刘恋,董育宁. 基于联合预测和双层模型的未知网络流量检测[J]. 智能计算机与应用,2026,16(1):81-88. DOI:10.20169/j.issn.2095-2163.24032901

基于联合预测和双层模型的未知网络流量检测

刘 恋,董育宁
(南京邮电大学 通信与信息工程学院,南京 210003)

摘 要:现有的开集流识别方法聚焦于新类检测,但在整体性能指标和分类速度方面仍有待提升。针对这一问题,本文提出了一种基于联合预测和双层分类的未知流量检测模型。由已知类训练出 2 个多分类器模型作为上层分类器,定义置信度差和标签距离用来评估上层分类器的结果。将伪新类与已知类通过上层分类器的输出值训练下层分类器,以有效检测出难以区分的新类。实验表明,本文方法的总体准确率比现有方法提高了 2%至 4%,分类时间也大大减少。

关键词:机器学习;开放集识别;新类检测;未标记样本
中图分类号:TP391 **文献标志码**:A **文章编号**:2095-2163(2026)01-0081-08

Unknown network traffic detection based on joint prediction and two-layer model

LIU Lian, DONG Yuning
(School of Communications and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China)

Abstract: Existing open set traffic recognition methods focus on new class detection, but still need to be improved in terms of overall performance metrics and classification speed. To address this problem, this paper proposes an unknown traffic detection model based on joint prediction and two-layer classification. Two multi-classifier models are trained from known classes as upper layer classifiers, and the confidence difference and label distance are defined to evaluate the results of the upper layer classifiers. The fake new classes and known classes are used to train the lower layer classifiers by the output values of the upper layer classifiers in order to effectively detect the unknown classes difficult to distinguish. Experiments show that the overall accuracy of the method in this paper is improved by 2% to 4% over existing methods, and the classification time is greatly reduced.

Key words: machine learning; open set recognition; new class detection; unlabeled sample

0 引 言

网络流分类技术是流量控制、异常检测和实时流量分析等网络管理的一个重要技术手段[1]。现有的闭集流分类方法假设训练集和测试集都是已知的,而网络环境中具有新应用产生的流量。开放集识别(Open Set Recognition, OSR)[2]解决的就是一种在测试过程中可能面临来自训练过程中不可见的类的识别问题。

为了更好地提高对新型流量类别的识别能力,Neira 等学者[3]采用集成分类器的思想,将不同的分类器和样本特征结合起来解决 OSR 问题。集成模型结合一些单模型的优点,能取得更好的整体预测性能[4]。Zhai 等学者[5]利用 bagging 算法构建集成模型,最后的分类决策利用简单的投票机制。但是 OSR 问题中会出现新类别,固定阈值和投票准则不能适应数据的动态变化。

针对上述问题,本文提出一种基于联合预测和双层模型的未知网络流量检测方法,用于完成对新类样本的在线检测和已知类样本的细分类。

1 相关工作

分类是在开放集网络流量世界中验证模型安全的一种重要方式[6]。传统的分类模型都在封闭的数据集里训练,假设训练样本和测试样本都来自同一数据集,但实际部署中总会遇到不属于本数据集里的样本。Haque 等学者[7]用无监督方法估计分类器的置信度,但要求事先了解置信度值的分布情况,

作者简介:刘 恋(1999—),女,硕士研究生,主要研究方向:网络流分类,多媒体通信与无线网络。
通信作者:董育宁(1955—),男,博士,教授,博士生导师,主要研究方向:网络流分类,多媒体通信与无线网络。Email:19900011@njupt.edu.cn。

收稿日期:2024-03-29

在实际场景中并不适用。

为了加强 OSR 的性能,生成式模型将生成的新类数据加入到训练中。因此,生成数据的方法显得尤为重要。文献[8]通过寻找与已知类样本相似的数据生成负样本来扩大数据集,但是生成的样本接近已知类,分类时容易和已知类混淆。与上述工作不同,Neal 等学者[9]使用编解码生成对抗网络生成与已知类相似的样本,但仅基于已知类别构建的模型,在识别新类别的能力上显然是有限的。

此外,Verma 等学者[10]比较了决策树和多层感知器等几种独立算法以及随机森林和 boosting 等不同的集成算法的准确性和响应时间,发现决策树和极端梯度提升树提供了最佳的性能折衷。但基于机器学习(Machine Learning,ML)的检测不能完全避免错误分类,特别是考虑到一些具有以前未见模式的新应用程序产生的流量。

