提升了 0.274 2，比 LSTM-CRF 模型提升了 0.187 8，比 GRU-CRF 模型提升了 0.236 2，比 BiLSTM-CRF 模型提升了 0.132 5，比 CNN-CRF 模型提升了 0.148 8，比 Bert-CRF 模型提升了 0.144 6。该实验充分说明了本文方法能通过 Bert 模型对小样本语料开展预训练，从而提升实体识别的效果。

表 5　各模型小样本标注的实体识别结果对比

模型	Precision	Recall	F_1-score
CRF	0.467 0	0.345 5	0.397 2
LSTM-CRF	0.553 6	0.429 2	0.483 6
GRU-CRF	0.557 0	0.357 1	0.435 2
BiLSTM-CRF	0.623 7	0.474 4	0.538 9
CNN-CRF	0.612 4	0.455 7	0.522 6
Bert-CRF	0.591 1	0.475 1	0.526 8
本文模型	0.780 0	0.589 4	0.671 4

接着，本文结合 APT 攻击流程和实体类别分布数量，在小样本标注情况下，详细对比了现有模型与本文模型在六大常见实体类别（APT 组织、攻击事件、攻击装备、攻击手法、恶意文件、利用软件）的 F_1 值，其实验结果如图 8 所示。通过雷达图能有效反映本文模型在不同类别的实体识别中取得的最佳效果。总之，本文模型能在少量样本标注的情况下实现实体识别，并取得更好的效果。



(a) CRF与本文模型对比　(b) LSTM-CRF与本文模型对比　(c) GRU-CRF与本文模型对比

(d) BiLSTM-CRF与本文模型对比　(e) CNN-CRF与本文模型对比　(f) Bert-CRF与本文模型对比

图 8　各模型小样本标注的 F_1 值对比

4.5　实体对齐与知识融合实验

APT 攻击领域通过实体识别任务能有效提取攻击实体，再通过本文提出的实体对齐算法能进一步形成各 APT 组织的结构化知识。表 6 显示了本文实验自动化抽取各类实体类别出现频率较高的命名实体，这些实体常常存在于 APT 攻击事件中。比如常见 APT 组织包括“APT29”“APT32”“APT28”和“Turla”等，常见攻击装备包括“PowerShell”“Cobalt Strike”和“Mimikatz”等，常见攻击手法包括“Spearphishing”“C2”“Watering Hole Attack”和“Backdoor”等，常见漏洞包括“CVE-2017-11882”“CVE-2017-0199”和“CVE-2012-0158”等。

经过实体识别处理后，为更好地抽取 APT 攻击领域知识，并为后续 APT 知识图谱或特征指纹库构建提供支撑，本文对 APT 组织名称进行了实体对齐与知识融合实验。本文通过基于语义相似度和 Birch 聚类的实体对齐算法将不同来源的组织命名实体进行匹配，判断其是否指向同一个目标实体，比如“APT28”又称为“Sofacy”“Fancy Bear”“Strontium”“Sednit”，这些 APT 组织对应的目标实体均相同。

表 6　　**本文模型所提取 APT 攻击领域的常见命名实体**

实体类别	被成功识别的命名实体
APT 组织	APT29; APT32; APT28; Turla; Sandworm; MuddyWater; OilRig; APT39; Kimsuky; FIN7; TA505
攻击装备	PowerShell; Cobalt Strike; Mimikatz; LaZagne; Cannon; Dropper; Empire; NBTscan; TrickBot; FireMalv
攻击手法	Spearphishing; C2; Anti-censorship; Backdoor; Payload; Persistence; SQL injection; Watering Hole Attack
攻击漏洞	CVE-2017-11882; CVE-2017-0199; CVE-2012-0158; CVE-2019-19781; CVE-2014-4114; CVE-2018-0802
攻击事件	DarkSeoul; Operation Blockbuster; Operation Flame; SolarWinds; Clinton Campaign; Stuxnet
攻击目标	NATO; Nuclear Facility; OPCW; Sony; ASEAN; World Health Organization; High-tech Company
攻击行业	Government; Espionage; Industry; Military Institutions; Financial Company; Manufacturing; Telecommunication
恶意文件	mshta.exe; wmiexec.vbs; rundl132.exe; backup.pst; csrss.exe; regsvr32.exe; sqlceip.exe; msfte.dll; pubprn.vbs
恶意软件家族	Trojan; Agent; Denes; Gh0st; Beacon; MSOffice.Alien.gen; CoreShell; Win32.Mimikatz; Win32.Cobalt
区域位置	America; Russia; North Korea; Iran; South Asia; Europe; U.S.; India; Germany
操作系统	Windows; Linux; Android; Mac OS; Unix; IOS; Kernel Operating System
利用软件	Office; Firefox; Word; RDP; Microsoft Exchange; Outlook; Adobe; WinRAR; PDF; Defender; Gmail

