doi: 10.3969/j.issn.0490-6756.2019.03.015

# 基于 BLSTM-CRF 模型的安全漏洞领域命名实体识别

张若彬[1], 刘嘉勇[2], 何 祥[1]
(1. 四川大学电子信息学院, 成都 610065; 2. 四川大学网络空间安全学院, 成都 610065)

**摘 要**: 非结构化文本资源提供了大量与漏洞相关的信息,传统的特定领域实体识别依赖特征模板和领域知识来识别相关实体,其识别性能很大程度上依赖于人工选取的特征函数质量. 如何利用机器挖掘文本隐含的特征,而不需要人工详细地制定领域术语的特征表达是一项具有挑战性的任务. 该文针对安全漏洞领域,提出一种双向长短期记忆网络 BLSTM 与条件随机场 CRF 相结合的安全漏洞领域实体识别模型,并使用基于词典的方法对结果进行校正,F 值可达到 85%以上. 实验表明,该方法在提高实体识别的准确率和召回率的同时,能够显著地降低人工选取特征的工作量.

**关键词**: 安全漏洞; 实体识别; BLSTM; CRF
**中图分类号**: TP391.1 **文献标识码**: A **文章编号**: 0490-6756(2019)03-0469-07

# Named entity recognition for vulnerabilities based on BLSTM-CRF model

*ZHANG Ruo-Bin*[1], *LIU Jia-Yong*[2], *HE Xiang*[1]
(1. College of Electronics and Information Engineering, Sichuan University, Chengdu 610065, China;
2. College of Cybersecurity, Sichuan University, Chengdu 610065, China)

**Abstract**: Unstructured text resources provide a large amount of information related to vulnerability. Traditional domain-specific entity recognition relies on feature templates and domain knowledge to recognize related entities. The recognition performance depends largely on the quality of manually selected feature functions. It is a challenging task to mine the features implied by the text automatically, rather than manually formulate the characterization of the domain terminology. In this paper, a BLSTM and CRF security vulnerability domain entity recognition model (BLSTM-CRF model) isproposed and a dictionary is used to correct the results generated by the model. The F value can reach 85%. Experiments show that this method can significantly reduce the workload of manually selecting features while improving the precision and recall.

**Keywords**: Cyber vulnerabilities; Named entity recognition; BLSTM; CRF

## 1 引 言

已知漏洞的预警信息或是新发现的漏洞信息通常首先出现在非结构化文本资源中,例如信息安全从业者活跃的黑客论坛中的博客文章,安全机构发布的公告等. 而且其中大部分信息都是以适合安

**收稿日期**: 2018-12-04
**基金项目**: 国家重点研发计划网络空间安全重点专项 (2017YFB0802900)
**作者简介**: 张若彬(1994-), 女, 硕士研究生, 研究方向为信息系统安全. E-mail: zrb@stu.scu.edu.cn
**通讯作者**: 何祥. E-mail: rivsearcher@gmail.com,

全专家理解的文本形式提供的,不易被自动安全系统理解或直接使用.因此合理利用非结构化文本中的漏洞信息,可以追踪并还原信息安全事件,以提供新的漏洞和攻击的早期预警,跟踪现有漏洞和攻击的演变,为归因提供证据并估计已知问题的发生率和地理分布.其中,命名实体识别是最主要的任务之一.

目前命名实体识别模型使用有监督的方法对大量标注文本语料进行训练,在特定领域的命名实体识别中均取得了不错的成绩,冯蕴天等[1]利用条件随机场CRF(Conditional Random Field)使用多种特征模板识别出军事领域实体.郭剑毅等[2]利用层叠CRF模型实现了简单旅游命名实体识别.利用领域知识基于CRF模型的方法在特定领域的实体识别取得了一定效果,但其识别性能在很大程度上依赖于大量的人工标注以及人工选取的特征函数,对大量专业性高的文本进行手动注释通常成本太高,无法成为有效的解决方案.