部分 OSR 场景中依赖于使用阈值的分类方案。在传统最近邻分类器(Nearest Neighbor Classifier,NNC)的基础上,Júnior 等学者[11]引入了一个在开放集环境下的 NNC 来处理 OSR 问题。与使用阈值直接对最相似的类别进行相似度评分的方法不同,利用一个阈值对最相似的 2 个类别的相似度评分比进行判断。

由于缺乏新类的信息,对已知类别样本的分布规律进行分析,并据此设定阈值来区分已知类和新类的方法不具有普适性,阈值的选择比较敏感。DACS 使用训练样本交叉模拟零日流,并结合了分布式训练的思想,利用随机森林和 SVM 算法实现[12]。通过 2 层结构和特殊的投票机制,能够在混合未知流量中执行全面的流量分类任务。但是模型依赖于已知类的组合方式,不够稳定。

上述文献方法在 OSR 领域内取得了成效,但是都存在不足。在 OSR 场景中,分类器不仅要准确地对已知类进行分类,而且能对新类进行有效探测。因此,本文提出了一种层级结构,利用上、下两层分类器识别出已知类和新类,利用上层分类器的结果对已知类别细分类。

2 本文方法

总体模型包括 2 个模块:训练阶段模块和测试阶段模块。在训练阶段,对数据集中的样本进行预处理、即特征提取(Feature Extraction, FE)和特征选择(Feature Selection, FS),再提取出伪新类样本;在测试阶段,计算预测指标值和进行新类检测。利用一个集分类器模型在筛选新类的同时,对已知类别样本进行细分类。下面分别予以介绍。

2.1 训练阶段模块

利用实际网络中的未标记网络流样本,依次经过多个由已知类样本训练的 OC-SVM 分类器后筛选出伪新类样本,过滤掉未标记样本中与已知类别相似的样本。从已知类和新类的置信度差和标签距离的分布差异入手,上层分类器 C_1 由 2 种极端梯度提升(eXtreme Gradient Boosting, XGBoost)[13] 多分类器构成(见图 1)。已知类和伪新类样本通过上层分类器得到的置信度差和标签距离指标,作为新的特征来训练下层分类器 C_2(XGBoost 二分类器),以筛选出已知类和新类;同时利用上层分类器的输出结果对下层分类器筛选出的已知类进行细分类。具体训练过程如图 1 所示。

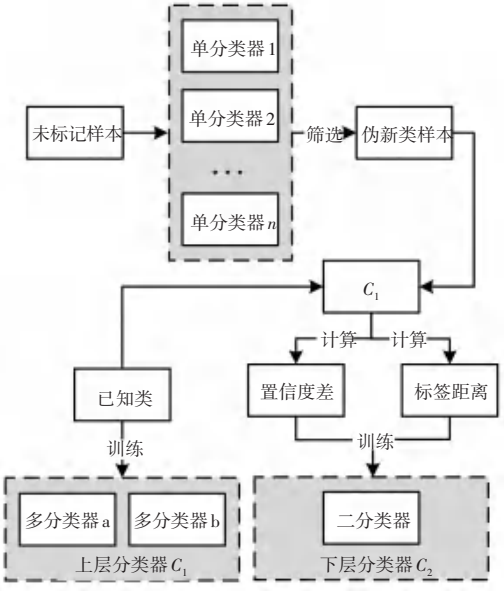

图 1 训练阶段框图

Fig. 1 Block diagram of the training stage

2.2 特征提取与选择

为了实现在线流分类,将采集到的流数据按照 1 s 时间段划分为流段。为了减少资源消耗,再使用每条流段中的前 10 个数据包来计算出对应的特征[14],共提取 131 个流特征。

首先计算特征与标签之间的皮尔森相关系数(Pearson Correlation Coefficient, PCC)[15],然后对每 2 个特征间的 PCC 进行计算,删除 PCC 值大于 0.9 的特征对中对应标签与特征相关性更小的那个特征,再通过极限树[16]特征重要性排序进行降维。按照此排序逐渐加入特征,观察模型性能的变化,寻找性能拐点,得到最优的特征子集。

2.3　提取伪新类样本

通过对已知类和未知类的置信度差和标签距离分布规律进行分析，发现设定一个阈值来区分已知类和新类的方法在部分网络流量分类场景中不能很好地工作。同时，生成样本的方法取决于生成器的性能[17]，而现实网络中的未标记流量样本更容易获得[18]。利用 n 类已知类样本训练 n 个 OC-SVM 分类器，将未标记流样本依次输入到每个单分类器后得到所需要的负类样本。用这些负类样本训练下层二分类器，其目的是根据不同的预测度量值来更好地区分已知类和新类。筛选伪新类样本的流程如图 2 所示。

