

基于词向量和图卷积的攻击模式与技术实体关联方法

裘炜程^{1,2}, 陈秀真^{1,2}, 马颖华^{1,2}, 马进^{1,2}, 周志洪^{1,2}

(1. 上海交通大学网络安全技术研究院, 上海 200240; 2. 上海市信息安全综合管理技术重点实验室, 上海 200240)

摘要: 安全威胁分析以包含大量安全实体的网络安全知识库为基础, 对威胁源、攻击能力、攻击动机、威胁路径等建模, 并结合资产的脆弱性及已部署的安全措施, 评估威胁的影响范围、程度以及带来的安全风险。但是基础知识库部分实体之间的关联关系缺失, 不利于实现安全事件追踪、攻击路径关联。针对 CAPEC 攻击模式和 ATT&CK 技术的实体关系缺失问题, 提出一种基于词向量和图卷积的攻击模式与技术实体关联方法, 即通过分析实体描述文本之间的关联度判断实体关系, 达到不同知识库间实体关系的补充、丰富威胁路径的目的。提出的方法使用预训练的安全领域 Word2Vec 模型提取特定领域的语义信息, 通过图卷积神经网络模型提取实体描述间的共现关系, 进一步将两种特征输入孪生网络预测实体间的关联关系。为解决现有知识库中实体关系较少的小样本学习问题, 通过词向量补充外部语义信息、模型训练时使用动态负采样和添加正则项避免过拟合。针对 MITRE 组织提供的 CAPEC 和 ATT&CK 知识库进行实验测试, 结果表明, 所提算法可以通过分析实体描述的语义信息, 将有关系的实体对在样本空间内与无关系实体对分离, 从而有效预测新的实体对的关联关系。在小样本条件下, 所提方法的预测准确率高于基于 Bert 的文本相似度预测方法, 同时训练时间更短、所需计算资源更少。实验结果证明, 基于词向量和图卷积的攻击模式与技术实体关联方法可以挖掘出新的攻击模式与技术实体关联关系, 安全威胁分析时有助于提高从安全漏洞和脆弱点等底层概念抽象出攻击技术和战术的能力。

关键词: 安全实体关联; 自然语言处理; 图卷积神经网络; 小样本学习

中图分类号: TP393

文献标志码: A

DOI: 10.11959/j.issn.2096-109x.2023052

Predicting correlation relationships of entities between attack patterns and techniques based on word embedding and graph convolutional network

QIU Weicheng^{1,2}, CHEN Xiuzhen^{1,2}, MA Yinghua^{1,2}, MA Jin^{1,2}, ZHOU Zhihong^{1,2}

1. Institute of Cyber Science and Technology, Shanghai Jiao Tong University, Shanghai 200240, China

收稿日期: 2022-06-07; **修回日期:** 2023-05-06

通信作者: 陈秀真, chenxz@sjtu.edu.cn

基金项目: 国家自然科学基金 (U2003206), 上海市科学技术委员会科技创新行动计划 (22511101202)

Foundation Items: The National Natural Science Foundation of China (U2003206), Shanghai Science and Technology Committee Science and Technology Innovation Action Plan (22511101202)

引用格式: 裘炜程, 陈秀真, 马颖华, 等. 基于词向量和图卷积的攻击模式与技术实体关联方法[J]. 网络与信息安全学报, 2023, 9(4): 40-52.

Citation Format: QIU W C, CHEN X Z, MA Y H, et al. Predicting correlation relationships of entities between attack patterns and techniques based on word embedding and graph convolutional network[J]. Chinese Journal of Network and Information Security, 2023, 9(4): 40-52.

2. Shanghai Municipal Key Lab of Integrated Management Technology for Information Security, Shanghai 200240, China

Abstract: Threat analysis relies on knowledge bases that contain a large number of security entities. The scope and impact of security threats and risks are evaluated by modeling threat sources, attack capabilities, attack motivations, and threat paths, taking into consideration the vulnerability of assets in the system and the security measures implemented. However, the lack of entity relations between these knowledge bases hinders the security event tracking and attack path generation. To complement entity relations between CAPEC and ATT&CK techniques and enrich threat paths, an entity correlation prediction method called WGS was proposed, in which entity descriptions were analyzed based on word embedding and a graph convolution network. A Word2Vec model was trained in the proposed method for security domain to extract domain-specific semantic features and a GCN model to capture the co-occurrence between words and sentences in entity descriptions. The relationship between entities was predicted by a Siamese network that combines these two features. The inclusion of external semantic information helped address the few-shot learning problem caused by limited entity relations in the existing knowledge base. Additionally, dynamic negative sampling and regularization was applied in model training. Experiments conducted on CAPEC and ATT&CK database provided by MITRE demonstrate that WGS effectively separates related entity pairs from irrelevant ones in the sample space and accurately predicts new entity relations. The proposed method achieves higher prediction accuracy in few-shot learning and requires shorter training time and less computing resources compared to the Bert-based text similarity prediction models. It proves that word embedding and graph convolutional network based entity relation prediction method can extract new entity correlation relationships between attack patterns and techniques. This helps to abstract attack techniques and tactics from low-level vulnerabilities and weaknesses in security threat analysis.

Keywords: security entity correlation, natural language processing, graph convolution neural network, few-shot learning

0 引言

安全威胁分析是识别、检查和减弱安全威胁的过程,使得安全专家可提前采取安全措施减少安全事件的发生,是风险评估的重要一环,其关键步骤是网络安全威胁模型构建。MITRE 公司提出杀伤链、STRIDE、ATT&CK(对抗性战术、技术与通用知识库)等安全威胁模型,建立了 CVE、CAPEC(通用攻击模式枚举和分类)、CWE、CPE 等基础知识库。其中,CAPEC 关注安全应用、描述并归类攻击者利用已知脆弱点所使用的通用技术,ATT&CK 关注网络防御,描述攻击周期、攻击前和攻击后的实施阶段,详细定义了高级持续性威胁(APT)在实施过程中的战术、技术和程序。同时,根据粗细粒度将当前威胁模型、威胁知识库划分为底层概念、中层模型和高层模型。

1) 底层概念:最细粒度的单位,包括 CVE、CWE、CAPEC 等,是具体攻击或漏洞的实例化描述。

2) 中层模型:即 ATT&CK^[1],从底层实例中抽取出攻击的技术、战术等,实现底层与高层之间的关联。

3) 高层模型:如杀伤链、STRIDE 等抽象模型,从宏观角度对威胁进行评估和分析。

分析三者的联系,有助于将底层概念映射到高层模型,从而对威胁事件有更全面的理解和定性。安全威胁分析往往从底层的 CVE、CWE、CAPEC、安全资产等开始,希望预测出目标系统可能或已经遭受的威胁和攻击、分析攻击的战术阶段,进而识别 APT 攻击等。图 1 给出一个安全实体关系预测案例,假设某安全资产中有设备存在使用短密码(CWE-326)、使用弱安全性的密码算法(CWE-327)等脆弱点,则其可能正在遭受暴力破解攻击(CAPEC-20)。然而,数据库中没有其相关的攻击技术,因此无法再向高层模型推理。

如果可以通过某些技术分析出 CAPEC-20 与 OS 凭证破解(T1003)或暴力破解密码(T1110.002)的相关性,则可进一步推理出系统可能或已经处于 ATT&CK 矩阵中攻击战术的凭证获取阶段(TA0006),从而完善安全威胁分析中底层概念到 ATT&CK 的映射分析。