神经网络具有更强的泛化性和更低的特征工程依耐性[3],其在自然语言处理领域备受关注.Huang等[4]首次将BLSTM-CRF应用于自然语言处理中基准标记数据集,获得更高的准确率的同时,对字词嵌入的依赖更少.Li等人[5]利用基于CRF的双向LSTM深层神经网络模型实现了对生物医学文本中的不规则实体识别,$F$值达到了81.09%.马建红等[6]面向新能源汽车专利文本领域提出了一种基于Attention的双向长短期记忆网络BLSTM(Bidirectional Long Short-Term Memory)与CRF相结合的领域术语抽取模型.利用BLSTM模型,可以解决目前机器学习中过度依赖领域知识以及人工定义特征问题,实现了端到端的命名实体识别模式.但BLSTM模型在处理输出标签时无法很好地处理有强烈依赖关系的数据,CRF相对于其他模型可以更有效地关注上下文标注信息,所以在BLSTM模型的基础上结合CRF模型可以有效提升识别准确率[7].

## 2 相关研究

Mulwad等[8]使用SVM分类器提取漏洞相关的文本,并查询名为Wikitology知识库,以生成有关漏洞、攻击和威胁的断言.Weerawardhana等[9]在大量标注语料的基础上,构建特征工程,完成漏洞相关实体识别.Joshi等[10]利用通用领域CRF模型,制定大量相应特征模板,识别漏洞相关实体.目前针对安全漏洞领域实体识别方法要么需要依据人工制定的特征模板要么查询知识库进行识别,而且实体识别多关注作用环境及攻击方法,少有学者关注漏洞类型实体提取.而漏洞类型有助于漏洞分类,便于检测和防止新漏洞攻击,提出早期预警.

针对安全漏洞领域实体识别难点,减少对人工特征和领域知识的依赖,考虑BLSTM模型可以有效利用上下文信息,自动挖掘特征,CRF模型可以根据实体的相关性对标注结果进行约束的基础上,本文提出BLSTM-CRF+校正模型,识别安全漏洞领域中漏洞名,漏洞类型,应用环境和攻击方法等4个类别的命名实体.

## 3 面向安全漏洞领域的命名实体识别模型

### 3.1 安全漏洞领域的实体

安全漏洞命名实体是指与安全漏洞相关的各种命名实体的统称,本文分析了几个与网络安全相关的博客,安全公告和CVE描述,并确定了一组与漏洞数据表示相关的实体.

(1) 漏洞编号:漏洞编号是漏洞唯一的标识.每个漏洞都有独一无二的编号,如"CVE-1999-1046".

(2) 漏洞名:漏洞名是人们在谈论经典漏洞时为其取得便于书写和记忆得名字,如"永恒之蓝(EternalBlue)"、"心脏出血(HeartBleed)"等.

(3) 漏洞类型:漏洞类型是一个漏洞所属的类型,如"反序列化漏洞","SSRF漏洞".

(4)漏洞利用条件:漏洞利用条件也称为漏洞所能作用的环境,是指它所能作用的应用商软件,包括软件供应商、操作系统、应用软件.

(a) 软件供应商:如"微软(Microsoft)","甲骨文(oracle)","苹果(Apple)".

(b) 操作系统:如"Windows","Ubuntu"等.

(c) 应用软件:包括浏览器以及各类软件应用,如"微信","福昕阅读器"等.

(5) 攻击方式:攻击方式是指黑客利用漏洞来展开攻击时使用的方法.如"内存溢出"、"拒绝服务"等.

鉴于漏洞编号有独特的命名方式:通常包含着具体的前缀,如"CVE-"、"CNVD-",并且呈现三段式的结构,因此可直接通过正则表达式对漏洞编号实体进行识别,不必利用深度学习模型来识别.

针对以上除漏洞编号实体以外的安全漏洞领域

实体，本文定义以下实体体系：$A=\{$ name，type，env，method$\}$，分别表示漏洞名，漏洞类型，应用环境和攻击方法. 采用 BIEO 标注法[11]，即“B_”代表一个实体开头，“I_”代表实体中间字符，“E_”代表实体中最后一个字符，“O”代表其他非实体字符，对安全漏洞领域实体进行标注，标注格式如表 1 所示.

**表 1　安全漏洞领域实体标注格式**

Tab. 1　Security vulnerability domain entity annotation format

| 标注符号 | 代表实体 |
|---|---|
| B_ name | 漏洞名实体开头字符 |
| I_ name | 漏洞名实体中间字符 |
| E_ name | 漏洞名实体结尾字符 |
| B_ type | 漏洞类型实体开头字符 |
| I_ type | 漏洞类型实体中间字符 |
| E_ type | 漏洞类型实体结尾字符 |
| B_ env | 漏洞应用环境实体开头字符 |
| I_ env | 漏洞应用环境实体中间字符 |
| E_ env | 漏洞应用环境实体结尾字符 |
| B_ method | 漏洞攻击方法实体开头字符 |
| I_ method | 漏洞攻击方法实体中间字符 |
| E_ method | 漏洞攻击方法实体结尾字符 |
| 0 | 其他非实体字符 |

## 3.2　BLSTM-CRF+校正联合模型框架

本文提出的 BLSTM-CRF＋校正模型是在 BLSTM 模型的基础上结合 CRF 模型对文本进行序列标注后再利用词典对标注结果进行校正最后识别实体的模型. 完整的识别流程如图 1 所示.

