计算机科学
COMPUTER SCIENCE

基于开放集的入侵检测方法研究

王春东, 张嘉凯

引用本文

王春东, 张嘉凯. 基于开放集的入侵检测方法研究[J]. 计算机科学, 2024, 51(11A): 231000033-6.

WANG Chundong, ZHANG Jiakai. Study on Open Set Based Intrusion Detection Method[J]. Computer Science, 2024, 51(11A): 231000033-6.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.231000033

基于开放集的入侵检测方法研究

王春东　张嘉凯
天津理工大学计算机科学与工程学院　天津 300384

摘　要　入侵检测是网络安全中的一项重要任务,旨在检测异常行为和潜在攻击。近几年,深度学习方法在入侵检测任务中取得了很大突破。但随着近几年互联网行业的迅猛发展,新型攻击类型不断增加,深度学习方法在测试中面对新型类别时,往往会以高置信度给出一个已知类别中的预测结果,导致无法识别未知攻击。基于此,提出一种基于不确定性建模的开放集识别方法,即将 MC-Dropout 应用于深度学习分类器中以捕获不确定性,从而获得高质量预测概率。该开放集合识别方法不仅能够对已知类别进行分类,同时还能够对未知类别进行判别。通过在 CICIDS2017 数据集上验证,所提出的方法能够实现对未知类别的检测,和其他现有方法相比具有一定的先进性,各项指标与基准模型对比均取得最好表现,能有效地应用于现实的网络环境。

关键词:入侵检测;开放集识别;深度学习;MC-Dropout
中图分类号　TP393

Study on Open Set Based Intrusion Detection Method

WANG Chundong and ZHANG Jiakai
School of Computer Science and Engineering,Tianjin University of Technology,Tianjin 300384,China

Abstract　Intrusion detection is an important task in network security,which aims to detect anomalous behaviors and potential attacks. In recent years,deep learning methods have made great breakthroughs in intrusion detection tasks. However,with the rapid development of the Internet industry in recent years,new types of attacks are increasing,and deep learning methods tend to give a prediction result in a known category with high confidence when faced with a new type of category in testing,resulting in the inability to recognize unknown attacks. Based on this,this paper proposes an open set identification method based on uncertainty modeling,i.e.,MC-Dropout is applied to deep learning classifiers to capture uncertainty and thus obtain high-quality prediction probabilities. This open set identification method is not only able to classify known categories,but also able to discriminate unknown categories. The proposed method is validated on the CICIDS2017 dataset,and is able to achieve the detection of unknown categories,and has a certain degree of sophistication compared with other existing methods,and achieves the best performance in all the metrics compared with the benchmark model,which can be effectively applied to the real-world network environment.

Keywords　Intrusion detection,Open set identification,Deep learning,MC-Dropout

1　引言

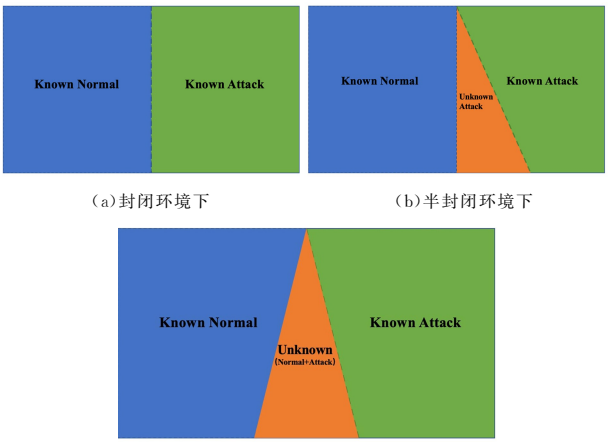
随着互联网技术的发展,网络空间的规模逐渐扩大,结构日趋复杂,在给人们生活带来极大便利的同时,网络安全问题也日趋严重。Skybox Security 在《2022 年漏洞与威胁趋势报告》中指出,2021 年新增零日漏洞数量创历史新高,达 166938 个。攻击者利用零日漏洞的速度和攻击能力进一步增强,这意味着网络空间安全面临着更多未知的网络威胁和恶意入侵行为。

恶意网络入侵行为的复杂性和多样性导致传统的基于规则的安全方法难以捕捉所有潜在的攻击方式。在该背景下,机器学习逐渐崭露头角,成为网络入侵检测与防御的有力工具。机器学习能够分析和识别大规模网络数据中的模式和异常,从而有效地检测潜在的入侵行为。然而,基于传统机器学习的入侵检测研究大部分基于闭集假设或半封闭假设。闭集假设是训练集和测试集中包含的类别相同。当新类别在测试集中出现时,深度学习模型会将该类别以较高置信度归类为已知标签中的一类。半封闭假设是在测试阶段通常会出现新的攻击,但是没有未曾见过的正常行为。现实的网络环境是多变的,互联网每天都会出现以前从未见过的应用程序、服务和恶意攻击,因此,只针对封闭环境和半封闭环境下的模型进行研究难以应对复杂的实际场景。