迄今为止,研究人员提出中层模型到高层模型的映射方法,但是缺少对底层概念与中层模型

对应关系的研究。MITRE 公司建立的 CAPEC、ATT&CK 知识库中, 525 个 CAPEC 实例和 552 个 ATT&CK 技术实例间有 157 条关系, 仅涉及 20% 的 CAPEC 实例, 不利于底层概念到 ATT&CK 的映射分析。因此, 亟须一种自动快速的通用攻击模式与攻击技术的关联方法。

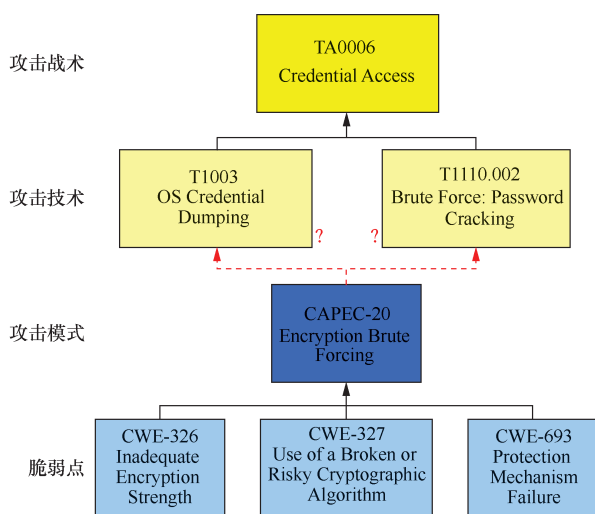


图1 安全实体关系预测案例
Figure 1 Use case of predicting security entities' relationship

本文针对知识库中 CAPEC 攻击模式与 ATT&CK 攻击技术关联关系少、不利于自动化分析的问题, 提出两类实体间的关联补充方法。所提方法基于有限数量的已知连接, 采用自然语言处理技术, 分析补充两者实例间的关系。该方法使用图神经网络对小样本数据建模, 避免了 Bert 类模型需要大量数据和算力进行预训练的要求; 预训练一个词向量模型 Word2Vec, 相较于使用通用语料库训练的 Bert 类模型, 能更好补充安全领域的文本特征和领域知识; 最后, 使用孪生网络将两种文本表示融合, 对实体间相似度直接建模。所提方法完善底层概念到中层模型的映射, 实现自动快速的通用攻击模式与攻击技术的关联, 克服人工确定两个知识库实体之间关联关系的专家知识依赖。

1 相关工作

本文基于自然语言处理技术, 通过分析攻击模式 CAPEC 和攻击技术 ATT&CK 的实体描述间的关联度, 预测两实体间的关系。网络安全实体关系是构建安全本体、安全知识图谱的基础, 研究者针对不同实体提出关系预测方法, 下面给出

该领域的相关工作。

Kiesling 等^[2]设计了多源网络安全知识图谱 SEPSES, 分析 CVE、CAPEC、CPE、CWE、CVSS 等不同知识库的数据采集、格式解析、关系抽取、存储方法等, 从中提取实体名称、实体属性和实体间的关系, 构建安全知识图谱, 并通过漏洞评估和入侵检测两个案例说明该图谱的价值。Guo 等^[3]从 CVE 知识库和安全报告文本中抽取实体和实体关系, 构建安全知识图谱, 将其建模为文本数据的多序列标注问题, 使用 Bert 得到文本初始向量、BiGRU 编码器和注意力机制提取特征, BiGRU 和 CRF 解码得到序列标注结果, 提取实体关系三元组并构建图谱。实验证明, 该方法可以达到 80% 的预测准确率。Xiao 等^[4]提出一种基于知识图谱的 CVE、CWE 和 CAPEC 实体嵌入和关系分析方法。该方法从公开知识库中提取 3 种实体的英语描述和实体间的关系, 使用卷积神经网络编码器和 Word2Vec 获取关系向量、实体向量和实体描述向量, 输入 TransH 模型中训练并预测新的连接。实验证明, 图谱中的实体种类越多、连接数量越多、实体关系种类越少, 则模型效果越好, 但总体的预测准确率越低。Yuan 等^[5]指出仅使用 TransH^[4]等基于图谱分析的方法, 忽略了实体周围邻居对该实体关系的影响, 预测效果不佳, 作者在 TransH 的基础上增加图注意力网络, 考虑图谱中实体的二阶邻居对其的影响。实验中, 同样使用 CVE、CWE 和 CAPEC 知识库, 结果表明其大幅提高了预测准确率。Das 等^[6]提出一个基于 Bert 的漏洞分类器 V2W-BERT, 通过分析 CVE 和 CWE 实体的描述, 预测 CVE 实体的 CWE 类别, 该方法设计重构解码器记录连接关系较少的 CVE 和 CWE 实体, 解决小样本问题, 实验表明其可准确预测 CVE 实体的 CWE 类别, 并有预测新 CWE 实体的能力。Milajerdi 等^[7]提出一种实时 APT 检测系统 HOLMES, 从系统溯源图中关联安全事件, 将底层系统日志映射到战术、技术和过程 (TTP), 再由 TTP 映射到 APT 攻击阶段, 最终提炼为攻击场景图。其中 TTP 到 APT 阶段的映射基于预先定义的规则, 尽管系统取得了较高的检测准确率, 但规则设计需要大量人工判断。

图神经网络 (GNN) 和孪生网络已广泛应用在自然语言处理的文本相似度、文本分类任务中^[8]。

TextGCN^[9]提出一种半监督的文档分类方法,将数据集中的所有文档和词构成图,通过图卷积神经网络标签传递的方式将有限的标签扩散到全图节点中,实验证明该方法效果优于监督学习模型。Liu 等^[10]指出当前工作缺少针对长文本相似度的分析,因此提出一种基于图分解和卷积的文本匹配方法。该方法首次提出用概念交互图表示文档,实验证明其提出的概念交互图可以有效表示文档、图卷积神经网络可以有效提取特征并完成相似度判断。但是该方法的效果依赖于概念交互图的生成质量,短文本可能无法提供有效的交互图。Reimers 等^[11]提出的 Sentence-bert 使用孪生网络架构,将句子对分别输入参数共享的 Bert,证明其在文本相似度任务上表现优于 Bert,并证明孪生网络在该任务的可行性。

现有的网络安全实体关联方法多集中于研究底层概念实体间的关系^[2-6]或中层模型到高层模型的关系^[7],未考虑底层概念到中层模型的映射关系,这对从安全事件中提取抽象攻击技术、战术等带来困难,因此需要一种分析底层和中层概念关系的方法。本文利用当前工作^[3-6]已证明的通过自然语言技术分析实体的描述信息有助于提升实体关系预测准确率的结论,改进现有基于自然语言处理的分析模型,提出一种新的网络安全实体关联方法,实现知识库中 CAPEC 攻击模式与 ATT&CK 攻击技术两类实体关系的预测。

2 安全实体关联方法

本文中安全实体的关联分析被建模为实体描

述的文本相似度判断问题,但是要处理的攻击模式 CAPEC 和 ATT&CK 技术实体文本数据存在数据量小、正负样本严重不均衡、文本长度介于短文本与长文本之间、领域专业词汇集中等特点。借鉴现有的主流文本相似度判断方法,如基于 Transformer 模型^[12-16]、基于孪生网络模型^[11]、基于图神经网络模型^[17]等,本文提出使用 Word2Vec 模型、图卷积神经网络和孪生网络,设计攻击模式与技术关联方法,通过分析实体描述的相似性和实体间的层次结构,判断两实体间是否存在关联。下面给出关联方法框架及核心方法。