图 2　多个 OC-SVM 分类器提取伪新类样本

Fig. 2　Extracting fake new class samples of multiple OC-SVM

2.4　测试阶段模块

在测试阶段，未知样本输入训练好的两层分类器 C_1 和 C_2，流程见图 3。若被下层分类器 C_2 识别为新类，则放入新类缓存，可用于模型增量更新；若被识别为已知类，则根据由上层分类器 C_1 中的 2 个多分类器分别计算的置信度差和标签距离指标，对已知类样本进行细分类，输出类别标签。

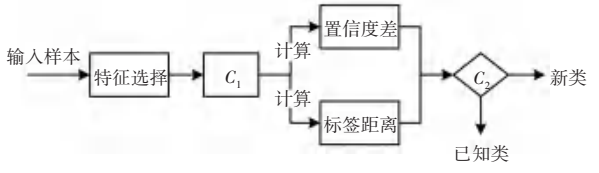

图 3　测试阶段框图

Fig. 3　Block diagram of testing phase

2.5　预测指标

2.5.1　置信度差定义

(1) 设置 XGBoost 算法输出结果为 n data × m class 的矩阵。每一行预测值表示样本所属各个类别的概率。其中，n 表示每一行样本，m 表示某类标签。

(2)定义置信度差(de)为样本预测概率的最大值与最小值的差值，预测标签就是最大置信度对应的下标。第 i 行的输出向量为 P，计算公式如下：

$$\mathrm{de} = \max(P) - \min(P) \tag{1}$$

2.5.2　标签距离定义

(1)对输入样本的标签进行独热编码，即转换成从 0 开始的数字类标签。

(2)设置 XGBoost 算法输出结果为 n data × 1 的向量。每个预测值是具体的类别。其中，n 表示每一行样本。

(3)定义标签距离(le)为输出值 y_{pre} 与预测值 y_{int}(对输出值四舍五入)的差值，预测标签就是对输出值取整后的结果。标签距离(le)的定义公式如下：

$$\mathrm{le} = | y_{\mathrm{pre}} - y_{\mathrm{int}} | \tag{2}$$

2.6　新类检测

现实场景中某些新类与已知类别较相似，仅用置信度在分类中表现一般，而置信度差和标签距离两者的组合鲁棒性会更强。本文利用已知类和新类的分布差异，根据置信度差和标签距离这 2 种预测度量的值来训练一个二分类器。分类算法描述具体如下。

算法 1　新类检测算法

输入　C_1 表示上层分类器，C_2 表示下层分类器，x 表示输入样本

输出　yk 表示已知类预测标签，yn 表示未知类预测标签

```
for x_i in x:
    对 x_i 进行 FE 和 FS 后通过 C_1
    计算得到置信度差 p_i 和标签距离 d_i
    将 p_i 和 d_i 输入到 C_2 得到预测标签 y_i
    if y_i == 0:
        if y(p_i) == y(d_i):   yk ← y_i
        else:   yn ← y_i
        end if
    else: yn ← y_i
    end if
return yk, yn
```

3　实验结果与讨论

3.1　数据集

本实验使用由 Wireshark[19] 软件在南京邮电大学校园网内采集的流量数据集(NYDS)、ISCX 公共数据集[20](ISCX)，以及这 2 个数据集组成混合数据集(MIXD)来进行实验验证。

3.1.1　NYDS 数据集

本数据集是在南京邮电大学校园网内采集的网络视频流量数据集，包括点播和直播两大类，根据不同清晰度分类(480 p、720 p 和 1 080 p)，共分为 6 种流量类型。NYDS 数据集见表 1。

3.1.2　ISCX 数据集

ISCX 数据集是一个公共数据集，本实验选用了部分数据，包括文件传输、语音通话、邮件、视频和聊

天这 5 种应用类型,共 8 种流量类别。ISCX 数据集见表 2。

表 1 NYDS 数据集

Table 1 NYDS dataset

流量类型	应用	样本数
点播_480 p	tencent、bilibili	1 000
点播_720 p	tencent、bilibili、youku	1 000
点播_1 080 p	tencent、bilibili、youku	1 000
直播_480 p	douyu、huya	1 000
直播_720 p	douyu、huya	1 000
直播_1 080 p	douyu、huya	1 000

表 2 ISCX 数据集

Table 2 ISCX dataset

流量类型	应用	样本数
文件传输	BitTorrent、Ftp、Skype_file	1 000
语音通话	Facebook_audio、Hangouts_audio、Skype_audio	1 000
邮件	Email	1 000
视频	Skype_video、YouTube	1 000
聊天	Facebook_chat	1 000

