第3卷 第3期 2021年6月
Vol.3 No.3 Jun.2021

ISSN 2096-742X
CN 10-1649/TP

文献DOI：
10.11871/jfdc.issn.
2096-742X.2021.
03.002

文献PID：
21.86101.2/jfdc.
2096-742X.2021.
03.002

页码：9-18

获取全文

网络安全知识图谱关键技术

李序[1,2*]，连一峰[2]，张海霞[2]，黄克振[2]

1.中国科学院大学，北京 100049
2.中国科学院软件研究所，可信计算与信息保障实验室，北京 100190

摘　要：【目的】复杂多变的网络攻击活动对网络安全工作带来了严峻挑战。将知识图谱引入网络安全领域，有助于刻画展现安全态势，支持安全决策和预警预测。【方法】本文综述了目前国内外知识图谱相关技术的研究进展及其在网络安全领域的应用现状。【结果】在此基础上，阐述了构建网络安全知识图谱的技术架构，定义了网络安全本体模型，采用深度学习的方法进行实体抽取和关系抽取，利用基于规则和基于知识表示学习的方法进行图谱推理，实现网络安全知识补全和分析挖掘。

关键词：网络安全；知识图谱；深度学习；威胁情报

Key Technologies of Cyber Security Knowledge Graph

LI Xu [1,2*], LIAN Yifeng [2], ZHANG Haixia[2], HUANG Kezhen [2]

1. University of Chinese Academy of Sciences, Beijing 100049, China
2. Trusted Computing and Information Assurance Laboratory, Institute of Software Chinese Academy of Sciences, Beijing 100190, China

Abstract: **[Objective]** Complex and changeable network attack activities bring severe challenges to network security. Introducing the knowledge graph into the field of network security is helpful to security situation depiction, security decision-making support, and early warning prediction. **[Methods]** This paper summarizes the research progress of knowledge graph technology at home and abroad and its application in the field of network security. **[Results]** On this basis, this paper expounds the technical framework of constructing the network security knowledge graph, defines the network security ontology model, uses the method of deep learning to extract entities and relations, uses rule-based and knowledge-based representation methods to carry out graph reasoning, and achieves the network security knowledge complement and analysis mining.

Keywords: cyber security; knowledge graph; deep learning; threat intelligence

基金项目：国家重点研发计划“网络空间地理图谱构建与智能认知关键技术研究”（2020YFB806500）课题四“基于网络空间地理图谱的网络安全行为智能认知技术研究”（2020YFB806504）
*通讯作者：李序（E-mail:lixu2019@iscas.ac.cn）

引　言

近年来，网络安全事件频发，网络攻击手段日益呈现复杂多变的特征，新型攻击工具层出不穷，单纯依靠入侵防御系统等被动防御手段已经无法有效地维护网络空间安全，特别是近年来频发的针对关键信息基础设施的攻击活动，对国家网络空间安全保障工作带来了巨大挑战[1]。同时，大数据、人工智能等技术的发展，也为网络安全防护提供了新的解决方案。互联网中存在大量的网络安全相关数据，例如防火墙、入侵检测系统等监测到的网络安全告警数据、网络安全研究机构或厂商建立的漏洞信息库（如 CNNVD），以及互联网安全论坛和厂商发布的安全通告等。安全分析人员通过挖掘此类数据中的信息，可以为网络安全态势感知提供支撑，实现安全预警预测，支持网络安全决策。然而，网络安全数据存在海量化、分散化、碎片化以及关系隐蔽化的特点，如何及时、精准地对海量数据进行分析处理，提取关键要素和关联关系，挖掘潜在的有价值信息，是网络安全领域面临的重要问题。

1988 年，Berners-Lee 率先提出了语义网（Semantic Web）的概念[2]，核心思想是在网页数据中添加能够被计算机所理解的语义信息，从而提升机器的理解能力。作为语义网的数据支撑，知识图谱（Knowledge Graph）的概念由谷歌公司于 2012 年提出，旨在实现更智能的搜索引擎，并于 2013 年开始在学术界和业界普及。知识图谱可以通过统一的框架将多源异构的数据组织起来，利用图结构表达数据之间的语义关系，为数据的分析和挖掘提供了支持。随着深度学习等人工智能技术的发展，知识图谱技术在金融风控、证券投资、医疗和地理信息等领域得到了广泛的应用。在网络安全领域，通过对海量安全数据进行知识抽取、融合和推理，能够实现多源异构数据的关联挖掘，从而在目标画像、APT 检测、攻击溯源等方面发挥作用。

目前，网络安全知识图谱的研究尚处于起步阶段，对于构建和应用网络安全领域图谱的整体技术框架的研究很少，本文重点对网络安全领域知识图谱的各类关键技术进行研究，提出了网络安全知识图谱的技术架构。