将入侵检测任务转化为开放集识别任务能够解决模型无法处理新型攻击的问题。开放集识别任务旨在对已知类别样本进行正确分类的同时,对出现的未知类别样本进行准确判别。它假定训练过程中模型获取到的知识是不完整的,因此测试样本既来自已知类别,也来自未知类别。这要求训练方案能根据已知类别的知识来拒绝这些未知类别的样本。图 1

基金项目:国家自然科学基金联合基金项目(U1536122);天津市科委重大专项(15ZXDSGX00030)
This work was supported by the Joint Funds of the National Natural Science Foundation of China(U1536122) and Tianjin Committee of Science and Technology Major Project,China(15ZXDSGX00030).
通信作者:王春东(michael3769@163.com)

直观地展示了3种假设下的样本分布情况。



(a)封闭环境下 (b)半封闭环境下

(c)开放集环境下

图1 数据集中样本分布假设

Fig. 1 Sample distribution assumptions in datasets

目前据我们所知,只有少数解决方案能够解决网络入侵检测的开放集问题。但它们仍然是基于半封闭的假设设计的,即在测试阶段会有新的攻击但没有新的正常行为,特别是在当前物联网设备爆炸性增长的时代,不同设备的底层基础设施变化总会导致良性数据分布发生变化而产生偏差,入侵检测系统检测到的许多异常很可能是误报。

为解决上述问题,本文提出一种新的面向未知类别的入侵检测方法。该方法将入侵检测任务转化为开放集识别任务,抛弃封闭假设和半封闭假设。因为在开放集识别任务的模型训练过程中未学习到新类别,模型对新类别的样本检测置信度具有高度不确定性,导致模型将新样本误判为已知类别,所以我们在分类器的权重层引入MC-Dropout(Monte Carlo Dropout)捕捉样本检测时的不确定性。带有不确定性的分类器能够捕捉到检测置信度的真实置信度,从而更好地区分已知类别与未知类别,也能更好适应数据的多样性和模型的复杂性。

本文的主要贡献总结如下:

1)将入侵检测任务转化为开放集识别任务。与传统的封闭式和半封闭式相比,在测试阶段额外考虑了潜在的新攻击和未见的正常行为。

2)提出了一种新的面向未知类别的入侵检测方法。该方法利用互信息进行特征选择,引入Focal Loss增强模型学习少数已知类别样本的能力,引入MC-Dropout方法通过对不确定性进行建模,来得到更加可靠的样本检测置信度,从而进行更精确的未知类判别。

3)在CICIDS-2017数据集上进行了实验,评估本文提出的开放集入侵检测方法。实验结果表明,所提方案相比其他算法有一定改进。

2 相关工作

2.1 传统入侵检测方法

在过去几年中,大量的研究人员对基于机器学习或深度学习的入侵检测算法进行了广泛的研究,这些研究主要集中在流量特征选择、学习模型选择和模型优化方面。在研究人员的努力下,近年来入侵检测系统的检测能力有了明显的提高。

Gu等[1]提出了一种利用SVM集成和特征增强的有效入侵检测框架。该方法在原始特征上实施对数边际密度比变换,从而获得新的质量更好的变换训练数据,然后使用SVM进行集成建立入侵检测模型。实验结果表明,该方法具有良好的鲁棒性能。Mustapha等[2]基于机器学习框架,使用Apache Spark评估了SVM、朴素贝叶斯、决策树和随机森林4种分类算法的性能。从检测准确率、构建时间和预测时间3个方面对总体性能进行评估。实验使用UNSW-NB15进行算法评估,结果表示随机森林分类器在检测准确率和预测时间方面表现效果较好。

Nasr等[3]提出了一个基于CNN的入侵检测系统DeepCorr。与传统的入侵检测系统相比,该系统将通用的统计相关算法转换为深度学习算法,抓住了噪声动态变化的复杂本质,对长度较短的流量数据具有较强的检测能力,可以获得更高的检测精度。Li等[4]提出了一个基于随机森林算法的AE入侵检测系统。该系统通过特征选择和特征分组构建训练集,使用浅层AE神经网络进行攻击检测,降低了计算的复杂性,并在AE特征的帮助下解决了传统方法中的样本不平衡问题,有效提高了检测精度。Xiao等[5]提出了一个基于AE和CNN的网络入侵检测模型。该模型通过AE降低处理后的数据集的维度,将降低后的数据转换为灰度图像输入到CNN,并提取和分析数据特征,由CNN进行分类。实验结果表明,该模型的准确性、检测率和FAR都优于大多数深度学习模型。上述传统入侵检测方法均是基于封闭环境对已知类别进行模型训练,并通过特征选择、模型构建等方法针对模型精度进行进一步提升。

