DOI: 10.3969/j.issn.1671-0673.2021.01.008

基于 BBNN 的网络攻击文本自动化分类方法

欧昀佳,周天阳,朱俊虎,臧艺超

(信息工程大学,河南 郑州 450001)

摘要: 基于描述文本的网络攻击自动化分类是实现 APT 攻击知识智能抽取的重要基础。针对网络攻击文本专业词汇多、难识别,语义上下文依赖强、难判断等问题提出一种基于上下文语义分析的文本词句特征自动抽取方法,通过构建 BERT 与 BiLSTM 的混合神经网络模型 BBNN(BERT and BiLSTM Neural Network),计算得到网络攻击文本的初步分类结果,再利用方差过滤器对分类结果进行自动筛选。在 CAPEC(Common Attack Pattern Enumeration and Classification)攻击知识库上的实验结果显示,该方法的准确率达到了 79.17%,相较于单一的 BERT 模型和 BiLSTM 模型的分类结果分别提高了 7.29%和 3.00%,实现了更好的网络攻击文本自动化分类。

关键词: 神经网络; APT 网络攻击; 文本分类

中图分类号: TP393 **文献标识码:** A **文章编号:** 1671-0673(2021)01-0044-07

Document Classification in Cyberattack Text Based on BBNN Model

OU Yunjia, ZHOU Tianyang, ZHU Junhu, ZANG Yichao

(Information Engineering University, Zhengzhou 450001, China)

Abstract: The document classification in cyberattack text is fundamental to automatic knowledge extraction from APT attack information. In this paper, an automatic method based on context analysis is proposed to tackle the problems rooted in cyberattack, such as having too many terminologies, being hard to distinguish and classify, over-relying on context, etc., by extracting text features in words level and sentences level respectively. This method, BBNN (BERT and BiLSTM Neural Network) model, synthesized BERT and BiLSTM Neural Network, can compute the preliminary classification results of cyberattack, and automatically filter the classification results of the text via variance. The experiment results from the attack knowledge base of CAPEC (Common Attack Pattern Enumeration and Classification) suggests this method can reach a 79.17% accuracy, which is increased by 7.29% and 3.00% compared to singular BERT or BiLSTM models, and thus achieve a better automatic classification of cyberattack.

Key words: neural network; cyber attack; document classification

随着信息技术的飞速发展,网络攻击手段层出不穷,在互联网上的黑客论坛、技术博客以及安全厂商发布的研究报告中含有大量关于高级可持续威胁(Advanced Persistent Threat,APT)攻击的研究成果。这些原始的攻击技术描述信息和攻击案例的流程分析文本是安全研究人员构建网络攻击安全知识体系,深化攻击研究的重要数据来源。

从复杂多样的攻击描述文本中抽取 APT 攻击

收稿日期: 2020-08-31; 修回日期: 2020-09-08

基金项目: 国家自然科学基金资助项目(61502528)

作者简介: 欧昀佳(1989-),男,硕士生,主要研究方向为网络安全。

(C)1994-2021 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

知识,构建完整的网络攻击知识图谱,可以更好地确定安全威胁程度,推测攻击者意图,拟定有针对性地防御措施。通过网络攻击文本分类,将攻击知识映射到 ATT&CK (Adversarial Tactics, Techniques, and Common Knowledge) 、CAPEC 等成熟攻击知识框架,是认知网络攻击,抽取攻击知识的重要基础。传统的人工阅读专家判定的分类方法存在人员素质要求严苛、效率低下、成本过高等问题。如何自动、快速、准确地识别大量网络安全文本中攻击手段的所属类别是当前网络安全研究的新热点。

开源的网络攻击描述文本具有来源多样、表达各异、语法不规范等特点。例如,技术性短文本偏重于对某一具体的攻击技术实现进行描述,而大多 APT 报告则通过长期跟踪着重对捕获案例的整个攻击流程进行描述。原始网络攻击文本的这些特点极大地增加了自动化识别与分类的难度。针对该问题,本文构建了一种基于新型混合神经网络模型的网络攻击文本分类方法,通过分析网络攻击文本的上下文语义,自动抽取文本词句特征,并以此作为自动化攻击判定与分类的计算依据。该方法利用模型输出的概率分布方差评估预测项的确定程度,对特定的低概率分类项进行随机选取预测,可以有效提高网络攻击文本分类的准确率。