2.1 基于自然语言处理的实体关联方法框架

基于自然语言处理的攻击模式与技术关联方法架构如图 2 所示(为后续叙述方便,将该方法命名为 WGS 方法,其中 W、G、S 分别为 Word2Vec、图神经网络、孪生网络的简写),该框架包含数据预处理、句向量获取、实体关联分析 3 个步骤,其中句向量获取包含了无监督学习模型词向量、有监督学习模型图神经网络以及孪生网络,达成较为精确的文本表达向量。而且,Word2Vec 词向量由安全语料库训练得到,图神经网络根据正负样本训练得到。

选定 CAPEC 实体 C_i 、ATT&CK 实体 T_j ,首先将实体描述文本经过停用词删除、特殊字符删除、低频词删除等预处理,分别输入 Word2Vec 词向量表示和 GCN 向量表示模块获取句向量,合并后输入孪生网络,最终得到两实体各自的表示向量;然后计算 CAPEC 实体和 ATT&CK 技术实

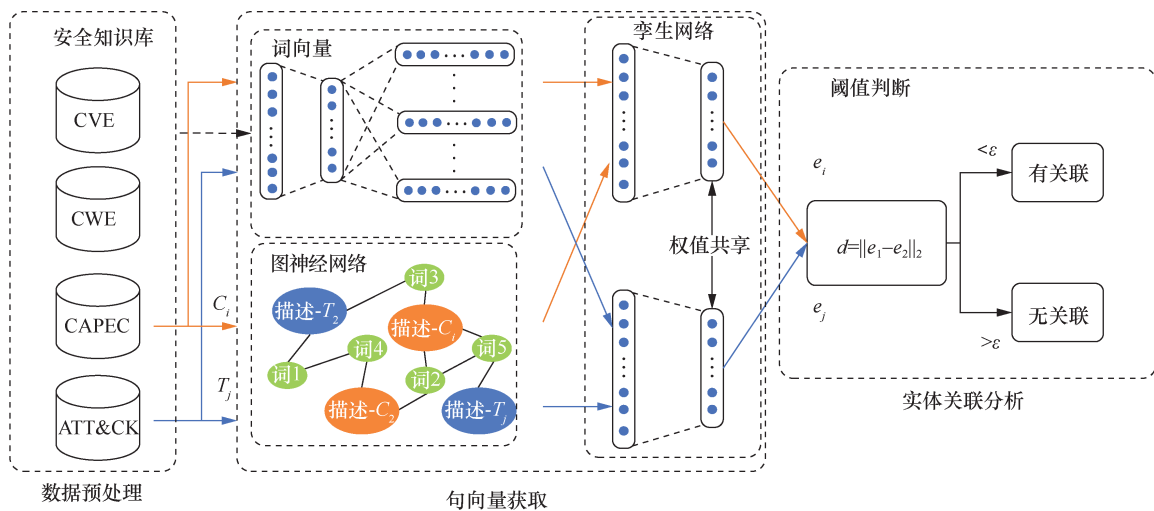


图2 基于自然语言处理的攻击模式与技术关联方法架构

Figure 2 Framework for relation prediction between attack patterns and techniques based on natural language processing

体的表示向量之间的欧几里得距离，并与设定的关联阈值进行比较，判断两实体间是否存在关联。

2.2 方法描述

本节详细介绍 WGS 方法框架中的核心方法，包括数据预处理、句向量获取和实体关联分析。

(1) 数据预处理

CVE、CWE、CAPEC 和 ATT&CK 实体描述为英语文本，因此无须分词。首先使用正则匹配的方法删除语句中的非字母字符，并将所有英语字符转为小写。然后，根据 Python 的 NLTK 库中已提供的常见停用词表删除停用词。最后，将在语料库中出现频率小于阈值 f_r 的单词删除，避免低频词影响文本分析的泛化能力。

将 MITRE 组织的 CAPEC 和 ATT&CK 知识库中已整理的 500 多个攻击模式和攻击技术实体，作为关联预测的目标实体。对于选定的 CAPEC 实体 C_i 、ATT&CK 实体 T_j ，经过数据预处理后的文本分别表示为 D_i 、 D_j ，输入句向量获取模块，以获得向量表示。CVE、CWE、CAPEC 和 ATT&CK 实体描述经过数据预处理，输入 Word2Vec 模型以训练安全领域词向量。

(2) Word2Vec 向量生成

Word2Vec^[18]是一种使用词向量表达语义的浅层神经网络模型，本文采用 Skip-Gram 训练方式，即给定中心词预测其上下文。输入单个单词的独热编码 X_{n_c} ， n_c 为语料库中不同单词的个数；输入层与隐层间全连接，隐层神经元个数为 n_w ，两层间权重 $W_{n_c n_w}$ 作为最终的词向量；隐层与输出层间全连接，权值矩阵 $W_{n_w n_c}$ ，输出目标单词的上下文单词的概率分布。训练时使用分层 softmax 和负采样减少计算量，并加快训练速度。

网络安全是一个特定领域，无法使用维基百科训练的 Word2Vec 模型，这是因为通用领域的模型难以理解特定领域的用法。例如，elevator 和 escalate 在通用领域中用法有区别，在词向量上相似度不高；但是在网络安全领域针对权限提升的描述中，两者是同一个含义。因此，本文训练了网络安全领域的 Word2Vec 模型，数据来源为安全知识库中 CVE、CWE、CAPEC 和 ATT&CK 的描述文本，语料库的大小为 15 万个段落，单词数

量约为 3 万个。训练模型后可以得到实体描述文本的词向量，假设实体描述 D_m 由 k 个单词 $x_1, x_2, \dots, x_k$ 组成，词向量分别为 $w_1, w_2, \dots, w_k$ ，则文本 D_m 的句向量为 e_w 。

$$e_w = \frac{1}{k} \sum_{i=1}^k w_i \quad (1)$$

即句向量由句子中单词向量的均值向量表示，其维数即词向量维数 n_w 。

(3) GCN 表征向量生成

本文参考近年来对英语文本建模较为成功的 TextGCN^[9]，其已被证明在文本分类领域有良好的效果、无须初始化向量，并且可以处理小样本问题、具有更好的泛化能力，采用图卷积神经网络^[19] (GCN) 以更好地捕捉段落中单词的共现关系。GCN 需要输入图数据，本文生成图的方法与 TextGCN 类似，安全语料库与文本图数据的转化如图 3 所示。首先将 CAPEC 和 ATT&CK 技术的所有实体作为节点，两实体的现有关系表示边，由此组成实体层图数据，关联分析的目标即判断不存在边的节点对是否有关联，存在则加入图。每个实体对应的描述组成文本层图数据，描述文档及其单词作为节点，词与文档的词频-逆文档频率 (TF-IDF) 值作为词节点与文档节点的边权重，词与词的点互信息 (PMI) 值作为词节点间的权重。记单词集合为 W ，文档集合为 D ， $n_{k,j}$ 表示第 k 个单词在第 j 个文档中出现的频次，单词 w_i 对文档 d_j 的 TF-IDF 计算式为

$$\text{TF-IDF} = \frac{n_{i,j}}{\sum_k n_{k,j}} \log \left(\frac{|D|}{1 + |j : w_i \in d_j|} \right) \quad (2)$$