2.2 开集识别入侵检测方法

Scheirer等[6]首先提出了开放集识别的本质,并将其定义为约束最小化问题,即经验风险和开放空间风险最小化问题。文章提出了1-vs-set的解决方法,即在原有的SVM决策平面之外,在训练样本的另一侧增加一个决策平面,并通过调整这两个平面来实现风险的小型化。

Cruz等[7]首次将开放集识别应用于入侵检测领域,并将W-SVM应用于KDDCUP'99数据集进行细粒度分类。Rudd等[8]提出EVM方法,它的工作原理是根据每个类的边缘距离找到已知空间的紧凑决策边界。Henrydoss[9]将EVM应用于入侵检测中,结果表明该方法在KDDCUP'99数据集上取得了和W-SVM相当的性能。

然而,鉴于机器学习模型在处理高维数据上的局限性,之后提出了各种深度学习方法。Shu等[10]提出了DOC方法,该方法利用Sigmoid函数代替Softmax,为每个已知类分配一个唯一的高斯阈值以进行未知检测。之后,Hassen等[11]引入了中心损失来提高特征嵌入的有效性,并利用以类为中心的距离阈值进行未知确定。Shieh等[12]通过高斯混合模型学习了属于已知聚类的样本的可靠性,然后分配了一个置信阈值来确定未知类。Lai等[13]提出了一种基于单类支持向量机(OCSVM)的评分校准和阈值化方法OpenSMax的域生成算法(DGA)检测,其中OCSVM建模了分类器最后两层的激活分布。其他判别模型继承了机器学习解决方案的优点,采用极值理论阈值来实现未知行为推断。根据这些想法,Zhang等[14]引入了开放式深度网络OpenMax,用于未知攻击检测。它提取深度网络的倒数第二层的激活,以在训练阶段拟合已知类的Weibull分布。在测试阶段,它根据Weibull分布重新

分配 Softmax 层的值，并为已知或未知类分配重新分布的概率。除了上述方法外，还有基于生成模型的判别方法。Liu 等[15]提出了 SFE-GACN 来生成用于入侵检测的未知嵌入，而 Guo 等[16]提出了 MOSR 来生成用于恶意软件识别的未知样本。上述开集识别检测方案在一定程度上解决了入侵检测中推断识别未知恶意流量的问题。但是上述方案均未考虑未知的正常样本。

3 开集识别入侵检测框架

图 2 给出了本文提出的开放集识别模型的整体框架。

1）开放集数据处理。我们根据开放集识别的需求对数据集进行划分。首先，针对正常流量数据进行 PCA 降维可视化，确定聚类数量。然后使用 GMM 聚类算法对正常流量进行聚类。最后通过统计设定未知类别加入测试集中。

2）开放集分类器训练。首先，对已知类别的流量基于互信息进行特征选择，从给定的特征集中选出特征之间相关性最高的特征子集，以便在后续任务中构建更有效的模型。然后利用划分好的训练集多次训练模型，训练过程中引入 Focal Loss 处理数据不平衡问题，并引入 MC-Dropout 处理样本不确定性。

3）开放集测试。首先，利用训练好的分类模型检测测试集 q 次，计算测试集中每个样本的类别置信度的标准偏差 s_i。将 s_i 和样本在训练集中的类别置信度的标准偏差 S_i 进行比较。如果小于指定阈值，则标记已知类，否则标记为未知类。