本文第 1 节介绍相关技术的国内外研究现状，第 2 节提出网络安全知识图谱技术架构，从本体模型、实体抽取、关系抽取、图谱构建与推理方法等方面详细阐述知识图谱关键技术，最后第 3 节对全文进行总结。

1　国内外研究现状

知识图谱的核心是本体结构[3]。本体是对一个特定领域中的概念及其之间关系的一种描述。知识图谱描述的是真实世界中存在的实体或概念，强调实体和属性值。一个本体可以用五元组来表达：O = (C,R,F,A,I)，C 是本体概念的集合，描述领域内的实际概念；R 是关系集合，描述概念之间的关系；F 是上下文关系的集合；A 是公理集合，代表本体内存在的事实，可以对本体内的概念或关系进行约束；I 表示实例的集合。

网络安全知识图谱在语义网技术作为知识表示的基础上，最重要的是本体结构[4]。Undercoffer 等人[5]提出了一个针对网络攻击的本体结构并应用到了分布式入侵检测系统中，作者分析了 4 000 多种网络攻击，从目标和攻击两个维度进行建模；Herzog 等人[6]定义网络安全本体模型的核心概念包括资产、威胁、漏洞和对策，并描述了资产与漏洞、威胁与目标资产之间的关联关系；Iannacone 等人[7]面向网络安全整体领域构建了一种本体，包含了 15 种实体及 115 个属性；SYED 等人[8]扩展了 Undercoffer 提出的面向入侵检测系统的本体，提出了一个更为通用的网络安全知识本体——UCO，可以将网络安全本体映射为 STIX 格式，对应 CVE 等网络安全知识库以及 DBPedia 等通用知识库。除此之外，国内很多学者也对网络安全领域的本体构建进行了研

究，贾焰等人[9]基于现有的漏洞数据库和攻击规则库，构建了包含漏洞、资产、软件、操作系统和攻击在内的网络安全实体；王通等人[10]根据威胁情报目标需求，参考威胁情报模型 STIX 和攻击模式模型 CAPEC 构建了网络威胁情报本体模型。

信息抽取是从自然语言文本中抽取指定类型的实体、关系、事件等事实信息，并形成结构化数据输出的文本处理技术，是构建知识图谱的关键技术。在知识图谱构建中，信息抽取主要包括实体抽取和关系抽取两项任务。

实体抽取又称为命名实体识别，目前的命名实体识别技术主要包括基于规则的方法、基于统计学习的方法和基于深度学习的方法。基于规则的方法一般由领域专家手工构建规则模板，选择词语的统计信息、指示词等作为特征，以模式匹配为主要手段，例如 Balduccini 等人[11]提出将本体与正则表达式相结合来抽取网络日志中的实体，该方法采用遗传算法生成正则表达式对日志段落中的信息进行标记，然后通过本体将标记信息匹配为实体；Liao 等人[12]采用语法树和正则表达式相结合的方法来识别网络安全博客文本中的失陷指标（Indicators Of Compromise）。基于规则的方法对于实体识别的准确率较高，但是需要耗费大量人力来构建规则，并且规则的移植性较差。基于统计学习的方法是将命名实体作为序列标注或多分类任务来处理，主要采用最大熵、条件随机场、隐马尔可夫等模型。随着机器学习技术的发展，出现了很多命名实体识别工具，例如 Stanford NLP、Stanform NER 等，但这些工具都是基于通用知识语料库进行训练的，直接应用到网络安全领域的信息抽取中并不能取得较好的结果。贾焰等人[9]使用现有漏洞数据库中的“influence platform”字段进行汇总，构建了实体字典，选择 Standform NER 中的字典特征进行训练，取得了较好的效果；Joshi 等人[13]在条件随机场（CRF）模型的基础上采用网络安全语料进行训练。基于统计学习的方法可以自动抽取实体，但需要大量的人工标注数据。随着深度学习技术的发展，神经网络方法被广泛应用到了命名实体识别任务中，并成为目前的主流方法，其中 Huang 等人[14]首次将 BiLSTM-CRF 模型应用到了命名实体识别中，利用双向长短时记忆网络（LSTM）进行特征提取和 CRF 进行实体标注；Houssem 等人[15]利用 LSTM 进行网络安全实体识别，也取得了较好的效果。