记 $p(w_i, w_j)$ 表示词 w_i 和 w_j 共同出现的频率， $p(w)$ 表示词 w 的出现频率，PMI 计算式为

$$\text{PMI} = \log \left(\frac{p(x, y)}{p(x)p(y)} \right) \quad (3)$$

记实体描述文档集合为 $D = \{d_1, d_2, \dots, d_m\}$ ，描述中单词集合为 $W = \{w_1, w_2, \dots, w_n\}$ ，节点集合即 $D+W$ 。由邻接矩阵 $A_{(m+n)(m+n)}$ 表示词与词、词与文档间的共现关系作为边，文档与文档间的边置 0。

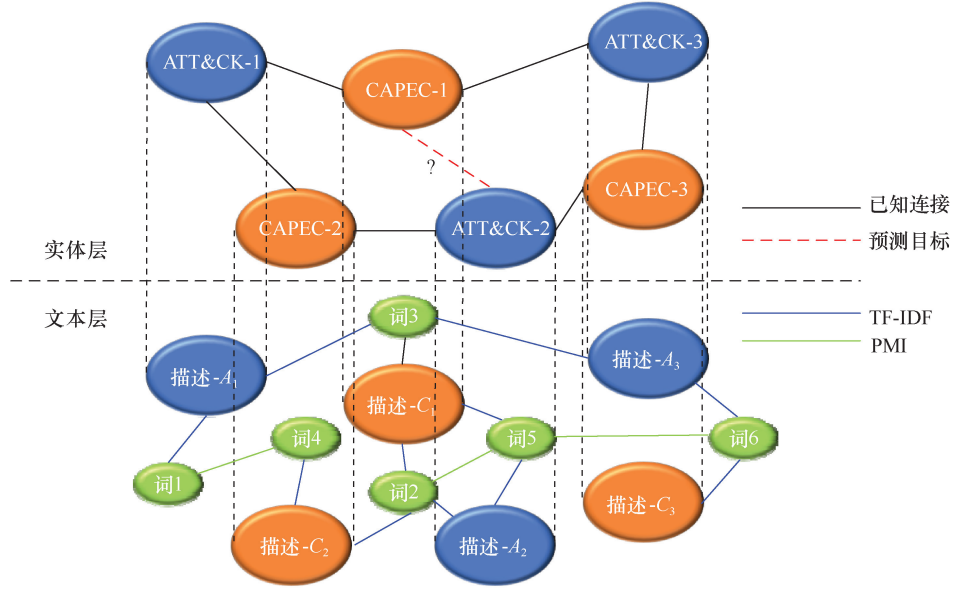


图3 安全语料库与文本图数据的转化
Figure 3 Transformation between security corpus and text graph

$$A_{ij} = \begin{cases} \text{TFIDF}(w_i, d_j) \\ \text{PMI}(w_i, w_j) \end{cases} \quad (4)$$

文档节点和词节点使用独热编码初始化，该方法常用于 GCN 的特征矩阵初始化。记各节点的顺序编号为 $\{1, 2, \dots, m+n\}$ ，初始化特征矩阵为 $X_{(m+n)(m+n)}$ 。

$$X_{ij} = \begin{cases} 1, i = j \\ 0, i \neq j \end{cases} \quad (5)$$

独热编码得到的特征矩阵尽管维数较大，但是该方法便于 GCN 学习词和文档的共现关系。独热编码即表示各个节点在初始状态下均匀分布在空间中，GCN 通过学习间隔损失函数，拉近有关系实体间的距离，并从中发现新的实体关联。这一过程与人工从大量实体对数据中找到部分有关联实体对的过程类似。实验中取得了比使用预训练词向量初始化更好的预测效果。

邻接矩阵 A 和特征矩阵 X 输入图卷积神经网络。GCN 由两个图卷积层构成，第 l 层图卷积层的前馈计算式为

$$H^{(l+1)} = \text{relu} \left(\tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} H^{(l)} W^{(l)} \right) \quad (6)$$

其中， $\tilde{A}$ 为邻接矩阵 A 与单位矩阵的和， $\tilde{D}$ 为 $\tilde{A}$ 的度矩阵， $H^{(l)}$ 为第 l 层的输出矩阵， $W^{(l)}$ 为第 l

层的权重向量， $H^{(0)}$ 即特征矩阵 X 。

网络训练时将训练集输入网络，根据间隔损失函数计算样本误差。

$$L(\theta) = \sum \max \{S(d1, d2) - S(d1', d2') + \gamma_g, 0\} \quad (7)$$

其中， $S(d1, d2)$ 表示正样本距离， $S(d1', d2')$ 表示负样本距离， γ_g 为间隔损失参数。正样本指数数据库中已有连接的实体对，实体对的距离为两实体表示向量的欧几里得距离；负样本生成方法为随机替换正样本实体对中的一个实体，且替换后不在正样本集合中。计算损失函数后，使用 Adam 优化器更新模型参数。训练后得到所有实体的文档向量，作为各实体的 GCN 表征向量 e_g ，维数为 n_g 。

(4) 孪生网络表征向量生成

孪生网络可有效分析成对的样本，本文利用孪生网络聚合实体的两种表征向量。抽取两目标实体的 Word2Vec 表示向量 e_w 和 GCN 表示向量 e_g ，各自合并成 $n_w + n_g$ 维向量，分别输入孪生网络。孪生网络由两个相同的多层感知机模型组成，单个模型由两层全连接层组成，训练过程中两个模型参数共享。

训练时将训练集输入网络，正负样本采集方法与图卷积网络相同。根据对比损失函数计算样本误差。

$$L(\theta) = \sum_{n=1}^N Y d_w^2 + (1 - Y) \max(\gamma_s - d_w, 0)^2 \quad (8)$$

其中, Y 表示样本对是否有关系的标签、 d_w 指样本对两向量的欧几里得距离, γ_s 为对比损失函数参数。计算损失函数后, 用 Adam 优化器更新模型参数。在预测阶段, 输入目标实体对的 e_w 和 e_g , 各输出 n_s 维孪生网络句向量表示 e_s 。

(5) 基于阈值判断的实体关联分析

本文中的 GCN 和孪生网络最终都使用欧几里得距离计算损失函数, 因此用欧几里得距离表达相似度。训练孪生网络时记录正样本对的平均欧几里得距离, 关联判断的阈值取该平均欧几里得距离的 α 倍。这一过程可形式化表示为: 记训练集中正样本对 $(c_i, t_j) \in T$, 其孪生网络向量表示为 (e_i, e_j) , 则目标实体对 (c_m, t_n) 的预测结果为

$$\text{Rel}(c_m, t_n) = \begin{cases} 1, & \|e_m - e_n\|_2 \leq \varepsilon \\ 0, & \|e_m - e_n\|_2 > \varepsilon \end{cases} \quad (9)$$

$$\varepsilon = \frac{\alpha}{|T|} \sum \|e_i - e_j\|_2 \quad (10)$$

针对 WGS 方法判断有关联的实体对, 可进行人工检查, 此时需要人工判断的数据量已大大减少。参数 α 可根据实际情况调节, α 增大则系统会预测出更多有连接的实体, 需要更多人工判断; α 减少则系统的预测准确度提高, 但可能遗漏部分确实有关系的实体对。

3 实验测试与结果分析

本节将通过实验证明:

1) CAPEC 与 ATT&CK 技术数据集中正样本数量过少, 导致训练过拟合、影响模型效果的问题, WGS 模型通过样本扩充、动态负样本采集、使用正则等缓解过拟合 (3.2 节);