1　相关研究

文本分类是自然语言处理的一大热门研究领域。早期的研究大多基于规则和统计的方法,但该类方法存在效率低下、人工成本高、泛化能力弱、特征提取不充分等缺点。神经网络能够弥补上述缺点更好地解决分类问题。

文献 [1] 首先提出基于卷积神经网络的文本特征提取,并用于文本倾向性分类任务上。文献 [2] 在此基础上提出 VDCNN 模型,该方法采用深度卷积网络对文本局部特征进行提取,提高文本分类效果。但是卷积神经网络在考虑文本的上下文关系时有很大局限性,其固定的卷积核不能建模更长的序列信息,使得文本分类效果很难进一步提升。

RNN(Recurrent Neural Network) 模型不仅可以处理不同长度的文本,而且能够学习输入序列的上下文语义关系。双向 RNN [3] 可以让输出获取当前输出之前以及之后的时间步信息。但由于 RNN 是通过时间序列进行输入,随着输入的增多,RNN 对很久以前信息的感知能力下降,将会产生长期依赖和梯度消失问题 [4] 。在 RNN 基础之上改进而来的门控制单元(GRU) [5] 和长短时序记忆模型(Long Short Term Memory,LSTM) [6] 利用门机制可以解决 RNN 的长期依赖和梯度消失问题。文献 [7] 利用 RNN 设计了 3 种用于多任务文本分类问题的模型。文献 [8] 提出了一种基于 word2vec 与 LSTM 模型的健康文本分类方法。

为进一步提高文本分类的准确率,文献 [9] 首先提出了一种用于文档分类的分层注意力网络模型。该模型在单词和句子级别应用了两个不同的注意力机制,使得模型在构建文档时能够给予重要内容更大权重,同时也可以缓解 RNN 在捕捉文档序列信息时产生的梯度消失问题。文献 [10] 提出一种基于注意力机制的深度学习态势信息推荐模型,该模型能学习指挥员与态势信息之间的潜在关系,为用户推荐其关心的重要信息。在注意力机制基础上,Google 团队提出了 Transformer 模型 [11] ,它摒弃了常用的 CNN 或者 RNN 模型,采用 Encoder-Decoder 架构,对文本进行加密、解密处理。Transformer 通过利用大量原始的语料库训练,从而得到一个泛化能力很强的模型,只需微调参数进行训练,就可以将模型应用到特定的文本分类任务中 [12] 。

词嵌入(word embedding) 是单词的一种数值化表示方法,是机器学习方法用于文本分类的基础。通过此技术可以将文本中的词语转化为在向量空间上的分布式表示,并大大降低词语向量化后的维度。

预训练模型是在具体分类任务训练前使用大量文本训练词嵌入模型,使模型通过学习一般语义规律,获得单词的特定词嵌入表达。文献 [13] 在 2013 年提出了 word2vec 词嵌入训练模型,包括 CBOW 模型和 Skip-Gram 模型 [14] ,这两种模型可以通过学习得到高质量的词语分布式表达,能训练大量的语料且捕捉文本之间的相似性。但其缺点是只考虑到了文本的局部信息,未考虑整体信息。文献 [15] 提出 GloVe 模型,利用共现矩阵,同时考虑局部信息和整体信息。上述提及的词嵌入技术均属于静态词嵌入,即训练后词向量固定不变。但在实际应用中,同一词在不同语境下的语义是不一样的,其表达需要根据不同语境进行变化。针对该问题,文献 [16] 提出了 BERT 预训练模型,使用多层双向 Transformer 编码器对海量语料进行训练,结合所有层的上下文信息进行提取,实现了文本的