3.2 评价指标

本文通过多个评价指标来对检验模型性能进行评估,分别从分类准确率(Accuracy,Acc)、查准率(Precision, P)、查全率(Recall, R)、$F1$ 得分($F1$_score,$F1$)和标准化准确度[11](Normalized Accuracy, NA)多个方面进行综合考量。其中,NA 考虑的是已知样本的准确度(Accuracy on Known Samples, AKS)和未知样本的准确度(Accuracy on Unknown Samples, AUS)。

$$\mathrm{Acc} = \frac{T_P + T_N}{T_P + F_P + T_N + F_N} \tag{3}$$

$$P = \frac{T_P}{T_P + F_P} \tag{4}$$

$$R = \frac{T_P}{T_P + F_N} \tag{5}$$

$$F1 = \frac{2 \times P \times R}{P + R} \tag{6}$$

$$\mathrm{NA} = \lambda \mathrm{AKS} + (1 - \lambda)\mathrm{AUS} \tag{7}$$

其中,T_P 和 F_P 分别表示被分类正确和被分类错误的的正样本数;T_N 和 F_N 分别表示被分类正确和被分类错误的负样本数;λ 为开集样本中已知类所占的比例,且 $0 < \lambda < 1$。

3.3 实验环境

本文采用 XGBoost 模型和 OC-SVM 模型组成的集分类器,其中 XGBoost 的参数 n_estimators 设置为 50,OC-SVM 中规则化参数设为 0.07。采用五折交叉验证对模型进行评估,并与文献[12]和文献[21]进行性能比较。所有的实验均在硬件配置为 Intel(R) Core(TM) i5-8250U CPU @ 1.60 GHz,16 GB 内存联想笔记本上电脑上完成,操作系统为 Windows 11。

3.4 指标分布

样本通过上层分类器后,已知类和新类根据不同的输出结果计算出不同的置信度差和标签距离。为了证明分布情况的普适性,从 MIXD 数据集中随机选择 7 个已知类和 7 个新类的样本各 500 条进行测试,结果见图 4。

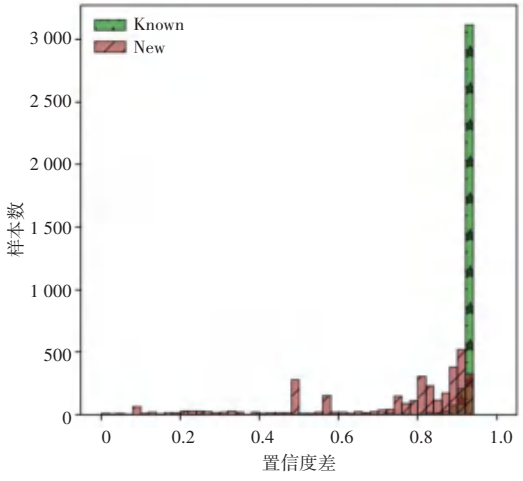

图 4 已知类和新类的置信度差分布图

Fig. 4 Distribution of confidence difference between known and new classes

已知类和新类的标签距离分布如图 5 所示。由图 4、图 5 可以看出,已知类集中在置信度差较大、标签距离较小的地方,而且一部分新类则是分布在接近已知类的地方,甚至在图 4 中有小部分新类置信度差的结果与已知类极其相似。

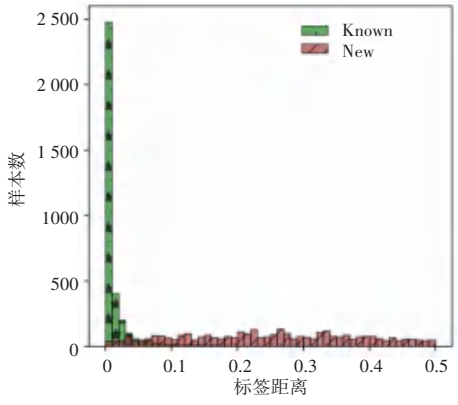

图 5 已知类和新类的标签距离分布图

Fig. 5 Distribution of labeling distances between known and new classes

在已知类训练出上层分类器后，若是已知类，样本输出的置信度差的值较大、标签距离的值较小；而若是未知类，则得到的置信度差的值更小、标签距离的值更大。已知类和新类的二维分布如图 6 所示。由图 6 也可以看出，新类分布比较分散，此时仅利用置信度差一个预测度量来判断，新类误分为已知类的可能性较大。采用置信度差和标签距离来联合判断、共同决策就能更好地筛选出新类样本。