本文结合语料标题和关键词对 APT 组织名称开展实体融合，最终构建了该数据集常见 APT 组织的知识消息盒，形成各 APT 组织的结构化知识。

表 7 和表 8 分别呈现 APT28 和 APT32 经过实体识别和实体对齐实验后的攻击领域知识。它们既包括本文模型所识别的实体知识，如攻击装备、攻击手法、攻击漏洞、攻击目标、恶意文件等，又包括经过实体对齐后的 APT 组织名称，有效融合 APT 攻击领域知识，为后续知识图谱构建提供支撑。

表 7　　APT28 常见的实体知识展示

实体类别	实体知识
APT 组织	APT28; Fancy Bear; Sofacy;Sednit; Strontium
攻击装备	PowerShell; Mimikatz; Koadic; JHUHUGIT; Dropper
攻击手法	Spearphishing; C2; Persistence; Script; DDoS; Backdoor
攻击漏洞	CVE-2015-1701; CVE-2017-0263;CVE-2017-0262
攻击事件	the Hillary Clinton Campaign; VPNFilter
攻击目标	NATO; WADA; OSCE; OPCW; Nuclear Facility
攻击行业	Government; Industry; Organization; Education
恶意文件	rundll32.exe; explorer.exe; twain_64.dll; srhost.exe
恶意软件家族	ChopStick; Trojan; Win32.Dynamer; Zebrocy
区域位置	Russia; U.S.; Europe; India; Germany; U.K.; Israel
操作系统	Windows; Android
利用软件	Office; Microsoft Exchange; Gmail; PDF; NetBIOS; Delphi

表 8　　APT32 常见的实体知识展示

实体类别	实体知识
APT 名称	APT32; SeaLotus; OceanLotus; APT-C-00
攻击装备	Cobalt Strike; PowerShell; Mimikatz; RC4; DKMC
攻击手法	Backdoor; C2; Scheduled task; Spearphishing; Script
攻击漏洞	CVE-2017-11882; CVE-2016-7255; CVE-2017-8759
攻击事件	Cobalt Kitty; OceanLotus Blossoms
攻击目标	ASEAN; Asian Nations; the Media; Civil Society
攻击行业	Government; Military Institutions; Industry
恶意文件	pubprn.vbs; rundll32.exe; regsvr32.exe; kb-10233.exe
恶意软件家族	Denis; Gh0st; Trojan; Beacon; Win32.Agent
区域位置	Vietnam; Cambodia; Philippine; China; Laos
操作系统	Windows; Mac OS
利用软件	Office; Outlook; COM; RTF; Dropbox; Amazon S3

5　结束语

本文设计并实现了一种融合实体识别和实体对齐的 APT 攻击知识自动抽取方法。该方法结合 APT 攻击特点设计了 12 种实体类别，构建了融合 Bert 和 BiLSTM-CRF+Attention 的 APT 实体识别模型，再结合实体对齐生成了不同 APT 组织的结构化知识。实验结果表明，本文模型能有效识别 APT 攻击实体，在少样本标注的情况下自动抽取高级可持续威胁知识，并生成常见 APT 组织的结构化特征画像。这将为后续 APT 攻击知识图谱构建和攻击溯源分析提供帮助。在下一步工作中，笔者将针对中文 APT 分析报告开展知识自动抽取研究，并结合图神经网络开展攻击关系抽取及攻击事件推理。

参考文献：

[1] STOJANOVIĆ B, HOFER-SCHMITZ K, KLEB U. APT datasets and attack modeling for automated detection methods: a review[J]. Computers & Security, 2020, 92: 101734.

[2] WANG W, ZHU M, ZENG X W, et al. Malware traffic classification using convolutional neural network for representation learning[C]// Proceedings of 2017 International Conference on Information Networking (ICOIN). Piscataway: IEEE Press, 2017: 712-717.

[3] LUO Y, XIAO Y, CHENG L, et al. Deep learning-based anomaly detection in cyber-physical systems: progress and opportunities[J]. ACM Computing Surveys, 2021, 54(5): 106: 1-36.

[4] MILAJERDI S M, GJOMEMO R, ESHETE B, et al. HOLMES: real-time APT detection through correlation of suspicious information flows[C]//Proceedings of 2019 IEEE Symposium on Security and Privacy. Piscataway: IEEE Press, 2019: 1137-1152.

[5] MARCHETTI M, PIERAZZI F, COLAJANNI M, et al. Analysis of high volumes of network traffic for advanced persistent threat detection[J]. Computer Networks, 2016, 109: 127-141.

[6] HAN X Y, PASQUIER T, BATES A, et al. Unicorn: runtime provenance-based detector for advanced persistent threats[C]//Proceedings 2020 Network and Distributed System Security Symposium. Reston: Internet Society, 2020: 1-19.

[7] LANGNER R. Stuxnet: dissecting a cyberwarfare weapon[J]. IEEE Security & Privacy, 2011, 9(3): 49-51.

[8] MUCKIN M, FITCH S C. A threat-driven approach to cyber security[J]. Lockheed Martin Corporation, 2015, 3(1): 1-8.