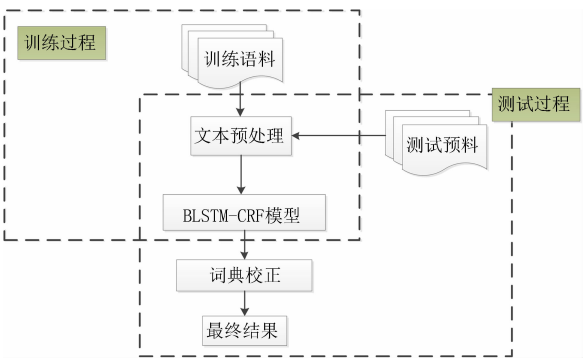


图 1　安全漏洞领域命名实体识别模型框架

Fig. 1　Security vulnerability domain named entity recognition model framework

(1) 首先对输入文本进行预处理. 即拆分句子为单个字符，然后对其进行向量化处理.

(2) 然后把预处理后的文本序列输入 BLSTM-CRF 联合模型. 输出文本序列的标注结果.

(3) 接着根据词典对模型标注好的语料进行校正，识别出模型没有识别出的实体.

(4) 最后输出最终结果.

3.2.1　文本预处理　在标记过程中可能会出现大量未登录词，为避免因错误的中文分词而导致新词识别率低下，本文对文本进行预处理时采用单字拆分方法：将语句以单独的字符划分，而不是以单词划分.

对文本进行分字处理后，进行 Character Embedding 向量化，Character Embedding 向量化是一种使用字典将字符映射到低维实数向量的方法，可为字符寻求更加深层次的特征表示[12]，同时避免 one hot 向量因维度过高而引起的向量稀疏问题. 处理过程如下，将一个含有 $n$ 个字的句子记作 $X=(x_1,x_2,\ldots,x_n)$，$x_i$ 是由 onehot 表示的字符向量. 利用 embedding 矩阵 $E=(e_1,e_2,\ldots,e_n)$ 将句子中的每个字符 $x_i$ 映射为低维稠密的字向量 $x_i'$，其中 $x_i'=(E\times x_i)\in R^d$，$e_i\in R^d$ 是预训练生成的词特征向量，$d$ 是 embedding 的维度. 在本文中，character embedding 由公开的 Glove[13] 程序在训练集上训练生成，对于未在训练集中出现的汉字，在测试集中采用随机初始化生成.

3.2.2　BLSTM-CRF 模型框架　安全漏洞相关文本经过预处理和 Character Embedding 向量化处理之后，进入 BLSTM-CRF 联合模型进行实体标注. 模型框架如图 2 所示.

(1) BLSTM 层.

LSTM 是一种特殊的循环神经网络（Recurrent Neural Networks，RNN），它通过引入记忆单元和门限机制很好的解决了 RNN 梯度消失问题. 对于许多实体识别任务，需要同时捕获上下文信息，然而 LSTM 模型仅从“过去”获取信息，对“未来”一无所知. 此时 BLSTM[14] 应运而生. BLSTM 在 LSTM 的基础上，提取每个序列前向状态和后向状态，以分别捕获过去和未来信息，不仅有效克服了梯度消失问题，还因为能捕获上下文，提供了更为全面的语义信息.