(C)1994-2021 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

深度双向表示。

网络攻击的描述文本相对特殊，大多采用人工分析，专家定义的方式。该类文本具有长度短、专业词汇多、文字特征稀疏、上下文依赖性强等特点。神经网络的优势在于对文本的特征提取上，但短文本的文字特征稀疏，单一的神经网络难以进行有效处理。为此，可以在此基础上引入迁移学习的思想，即利用经过预训练的深度神经网络对文本进行编码，经过这样编码的文本蕴含了在大量文本上提取的一般特征，然后再针对具体分类任务将这些特征用于神经网络的训练。由于具体分类任务的关注点不同，导致文本信息对分类预测的重要程度不同。注意力机制可以使分类模型关注到重点信息，从而有效提升模型性能。本文将神经网络、预训练和注意力机制进行有机融合，开展网络攻击分类方法研究。

2 网络攻击文本分类模型

网络攻击文本分类任务不同于一般的新闻文本分类任务，其存在描述范围有限、各类攻击关联性强、区分度不大且专业性词汇多的特点。例如:

句1: Client side injection-induced buffer overflow, this type of attack exploits a buffer overflow vulnerability in targeted client software through injection of malicious content from a custom-built hostile service.

句2: Log injection-tampering-forging, this attack targets the log files of the target host.

句1、句2都有对注入injection的描述，但两者描述的攻击类别分别属于Manipulate Data Structures类和Manipulate System Resources类。句子中的exploits属于专业词汇，与平时所指代的“开发、开拓”意义不同，这里指的是“对软件的脆弱性利用”。这导致了普通的静态预训练模型在该任务上性能不佳。

针对以上问题，本文提出从词语级和句子级分别对攻击文本进行特征提取。利用BERT模型的动态编码机制对输入攻击文本进行动态编码表示，使计算出的句子向量能包含具体攻击描述文本中的语义信息，解决文本中存在的语义偏移和专业词汇识别难等问题。同时，采用BiLSTM模型提取攻击文本中词与词的前后关联关系特征，学习的词向量利用注意力机制强化特殊词汇对于攻击文本分类的权重比值，解决文本中存在的描述类似导致区分度小等问题。然后，结合以上两个模型学习到的特征，分别计算出攻击文本分类的预估值。继而对获取的两类预估值进行分析判定，求出最后的分类预测值。如图1所示，BBNN攻击文本分类模型分为4层，输入层、模型计算层、预值分析层、输出层。下面将依次对模型的各层结构进行详细介绍。

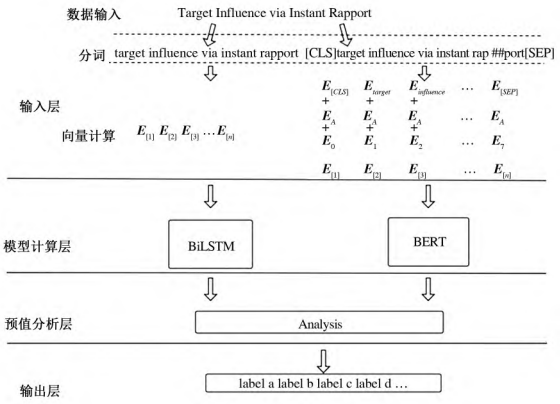

图1 BBNN攻击文本分类模型

输入层的数据处理要经过数据输入、分词、向量计算3个阶段。对于要输入给BiLSTM模块的数据只需经过简单的分词和词嵌入处理。而对于要输入给BERT模块的数据则需要由词向量、分段向量和位置向量组成。BERT模块的分词阶段主要包含两个部分: ①将输入的文本进行wordpiece处理，wordpiece处理是将词划分为更小的单元，如图1中rapport一词就被划分为rap和##port，这样做的目的是压缩了词典大小，并且增大了可表示词数量; ②在句子首尾分别嵌入【CLS】和【SEP】两个特殊token。对于BERT模块来说，由于文本分类任务的输入对象是一个单句，所以分段向量 E_A 全部以0进行填充。位置向量(E_1 至 E_7) 则通过正弦余弦编码来获取，表征该句子序列的位置信息。将上述3部分向量相加得到BERT模块的输入向量。