[9] 宋文纳, 彭国军, 傅建明, 等. 恶意代码演化与溯源技术研究[J]. 软件学报, 2019, 30(8): 2229-2267.
SONG W N, PENG G J, FU J M, et al. Research on malicious code evolution and traceability technology[J]. Journal of Software, 2019, 30(8): 2229-2267.

[10] GIURA P, WANG W. A context-based detection framework for advanced persistent threats[C]//Proceedings of 2012 International Conference on Cyber Security. Piscataway: IEEE Press, 2012: 69-74.

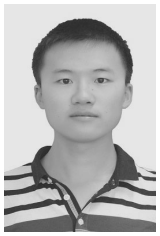
[11] KIM Y H, PARK W H. A study on cyber threat prediction based on intrusion detection event for APT attack detection[J]. Multimedia Tools and Applications, 2014, 71(2): 685-698.

[12] 付钰, 李洪成, 吴晓平, 等. 基于大数据分析的 APT 攻击检测研究综述[J]. 通信学报, 2015, 36(11): 1-14.
FU Y, LI H C, WU X P, et al. Detecting APT attacks: a survey from the perspective of big data analysis[J]. Journal on Communications, 2015, 36(11): 1-14.

[13] YANG H P. Method for behavior-prediction of APT attack based on dynamic Bayesian game[C]//Proceedings of 2016 IEEE International Conference on Cloud Computing and Big Data Analysis. Piscataway: IEEE Press, 2016: 177-182.

[14] 张小松, 牛伟纳, 杨国武, 等. 基于树型结构的 APT 攻击预测方法[J]. 电子科技大学学报, 2016, 45(4): 582-588.
ZHANG X S, NIU W N, YANG G W, et al. Method for APT prediction based on tree structure[J]. Journal of University of Electronic Science and Technology of China, 2016, 45(4): 582-588.
[15] MILAJERDI S M, ESHETE B, GJOMEMO R, et al. POIROT: aligning attack behavior with kernel audit records for cyber threat hunting[C]// Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security. New York: ACM Press, 2019: 1813-1830.
[16] HUMPHREYS K, GAIZAUSKAS R, AZZAM S, et al. University of sheffield: description of the LaSIE-II system as used for MUC-7[C]// Proceedings of the Seventh Message Understanding Conferences. Stroudsburg: ACL Press, 1998: 1-20.
[17] BLACK W J, RINALDI F R, MOWATT D. Facile: description of the NE system used for MUC-7[C]//Proceedings of the Seventh Message Understanding Conference. Stroudsburg: ACL Press, 1998: 1-10.
[18] COLLINS M, SINGER Y. Unsupervised models for named entity classification[C]//Proceedings of the Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora. Stroudsburg: ACL Press, 1999: 100-110.
[19] FREITAG D, MCCALLUM A. Information extraction with HMMs and shrinkage[C]//Proceedings of the AAAI-99 Workshop on Machine Learning for Information Extraction. Palo Alto: AAAI Press, 1999: 31-36.
[20] CHIEU H L, NG H T. Named entity recognition: a maximum entropy approach using global information[C]//Proceedings of the 19th International Conference on Computational Linguistics. Stroudsburg: ACL Press, 2002: 1-7.
[21] LI Y Y, BONTCHEVA K, CUNNINGHAM H. SVM based learning system for information extraction[C]//International Workshop on Deterministic and Statistical Methods in Machine Learning. Berlin: Springer, 2005: 319-339.
[22] MCCALLUM A, LI W. Early results for named entity recognition with conditional random fields, feature induction and web-enhanced lexicons[C]//Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL. Stroudsburg: ACL Press, 2003: 188-191.
[23] HAMMERTON J. Named entity recognition with long short-term memory[C]//Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL. Stroudsburg: ACL Press, 2003: 172-175.
[24] STRUBELL E, VERGA P, BELANGER D, et al. Fast and accurate entity recognition with iterated dilated convolutions[C]//Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. Stroudsburg: ACL Press, 2017: 2670-2680.
[25] ZHANG Y, YANG J. Chinese NER using lattice LSTM[C]//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. Stroudsburg: ACL Press, 2018: 1554-1564.
[26] 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.
[27] DEVLIN J, CHANG M W, LEE K, et al. BERT: pre-training of deep bidirectional transformers for language understanding[C]//Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Stroudsburg: ACL Press, 2019. 4171-4186.

[作者简介]



杨秀璋（1991− ），男，贵州凯里人，武汉大学博士生，主要研究方向为网络与信息系统安全。

彭国军（1979− ），男，湖北荆州人，博士，武汉大学教授、博士生导师，主要研究方向为网络与信息系统安全。

李子川（1999− ），男，河北邯郸人，武汉大学硕士生，主要研究方向为 IoT 安全、漏洞自动化挖掘与利用。

吕杨琦（1997− ），女，湖北孝感人，武汉大学硕士生，主要研究方向为网络与信息系统安全。

刘思德（1997− ），男，湖北荆州人，武汉大学博士生，主要研究方向为恶意代码检测与系统安全。

李晨光（1999− ），男，湖北十堰人，武汉大学硕士生，主要研究方向为网络与信息系统安全。