2) WGS 在小样本条件下的准确率高于现有文本相似度预测模型, 且训练时间和资源消耗更少 (3.3 节);

3) WGS 可以预测出新的攻击模式与技术实体关联 (3.4 节)。

3.1 数据集及实验环境

本文方法的测试使用 MITRE 组织提供的 CAPEC 和 ATT&CK 知识库。文本预处理中设定低频词阈值 $f_r = 5$, 删除出现频次小于该值的单词。CVE 和 CWE 知识库仅用于 Word2Vec 训练,

数据预处理同 2.2 节给出的过程。4 个知识库中的实体及其描述的统计信息如表 1 所示。

表 1 知识库中的实体及其描述的统计信息
Table 1 Statistics on entities and descriptions in knowledge bases

知识库	实体个数/个	描述平均单词数/个	词汇量/个
CAPEC	525	57.71	1 346
ATT&CK	552	121.27	1 805
CVE	148 264	31.70	12 552
CWE	922	48.92	1 604

CAPEC 知识库中记录了与其实体对应的 ATT&CK 技术实体, 把数据集划分为正负样本两类集合, 具体如下。

1) 正样本: 知识库中已有 157 对 CAPEC 与 ATT&CK 技术关系, 构成训练的正样本集合 T 。分析两类实体的类内关系, 可补充正样本。CAPEC 和 ATT&CK 技术都可看作树状结构, 父节点实体一般为子节点的抽象实体, 如 CAPEC-100 (缓冲区溢出) 有子节点 CAPEC-10 (通过环境变量的缓冲区溢出)。假设已有关系 $(c_i, t_j) \in T$, c_i 存在父节点 $\dot{c}_i$, t_j 存在父节点 $\dot{t}_j$, 则 $(\dot{c}_i, \dot{t}_j) \in T$ 、 $(c_i, \dot{t}_j) \in T$, 由此可以将正样本补充至 485 个。

2) 负样本: 525 个 CAPEC 实体和 552 个 ATT&CK 技术实体可形成 25 万个候选实体对, 不属于正样本集合 T 的实体对都作为负样本集合 T' 。

在训练和评估 GCN 表示模块和孪生网络表示模块时, 数据集中的 485 个正样本以 7:1:2 的比例划分训练集、验证集和测试集。负样本数量与正样本数量比例为 1:1。实验环境操作系统和硬件配置如表 2 所示。

表 2 实验环境操作系统和硬件配置
Table 2 OS and hardware used in experiments

操作系统	CPU	GPU
Ubuntu 18.04-64 bit	Intel Core i5-8265U,	Nvidia GeForce RTX
Windows 10-64 bit	8 GB 内存	2080 Ti, 10 GB 内存

3.2 模型训练与性能测试

本节展示 WGS 方法在数据集上的训练和测试效果。其中, Word2Vec 表示模块的训练为无监

督模式，GCN 和孪生网络模块为有监督训练模式。

(1) Word2Vec 模型训练

网络安全领域 Word2Vec 模型使用 Python 的 gensim 库训练模型，设置输出 $n_w = 100$ 维词向量、训练 10 个轮次、负采样比例为 3、设置分层 softmax。训练完后可简单测试效果，查看 elevate 和 escalate 的相似度，本文的 Word2Vec 模型得出相似度 0.88，而通用模型得出的相似度为 0.53，说明本文训练的词向量可以更好地反映安全领域词汇的特点。

(2) GCN 模型训练

考虑到本文数据集中负样本数量远远大于正样本，GCN 使用的间隔损失函数每一次采样取 2 个负样本（分别改变正样本中 CAPEC 实体和 ATT&CK 实体），每个样本权重为 0.5。每个正样本在每一轮训练中随机采样负样本，以此扩大涉及的负样本数量，增强了模型的泛化能力。实验中尝试 GCN 每轮次固定负样本，发现模型在验证集上无法收敛。由于负样本被赋予归一化权重，因此训练时总体的正负样本比例仍保持 1:1，防止数据类别偏斜。训练数据中包括词汇 2 265 个、文档 1 077 个，因此邻接矩阵和独热特征矩阵维数为 3 342。模型由 2 个图卷积层组成，中间层维数为 300 维。输入邻接矩阵和特征矩阵，最终输出 $n_g = 100$ 维 GCN 表示向量 e_g ，损失函数中取 $\gamma_g = 1$ 。

图 4 展示了 GCN 训练集和验证集损失，在训练 200 个轮次后损失趋于平稳。此时训练误差已降至 0.01 以下，说明模型进行了有效训练；验证集误差的变化较大，最终在 0.28 附近波动且远大于训练误差，存在过拟合现象，这与样本数据量过少有关。

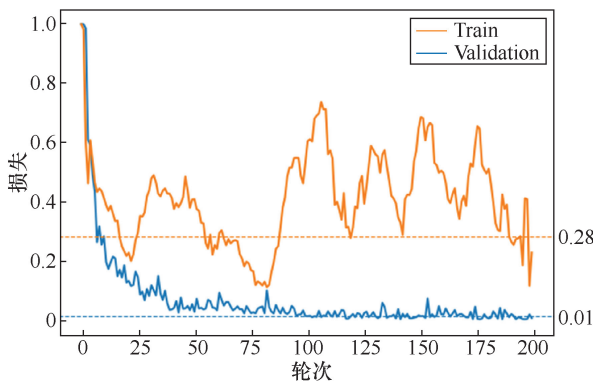


图4 图卷积神经网络训练集和验证集损失
Figure 4 Training and validation loss of GCN

(3) 孪生网络模型

实体的 Word2Vec 表示向量和 GCN 表示向量合成一个向量，维数为 $n_w + n_g = 200$ 维，实体对的两个向量分别输入孪生网络模块，最终输出 $n_s = 50$ 维的孪生网络表示向量 e_s 。孪生网络由 2 个全连接层组成，中间层神经元个数为 100，损失函数中取 $\gamma_s = 1$ 。负样本动态采样方式与 GCN 训练时相同。

图 5 展示了孪生网络训练集和验证集损失，在训练 100 个轮次后损失趋于平稳。可以看到训练误差已降至 0.045 以下，模型已经收敛；验证集误差最终为 0.125 左右，大于训练误差，存在过拟合现象，这依旧与样本数据量过少有关，但相比 GCN 的训练结果有较大改善。

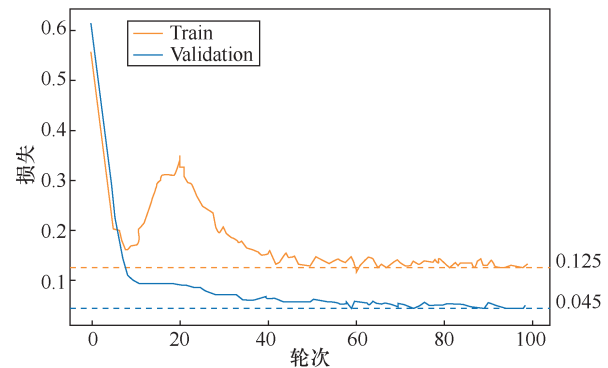


图5 孪生网络训练集和验证集损失
Figure 5 Training and validation loss of Siamese network

(4) WGS 性能测试结果

以孪生网络表示向量为系统最终输出向量，在测试集上评估 WGS 方法效果。评价指标为准确率 (Acc, accuracy)、召回率 (Rec, recall)、精度 (Pre, precision)、F1 值 (F1-score) 和 AUC 值。