模型计算层包含BiLSTM计算模块和BERT计算模块。这两个模块需要进行单独训练，再加入到整体的BBNN攻击文本分类模型结构中，如图2所示。

从图2可以看到，BiLSTM模块包含一个前向LSTM神经网络和一个后向LSTM神经网络，分别学习输入序列中各个词的前后信息。该模块接收输入层编码的词嵌入向量 *Embedding* 作为输入。E 代表经过词嵌入处理的词向量，t 代表第 t 个时间步。$\overrightarrow{h_t}$ 是第 t 个时间步的前向隐藏状态编码，它由上一个时间步的隐藏状态编码 $\overrightarrow{h_{t-1}}$ 和当前输

(C)1994-2021 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

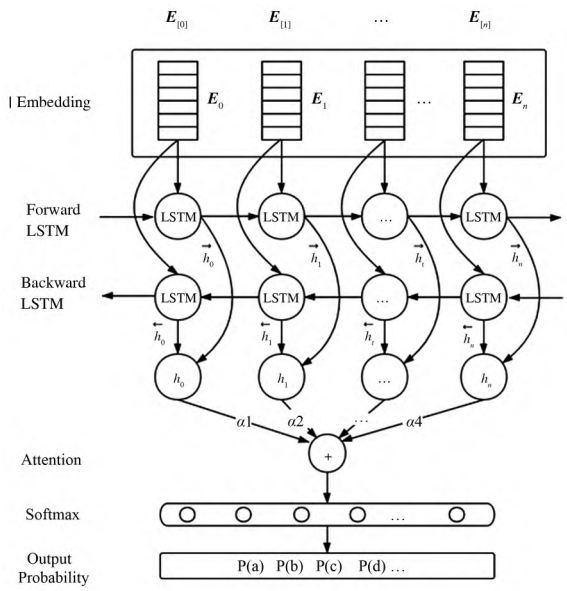

图 2　BiLSTM 模块结构

入向量 E_t 计算得出。$\overleftarrow{h_t}$ 是第 t 个时间步的后向隐藏状态编码,由下一个时间步隐藏状态编码 $\overleftarrow{h_{t-1}}$ 和当前输入向量 E_t 计算得出。连接前向后向隐藏状态向量得到含有上下文信息的 t 时刻隐藏状态向量 $h=[\overrightarrow{h_t},\overleftarrow{h_t}]$。综合各个时间步隐藏状态向量得到文本词向量矩阵 $H=\{h_0,h_1\cdots,h_n\}$ 作为输出,传递给实现注意力机制的 Attention 网络。Attention 网络接收文本向量矩阵 H 作为输入,为每个输入值分配不同的权重 ω 和偏置项 b,计算出文本中每个单词的权重 μ,得到权重矩阵 U。W 是随机初始化生成的矩阵:

$$u_t=\tanh(h_t*\omega_t+b) \tag{1}$$

$$U=\tanh(H*W+b) \tag{2}$$

V 是随机初始化生成的一维矩阵,使用 *softmax* 函数计算出每个时间步对当前时刻的权值向量 α_t,得到得分矩阵 A:

$$\alpha_t=\frac{e^{u_t}*V}{\sum_t e^{u_t}*V} \tag{3}$$

$$A=\mathrm{Softmax}(UV) \tag{4}$$

将每个得分权重和对应的隐藏状态编码进行加权求和,得到 Attention 网络的输出 s:

$$s=\sum_{t=0}^{n}\alpha_t*h_t \tag{5}$$

最后将得到的向量 s 进行 softmax 归一化处理,计算各个分类的概率情况作为 BiLSTM 模块的输出。

BERT 模块接收输入层的输出向量作为输入,其模型结构如图 3 所示。双向 Transformer 模型能利用内部的 self_attetnion 机制计算出该句其他词和当前编码词的语义关联程度,从而增强上下文语义信息的提取。词嵌入向量通过双向 Transformer 模型结构进行编码学习,其第一位词嵌入向量(即【CLS】对应词向量编码)包含该句所有信息。抽取第一位词嵌入向量作为该句子的向量表示,利用 tanh 函数激活后做 softmax 处理,获得该句子的各类概率分布作为 BERT 模块的输出。