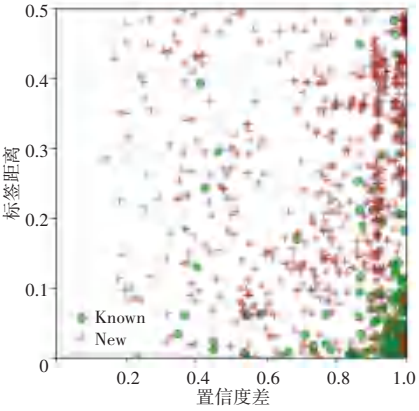

图 6　已知类和新类的二维分布图

Fig. 6　Two－dimensional distribution between known and new classes

3.5　消融实验

为了验证下层二分类器的有效性，现与使用先验知识挑选阈值（TH_CLD）的方法进行比较，在 3 个不同数据集上对实验方法进行分析。NYDS 上分类性能对比结果见表 3，ISCX 上分类性能对比结果见表 4，MIXD 上分类性能对比结果见表 5。表 4 中 TH_CLD 方法的分类性能不如表 3 和表 5，是因为在不同数据集上保持同一阈值来进行分类实验时，不同的样本产生了不同的分类效果，选出的阈值不再具有普适性。而 JPTM 不利用先验知识挑选阈值，利用上层分类器的结果训练下层分类器筛选新类，模型自动化的程度大大提高。在 3 个数据集上的 NA 值比 TH_CLD 高出 2%～10%，已知类和新类的 $F1$ 得分也高出约 5%。其中，JPTM 的新类的查全率 R 在不同数据集中都保持在 95%以上，测试集中的新类样本被较好地识别出来。

表 3　NYDS 上分类性能对比

Table 3　Comparison of classification performance on NYDS

方法	类别	$F1$	P	R	NA
TH_CLD	已知类	0.897	0.965	0.838	0.979
	新类	0.914	0.861	0.974	0.979
JPTM	已知类	**0.951**	**0.988**	**0.917**	**0.995**
	新类	**0.963**	**0.931**	**0.998**	**0.995**

表 4　ISCX 上分类性能对比

Table 4　Comparison of classification performance on ISCX

方法	类别	$F1$	P	R	NA
TH_CLD	已知类	0.906	0.870	0.950	0.917
	新类	0.898	0.950	0.851	0.917
JPTM	已知类	**0.958**	**0.990**	0.929	**0.992**
	新类	**0.962**	0.934	**0.991**	**0.992**

表 5　MIXD 上分类性能对比

Table 5　Comparison of classification performance on MIXD

方法	类别	$F1$	P	R	NA
TH_CLD	已知类	0.912	0.946	0.903	0.963
	新类	0.922	0.907	0.939	0.963
JPTM	已知类	**0.962**	**0.968**	**0.956**	**0.981**
	新类	**0.963**	**0.957**	**0.969**	**0.981**

3.6　不同方法对比

为了更好地分析模型，本文 JPTM 方法分别在 3 个数据集上与参考文献方法（DACS[12] 和 APT[21]）进行性能对比分析，结果见表 6～表 8。其中，DACS 方法是利用随机森林和 SVM 的双层系统，将随机森林输出的置信度最大值和方差输入 SVM，进一步根据 SVM 的投票结果来检测新类。APT 方法是利用最近邻算法确定测试对象属于已知类的概率，再根据阈值区分已知类别和未知类别。对比这些方法，在不同数据集上本文方法的已知类和新类的 $F1$ 得分及 NA 值较其他方法都有所提升。

表 6　NYDS 上不同方法分类性能对比

Table 6　Comparison of classification performance of different methods on NYDS

方法	已知类			新类			NA
	$F1$	P	R	$F1$	P	R	
DACS	**0.957**	0.960	**0.956**	0.961	**0.959**	0.962	0.978
APT	0.890	0.908	0.876	0.897	0.883	0.911	0.947
JPTM	0.951	**0.988**	0.917	**0.963**	0.931	**0.998**	**0.995**

表 7　ISCX 上不同方法分类性能对比

Table 7　Comparison of classification performance of different methods on ISCX

方法	已知类			新类			NA
	$F1$	P	R	$F1$	P	R	
DACS	0.886	0.874	**0.961**	0.721	**0.941**	0.584	0.773
APT	0.958	0.965	0.953	0.895	0.881	0.909	0.969
JPTM	**0.958**	**0.990**	0.929	**0.962**	0.934	**0.991**	**0.992**