本文将预处理后安全漏洞领域非结构化文本的 char embedding 序列 $(x_1,x_2,\ldots,x_n)$ 作为 BLSTM 层输入，该层将前向 LSTM 输出的隐状态序列 $\overrightarrow{h_t}=(\overrightarrow{h_1},\overrightarrow{h_2},\ldots,\overrightarrow{h_n})\in R^{n\times m}$ 与反向 LSTM 输出的隐状态序列 $\overleftarrow{h_t}=(\overleftarrow{h_1},\overleftarrow{h_2},\ldots,\overleftarrow{h_n})\in R^{n\times m}$ 进行拼接，得到完整的隐状态序列 $h_t=[\overrightarrow{h_t};\overleftarrow{h_t}]\in R^{n\times 2m}$，$m$ 是隐藏层的单元数. 然后接入一个线性层，将隐状态映射到 $k$ 维，$k$ 是标注集的标签数，从而得到自动提取的特征，记作矩阵 $P$，$P_{ij}$ 表示将字 $x_i$ 分类

到第 $j$ 个标签的分值其中,然后经过 Softmax 层,即对各个位置进行 $k$ 分类,BLSTM 层输出每个字的最终标签. 但是仅仅利用 BLSTM 层对各个位置进行标注无法利用已经标注过的信息[8],所以接下来将接入一个 CRF 层来进行约束校正.

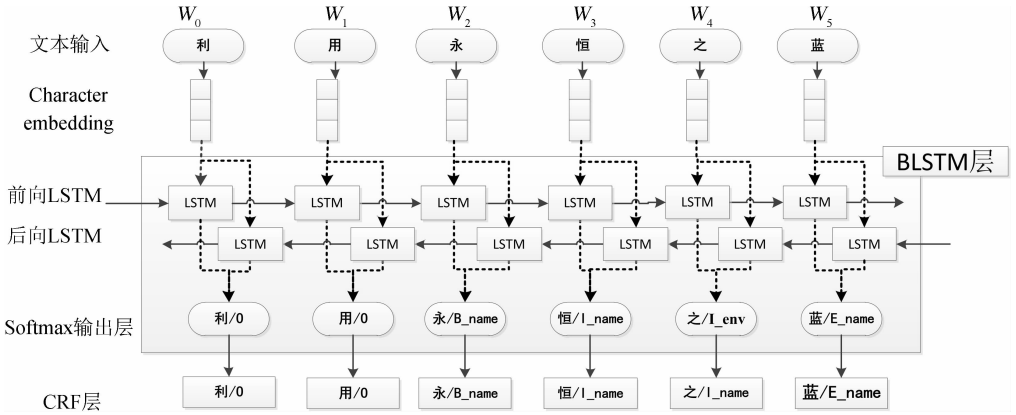


图 2 BLSTM-CRF 模型框架

Fig. 2 BLSTM-CRF model framework

(2) CRF 层.

CRF 是由 Lafferty 等[15,16]于 2001 年提出的一种判别式概率模型. 由于 BLSTM 层对语料进行自动标注返回的结果并不是每个字符都有积极的意义,如它不会考虑标签"B_name"与"I_name"之间的依赖关系,可能会在"B_name"标签之后接上"I_env"标签,此时就不能正确识别出一个漏洞名实体. 但将所有字符进行重新标记是对人力的严重浪费. CRF 模型可以有效的考虑上下文的依赖关系,基于此,在 BLSTM 模型后引入 CRF 模型,使得模型在结合上下文信息的同时可以有效地考虑标签前后的依赖关系,以确保它们是有效的[12].

在 BLSTM 的 Softmax 输出层之后加入 CRF 层,进行句子级的序列标注并输出. CRF 层引入转移矩阵 $A$ 作为 CRF 层的参数,其中 $A_{ij}$ 表示从第 $i$ 个标签转移到第 $j$ 个标签的概率. 本文采用最大似然估计作为代价函数,采用维特比算法解码,那么模型对于句子 $X$ 的标注序列 $Y=(y_1,y_2,\ldots,y_n)$ 的概率为

$$P(Y|X)=\frac{\exp(\text{score}(x,y))}{\sum_{y'}\exp(\text{score}(x,y'))} \tag{1}$$

$$\text{score}(X,Y)=\sum_{i=1}^{n}(P_{i,y_i}+A_{y_i,y_{i+1}}) \tag{2}$$

3.2.3 标注结果校正 经过 BLSTM-CRF 模型识别后还是会有一部分实体未被模型标注出,本文人工收集国家信息安全漏洞库等漏洞数据库以及 freebuf 等安全论坛中的安全漏洞领域相关实体,人工建立安全漏洞领域,包括漏洞名、攻击方法、漏洞类型以及应用环境这四个实体类型的命名实体词典,共 323 个词条. 安全漏洞领域命名实体词典特征词类型以及其示例如表 2 所示.