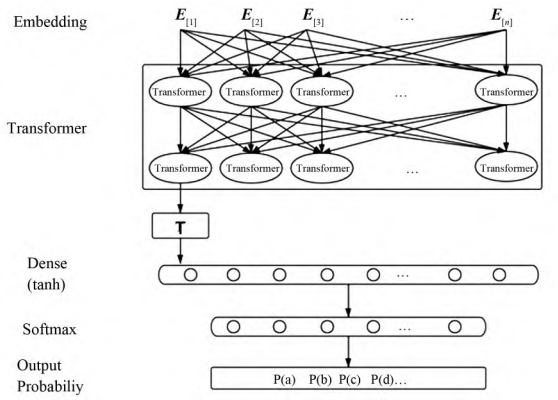

图 3　BERT 模块结构

分析层的输入为 BiLSTM 模块和 BERT 模块对文本的预测概率分布。分析层主要有两个阶段,①选取两个模型确定性更大的预测分布作为最终预测分布,②在最终预测分布上筛选不确定性更大的预测项,对该项分布的最大二值进行随机抽取确定最终输出标签。

第 1 个阶段中,两个概率分布的离散程度反应了该模型对预测分类的准确度,即离散程度越大表明该模型对此项分类判定的确定性越大。反之,概率分布的离散程度越小,代表各个分类的概率值相近,判断相对模糊。因此,通过对比两个模型得出的概率分布可以获知某个模型对该段文本的预测确定性更高。取确定性更高的分布作为 BBNN 攻击文本分类模型的最终概率分布输出。如算法 1 所示。

算法 1　基于概率分布的预测确定性分析算法

输入: BiLSTM 和 BERT 模块预测的各类别的概率分布 P_{bert}、P_{lstm}

输出: 最终判定的概率分布 P_{out}

```
v_bert = GetVariance(P_bert)
v_lstm = GetVariance(P_lstm)
if v_bert > v_lstm
    P_out = P_bert
else
    P_out = P_lstm
return P_out
```

(C)1994-2021 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

步骤1、步骤2分别计算BERT模块、BiLSTM模块预测概率分布的方差;步骤3~步骤6通过对比方差大小选定确定性更大的概率分布;步骤7输出最终概率分布。

第2个阶段中,认为最终输出概率分布中低概率值、低方差区域的预测项不确定性最大,这些分布的各项预测值相近。其中最大预测值项并不是真实分类项的概率最高。因此,可扩大选择范围,对相对较高的预测值进行随机选取以确定最后预测项。这里,采用最优二值随机选择的办法来进行。首先通过设置阈值选定随机区域,然后对方差和概率值属于该区域的预测项进行最优二项值随机选取。如算法2所示。

算法2 基于概率分布方差及概率值的低确定性随机选择算法

输入: 预测概率分布 P_{in},概率项阈值 α、方差项阈值 β

输出: 最终预测分类编号

$v=\mathrm{GetVariance}(P_{in})$

$P_{max1}=\mathrm{Max}(P_{in}(x_i))$

if $P_{max1}<\alpha$ and $v<\beta$

$P_{max2}=\mathrm{Max}(P_{in}(X=x_i)-P_{max1})$

$P_{final}=\mathrm{Random}(P_{max1},P_{max2})$

else

$P_{final}=P_{max1}$

$x_{out}=\mathrm{Classify}(P_{out})$

return x_{out}

步骤1获取输入分布的方差值 v,步骤2获取输入分布的最大概率值 P_{max1},步骤3判断最大概率值和方差是否均分别小于阈值 α、β,步骤4、步骤5表示若满足步骤3的条件则求出第二大概率值并在最大概率值和第二大概率值间做随机选择赋值给最终选择概率 P_{final},步骤6、步骤7表示若不符合步骤3条件则将最大概率值赋值给最终选择概率 P_{final},步骤8求出输入分布中概率为 P_{final} 的分类标签 x_{out},步骤9输出标签。