模型得到输出后，根据式(9)、式(10)，基于阈值判断实体关联关系。式(10)表示距离阈值 ε 等于正样本的平均间距乘以 α 倍。由于 α 越小，距离 ε 越小，被判定为正样本的数量越少，在评估中表现为召回率变小、精度增加，因此最佳的 α 应取召回率和精度的平衡点，即最大的 F1 值。实验中多次训练模型取均值，计算验证集上正样本的平均距离，改变 α 得到 ε ，并计算相应的召回率、精度和 F1 值，得到图 6 的阈值 α 与召回率、精度和 F1 值关系曲线。随着 α 增大，F1 值先增大后减小，在 $\alpha \approx 2.5$ 时取最大值 $F1 = 0.87$ ，因此将该值作为阈值，验证集上对应的距离阈值 $\varepsilon \approx 1.6$ 。

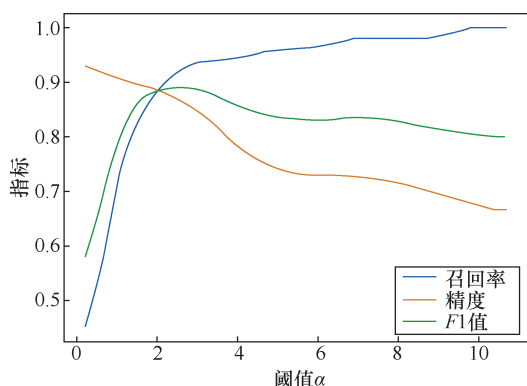


图6 阈值 α 与召回率、精度和 F1 值关系
Figure 6 Relations between threshold α and recall, precision and F1-score

确定阈值后在测试集上评估 WGS 方法, 最终得到评估指标, 其中 $\text{Acc}=87.10\%$, $\text{Rec}=88.10\%$, $\text{Pre}=86.41\%$, $F1=0.8724$, $\text{AUC}=0.9273$, 且各指标在训练集和测试集上差距小于 1%, 说明过拟合问题得到缓解。

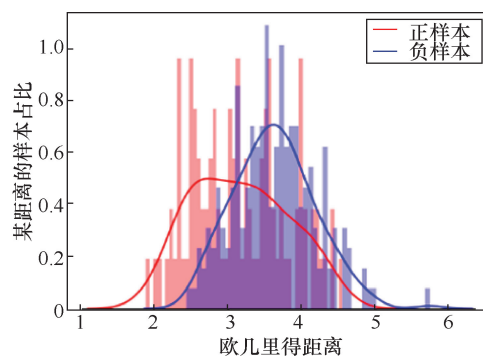
最终的关联预测取决于两个实体对应向量的欧几里得距离, 本文方法的目标即分离正负样本欧几里得距离分布, 图 7 展示了各级模型正负样本欧几里得距离分布。可以看到初始的 Word2Vec 向量几乎不能区分正负样本; GCN 表示向量效果较好, 但仍有大量负样本与正样本分布重叠; 孪生网络表示向量则可以有效区分正负样本。在图 7(c) 中, 式(8)中的距离阈值 ϵ 相当于一条平行于 y 轴的直线, 直线左边即判断为有关联的实体对, α 越大, 直线越靠右, 覆盖更多正样本的同时覆盖了部分负样本, 最佳阈值即求最佳平衡点。

实验表明, 样本过少会导致严重的过拟合。为解决该问题, WGS 方法训练安全领域 Word2Vec 补充外部语义信息; GCN 与孪生网络训练时损失函数中添加了 L2 正则项, 减少过拟合; 训练的每个轮次随机采样负样本, 增加了模型的泛化性。从测试结果中可以看出, 小样本带来的过拟合得到有效缓解。

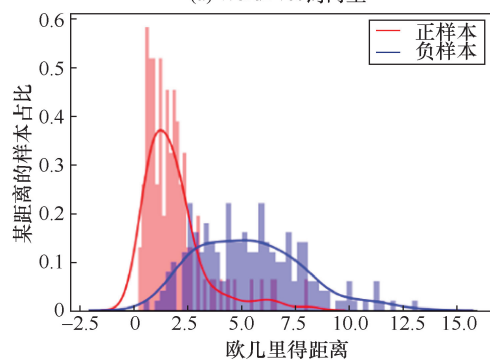
3.3 消融对比实验

文本相似度是自然语言处理领域的经典任务, 研究者已提出多种方法。本文选择的对比方法分类及训练细节列举见表 3, 表中“C-A 数据集”指本文使用的 CAPEC-ATT&CK 技术数据集。基线方法包括 TF-IDF、Word2Vec 和 Doc2Vec^[20] 等模型, 本文实验中导入已训练模型或直接计算

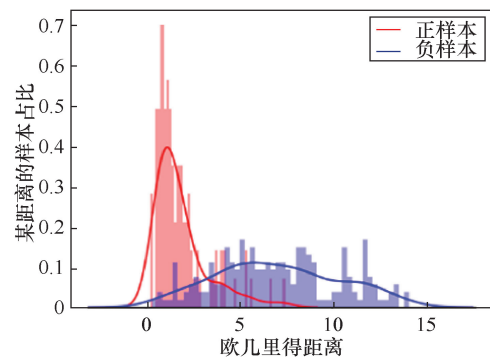
得到文本向量, 用余弦相似度判断是否有关系。Bert 已被证明在大部分自然语言处理任务取得了良好的效果, 多个变种如 RoBERTa、ALBERT 等优化 Bert 的语料库和训练方式, 都是获取文本相似度的经典模型。本文还选择在文本相似度研究中通用的 STS 数据集^[21]上表现最好的开源算法与所提方法进行比较, 包括 Sentence Bert^[11]和 XLNet^[16]。Bert 变种和 SOTA 模型都是在预训练模型基础上使用 C-A 数据集微调, 模型直接输出二分类结果表示有无关联。



(a) Word2Vec 词向量



(b) GCN 表示向量



(c) 孪生网络表示向量

图7 各级模型正负样本欧几里得距离分布

Figure 7 Euclidean distance distribution of positive and negative samples in each model

表 4 展示了本文方法与现有方法的预测效果

和训练时间对比。基线方法中的模型训练时间短或无须训练，但是准确率较低；基于 Bert 的模型和 SOTA 模型训练时间较长，实验中使用 GPU 训练，但训练单个模型的时间仍然超过两小时，平均准确率在 82% 左右，其中最佳的是 Roberta (Large) 模型，达到 83.16% 的准确率和 84.69% 的召回率；本文的 WGS 方法准确率高于所有模型，并且训练时间短，在 CPU 上训练时间也不超过 5 min。

表3 对比方法分类与训练细节

Table 3 Classification and training details of methods for comparison

方法分类	具体模型	训练细节	相似度评估
基线方法	TF-IDF	C-A 数据集上计算	余弦相似度
	Doc2Vec	使用已训练模型	余弦相似度
	Word2Vec	使用已训练模型	余弦相似度
Bert 及其变种微调	Bert	预训练模型+C-A 上微调	二分类
	RoBerta	预训练模型+C-A 上微调	二分类
	AlBert	预训练模型+C-A 上微调	二分类
SOTA 模型微调	Sentence Bert	预训练模型+C-A 上微调	二分类
	XLNet	预训练模型+C-A 上微调	二分类

表4 本文方法与现有方法的预测效果和训练时间对比
Table 4 Comparison on prediction results and training time with existing methods