**表 2 安全漏洞领域术语特征词典示例**

Tab. 2 Security vulnerability domain term feature dictionary example

| 实体词类型 | 实例 |
|---|---|
| 漏洞名 | 心脏滴血(HeartBleed)、永恒之蓝(EternalBlue)、…… |
| 攻击方法 | SQL 注入、XSS、…… |
| 漏洞类型 | 权限提升类型、越权访问类型、…… |
| 应用环境 | 谷歌浏览器(Chrome),微软、…… |

词典中每一个词条代表着一个命名实体,依据该词典可对模型标注的结果进行校正,校正处理流程如图 3 所示,步骤如下.

---

**输入:**当前词典 $D$,标注语句 $X$,词典中最长词条的长度 $n$

**输出:**标注实体

(1) 取标注语句前 $n$ 个字符构成的字符串 $S_n$

(2) 若 $S_n$ 长度等于 1,删除 $X$ 中最左侧字符,回到步骤 1),否则进入步骤 3);

(3) 遍历词典 $D$ 中的每一个实体 $i$,判断实体 $i$ 是否与字符串 $S_n$ 相等;

(4) 若 $i=S_n$,则 $S_n$ 是 $i$ 所表示的实体,从标注语句 $X$ 中从左侧删除字符串 $S_n$,并标注 $S_n$ 为 $i$ 的所属类型,回到步骤 1);

(5) 若 $i\neq S_n$,删除 $X$ 最左侧字符,回到步骤 2).

---

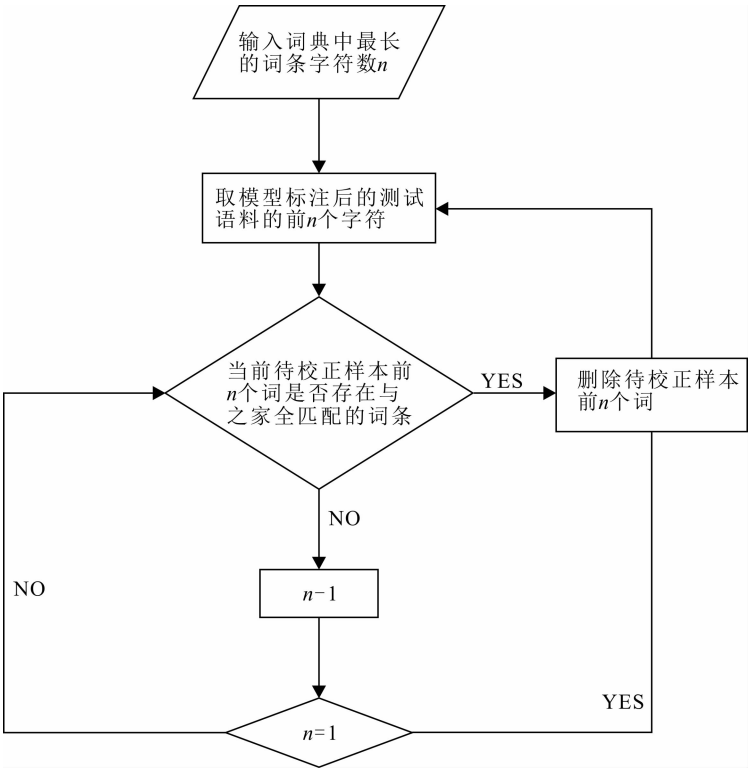


图 3　基于词典的校正流程

Fig. 3　Dictionary-based calibration process

## 4　实验结果及分析

### 4.1　实验语料

目前因缺乏权威的中文安全漏洞领域语料，因此采用人工收集的方式构建漏洞领域文本库. 实验样本为从安全论坛 Freebuf 上收集的漏洞模块文章. 文章均是与漏洞相关的非结构化文本，共计 1011 篇，人工标注出 1481 个安全漏洞实体，实验语料数据如表 3 所示，标注样例如表 4 所示.