信息抽取中的另外一项任务是关系抽取，不同的关系将独立的实体连接在一起形成知识图谱。目前关系抽取主要分为三种方法：基于规则的模式匹配方法、基于监督学习的方法和基于半监督或无监督的方法。早期的关系抽取主要采用基于规则的模式匹配方法，由领域专家定义各类关系的规则，然后使用规则和文本进行模式匹配，但是领域专家无法对所有关系的规则进行穷举。基于监督学习的方法把关系抽取作为多分类问题来处理，每一种关系都是一个类别，通过标签数据对分类器进行训练。这种方法依赖于标注数据的规模和特征的选择，获得大量标注数据的代价通常是非常高昂的。为了解决这个问题，出现了基于半监督或无监督的关系抽取方法，主要包括基于 Bootstrapping 的方法和远程监督的方法，其中 Bootstrapping 方法利用少量实例作为初始种子 (seed tuples) 集合，通过学习得到新的模式 (pattern)，进而基于新的模式发现更多的实例，不断迭代从非结构化数据中寻找和发现新的潜在关系三元组；Mintz 等人[16]提出了远程监督方法，通过将知识库与非结构化文本对齐来自动构建大量训练数据，然后构建特征用于训练分类器；Riede 对传统的远程监督学习方法进行改进，提出了增强的远程监督假设，即“如果两个实体之间存在某种关系，那么至少有一个提到两个实体的句子可以表达这种关系”，使用无向图模型预测实体之间的关系以及哪个句子表达了这个关系，与原始的远程监督方法相比，错误率降低了 31%；Zeng 等人[17]使用卷积神经网络来自动提取特征，解决了采用词性标注、依存句法树等技术构建特征时错误率偏高的问题；Miwa

等人 [18] 提出了使用双向 LSTM 和树形 LSTM 同时对实体和句子进行建模的方法。在网络安全领域的关系抽取中，Pingle 等人 [19] 在网络安全语料库上训练 Word2Vec 模型对实体进行词嵌入，采用前馈神经网络 FFNN 预测实体间的关系。

在网络安全知识图谱的构建和推理方面，绿盟科技 [20] 基于知识图谱进行 APT 组织的追踪分析，通过采集威胁情报、各机构发布的 APT 报告及安全通告等数据，定义 APT 攻击本体，建立 APT 攻击知识图谱，实现对 APT 攻击行为的追踪溯源。瑞星公司构建了威胁情报及网络安全知识图谱 [21]，包含 100 亿 + 实体以及 400 亿 + 关系，其中，实体包含文件、漏洞、IP、黑客组织等网络安全攻击事件中涉及到的所有元素，与普通的威胁情报平台相比，在恶意软件领域可以发挥特长，将一些恶意软件模糊搜索、自动归类的技术应用到了知识图谱的检索中。在学术界，也有很多研究人员对知识图谱在网络安全领域中的应用开展了研究工作，Yulu 等人 [22] 基于网络安全知识图谱对网络攻击进行溯源分析；Wei 等人 [23] 通过知识图谱来过滤不相关的警报日志；Narayanan 集成不同来源的威胁情报构建网络威胁情报图谱 [24]，实现了简单的网络安全事件预测；陶源等人利用知识图谱建立日志审计分析模型，以支持网络安全等级保护工作 [25]。

2　网络安全知识图谱技术架构

当前，知识图谱相关技术发展迅猛，网络安全作为新兴的应用领域，相关的知识图谱本体模型、实体抽取、关系抽取，以及图谱构建及推理技术逐渐引起研究人员的重视。网络安全知识图谱技术架构主要分为三个层次，其中：

（1）本体构建层负责定义网络安全领域的概念及其关系，例如网络攻击者、攻击工具、木马病毒、攻击活动、安全事件、漏洞隐患、防护措施等；

（2）信息抽取层负责从多源异构的网络安全数据中抽取相关实体及其关系，将信息抽取过程中得到的实体进行对齐和链接，并通过对抽取到的实体及关系进行评估校验后构建知识图谱；

（3）知识推理层负责在初步构建的知识图谱基础上，通过知识推理分析挖掘新的实体或隐含关系，对图谱进行补全，提供网络安全决策支持。

2.1　本体模型

网络安全本体模型的构建应根据具体的目标需求来完成，例如针对 APT 攻击，本体模型应重点围绕 APT 攻击相关的组织、技术、工具、历史攻击活动、掌握资源等要素定义实体、属性及其关系；针对勒索病毒，则本体模型应重点定义病毒、代码特征、利用漏洞、目标对象、软硬件版本、传播范围、阻