3 实验

3.1 实验设置

本文采用通用攻击模型枚举分类项目,CAPEC中的攻击分类描述文本作为实验数据集,该数据集包括了攻击定义描述、攻击实例描述、攻击条件准备等信息。

目前,从攻击机制的角度,CAPEC将所有攻击共分为9个基础类型[17](截止2019年9月30日)。本次实验数据通过对数据集的攻击名称、攻击定义描述、实例描述数据进行提取并标注,获得了1 452条样本信息。样本分布信息如图4所示,CAPEC的9类数据中Collect and Analyze Information类别的样本数最多,这是因为网络攻击过程中的多个步骤均涉及信息的搜集与分析,所占比例最大。Employ Probabilitstic Techniques和Manipulate Timing and State类的样本数不足50所占比例最少,表明该两类攻击使用领域较小。

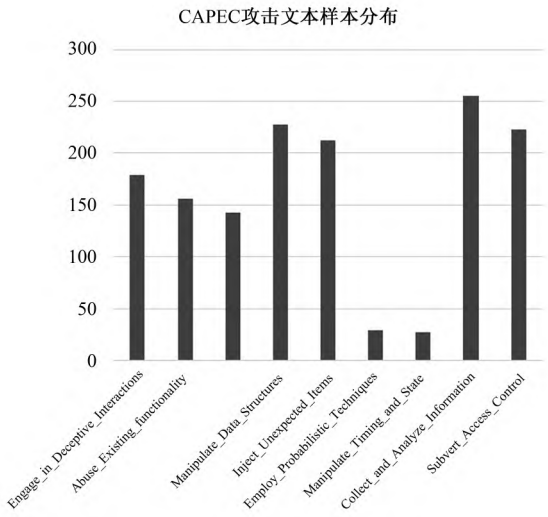

图4 样本分布

实验中将样本数据做随机抽取,选取其中70%的样本数据作为训练集,10%作为验证集,20%作为测试集。实验环境如表1所示。为验证BBNN模型的有效性,实验使用以下几个模型进行对比:

①通过GloVe预训练模型获得词嵌入向量输入微调训练后的BiLSTM+Attention模型;

②经数据微调训练后获得的BERT模型;

③无随机选择的BBNN模型;

④选定阈值随机选择的BBNN模型。

表1 实验环境

项目	版本
操作系统	Linux Mint 19.3 Cinnamon 4.4.5
处理器	2.6 GHz 18核 Intel Core i9-7980XE
内存	125.6 GiB
显卡	NVIDIA Corporation GV100 [TITAN V]
Python	3.6
Tensorflow	Gpu-2.2.0
Numpy	1.18.1
Pandas	1.0.3
Keras	2.3.1
Matplotlib	3.1.3

实验中GloVe预训练模型使用预训练好的

(C)1994-2021 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

glove. 6B. 300d,即编码后的词嵌入向量为 300 维,LSTM 隐藏单元数设置为 64,Attention 网络隐藏单元为 16。BERT 预训练模型结构使用 Google 发布的 uncased_L-12_H-768_A-12,经过该 BERT 模型结构的词向量维度为 768。为避免过拟合 BiLSTM 和 BERT 模型输出后的 Drpout 设置为 0. 5,模型最大句子长度为 512。BBNN 模型的随机选择区域阈值,概率项设为 0. 75 和方差项设为 0. 06,即概率值小于 0. 75 且方差小于 0. 06 的预测项进行最优二项选择。由于随机选择存在随机性,BBNN 模型的随机预测结果评估值取 100 次预测的平均评估值作为最后的参考指标。