方法	准确率	召回率	精度	F1 值	训练时间
TF-IDF	70.41%	72.58%	69.57%	0.7104	<1 min
Word2Vec	68.04%	76.08%	65.54%	0.7042	—
Doc2Vec	62.58%	70.52%	60.85%	0.6533	—
Bert (Base)	76.84%	81.33%	74.72%	0.7782	20 min
Roberta (Large)	83.16%	84.69%	82.18%	0.8342	>2 h
AlBert (Base)	81.82%	85.71%	79.75%	0.8273	>2 h
XLNet (Base)	78.57%	87.07%	74.46%	0.8025	>2 h
Sentence Bert	82.40%	84.18%	81.32%	0.8272	>1 h
WGS	87.10%	88.10%	86.41%	0.8724	<5 min

从实验中可以看出，C-A 数据集中训练样本少，基于 Bert 和 Transformer 的模型又有大量参数需要训练微调，导致其无法有效学习安全领域的文本特征；同理，这些模型提供的 Large 预训练模型不一定比 Base 效果好，因为前者只是比后者增加了参数量，反而加重了样本少带来的过拟合问题。WGS

模型不依赖于大语料库的预训练，使用安全领域 Word2Vec 更好地表征安全语义；GCN 以 TF-IDF 和 PMI 特征构成图，保留了 TF-IDF 这个在自然语言处理多个任务中都很有效的关键特征；孪生网络集成 Word2Vec 和 GCN 表征向量，前者偏向于理解段落中词与词之间的局部关系，对语义进行建模，后者更关注段落与段落之间全局关系，类似于词袋模型，两者结合可以更好地表征文本。因此，WGS 模型更适合于 C-A 数据集中限定领域、样本少的情况。

表 5 展示了 WGS 方法与其子模块的消融实验对比。如果仅使用 Word2Vec 或 GCN，准确率很低，GCN 相比 Word2Vec 略高，但受小样本问题影响还是无法准确预测。两种方法各自加上孪生网络后效果明显提升，可以达到 86% 左右的准确率；将三者结合起来可进一步提升准确率、召回率和精度。实验中还尝试了将 Word2Vec 换成 Doc2Vec，后者改变了前者通过取词向量均值的方式生成文档向量，因此能够更好地对文档建模，替换后整体方法的准确率略有提升。

表5 WGS 方法与其子模型的消融实验对比
Table 5 Comparison of ablation studies on WGS and its sub-models

方法	准确率	召回率	精度	F1 值
GCN	79.59%	81.63%	78.43%	0.8000
Word2Vec	68.04%	76.08%	65.54%	0.7042
GCN+SIA	86.22%	87.75%	85.15%	0.8643
W2V+SIA	86.48%	87.76%	85.57%	0.8665
GCN+W2V+SIA	87.10%	88.10%	86.41%	0.8724
GCN+D2V+SIA	87.70%	88.10%	87.48%	0.8776

总体而言，本文方法可有效应对小样本问题，通过 Word2Vec 补充语义信息、GCN 中的间隔损失函数、负样本的动态随机选取和加权计算方式，减少过拟合、正负样本不均衡带来的影响。相比基于 Bert 的类方法，本文方法取得了更高的预测准确率，且需要更少的训练时间、不依赖于庞大的语料库和预训练模型。这些特点可以使该方法扩展到其他领域的小样本实体关系预测问题。

3.4 实体关联分析

本文方法最终的目标为预测 CAPEC 与 ATT&CK 技术之间的新关联。实验中取模型预测结果中距离最小（或相似度最大）的 TopK 个实

体对作为模型认为“有关联”的实体对，各随机抽样 200 个进行人工分析。具体做法为，首先将 3.1 节中生成的 24 万多组负样本作为目标实体对，输入模型预测。对于本文的 WSG 方法，预测结果中筛选出距离小于等于 ε (≈ 1.6) 的实体对，排序后取前 600 个距离最小的样本，并随机抽样 200 对。用于比较的基准模型 Bert 直接对预测结果排序，同样取前 600 个相似度最大的样本，随机抽样 200 对。WSG 方法与 Bert 预测效果分析与对比如表 6 所示，200 个 WSG 方法预测为“有关联”的实体对中，41 对被人工判定为有关联，预测正确比例为 20.5%；Bert 可以正确预测 31 对，正确比例低于 WSG。20.5% 的预测正确率说明，会产生近 80% 的误报率。本文 WSG 正确率较低的原因在于：实际有关联的实体对较少，人工随机从 24 万多组目标实体对中抽取 50 组，有关联的仅 1~2 组，其比例远低于本文方法的 20.5%；而本文仅使用自然语言特征，融合更多信息可能会达到更好效果。

更进一步，可以通过预测概率与真实标签间的交叉熵分析二分类效果。WSG 输出为距离值，可以通过下式变换为相似度（即概率值）。

$$\text{sim} = \frac{1}{1.6 + \text{dist}}$$

WSG 在 200 对样本上的交叉熵为 0.834 5，而 Bert 为 6.299 3，这说明 WSG 的输出更符合标签的分布，即正样本的相似度更接近 1，而负样本相似度更接近 0。实验中发现 Bert 的相似度输出非常接近 0 或 1，相对而言 WSG 的输出分布更均匀，适合通过阈值控制漏报率和误报率。

表 6 WSG 方法与 Bert 预测效果分析与对比
Table 6 Analysis and comparison on predictions of WSG and Bert

方法	预测正确个数	预测正确率	交叉熵
WSG	41	20.5%	0.834 5
Bert	31	15.5%	6.299 3

表 7 列举了新实体关系预测结果分析，其中

表 7 新实体关系预测结果分析
Table 7 Analysis on predicting correlation results among new entities

表项	CAPEC	ATT&CK 技术	描述
1	修改注册表运行键值	启动或登录自启动执行：活动设置	CAPEC: ...在注册表中添加一个新的运行键值...在用户登录时启动... ATT&CK: ...通过添加注册表键值实现持久化...在用户登录计算机时执行...
2	HTTP 泛洪	端点拒绝服务：应用程序耗尽泛洪	CAPEC: ...使用 HTTP 执行泛洪攻击，目的是通过消耗应用层（如 Web 服务）的资源来拒绝合法用户访问服务... ATT&CK: ...针对 Web 应用程序的资源密集型功能，造成拒绝服务（DoS）...
3	现有进程中包含代码	进程注入：动态链接库注入	CAPEC: ...运行进程以在单独的实时进程的地址空间中执行任意代码...示例方法包括但不限于：动态链接库（DLL）注入... ATT&CK: ...DLL 注入是一种在单独的实时进程的地址空间中执行任意代码的方法...
4	JSON 劫持（又名 JavaScript 劫持）	信息网络钓鱼：鱼叉式网络钓鱼链接	CAPEC: ...攻击者让受害者访问其恶意页面，该页面包含脚本标记，其源指向易受攻击的系统... ATT&CK: ...发送带有恶意链接的鱼叉式网络钓鱼邮件，以引出可在定位过程中使用的敏感信息...
5	XML 泛洪	端点拒绝服务：应用程序或系统利用	CAPEC: ...使用 XML 消息执行泛洪攻击...XDoS 的主要弱点是服务提供商通常必须检查、解析和验证 XML 消息... ATT&CK: ...利用可能导致应用程序或系统崩溃并拒绝用户可用的软件漏洞...
6	加密暴力破解	操作系统凭据转储	CAPEC: ...对密钥空间执行暴力搜索，以确定解密密文并获取明文的密钥... ATT&CK: ...尝试转储凭据以获取帐户登录名和凭据材料，通常采用哈希或明文密码的形式...
7	窃听	网络嗅探	窃听与嗅探攻击不同，因为它不会发生在基于网络的通信通道（如 IP 流量）上