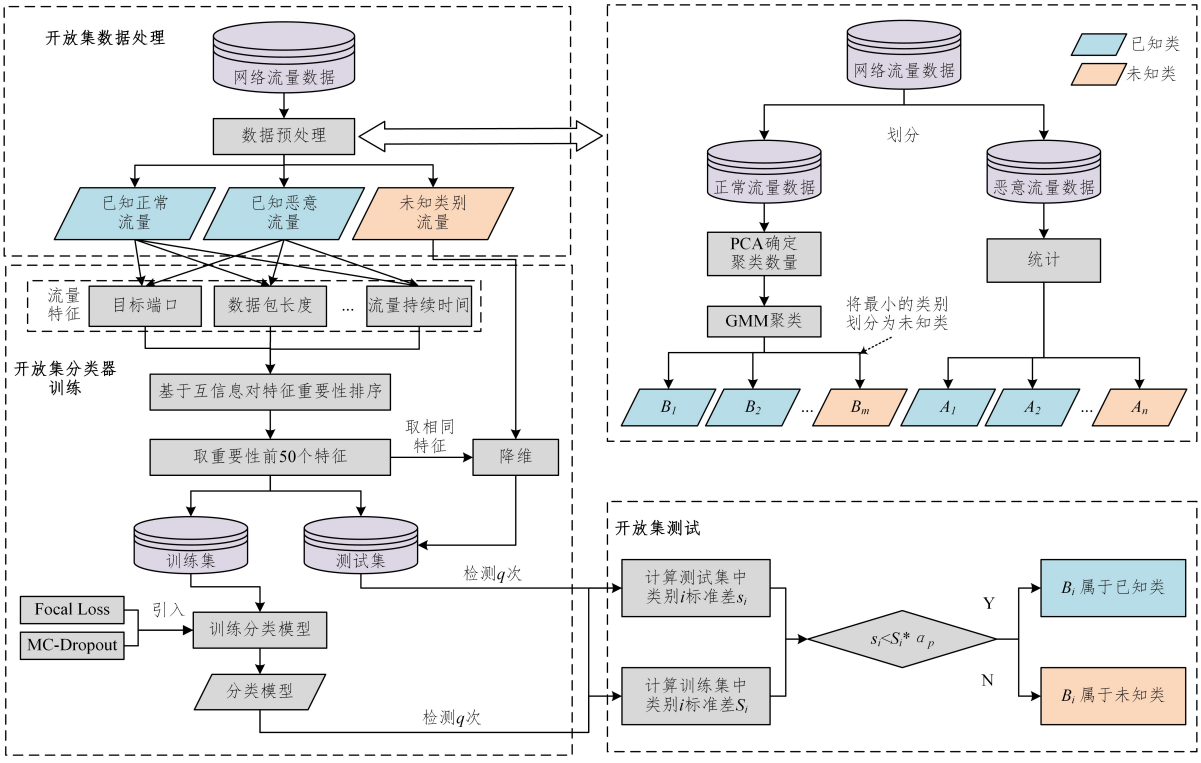


图 2 开集识别入侵检测框架

Fig. 2 Framework of open set recognition intrusion detection

3.1 开放集数据处理

不同于以往开集识别问题，我们认为正常流量也存在多种不同的行为模式，在特征空间上存在多种不同分布。为了使模型能够更加精细地学习正常流量的行为模式，在预处理阶段，我们通过 PCA 进行降维可视化，以确定后续聚类数量，每一个子类都表示一种正常流量的行为模式。本文采用 CICIDS2017 数据集进行实验，可视化结果如图 3 所示。

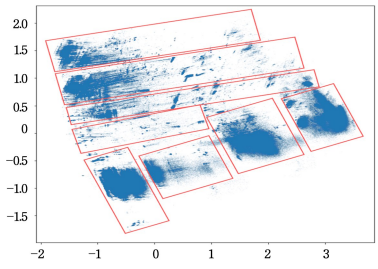


图 3 正常流量 PCA 降维

Fig. 3 PCA dimension reduction for normal flow

由图 3 可知，CICIDS2017 数据集中的正常流量可被清晰地分为 8 个不同类别，将其作为 GMM 算法的聚类类别个数。GMM 能够为每个数据点分配一个概率分数，表示该数据点属于每个聚类的可能性，能够更好地处理具有模糊边界的聚类问题。

最终选择样本数量最少的一类作为未知正常流量加入未知类别中。而在 CICIDS2017 数据集攻击流量中，我们同样选择样本数量最少的 5 类作为未知攻击样本加入未知类别。

3.2 开放集分类器训练

分类器训练阶段的主要目的是实现已知类别的入侵检测模型。由于训练集中没有未知样本，因此该训练阶段可视为封闭假设环境下的入侵检测模型训练。鉴于深度学习在入侵检测分类中出色的表现，本文构建深度学习模型来对已知类别进行分类。

3.2.1 特征筛选

有研究表明[17]，提升模型闭集分类精度可以提升开集检测的准确率。为了提升模型在封闭集上的性能，我们在构建网络前利用互信息（Mutual Information，MI）进行了特征选