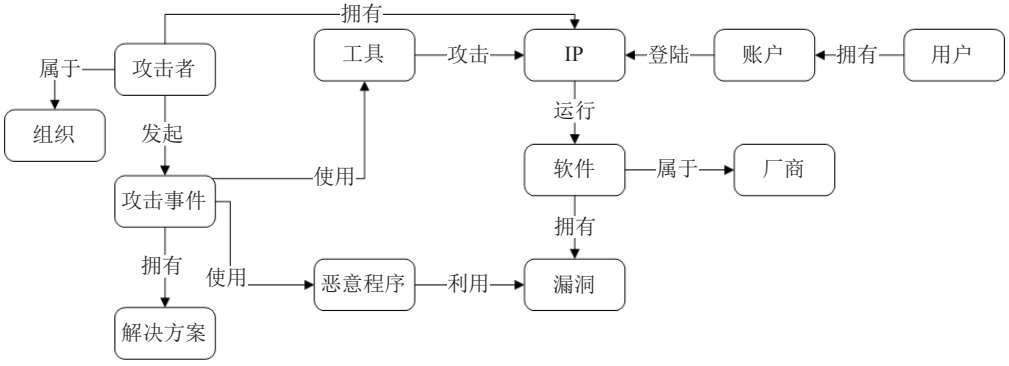


图 1 网络安全本体模型示例

Fig.1 An example of network security ontology model

断方式等要素。

图 1 给出了针对通用网络安全目标需求的本体模型示例。图中每个节点代表本体模型的一类实体，节点间的连接代表实体间关系。例如，归属于某组织的攻击者利用攻击工具或恶意程序，发起对某个 IP 主机的攻击事件，该攻击工具或恶意程序利用了某款软件存在的安全漏洞。

2.2 实体抽取

网络安全实体抽取任务主要面向的是网络安全相关的非结构化文本，例如网络安全网站、论坛和各类社交媒体上发布的内容。本文介绍基于经典的双向长短时记忆网络 - 条件随机场（BiLSTM-CRF）模型的实体抽取方法。其中双向长短时记忆网络（BiLSTM）负责学习句子的上下文关系，条件随机场（CRF）则负责处理实体类型之间的依赖关系，模型结构如图 2 所示。

模型的第一层是词嵌入层，通过 Word2Vec 工具，将单词序列 $(w_1, w_2, w_3, ..., w_T)$ 中的每个单词映射成低维向量 $x_i \in R^d$，d 为词向量的维度。

模型的第二层是双向 LSTM 层，负责自动提取句子特征。将单词序列的各个词向量 $(x_1, x_2, x_3, ..., x_T)$ 作为双向 LSTM 在各个时间点的输入，再将正向 LSTM 输出的隐状态序列与反向 LSTM 在各个位置输出的隐状态进行拼接，得到完整的隐状态序列 $(h_1, h_2, h_3, ..., h_T) \in R^{Txm}$，接入一个线性层，将隐状态向量从 m 维映射为 13 维向量（共有 13 种实体类别），从而得到自动提取的句子特征，记作