**表 3　数据分布表**

Tab. 3　Data distribution table

| 实体类型 | 实体个数(个) |
|---|---|
| 漏洞名 | 56 |
| 漏洞类型 | 410 |
| 应用环境 | 548 |
| 攻击方法 | 467 |

**表 4　安全漏洞领域文本标注样例**

Tab. 4　Security vulnerability domain text annotation sample

| BadKernel | 漏 | 洞 | , | 编 | 号 | -CNNVD |
|---|---|---|---|---|---|---|
| *B_name* | *I_name* | *E_name* | *O* | *O* | *O* | *E_id* |
| — | 201608 | — | 414 | , | 存 | 在 |
| *I_id* | *I_id* | *I_id* | *E_id* | *O* | *O* | *O* |
| 于 | Chrome | 中 | , | 攻 | 击 | 者 |
| *O* | *B_env* | *O* | *O* | *O* | *O* | *O* |
| 可 | 利 | 用 | 该 | 漏 | 洞 | 对 |
| *O* | *O* | *O* | *O* | *O* | *O* | *O* |
| 微 | 信 | 进 | 行 | 远 | 程 | 攻 |
| *B_env* | *E_env* | *O* | *O* | *B_method* | *I_method* | *I_method* |
| 击 | | | | | | |
| *E_method* | | | | | | |

### 4.2 实验评价标准

本文采用常用的命名实体识别指标,如式(3)~(5)所示,来衡量实验结果:准确率 $P$,召回率 $R$ 和 $F_1$ 值.

$$P=\frac{\text{识别出的正确实体个数}}{\text{识别出的所有实体个数}}\times 100\% \tag{3}$$

$$R=\frac{\text{识别出的正确实体个数}}{\text{所有标注的实体个数}}\times 100\% \tag{4}$$

$$F_1=\frac{2PR}{P+R}\times 100\% \tag{5}$$

### 4.3 实验设置

为检验本文模型对于安全漏洞领域术语识别效果,本文采用 5 折交叉验证设计了以下对比实验来进行分析:将数据均分为 5 组,轮流选择 1 组作为测试集,其余 4 组作为训练集,对传统的 CRF 模型、BLSTM 模型、BLSTM-CRF 模型和本文提出的 BLSTM-CRF+校正模型这四个模型进行试验,共得到 5 次结果,将其均值作为模型性能指标,校验各个模型的性能.

### 4.4 实验结果

经过上述实验后,实验结果如表 5 所示. 方法 1 仅仅使用了 CRF 模型;方法 2 仅仅使用了 BLSTM 模型,它们在非结构化安全漏洞文本中识别的 F 值分别为 69.49%, 78.91%. 对比方法 1 和方法 2 可以看出 CRF 模型和 BLSTM 模型在一定程度上可以识别出特定领域的实体,但整体效果不佳,而且单用 BLSTM 模型比单用 CRF 模型效果更好,这是因为 CRF 模型要给予大量手工标注语料,但 BLSTM 可以结合上下文的深度学习,可以在少量标注语料的情况下取得较好的正确率和召回率.

方法 3 采用 BLSTM 和 CRF 相结合的模型,$F$ 值达到了 81.07%,通过方法 3 和方法 1、2 的对比可以看出,本文设计的 BLSTM-CRF 模型相比 CRF 模型以及 BLSTM 模型能够有效地提高安全漏洞领域术语抽取的效果,比仅使用 CRF 模型 $F$ 值提高了 11.58%. 比仅使用 BLSTM 模型 $F$ 值提高了 2.16%. 这是因为 BLSTM -CRF 模型在考虑上下文信息的同时由于 CRF 特征模板的合理制定约束了句子前后的标签序列,不会出现类似于"B_name"后面跟"I_env"这样的非法序列,所以该模型能够取得不错的效果.

方法 4 在 BLSTM 和 CRF 模型结合的基础上加上了校正,$F$ 值对比 BLSTM-CRF 模型提高了 4.23%. 引入词典和规则可以准确识别出更多的安全漏洞领域术语,明显提升正确率和召回率. 这是因为基于词典的校正模型是基于对句子结构的深入分析,并且考虑漏洞文本的表达习惯而提出的. 综上所述,本文设计的 BLSTM-CRF+校正模型可以取得比一般深度学习模型更好的实验效果.