3. 2　实验结果与分析

几种模型在测试集上的预测分布方差和预测概率值如图 5 所示,图 5(a) 为 BERT 模型预测结果,图 5(b) 为 BiLSTM 模型预测结果,图 5(c) 为 BBNN 模型低命中区域无随机选取的预测结果,图 5(d) 为 BBNN 低命中区域随机选取的预测结果。图 5 中,横轴代表预测的概率值,纵轴代表预测分布的方差。点标代表预测正确的文本,叉标表示预测错误的文本。对比图 5(a) 和图 5(b) 可以看出,大多数预测错误项的方差和概率值小于预测正确项的方差和概率值,预测正确的值集中在方差与预测概率值较高的区域。并且,方差越小概率值越小。图 5(c) 对比图 5(a) 和图 5(b) 可以发现经过 BBNN 模型优选后的输出结果在方差和概率值低的区域预测错误项有明显下降。对比图 5(c) 和图 5(d) ,设置阈值并选取低概率值、低方差区域做随机猜选能够提高低命中率区域的命中率。

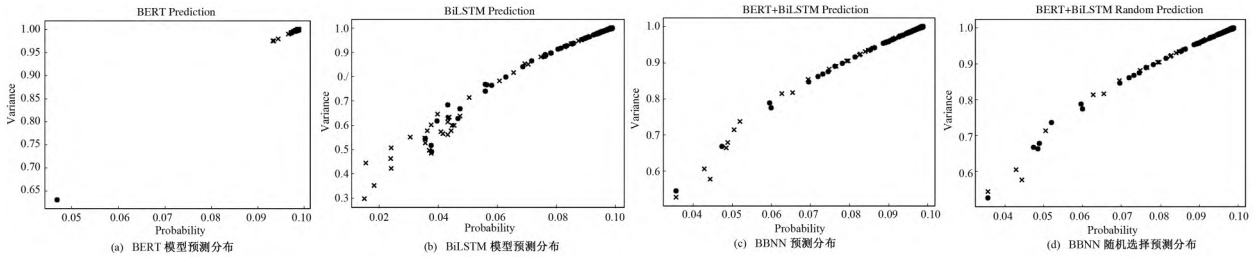

图 5　模型预测结果分布

模型在测试集上的预测表现如表 2 所示,BBNN 模型在测试集上的各项指标优于其他模型。对比 BERT 和 BiLSTM + Attention 模型可以发现 BBNN 模型的优选机制对模型的准确率有一定提升,分别提高了 7. 29 %和 3. 00%。同时,$F1$ 值分别提升了 7. 62%、1. 73%,精确率分别提升了 8. 29%、1. 72 %,召回率分别提升了 6. 22%、1. 90%。

表 2　测试集预测评估结果对比表

模型	Accuracy(%)	$F1$(%)	Precision(%)	Recall(%)
BERT	71. 88	72. 05	72. 70	72. 94
BiLSTM	76. 17	77. 94	79. 27	77. 26
BBNN	78. 52	78. 90	79. 92	78. 69
BBNN(Random)	79. 17	79. 67	80. 99	79. 16

实验中使用 BBNN 模型进行了 100 次随机选择预测,结果如图 6 所示。图中,横轴代表预测次数,纵轴代表百分比。4 项指标中的最好预测结果为准确率 80. 46%、精确率 82. 21%、$F1$ 值 80. 63%、回归率 80. 08%; 最差预测结果为准确率 77. 73%、精确率 79. 45%、$F1$ 值 78. 33%、回归率 78. 14%。最差预测结果的各项指标相比单一的模型预测结果要好。随机选择机制使得模型的各项指标在 3%以内上下浮动,存在一定的不稳定性。

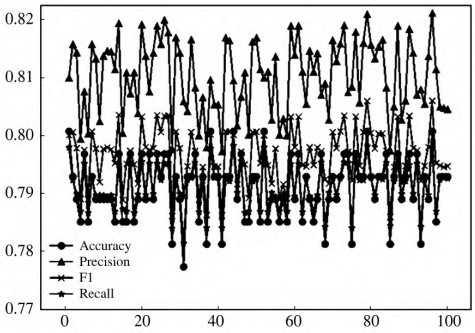

图 6　100 次随机预测评估结果

4　结束语

网络中存在大量攻击描述文本,对该类文本进行快速分类有助于研究人员从原始数据中迅速提炼知识,形成有体系的知识架构。本文通过对网络攻击描述文本进行分析,提出一种 BBNN 混合优选型模型,用于自动识别网络攻击文本中描述的攻击方式。与传统的机器学习模型不同,该模型能从词语级和句子级的层面对文本特征进行学习,并通过优选机制得到更确定的预测分类。实验表明,