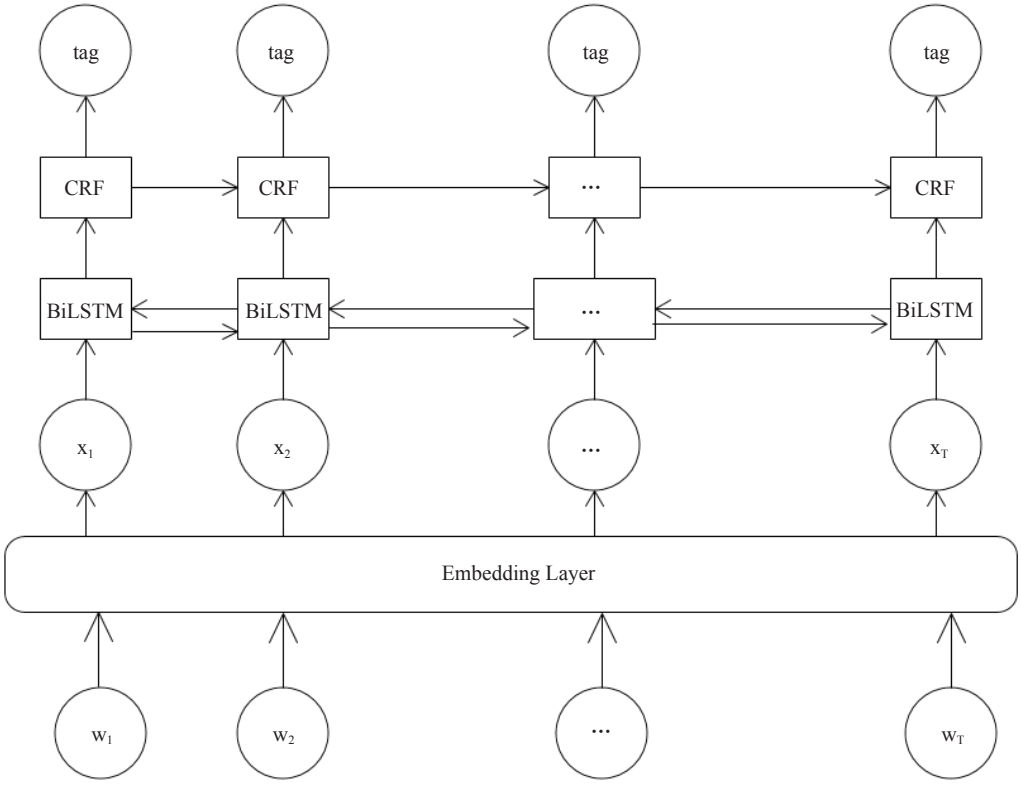


图 2 BiLSTM-CRF 模型

Fig.2 BiLSTM-CRF model

$L=(L_1,L_2,L_3,...,L_T)\in R^{Tx13}$，$L_i\in R^{13}$ 的每一维 L_{ij} 是把单词 w_i 分类为第 j 类实体的得分：

$$score(x,y)=\sum_{i=1}^{T}L_{iy_i}+\sum_{i=1}^{T+1}A_{y_{i-1}y_i} \quad (1)$$

进而得到归一化之后的概率如公式（2）所示，$P(y|x)$ 表示将单词序列 x 的实体类别预测为 y 的概率，Y 表示单词序列 x 对应所有可能的实体类别序列构成的集合，$|Y|=13^T$：

$$P(y|x)=\frac{\exp(score(x,y))}{\sum_{y'\in Y}\exp(score(x,y'))} \quad (2)$$

模型通过最大化似然函数进行训练，一个训练样本 (x,y^x) 的似然函数计算如公式（3）所示，其中 $P(y^x|x)$ 表示单词序列 x 的实体类别序列为 y^x 的概率：

$$\log P(y^x|x)=score(x,y^x)-\log(\sum_{y'\in Y}\exp(score(x,y'))) \quad (3)$$

最后由条件随机场（CRF）层使用动态规划 Viterbi 算法来得到预测值。

2.3 关系抽取

针对网络安全关系抽取任务，由于缺乏中文标注的网络安全实体关系数据集，因此传统的模式匹配和监督学习方法并不适用。考虑使用远程监督方法，在只需要少量标注数据集的基础上进行模型训练。本文介绍分段卷积神经网络（Piecewise Convolutional Neural Networks，PCNN）模型 [26]，将远程监督学习看作是一个多实例学习问题，使用卷积神经网络（CNN）模型自动学习文本特征，在最后的池化操作中使用分段池化的方法，利用该模型进行网络安全实体关系的识别。PCNN 模型结构如图 3 所示。

PCNN 模型的第一层是词嵌入层，将输入的单词转化为词向量。PCNN 模型根据每个单词相对两个实体的位置信息进行拼接形成位置向量，然后在卷积层通过 CNN 模型来提取文本特征。常用的最大池化操作因为对句子长度特征池化，不适合关系抽取任务。PCNN 模型将句子按照实体位置分为三段，分别对每段进行池化，最后通过 softmax 层计算句子属于每类关系的得分。

PCNN 使用多实例学习方法来降低错误标注带来的影响。多实例学习每次使用一袋包含同一对实体的样本，袋的标签为实体对在知识图谱中的关系，袋中的数据相互独立。每次对 M 袋数据进行训练，首先从每一袋数据中选取最具代表性的样本，计算方式如下：

$$y^*=\arg\max_j p(y_i|m_i^j;\theta),1\le j\le q_i \quad (4)$$

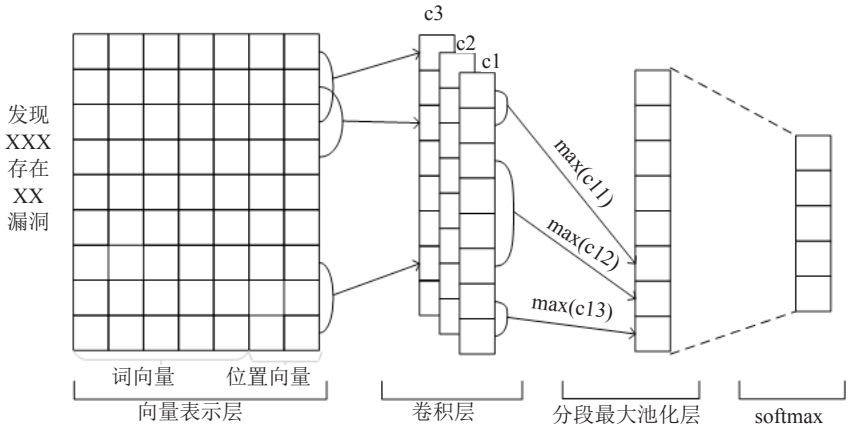


图 3 PCNN 模型

Fig.3 PCNN model

其中，q_i 表示第 i 袋样本的数量，y_i 为第 i 袋数据的标签，m_i^j 表示第 i 袋数据中的第 j 个样本；然后，将该样本的标签视为此袋数据的预测标签，计算交叉熵损失：