DACS 是一个双层系统，把已知类样本按照不

同的标签类别分为 2 组对应 2 个模型对,2 组数据在一个模型对里互为正反类样本来训练模型。在不同数据集上,DACS 的模型性能表现不一。表 6 和表 8 中,DACS 得到了较好的分类性能,而表 7 上的 DACS 方法性能效果最差,因为 DACS 要对已知类划分不同组合,不同的组合可能产生不一致的结果,说明 DACS 方法也不如 JPTM 稳定。对比 DACS,在 NYDS 数据集、即表 6 中 2 种方法的已知类 $F1$ 和新类 $F1$ 得分相似,但是 JPTM 的 NA 值提升了约 2%。究其原因在于 JPTM 的新类的查全率 R 更高,而已知类的查准率 P 更高,避免了更多的新类被误分到已知类中。在另外 2 个数据集中,JPTM 的已知类和新类的 $F1$ 得分以及 NA 值都有所提升,在对比方法表现更好的表 8 中,JPTM 的已知类和新类的 $F1$ 得分比对比方法 DACS 高出 2% 左右,NA 值高出约 3%。综上分析可知,本文方法在分类性能上优于此对比方法。

表 8 MIXD 上不同方法分类性能对比

Table 8 Comparison of classification performance of different methods on MIXD

方法	已知类			新类			NA
	*F*1	*P*	*R*	*F*1	*P*	*R*	
DACS	0.942	0.931	0.956	0.942	**0.958**	0.927	0.958
APT	0.865	0.873	0.859	0.870	0.864	0.877	0.927
JPTM	**0.962**	**0.968**	**0.956**	**0.963**	0.957	**0.969**	**0.981**

与 APT 方法对比,JPTM 的 NA 值提升了 2%~5%,其中已知类的 $F1$ 得分提高了约 5%,新类的 $F1$ 得分提高了约 7%。在不同数据集上,APT 的方法已知类和新类的查全率 R 和查准率 P 都比 JPTM 低,说明 APT 的已知类和新类的分类效果都比本方法要差。而本文目的是为了实现新类检测,对比方法 APT 将已知类样本和要分类的对象放入一个集合中,根据测试对象的已知类邻居获得概率值来区分已知类和新类,对识别出的已知类打上伪标签再迭代计算概率,最后再利用随机森林分类器对属于已知类的样本细分类。但是 APT 仅确保了已知类的检测性能,在已知类和新类的二分类中,容易将一些低于阈值下的已知类别误分到新类中。而 JPTM 利用伪新类样本训练出双层分类器模型,利用下层分类器区分已知类和新类,并根据上层分类器的结果对已知类别细分类。相比于对比文献的方法,JPTM 双重决策可以有效避免一个指标误分的影响;同时,下层分类器不用阈值区分,能提供更灵活的决策边界。综上,本文方法在分类性能上优于对比方法。

为了验证实验方法的普适性,本文在不同数据组合情况下进行新类检测实验、结果见表 9。对于不同实验方法的同一组合保持共同的已知类样本,更换已知类的组合对不同的新类组合进行实验对比每个方法的平均准确性 NA 和 F1 得分。实验结果如图 7 所示。图 7 表示不同组合下不同实验方法的平均准确性,每个方法都有最大 NA 值对应的组合。但是本文方法 JPTM 的 NA 值的稳定性优于对比方法 APT 优于对比方法 DACS,在组合 4 下对比方法 APT 较差,在组合 4 和组合 9 下 DACS 的 NA 值达到了最差的效果,分类性能也最不稳定。

表 9 MIXD 上不同已知类和新类的组合

Table 9 Combination of different known and new classes on MIXD

组合	已知类 *	新类 *
1	1、5、6、7、8、16	3、9、10、11、13
2	1、2、3、8、13、18	6、7、10、11、12
3	6、7、8、9、10、12	1、2、3、11、13
4	4、5、7、8、13、16	3、11、12、15、18
5	1、2、3、5、10、11	6、7、9、12、18
6	1、3、6、7、9、11	5、8、12、13、16
7	2、6、9、10、11、18	5、7、12、13、15
8	1、5、7、8、10、11	2、6、13、16、18
9	2、7、9、15、16、18	1、4、10、11、13
10	5、6、9、15、16、18	2、3、11、12、13

* 注:数字标签与样本类别的对应关系:1-BitTorrent、2-Email、3-Facebook_audio、4-Facebook_chat、5-Ftp、6-Hangouts_audio、7-Skype_audio、8-Skype_file、9-Skype_video、10-YouTube、11-douyu(480 p)、12-huya(480 p)、13-tencent(480 p)、15-tencent(720 p)、16-youku(720 p)、18-douyu(1 080 p)