(C)1994-2021 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

BBNN模型对攻击文本的识别效果最好，但是该模型更为复杂，训练和使用时相比其他模型耗时更长，占用资源更多，且预测结果存在一定的不稳定性。下一步工作中将进一步扩大数据集，提高样本覆盖率。同时有针对性地对模型结构进行简化和优化，提高攻击文本分类方法的效率和稳定性。

参考文献:

[1] KIM Y. Convolutional neural networks for sentence classification [C]//Empirical Methods in Natual Language Processing. Doha Ratar, 2014: 1746-1751.

[2] CONNEAU A, SCHWENK H, BARRAULT L, et al. Very deep convolutional networks for text classification [J]. Proceedings of the 15th Conference of the Europeam Chapter of the Associational Linguistics. Vacencia Spain, 2017: 193-207.

[3] 李洋，董红斌. 基于CNN和BiLSTM网络特征融合的文本情感分析[J]. 计算机应用，2018，38(11)：29-34.

[4] BENGIO Y, SIMARD P, FRASCONI P. Learning long-term dependencies with gradient descent is difficult [J]. IEEE Transactions on Neural Networks, 1994, 5(2): 157-166.

[5] CHO K, VAN MERRIENBOER B, GULCEHRE C, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation [C]//2014 Conference on Empirical Methods in Natural Language Processing. Doha Qatar, 2014: 1724-1734.

[6] HOCHREITER S, SCHMIDHUBER J. LSTM can solve hard long time lag problems [C]//Advances in Neural Information Processing Systems, 1997: 473-479.

[7] LIU P, QIU X, HUANG X. Recurrent neural network for text classification with multi-task learning [C]//International Joint Conference on Artifical Intelligence. New York, 2016: 2873-2879.

[8] 赵明，杜会芳，董翠翠，等. 基于word2vec和LSTM的饮食健康文本分类研究[J]. 农业机械学报，2017，48(10)：202-208.

[9] YANG Z, YANG D, DYER C, et al. Hierarchical attention networks for document classification [C]//Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics Human Language Technologies. San Diego, 2016: 1480-1489.

[10] 周春华，郭晓峰，沈建京，等. 基于注意力机制的深度学习态势信息推荐模型[J]. 信息工程大学学报，2019，20(5)：597-603.

[11] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [C]//Advances in Neural Information Processing Systems, 2017: 5998-6008.

[12] QIU X, SUN T, XU Y, et al. Pre-trained models for natural language processing: a survey [EB/OL]. [2020-08-24]. https: arXiv. org/abs/2003. 08271.

[13] MIKOLOV T, CHEN K, CORRADO G, et al. Efficient estimation of word representations in vector space [EB/OL]. [2013-09-07]. arXiv preprint, 2013 arXiv: 1301. 3781.

[14] MIKOLOV T, SUTSKEVER I, CHEN K, et al. Distributed representations of words and phrases and their compositionality [C]//Advances in neural information processing systems, 2013: 3111-3119.

[15] PENNINGTON J, SOCHER R, MANNING C D. Glove: Global vectors for word representation [C]//Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP). Doha Qatar, 2014: 1532-1543.

[16] DEVLIN J, CHANG M W, LEE K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding [C]//2019 Conference of the North Amerilan Chapter on the Association for Computational Linguistics Human Language Technologies. Minneapolis USA, 2019: 4171-4186.

[17] CAPEC TEAM. Schema Documentation [EB/OL]. (2019-09-30). [2020-06-01]. https://capec. mitre. org/documents/schema/index. html.

(编辑: 高明霞)

(C)1994-2021 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net