$$J(\theta)=\sum_{i=1}^{M}\log p(y_i \mid m_i^j \mid \theta) \quad (5)$$

其中，M 表示袋的数量，y_i 为第 i 袋数据的标签，m_i^j 为第 i 袋数据中选出的最具代表性的样本。

2.4 图谱构建与推理方法

经过实体抽取和关系抽取之后，网络安全数据中的实体和关系可以链接到本体模型中定义的概念及关系，通过 Neo4j 等图数据库可以存储初步形成的知识图谱。为保证图谱的质量，还需对图谱中的知识进行评估校验，去除多数据源中的冗余知识，并研判解决存在冲突的信息，避免在知识推理过程中错误传播。

由于很多网络安全数据的组织形式比较简单，信息抽取之后创建的知识图谱中主要包含句子中显式表达的关系，还需要在现有知识的基础上通过知识推理，挖掘潜在的隐含知识，丰富网络安全知识图谱。网络安全知识图谱的知识推理可以结合具体的任务需求，综合使用基于规则的推理和基于知识表示学习的推理方法。某些网络安全数据可以根据专家经验知识定义规则，例如对于某些具有鲜明特征的 APT 组织的攻击手段或技术方法，可以由专家定义规则知识库，将图谱知识与规则库进行模式匹配。

另一方面，知识表示学习可以将图谱中离散的关系和实体映射成低维的连续向量，同时不损失知识图谱中的原有语义。目前常用的方法主要是基于深度学习的知识表示学习，针对本文构建的网络安全知识图谱，将＜实体，关系，实体＞三元组映射成低维的向量，使用循环神经网络模型进行多步知识推理。目前在知识图谱推理的基础研究中，结合领域知识图谱的本体知识来构建图谱表示模型的研究成果较少，研究针对网络安全领域知识图谱的表示模型，可以在一定程度上提高图谱推理的准确率，实现更为精准、更具可操作性的安全决策推理。

3 小结

本文提出了网络安全知识图谱的技术架构，从本体模型定义、实体抽取、关系抽取、图谱构建及推理等方面阐述了网络安全领域知识图谱的关键技术。当前，知识图谱在信息检索、推荐系统等领域得到了广泛应用，在网络安全领域中也开始发挥越来越重要的作用。将知识图谱引入网络安全领域中，可以将互联网中零散的网络安全数据组织在一起，挖掘网络安全数据之间潜在的语义关系，帮助全方位掌握威胁信息，对当前的网络安全态势做出判断，进而预警、预测未来可能发生的威胁。

本文提出的网络安全知识图谱的技术架构中知识抽取、推理等关键技术主要还是基于深度学习技术，然而使用深度学习技术构建知识图谱仍然存在不准确、不全面的问题，首先深度学习技术依赖于大量的标注的语料库，目前通用知识语料库主要还是关注人物、事物等，将深度学习知识图谱引入到领域图谱中时会出现准确率大大降低等问题，可移植性较低；其次，知识图谱涉及各个方面各个场景，并不像图片、语音可以在单一的维度来训练模型，从而达到足够的精度和召回率；在知识推理方面，目前主流的方法还是基于深度学习与知识表示学习，单纯依赖大量的标注数据，在网络安全领域的知识图谱中，有诸多的先验知识无法有效使用并融合到深度学习的推理模型当中，以提高知识推理的精度。

后续可以围绕如何提升网络安全领域信息抽取的准确性，如何融合已有的专家知识构建网络安全领域知识图谱表示模型和推理模型，进一步开展更多的研究和探索工作，以提高网络安全主动防御能力。

利益冲突声明

所有作者声明不存在利益冲突关系。

参考文献

[1] 郭启全, 张海霞. 关键信息基础设施安全保护技术体系[J].信息网络安全, 2020, 20(11):1-9.

[2] Berners-Lee T, Handler J, Lassila O . The Semantic Web[J]. Scientific American, 2003, 284(5):34--43.

[3] 杨玉基, 许斌, 胡家威, 等. 一种准确而高效的领域知识图谱构建方法[J]. 软件学报, 2018, 029(010):2931-2947.

[4] Razzaq A, Anwar Z, Ahmad H F, et al. Ontology for attack detection: An intelligent approach to web application security[J]. Computers & Security, 2014, 45(sep.):124-146.