择。互信息是衡量两个随机变量之间依赖关系的一种指标，可以使用互信息来衡量特征与标签之间的依赖关系，从而筛选出与标签相关性较高的特征。

互信息定义如下：

$$MI(X,Y)=\sum_{x\in X}\sum_{y\in Y}p(x,y)\log\frac{p(x,y)}{p(x)p(y)} \tag{1}$$

其中，X 和 Y 分别表示两个随机变量，$p(x,y)$ 是 $X=x$ 且 $Y=y$ 的概率，$p(x)$ 和 $p(y)$ 分别是 $X=x$ 和 $Y=y$ 的概率。互信息的值越大，表示 X 和 Y 之间的关联程度越高。具体流程如算法 1 所示。

算法 1 基于互信息筛选特征

输入：特征集 $F=\{f_1,f_2,\cdots,f_{78}\}$，标签 Y

输出：筛选后特征子集 F_{Best}

1. For i in F do
2. 计算 f_i 和 Y 之间的互信息 MIi
3. 将 MIi 存入互信息列表 List
4. End For
5. 对互信息列表 List 中元素从大到小排序
6. 取前 50 个元素组成 F_{Best}
7. End

3.2.2 损失函数

在模型训练过程中，样本数据不平衡也是影响分类器效果的重要因素之一。针对这个问题，本文使用 Focal Loss 作为损失函数代替交叉熵函数，使分类算法更加关注少类样本，进而提高少类样本的分类精度，从而有效降低数据不平衡对分类器的影响。Focal Loss 的公式如下所示：

$$FL(p_t)=-\alpha(1-p_t)^{\gamma}\log(p_t) \tag{2}$$

其中，p_t 为待分类样本属于正类的概率；γ 为聚焦参数，能够减低多数类样本的损失贡献，本文取 2；α 为超参数，能够调节正负样本损失之间的比例，本文取 0.25。

3.2.3 处理检测置信不确定性

大多数深度神经网络的预测分布通常是以点预测为主，但是在开放集合识别这样判断高风险样本的任务中，我们不仅需要获得真实的预测概率，同时也需要获得相应的置信度，以便评估预测的风险并避免异常的置信度预测。在开集识别中，我们认为模型对于从未见过的样本置信度不确定性较高。在这种情况下，贝叶斯方法得到了广泛关注。贝叶斯神经网络通过将神经网络设定为近似贝叶斯模型，并学习参数的后验分布，为不确定性提供了有效的估计。

MC-Dropout 是一种基于深度神经网络的不确定性估计方法，可以被视为贝叶斯神经网络的一种近似推断方法，我们利用该方法处理检测置信度不确定性。MC-Dropout 使用 Dropout 技术来近似参数的后验分布。Dropout 是指在训练过程中随机地将一些神经元的输出设置为 0，从而强制神经网络在训练时学习多个互相矛盾的子模型，从而减少过拟合现象。而 MC-Dropout 算法的核心思想是在带有 Dropout 结构的贝叶斯网络上进行多次运算。由于每次运行时，参与信息传递的神经元不同，因此模型的预测结果也会有差异，类似于对预测结果进行了蒙特卡洛采样。通过进行大量的采样，可以利用概率论的方法计算模型预测结果的不确定性。

本文构建了一个由 Conv1D、GlobalMaxPooling1D 和多层 Dense 组成的 9 层分类模型，具体结构如表 1 所列。在 Conv1D 和 Dense 前加入 Dropout，Dropout 比例为 30%。在训练分类模型后，使用蒙特卡洛采样方法，利用激活 Dropout 层的分类模型多次检测样本，得到带有置信区间的检测结果。这意味着每次检测样本时都会随机丢弃某些权重来构建多个子分类模型。由于每个子分类模型都是从已训练好的分类模型的参数分布中随机采样得到的，因此它们的检测结果在一定程度上反映了参数分布的不确定性。在后续步骤中，我们能够利用这些检测结果在减轻不确定性影响的条件下判断未知样本。

表 1 模型结构

Table 1 Model structure

Layer	Output Shape	Param
InputLayer	(50)	0
Embedding	(50,40)	1 600
Dropout	(50,40)	0
Conv1D	(50,40)	6 400
GlobalMaxPooling1D	(40)	0
Dense	(32)	1 312
Dropout	(32)	0
Dense	(32)	1 056
Activation	(32)	0
Dense	(8)	264
Activation	(8)	0

3.3 开放集测试

在测试阶段，首先我们利用带有置信区间的检测结果统计训练集中已知类别中每一类的类标准差 $\{\sigma_1,\sigma_1,\cdots,\sigma_k\}$。之后对测试集中的第 i 个样本检测 q 次，统计第 i 个样本检测置信度标准差 $\sigma_{p,i}$，其中 $p\in\{1,2,\cdots,k\}$。具体流程如图 4 所示。