表项 1-6 被人工判别为有关系, 表项 7 无关系, 表中对比本文方法与 Bert 的预测结果。表项 1~3 的实体对在 WGS 和 Bert 中都被预测为有关系, 通过人工分析可以发现其描述之间有明显相似, 以表项 1 为例, 两者都是描述通过在注册表中添加开机自启动的键值达到恶意代码自动执行的目的, 两者的描述中都出现注册“表键值 (registry key)”“用户登录 (user logs in)”等词, 体现其相关性; 表项 3 中 CAPEC 的描述直接出现了技术名“DLL 注入 (DLL Injection)”, 因此模型能准确判断。编号 4~6 的实体对被 WGS 预测为有关系, 但被 Bert 预测为无关系。表项 4 中“鱼叉式链接 (Spearphishing Link)”可作为实现“Json 劫持 (Json Hijacking)”的方式之一, 通过发送钓鱼链接诱骗用户点击、截取服务器回复客户端的 Json 文件盗取敏感信息。表项 6 中加密文件的暴力破解可用于 OS 凭证破解, 两者描述中没有明确相同的词, 但可以看出其表达意思相近。编号 4~6 的实体对间的关系不如 1~3 明显, 需要逻辑推理或近义词对比后得到, 本文方法通过 Word2Vec 捕捉领域内近义词关系、GCN 模仿推理过程, 因此比 Bert 预测更准确。还有一些表项, 本文方法和 Bert 都无法正确预测, 如表项 7 的描述中明确指出窃听和网络嗅探有区别, 前者嗅探物理设备、后者截取网络流量。预测错误的原因无法准确分析出“不同 (different from)”等表示否定的词, 基于共现关系的文本相似度分析反而会因为两者描述中出现共同的短语而判别为有关系, 这是后续工作中需要解决的问题。

4 结束语

本文针对 CAPEC 和 ATT&CK 技术缺失的实体关系, 提出一种基于自然语言处理的安全实体关联方法。提出的方法通过训练网络安全领域 Word2Vec 模型, 解决通用领域模型无法处理领域内近义词问题; 使用 TextGCN 提取词与句、句与句之间的共现关系, 并采用动态负采样技术, 有效缓解小样本带来的过拟合问题。通过计算孪生网络输出的实体向量间的欧几里得距离, 预测实体间的关联。本文方法对 MITRE

组织提供的 CAPEC 和 ATT&CK 知识库进行测试, 可以达到 87% 的预测准确率, 优于文本相似度领域的 SOTA 模型, 且无须大语料库预训练、减少计算资源、缩短训练时间。通过本文方法预测结果的分析, 证明其可以很好地预测出新的 CAPEC 和 ATT&CK 技术实体关系。

参考文献:

- [1] STROM B E, APPLEBAUM A, MILLER D P, et al. Mitre att&ck: design and philosophy[R]. Technical Report, 2018.
- [2] KIESLING E, EKELHART A, KURNIAWAN K, et al. The SEPSES knowledge graph: an integrated resource for cybersecurity[C]//International Semantic Web Conference. 2019: 198-214.
- [3] GUO Y, LIU Z, HUANG C, et al. CyberRel: joint entity and relation extraction for cybersecurity concepts[C]//International Conference on Information and Communications Security. 2021: 447-463.
- [4] XIAO H, XING Z, LI X, et al. Embedding and predicting software security entity relationships: a knowledge graph based approach[C]//International Conference on Neural Information Processing. 2019: 50-63.
- [5] YUAN L, BAI Y, XING Z, et al. Predicting entity relations across different security databases by using graph attention network[C]//2021 IEEE 45th Annual Computers, Software, and Applications Conference (COMPSAC). 2021: 834-843.
- [6] DAS S S, SERRA E, HALAPPANAVAR M, et al. V2W-BERT: a framework for effective hierarchical multiclass classification of software vulnerabilities[C]//2021 IEEE 8th International Conference on Data Science and Advanced Analytics (DSAA). 2021: 1-12.
- [7] MILAJERDI S M, GJOMEMO R, ESHETE B, et al. Holmes: real-time apt detection through correlation of suspicious information flows[C]//2019 IEEE Symposium on Security and Privacy (SP). 2019: 1137-1152.
- [8] WU L, CHEN Y, SHEN K, et al. Graph neural networks for natural language processing: a survey[J]. arXiv Preprint arXiv:2106.06090, 2021.
- [9] YAO L, MAO C, LUO Y. Graph convolutional networks for text classification[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2019: 7370-7377.
- [10] LIU B, NIU D, WEI H, et al. Matching article pairs with graphical decomposition and convolutions[J]. arXiv Preprint arXiv: 1802.07459, 2018.
- [11] REIMERS N, GUREVYCH I. Sentence-bert: sentence embeddings using siamese bert-networks[J]. arXiv Preprint arXiv:1908.10084, 2019.
- [12] DEVLIN J, CHANG M W, LEE K, et al. Bert: pre-training of deep bidirectional transformers for language understanding[J]. arXiv Preprint arXiv:1810.04805, 2018.

- [13] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//Advances in Neural Information Processing Systems, 2017, 30.
- [14] LIU Y, OTT M, GOYAL N, et al. Roberta: a robustly optimized bert pretraining approach[J]. arXiv Preprint arXiv:1907.11692, 2019.
- [15] LAN Z, CHEN M, GOODMAN S, et al. Albert: a lite bert for self-supervised learning of language representations[J]. arXiv Preprint arXiv:1909.11942, 2019.
- [16] YANG Z, DAI Z, YANG Y, et al. Xlnet: generalized autoregressive pretraining for language understanding[C]//Advances in Neural Information Processing Systems, 2019, 32.
- [17] LIU B, NIU D, WEI H, et al. Matching article pairs with graphical decomposition and convolutions[J]. arXiv Preprint arXiv:1802.07459, 2018.
- [18] MIKOLOV T, CHEN K, CORRADO G, et al. Efficient estimation of word representations in vector space[J]. arXiv Preprint arXiv:1301.3781, 2013.
- [19] KIPF T N, WELING M. Semi-supervised classification with graph convolutional networks[J]. arXiv Preprint arXiv:1609.02907, 2016.
- [20] LE Q, MIKOLOV T. Distributed representations of sentences and documents[C]//International Conference on Machine Learning. PMLR, 2014: 1188-1196.
- [21] CER D, DIAB M, AGIRRE E, et al. Semeval-2017 task 1: semantic textual similarity-multilingual and cross-lingual focused evaluation[J]. arXiv Preprint arXiv:1708.00055, 2017.

[作者简介]



裘炜程（1997-），男，上海人，上海交通大学硕士生，主要研究方向为车联网信息安全、网络安全、人工智能。



陈秀真（1977-），女，山东聊城人，博士，上海交通大学副教授，主要研究方向为车联网信息安全、网络信息系统安全检测与评估、社交网络大数据分析。



马颖华（1973-），女，山东泰安人，博士，上海交通大学讲师，主要研究方向为网络内容安全、自然语言处理。



马进（1977-），女，山东滕州人，博士，上海交通大学高级工程师，主要研究方向为大数据与人工智能应用、车联网信息安全、网络空间安全综合管理新技术。



周志洪（1979-），男，江西九江人，上海交通大学讲师，主要研究方向为密码应用、安全测评、车联网安全。