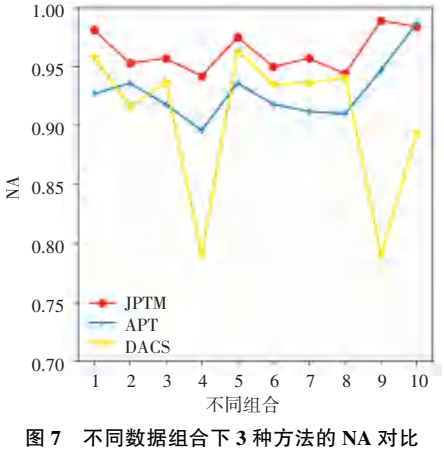

图 7 不同数据组合下 3 种方法的 NA 对比

Fig. 7 Comparison of NA of the three methods with different data combinations

不同数据组合下 3 种方法的平均 $F1$ 对比结果

如图 8 所示。由图 8 可以看出，JPTM 的稳定性优于对比方法 DACS。在 NA 不好的组合下，平均 $F1$ 得分也较差。由于 JPTM 采取双层分类器系统，分类边界更灵活，在不同的数据组合下的普适性更强。结合表 10 的实验结果，JPTM 的平均 NA 值和平均 $F1$ 得分都高于对比方法。由表 10 也可以发现 APT 方法的 NA 值高出 $F1$ 得分较多，因为 APT 的方法中利用阈值来区分已知类和新类，可能概率小的已知类样本会误分到新类中，概率大的新类被分到已知类中，多次迭代后误差放大，总已知类和新类互相混淆；利用传统闭集分类器随机森林来区分多种已知类，总已知类的准确性较好。而 NA 值计算的是总已知类和新类的准确性，其中任一方的值都会影响最终的 NA 值，所以 APT 方法的 NA 值高，已知类和新类的 $F1$ 得分较低。

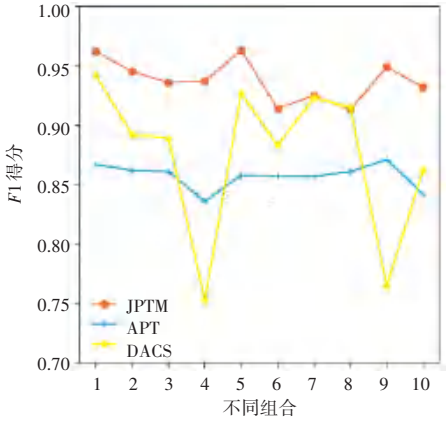

图 8　不同数据组合下 3 种方法的平均 $F1$ 对比

Fig. 8　Comparison of the average $F1$ of the three methods under different data combinations

表 10　不同方法平均分类性能对比

Table 10　Comparison of average classification performance of different methods　%

方法	平均 NA	平均 $F1$ 值
DACS	90.6	87.5
APT	92.9	85.7
JPTM	**96.3**	**93.8**

除了对比不同方法的分类性能外，本次研究还比较了不同方法的时间性能。不同方法分类时间对比见表 11。由表 11 可知，本方法的分类时间约为 0.156 3 ms/sample，时间复杂度最小，而 APT 的时间复杂度最高。分析发现为了获得测试对象属于已知类的概率要找出多个与当前对象最相似的样本，分类的计算复杂度和训练集中的样本数成正比，计算时间大大增加。DACS 先对已知类分组、再利用 2 个模型对，每个模型对包含一个随机森林分类器和一个 SVM 分类器，而本文使用两层分类器，下层分类器仅有一个二分类模型，时间复杂度小于 DACS 模型。在分类时间上，JPTM 的分类耗时约为对比方法 DACS 的 1/3，为对比方法 APT 的 1/6。综合来看，本文方法要优于对比方法。

表 11　不同方法分类时间对比

Table 11　Comparison of classification time of different methods

ms · sample^{-1}

方法	NYDS	ISCX	JPTM
DACS	0.472 0	0.489 2	0.472 1
APT	0.982 6	0.930 5	0.949 1
JPTM	**0.160 0**	**0.152 8**	**0.156 1**

4　结束语

在现实场景中，不仅要正确识别已知类，也要准确地将未知的类别检测出来。本文设计了一个新的开放集未知流量检测模型，基于置信度差和标签距离联合决策，通过双层分类器在保持已知类别分类精度的同时，增强了对未知类别流量的检测能力。首先，由已知类训练出上层分类器，接着从未标记样本中提取出伪新类样本，将其与已知类样本进行组合。借助上层分类器进行预测，并将计算的结果作为新的特征训练出下层分类器，通过拒绝不确定的预测筛选出未知的新类样本。在 2 个真实的数据集及组成的混合数据集上进行实验并与代表性的文献方法对比，本文方法的总体准确率都大于 95%，不仅在多个分类性能指标上均优于对比文献，所使用的分类时间也明显减少。

下一步工作是研究模型的增量更新能力，以便在动态变化的场景中持续识别出不同类别的未知流量。

参考文献

[1] 于治平，刘彩霞，刘树新，等. 基于机器学习的网络流量分类综述[J]. 信息工程大学学报，2023，24(4)：447-453.