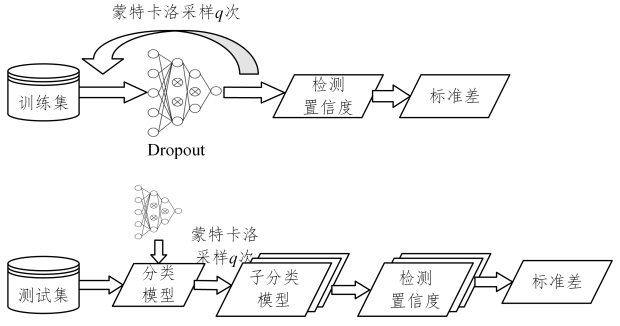


图 4 开放集测试流程

Fig. 4 Open set testing process

最终将预测标签的类标准差与该数据标准差进行比较，计算式如下：

$$y=\begin{cases}\text{样本 } i\in\{\text{已知类样本}\}, & \text{if } \sigma_{p,i}<\alpha_p\cdot\sigma_i\\ \text{样本 } i\in\{\text{未知类样本}\}, & \text{otherwise}\end{cases} \tag{3}$$

其中，α_p 为阈值超参数。α_p 越小，模型更容易将样本检测为未知类样本。

4 实验

4.1 数据集介绍

本文使用 CICIDS2017 作为基准数据集。该数据集包含良性流量和最新的常见攻击，共有 78 维特征。该数据集从 5 天的网络流量中提取了 280 万条流量记录，其中包含正常流量和 14 种最新的攻击方式，数据集分布情况如表 2 所列。

表 2　数据集分布情况

Table 2　Distribution of datasets

标签名称	标签编码	数量
BENIGN	0	2 273 097
Dos	1	252 661
DDoS	2	128 027
PortScan	3	158 930
FTP-Patator	4	7 938
Web Attack	5	2 180
SSH-Patator	6	5 897
Heartbleed	7	11
Bot	8	1 966
Infiltration	9	36

首先根据 2.1 节的方法将数据集按照标签分为已知类别和未知类别。依据现实场景中未知类别通常占比较少这一假设，我们选择数量最少的 5 种恶意流量类别作为未知类别。其中，为了区分不同行为模式的正常流量，我们将标签为 BENIGN 的流量进行聚类，分成多个簇 $\{B_1, B_2, \cdots, B_8\}$。将样本数最少的簇 $\{B_{\min}\}$ 设为未知类别。最终，开集入侵检测问题的标签空间变为 $\{B_1, B_2, \cdots, B_7, A, U\}$，其中 A 表示已知类别的恶意流量集合，U 表示未知类别流量。我们使用重新标记后的标签训练检测模型，对每一类正常流量学习其独立的流量特征，即训练过程中数据标签为 $\{B_1, B_2, \cdots, B_7, A\}$。需要注意的是，尽管在训练和测试过程中，我们会将正常流量归类于某个子类，但是由于该子类是由聚类算法划分而得的，不具备实际层面的意义，因此在最后评估模型性能时，我们将所有正常流量子类均看成一类来进行相关指标的计算。

表 3　数据标签划分情况

Table 3　Division of data labels

已知类别	未知类别
$\{B_1, B_2, \cdots, B_7\}$, Dos, DDoS, PortScan, FTP-Patator	Web Attack, SSH-Patator, Heartbleed, Bot, Infiltration, $\{B_{\min}\}$

图 5 为经过互信息计算后，所有特征互信息值从大到小排列的结果。本文选择前 50 个特征作为特征筛选后的结果。

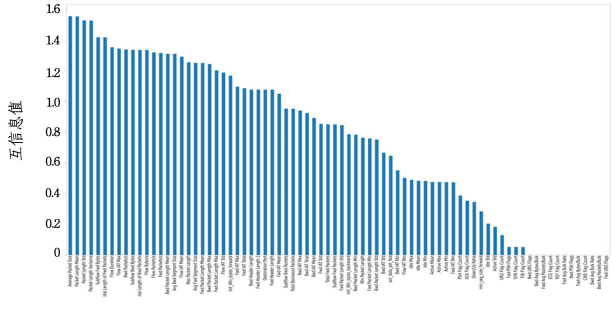


图 5　特征互信息值

Fig. 5　Feature mutual information values

4.2　评价指标

我们使用混淆矩阵、准确度 ACC、Macro-F1 和未知类别 U-F1 值进行模型评估。准确率主要计算所有类别中正确分类的比例；F1 分数综合考虑了模型的精确率(Precision)和召回率(Recall)，通过平衡这两个指标来提供一个单一的评估指标，F1 值越接近 1 表示模型的性能越好。假设，TP 表示真正例(True Positive)，即模型正确预测为正例的样本数量；FP 表示假正例(False Positive)，即模型错误预测为正例的样本数量；FN 表示假反例(False Negative)，即模型错误预测为反例的样本数量。各评价指标公式如下所示：

$$\begin{cases} P=\dfrac{TP}{TP+FP} \\ R=\dfrac{TP}{TP+FN} \\ F1=\dfrac{2\times P\times R}{P+R} \\ Macro\text{-}F1=\dfrac{1}{n}\sum_{i=1}^{n}F1_i \end{cases} \tag{4}$$