[5] Undercoffer J, Pinkston J, Joshi A, et al. A Target-Centric Ontology for Intrusion Detection[C]. Proceedings of the IJCAI-03 Workshop on Ontologies and Distributed Systems, 2003: 101-108.

[6] HERZOG, Almut; SHAHMEHRI, Nahid; DUMA, Claudiu. An ontology of information security[J]. International Journal of Information Security and Privacy (IJISP), 2007, 4: 1-23.

[7] Michael Iannacone, Shawn Bohn, Grant Nakamura, et al. Developing an Ontology for Cyber Security Knowledge Graphs[C]. Proceedings of the 10th Annual Cyber and Information Security Research Conference on - CISR '15, 2015:1-4.

[8] Syed Z, Padia A, Finin T, et al. UCO: A Unified Cybersecurity Ontology[C]// AAAI Workshop on Artificial Intelligence for Cyber Security, 2016: 14-21.

[9] 贾焰, 亓玉璐, 尚怀军, 等. 一种构建网络安全知识图谱的实用方法[J]. Engineering, 2018,4(01):117-133.

[10] 王通, 艾中良, 张先国. 基于深度学习的威胁情报知识图谱构建技术[J]. 计算机与现代化, 2018, 280(12):25-30.

[11] Balduccini M, Kushner S, Speck J. Ontology-Driven Data Semantics Discovery for Cyber-Security[C]// International Symposium on Practical Aspects of Declarative Languages. Springer International Publishing, 2015: 1-16.

[12] Liao X, Kan Y, Wang X F, et al. Acing the IOC Game: Toward Automatic Discovery and Analysis of Open-Source Cyber Threat Intelligence[C]// Acm Sigsac Conference on Computer & Communications Security. ACM, 2016: 755-766.

[13] Joshi A, Lal R, Finin T, et al. Extracting Cybersecurity Related Linked Data from Text[C]// Semantic Computing (ICSC), 2013 IEEE Seventh International Conference on. IEEE, 2013: 252-259.

[14] Huang Z , Wei X , Kai Y . Bidirectional LSTM-CRF Models for Sequence Tagging[J]. IEEE Intelligent Systems, 2017,32(6): 74-80.,

[15] Houssem Gasmi, Abdelaziz Bouras, Jannik Laval. LSTM recurrent neural networks for cybersecurity named entity recognition[C]. The Thirteenth International Conference on Software Engineering Advances, 2018: 1-6.

[16] Mintz M, Bills S, Snow R, et al. Distant supervision for relation extraction without labeled data[C]// ACL 2009, Proceedings of the 47th Annual Meeting of the Association for Computational Linguistics and the 4th International Joint Conference on Natural Language Processing of the AFNLP, 2-7 August 2009, Singapore. Association for Computational Linguistics, 2009: 1003-1011.

[17] Zeng D, Liu K, Lai S, et al. Relation classification via convolutional deep neural network[C]//Proceedings of COLING 2014, the 25th international conference on computational linguistics: technical papers. 2014: 2335-2344.

[18] Miwa M, Bansal M. End-to-End Relation Extraction using LSTMs on Sequences and Tree Structures[C]// Proceedings of the 54th Annual Meeting of the Association

for Computational Linguistics (Volume 1: Long Papers). 2016: 1105-1116.

[19] Pingle A, Piplai A, Mittal S, et al. RelExt: Relation Extraction using Deep Learning approaches for Cybersecurity Knowledge Graph Improvement[C]//2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM). IEEE, 2020: 879-886.

[20] 绿盟科技．基于知识图谱的 APT 组织追踪治理[EB/OL]．[2020-12-30]．https://mp. weixin. qq. com/s/CluHeu1oy7DneBuR0cXZSQ.

[21] 瑞星.瑞星发布威胁情报及网安知识图谱[EB/OL]．[2019-05-21]．https://wenku.baidu.comview/ fd935990bc23482fb4daa58da0116c175e0e1e44.

[22] Qi Y, Jiang R, Jia Y, et al. Association Analysis Algorithm Based on Knowledge Graph for SPACE-Ground Integrated Network[C]//2018 IEEE 18th International Conference on Communication Technology (ICCT). IEEE, 2018: 222-226.

[23] Wei W, Rong J, Yan J, et al. KGBIAC: Knowledge Graph Based Intelligent Alert Correlation Framework[J]. Springer, Cham, 2017: 523-530.

[24] Narayanan S, Ganesan A, Joshi K, et al. Cognitive Techniques for Early Detection of Cybersecurity Events[Z]. arXiv preprint arXiv:2018,1808.00116.

[25] 陶源, 黄涛, 李末岩, 等.基于知识图谱驱动的网络安全等级保护日志审计分析模型研究[J].信息网络安全, 2020, 20(1): 46-51.