[2] GENG Chuanxing，HUANG Shengjun，CHEN Songcan. Recent advances in open set recognition：A survey[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence，2021，43(10)：3614-3631.

[3] NEIRA C M A，RIBEIRO M J P，ROCHA A，et al. Data-fusion techniques for open-set recognition problems[J]. IEEE Access，2018，6：21242-21265.

[4] KRAWCZYK B，MINKU L L，GAMA J，et al. Ensemble learning for data stream analysis：A survey[J]. Information Fusion，2017，37：132-156.

[5] ZHAI Weixin，XIONG Xiya，MO Guozhao，et al. A Bagging-SVM field-road trajectory classification model based on feature

enhancement [J]. Computers and Electronics in Agriculture, 2024, 217: 108635.
[6] AL-OMARI M, RAWASHDEH M, QUTAISHAT F, et al. An intelligent tree-based intrusion detection model for cyber security [J]. Journal of Network and Systems Management, 2021, 29: 20.
[7] HAQUE A, KHAN L, BARON M. SAND: Semi-supervised adaptive novel class detection and classification over data stream [C]// Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI'16). Arizona: AAAI, 2016: 1652-1658.
[8] YU Yang, QU Weiyang, LI Nan, et al. Open category classification by adversarial sample generation [C]// Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI'17). New York: ACM, 2017: 3357-3363.
[9] NEAL L, OLSON M, FERN X, et al. Open set learning with counterfactual images [C]//Proceedings of the European Conference on Computer Vision (ECCV). Cham: Springer, 2018: 613-628.
[10] VERMA A, RANGA V. Machine Learning Based Intrusion Detection Systems for IoT Applications [J]. Wireless Personal Communications, 2020, 111(4): 2287-2310.
[11] JÚNIOR M P R, SOUZA D R M, WERNECK R D O, et al. Nearest neighbors distance ratio open-set classifier [J]. Machine Learning, 2017, 106(3): 359-386.
[12] LIANG Y, WANG F, CHEN S. DACS: A double-layer application classification scheme for hybrid zero-day traffic [C]// Proceedings of 2022 IEEE 22nd International Conference on Communication Technology (ICCT). Piscataway, NJ: IEEE, 2022: 1380-1387.
[13] CHEN Tianqi, GUESTRIN C. XGBoost: A scalable tree boosting system [C]//Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2016: 785-794.
[14] 项阳, 董育宁, 魏昕. 一种基于机器学习的网络流早期分类方法[J]. 南京邮电大学学报(自然科学版), 2022, 42(4): 96-104.
[15] PERNA G, MARKUDOVA D, TREVISAN M, et al. Online classification of RTC traffic [C]//Proceedings of 2021 IEEE 18th Annual Consumer Communications & Networking Conference (CCNC). Piscataway, NJ: IEEE, 2021: 1-6.
[16] LIN R, DONG C, MA P, et al. A fused multidimensional EEG classification method based on an extreme tree feature selection [J]. Computational Intelligence and Neuroscience, 2022, 2022: 7609196.
[17] GUI Jie, SUN Zhenan, WEN Yonggang, et al. A review on generative adversarial networks: Algorithms, theory, and applications [J]. IEEE Transactions on Knowledge and Data Engineering, 2023, 35(4): 3313-3332.
[18] ZHANG Junting, ZHANG Jie, GHOSH S, et al. Class-incremental learning via deep model consolidation [C]// Proceedings of 2020 IEEE Winter Conference on Applications of Computer Vision (WACV). Piscataway, NJ: IEEE, 2020: 1120-1129.
[19] CHO Y B. Diagnosis of communication security vulnerability of network printer using wireshark [J]. Journal of Digital Contents Society, 2020, 21(3): 601-607.
[20] UNB. Available dataset [EB/OL]. [2025-01-01]. https://www.unb.ca/cic/datasets/index.html.
[21] LIU Zhunga, HUI Xuemeng, FU Yimin. A new progressive open-set recognition method with adaptive probability threshold [J]. Chinese Journal of Aeronautics, 2023, 36(11): 297-310.