4.3　实验结果

在对已知类别进行入侵检测阶段，模型输出不考虑未知行为推断，因此只分为恶意流量或正常流量。在这一阶段，我们以混淆矩阵的形式对结果进行展示，如图 6 所示，模型对正常和恶意流量识别准确率均可达 99%，其对于封闭集上的分类效果较好，但未知流量均被错误分类。

1)超参数选择实验

为了评估超参数 α 对模型训练的影响，我们使用 F1 值作为评价标准，从 0 到 5，以 0.5 为步长进行实验，图 6 展示了实验结果。当 α 取值大于 3 时，已知正常类别和已知攻击类别的 F1 值变化趋于平缓，而未知类别的 F1 值开始下降。因此，在后续实验中，将超参数 α 值取定为 3。

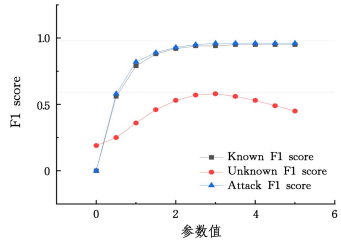
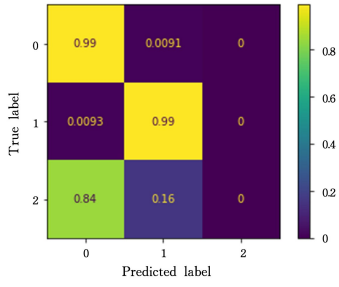


图 6　不同 α 下的 F1 值

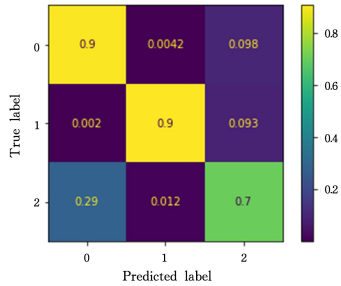
Fig. 6　F1 values with different α

2)开放集有效性实验

在对未知流量进行检测阶段，我们应用 MC-Dropout 方法，为每个输入生成 20 个 Monte Carlo 样本，用于不确定性建模。图 7 分别展示了在封闭集和开放集上的实验结果，标签 0，1，2 分别表示已知正常类、已知攻击类和未知类别。



(a)封闭集下入侵检测结果



(b)开放集下入侵检测结果

图 7　实验结果对比

Fig. 7　Comparison of experimental results

图 7(a)即为未采用开放集测试算法的实验结果，经对比可以发现，所提方法可成功识别 70%未知样本，并且保持对已知类样本检测率在 90%左右。比较图 7 中结果可以发现，虽然已知类识别率略有下降，但是所提出的方法可以识别误分类为已知类的未知样本。例如，被错误分类为已知正常和已知恶意的未知流数量比例分别减少了 65.4%和 92.5%。

3)性能对比实验

本文选择 5 个常用的开放集识别方法验证所提方法有效性，包括 OpenMax[14]，OpenSMax[13]，DOC[10]，GMM[12]，CE+ii[18]。其中，OpenMax，OpenSMax，DOC 方法在引言中已有描述；GMM 是一种使用高斯混合模型确定未知 DDos 攻击的方法；CE+ii 是一种最小化交叉熵损失和类内距离并最大化类间差异的表示学习方法。在对比实验中，我们实现了 5 个开放集识别模型，针对每个模型，我们均对论文中的方法进行了复现，表中列出了对比实验的结果，其中最佳性能以粗体表示。

表 4 对比实验结果

Table 4 Comparative experimental results

方法	Macro-F1	ACC	U-F1
OpenMax	0.77*	0.83	0.55*
OpenSMax	0.75	0.85	0.50
DOC	0.68	0.89*	0.14
GMM	0.66	0.82	0.24
CE+ii	0.67	0.79	0.32
Our Method	**0.84**	**0.91**	**0.58**

除了 Macro-F1 和 ACC 外，还记录了 U-F1 的性能，以考虑模型执行未知行为推断的能力。可以看出，在 Macro-F1 分数和 ACC 上，本文提出的方法取得了最好的结果，分别提升了约 9%和 2%，这验证了本文提出的方法在检测能力上的有效性。此外，本文提出的方法还获得了最高的 U-F1，这证明了其在区分已知类别和未知类别方面的有效性。