[26] Zeng, Kang L, Chen Y, et al. Distant Supervision for Relation Extraction via Piecewise Convolutional Neural Networks[C]//Conference on Empirical Methods in Natural Language Processing, 2015: 1753-1762.

收稿日期：2021 年 5 月 27 日

李序，中国科学院软件研究所，硕士研究生，主要研究兴趣为网络安全态势感知。

本文中负责总体统稿、知识图谱技术调研和综述。

LI Xu is currently a postgraduate of Institute of software, Chinese Academy of Sciences. Her current research interest is cyber security situation awareness.

In this paper, she is responsible for the overall draft as well as the survey and summary of knowledge graph technology.

E-mail: lixu2019@iscas.ac.cn

连一峰，中国科学院软件研究所，研究员，博士生导师，主要研究方向包括网络安全态势感知技术、网络攻防技术、安全测评技术、等级保护关键技术等，主持承担国家 863 计划、国家自然科学基金、国家高技术产业化等 20 余项重要科技项目，发表学术论文 50 余篇，出版专著 4 部，发明专利 14 项，编制国家技术标准 3 项。

本文中负责网络安全知识图谱技术架构分析。

LIAN Yifeng, researcher and doctoral supervisor of Institute of software, Chinese Academy of Sciences, focuses on cyber security situation awareness technology, cyber attack and defense technology, security evaluation technology, key technology of classified protection, etc. He has presided over and undertaken more than 20 important scientific and technological projects such as national 863 plan, National Natural Science foundation of China and national high-tech industrialization, and published more than 50 academic papers, four monographs, 14 invention patents, and 3 national technical standards.

In this paper, he is responsible for the cyber security knowledge graph technology architecture analysis.

E-mail: lianyifeng@iscas.ac.cn

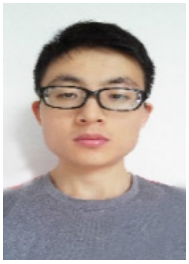
张海霞，中国科学院软件研究所，高级工程师，博士，长期从事网络及信息安全技术研究、规划设计与工程建设工作，主要研究方向包括信息安全测评技术、等级保护关键技术、网络安全监测预警技术等，先后承担国家863计划、国家发改委、公安部、国家测评中心、认证中心等国家级、部委级重要科技项目20余项，在核心学术期刊发表论文多篇，申请国家发明专利多项。

本文中负责网络安全知识图谱本体构建指导。

ZHANG Haixia, doctor of engineering, is a senior engineer of Institute of software, Chinese Academy of Sciences. She has long been engaged in cyber and information security technology research, planning, design and engineering construction. Her key research directions include information security evaluation technology, classified protection key technology, network security monitoring and Pre-warning Technology, etc. She has successively undertaken the National 863 program, and more than 20 other national and ministerial level important science and technology projects from national development and Reform Commission, Ministry of public security, and national evaluation center Certification Center. She has published many papers in core academic journals, and applied for a number of national invention patents.

In this paper, she is responsible for guiding the noumenon definition of cyber security knowledge graph.

E-mail: zhanghaixia@iscas.ac.cn



黄克振，中国科学院软件研究所，工程师，主要研究方向包括网络安全态势感知技术、网络攻防技术等，参与国家863计划、国家自然科学基金、国家高技术产业化等20余项重要科技项目，发表学术论文多篇，申请国家发明专利多项。主要研究兴趣为网络安全威胁感知技术，在国内重要期刊及会议上发表学术论文10余篇。

本文中承担信息抽取技术分析。

HUANG Kezhen, the engineer of Institute of software, Chinese Academy of Sciences, focuses on cyber security situation awareness technology, network attack and defense technology, etc. He has participated in more than 20 important scientific and technological projects such as national 863 plan, National Natural Science Foundation of China and national high-tech industrialization, published many scientific papers and applied for many national invention patents. His main research interest is cyber security threat perception technology, and he has published more than 10 academic papers in important domestic journals and conferences.

In this paper, he is responsible for the analysis of information extraction technology.

E-mail: huangkezhen@iscas.ac.cn

引文格式：李序,连一峰,张海霞,黄克振. 网络安全知识图谱关键技术[J].数据与计算发展前沿,2021,3(3):9-18.DOI:10.11871/jfdc.issn.2096-742X.2021.03.002.PID:21.86101.2/jfdc.2096-742X.2021.03.002.

LI Xu, LIAN Yifeng, ZHANG Haixia, HUANG kezhen. Key Technologies of Cyber Security Knowledge Graph [J]. *Frontiers of Data & Computing*,2021,3(3):9-18.DOI:10.11871/jfdc.issn.2096-742X.2021.03.002.PID:21.86101.2/jfdc.2096-742X.2021.03.002.