**表 5 实验结果**

Tab. 5 Experimental result

| 方法标号 | 模型名称 | 实体总数 | 识别个数 | 正确个数 | 正确率 P/% | 召回率 R/% | F 值/% |
|---|---|---|---|---|---|---|---|
| 1 | CRF 模型 | 422 | 381 | 279 | 73.23 | 66.11 | 69.49 |
| 2 | BLSTM 模型 | 422 | 389 | 320 | 82.26 | 75.83 | 78.91 |
| 3 | BLSTM−CRF 模型 | 422 | 397 | 332 | 83.63 | 78.67 | 81.07 |
| 4 | BLSTM−CRF 模型+校正 | 422 | 401 | 351 | 87.53 | 83.18 | 85.30 |

## 5 结 论

本文对非结构化安全漏洞文本中的命名实体进行定义和研究,针对安全漏洞实体具有数量、种类、表述方式多,存在缩写、嵌套、大小写混合、中英文混杂等特点提出了一种面向安全漏洞领域的 BLSTM 和 CRF 相结合的命名实体识别方法,该方法解决了特定领域严重依赖人工特征和领域知识的问题. 在经过 BLSTM-CRF 模型标注的基础之上,文本深入挖掘领域术语的构成和表达特征,进一步制定了基于词典的校正模型,提高了识别的效果. 经过对比实验表明,本文模型具有较好的准确性. 接下来将在继续增加语料的基础上对方法继续优化,并进一步找出命名实体之间的关系,并创建能被自动安全系统使用的完备的安全漏洞通用库是下一步研究的重点.

**参考文献:**

[1] 冯蕴天, 张宏军, 郝文宁. 面向军事文本的命名实体识别 [J]. 计算机科学, 2015, 42: 15.

[2] 郭剑毅, 薛征山, 余正涛, 等. 基于层叠条件随机场的旅游领域命名实体识别 [J]. 中文信息学报,

2009，23：47.

[3] 杨可心，桑永胜. 基于 BP 神经网络的 DDoS 攻击检测研究 [J]. 四川大学学报：自然科学版，2017，54：71.

[4] Huang Z，Xu W，Yu K. Bidirectional LSTM-CRF models for sequence tagging [J]. Comput Sci，2015，2015：1.

[5] Li F，Zhang M，Tian B，*et al*. Recognizing irregular entities in biomedical text via deep neural networks [J]. Pattern Recogn Lett，2018，2018：105.

[6] 马建红，张亚梅，姚爽，等. 基于 BLSTM_Attention_CRF 模型的新能源汽车领域术语抽取 [J/OL]. 计算机应用研究，(2018-03-09) [ 2018-08-18]. http://www. arocmag. com/article/02-2019-05-013. html.

[7] Chiu J P C，Nichols E. Named entity recognition with bidirectional LSTM-CNNs [J]. Trans Assoc Comput Linguist，2016，4：357.

[8] Mulwad V，Li W，Joshi A，*et al*. Extracting information about security vulnerabilities from web text [C]//Proceedings of the 2011 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology-Volume 03. [s. l.]：IEEE Computer Society，2011.

[9] Weerawardhana S，Mukherjee S，Ray I，*et al*. Automated extraction of vulnerability information for home computer security [C]//International Symposium on Foundations and Practice of Security. Berlin：Springer International Publishing，2014.

[10] Joshi A，Lal R，Finin T，*et al*. Extracting cybersecurity related linked data from text [C]//Proceedings of the 2013 IEEE Seventh International Conference on Semantic Computing (ICSC). [s. l.]：IEEE，2013.

[11] Ma X，Hovy E. End-to-end sequence labeling via bidirectional lstm-cnns-crf [J/OL]. (2016-03-22) [2018-08-21]. https://arxiv. org/abs/1603. 01354.

[12] 周顺先，蒋励，林霜巧，等. 基于 Word2vector 的文本特征化表示方法 [J]. 重庆邮电大学学报：自然科学版，2018，30：272.

[13] Pennington J，Socher R，Manning C. Glove：Global vectors for word representation [C/OL]//Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP). (2014-08-05). [2018-08-21]. https://nlp. stanford. edu/projects/glove/

[14] Graves A，Schmid＋huber J. Framewise phoneme classification with bidirectional LSTM and other neural network architectures [J]. Neural Networks，2005，18：602.

[15] Lafferty J，McCallum A，Pereira F. Conditional random fields：probabilistic models for segmenting and labeling sequence data [C]//Proceedings of International Conference on Machine Learning，Massachusetts. [s. l.]：[s. n.] 2001.

[16] 陈波. 基于循环结构的卷积神经网络文本分类方法 [J]. 重庆邮电大学学报：自然科学版，2018，30：705.

**引用本文格式：**

中　文：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，2019，56：469.

英　文：Zhang R B，Liu J Y，He X. Named entity recognition for vulnerabilities based on BLSTM-CRF model [J]. J Sichuan Univ：Nat Sci Ed，2019，56：469.