结束语 本文针对入侵检测过程中未知威胁行为不断增多，无法有效检测的问题，使用基于 MC-Dropout 方法，在公共数据集上构建了开放集合，并进行了模型搭建，提出了一种开集检测的方法。实验结果表明，该方法可以有效对未知类别进行检测并优于其他对比方法。然而，尽管本文提出的方法在入侵检测方面取得了积极的进展，但仍然面临一些挑战。首先，数据集的质量和多样性对于训练鲁棒的模型至关重要，因此需要更多关于入侵行为的真实数据以及合适的数据增强技术。其次，模型的解释性和可解释性是一个重要问题，特别是误报率方面，需要开发更加可解释的模型和算法。此外，如何对入侵检测做细粒度开放集合分类也是一个重要的挑战。

参 考 文 献

[1] GU J, WANG L H, WANG H W, et al. A novel approach to intrusion detection using SVM ensemble with feature augmentation[J]. Comput. Secur., 2019, 86: 53-62.

[2] BELOUCH M, EL HADAJ S, IDHAMMAD M, et al. Performance evaluation of intrusion detection based on machine learning using Apache Spark[J]. Procedia Computer Science, 2018, 127: 1-6.

[3] NASR M, BAHRAMALI A, HOUMANSADR A, et al. DeepCorr: Strong Flow Correlation Attacks on Tor Using Deep Learning[C]// Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security. 2018.

[4] LI X K, CHEN W, ZHANG Q R, et al. Building Auto-Encoder Intrusion Detection System based on random forest feature selection[J]. Comput. Secur., 2020, 95: 101851.

[5] XIAO Y H, XING C, ZHANG T N, et al. An Intrusion Detection Model Based on Feature Reduction and Convolutional Neural Networks[J]. IEEE Access, 2019, 7: 42210-42219.

[6] SCHEIRER W J, DE REZENDE ROCHA A, SAPKOTA A, et al. Toward Open Set Recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(7): 1757-1772.

[7] CRUZ S, COLEMAN C, RUDD E M, et al. Open set intrusion recognition for fine-grained attack categorization[C]// 2017 IEEE International Symposium on Technologies for Homeland Security(HST). Waltham, MA, USA, 2017: 1-6.

[8] RUDD E M, JAIN L P, SCHEIRER W J, et al. The Extreme Value Machine[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(3): 762-768.

[9] HENRYDOSS J, CRUZ S, RUDD E M, et al. Incremental Open Set Intrusion Recognition Using Extreme Value Machine[C]// 2017 16th IEEE International Conference on Machine Learning and Applications(ICMLA). Cancun, Mexico, 2017: 1089-1093.

[10] SHU L, XU H, LIU B. Doc: Deep open classification of text documents[C]// Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. 2017: 2911-2916.

[11] HASSEN M, CHAN P K. Learning a neural-network-based representation for open set recognition[C]// Proceedings of the 2020 SIAM International Conference on Data Mining. SIAM, 2020: 154-162.

[12] SHIEH C S, LIN W W, NGUYEN T T, et al. Detection of unknown ddos attacks with deep learning and gaussian mixture model[J]. Applied Sciences, 2021, 11(11): 5213.

[13] LAI Y, PING G, WU Y, et al. Opensmax: Unknown domain generation algorithm detection[J]. Frontiers in Artificial Intelligence and Applications, 2020, 325: 1850-1857.

[14] ZHANG Y, NIU J, GUO D, et al. Unknown network attack detection based on open set recognition[J]. Procedia Computer Science, 2020, 174: 387-392.

[15] LIU A, WANG Y, LI T. SFE-GACN: A novel unknown attack detection under insufficient data via intra categories generation in embedding space[J]. Computers & Security, 2021, 105: 102262.

[16] GUO J, GUO S, MA S, et al. Conservative Novelty Synthesizing Network for Malware Recognition in an Open-Set Scenario[J]. IEEE Transactions on Neural Networks and Learning Systems, 34(2): 662-676.

[17] VAZE S, HAN K, VEDALOI A, et al. Open-Set Recognition: Good Closed-Set Classifier is All You Need[J]. arXiv: 2110.06207, 2022.

[18] HASSEN M, CHAN P K. Learning a Neural-network-based Representation for Open Set Recognition[C]// SDM. 2018.

WANG Chundong, born in 1969, Ph. D, professor, is a senior member of CCF (No. 16230M). His main research interests include network information security, mobile intelligent terminal secu-rity, public opinion analysis and control, Internet of Things security and security situation awareness.