文章编号:2096-1472(2023)11-0043-05
DOI:10.19644/j.cnki.issn2096-1472.2023.011.010

一种基于信息熵的级联式新类识别方法

曾文玺,董育宁

(南京邮电大学通信与信息工程学院,江苏 南京 210003)
✉ 673553642@qq.com; 19900011@njupt.edu.cn

摘 要:针对传统机器学习在新类识别中存在分类精度较低和分类耗时较长的问题,提出了一种基于信息熵的级联式新类识别方法。利用随机森林的投票机制,计算并统计分析各样本的信息熵,作为新类检测的依据,识别已知类和候选新类样本;通过滤除候选新类中的异常流样本,提高分类准确率。实验表明:所提方法在南邮数据集和ISCX数据集的两个实际网络数据集上均能实现约95%的分类准确率,并且单个样本的分类时长仅需0.079 ms;分类精度和时间性能明显优于代表性文献方法。

关键词:网络流分类;新类检测;信息熵
中图分类号:TP391 **文献标志码**:A

A Cascaded Novel Class Recognition Method Based on Information Entropy

ZENG Wenxi, DONG Yuning

(*College of Telecommunications & Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing* 210003, *China*)
✉ 673553642@qq.com; 19900011@njupt.edu.cn

Abstract: Aiming at the shortcomings of traditional machine learning in novel class recognition, such as low classification accuracy and long classification time, this paper proposes a cascaded novel class recognition method based on information entropy. This method utilizes the voting mechanism of a Random Forest to calculate and analyze the information entropy of each sample. The entropy is used as a basis for novel class detection to identify known classes and candidate novel class samples. The classification accuracy is improved by filtering out abnormal flow samples in candidate novel classes. Experiments show that the proposed method can achieve a classification accuracy of about 95% on both actual network datasets of NJUPT Dataset (NDset) and ISCX Dataset, and the classification time for a single sample is only 0.079 ms. It is significantly superior to representative literature methods in classification accuracy and time performance.

Key words: network traffic classification; novel class detection; information entropy

0 引言(Introduction)

在常见的闭集假设中,传统机器学习(Machine Learning,ML)已取得显著的成效[1]。但是,现实场景已不再是简单的静态环境,这大大削弱了现有方法的鲁棒性,因此新类检测(Novel Class Detection,NCD)问题成为网络流分类的重要挑战之一。

针对开放环境的问题,目前ML中有一种解决方案是基于极值理论(Extreme Value Theory,EVT)[2]的方法。BALASUBRAMANIAN等[3]将EVT与ML中的随机森林(Random Forest,RF)相结合,基于每个已知类Weibull分布的累积概率识别新类。本文在南邮数据集和ISCX数据集两个数据集上进行了实验验证,分类精度只有85%左右,并且由于需要对不同

收稿日期:2023-03-20

的已知类类别分别进行拟合，并判断是否拒绝拟合，导致预测时间较长。

上述方法未能很好地解决 ML 中的 NCD 问题，其分类准确率有待提高且不满足在线分类的速度要求。因此，本文提出一种基于信息熵的级联式新类识别（Entropy based Cascade NCD，EntC-NCD）方法用于改善以上问题，并将其与现有代表方法进行了对比。

1 相关工作（Related work）

目前，针对 NCD 问题，研究人员从生成模型（Generative Model，GM）和判别模型（Discriminative Model，DM）两个不同的角度进行探索，并取得一定成果。现有的方法主要有基于距离、基于支持向量机（Support Vector Machine，SVM）和基于 EVT 的方法。

在基于距离的方法研究中，MU 等[4]基于孤立树异常检测算法[5]的思想提出了基于完全随机树的无监督学习算法（SENCForest）；武炜杰等[6]则是在 SENCForest 基础上融入了 k 近邻，不仅提高了在异常区域内搜索新类的准确率，也降低了系统开销。

基于 SVM 的方法是由 SCHEIRER 等[7-8]首次应用到 NCD 中，首先提出 1-vs-Set 模型，再进一步使用非线性内核融入 EVT，提出了基于 Weibull 校正的 SVM（W-SVM）模型；针对 W-SVM 中所有的已知类具有相同阈值的问题，JAIN 等[9]又引入了概率开放集 SVM（Probabilistic Open Set SVM，POS-SVM），该分类器可以对每个已知类采用不同的拒绝阈值，从而达到提高分类准确率的效果。

基于 EVT 拟合分布的方法如今被广泛使用，除了前文提到的 W-SVM；BALASUBRAMANIAN 等[3]则是提出了基于投票的极值理论模型（Vote-Based EVT，V-EVT），通过结合 RF 拟合已知类别样本的投票分布，得到逐类的 Weibull 分布。通过对应的 Weibull 分布计算其累积概率，根据阈值判断是否为已知类。

受 V-EVT 思路的启发，本文选择传统 ML 中分类效果较好的 RF 模型，与评估不确定性的信息熵相结合，提出基于信息熵的新类检测方法，想要达成的目标是在保证较高分类准确率的同时，克服需要多次计算 Weibull 累积概率导致分类耗时较长的问题。

2 本文方法（The proposed method）

基于信息熵和 RF 的 NCD 方法的模型框架如图 1 所示，主要分为训练、校准和测试三大模块。其中：训练集只包含已知类样本，校准集包含已知类和少量伪新类样本，测试集包含全部已知类和新类样本；训练集按照 3∶7 的比例随机分为 D_1 和 D_2 两个部分，D_1 训练多分类器 RF_1；θ 为新类判别阈值；β 为异常流样本置信度阈值。

2.1 基于信息熵的新类发现方法

RF 投票的分布中含有较多信息，投票的分散程度反映出分类器对样本的不确定性。当训练样本的类别 $c_i \in C_k = \{c_1, c_2, \cdots, c_n\}$ 时，若测试样本的类别 $c_i \notin C_k$，分类器对其判决的不确定性会远高于类别 $c_i \in C_k$ 的测试样本。据此引入信息熵作为评估不确定性的标准，并作为已知类和新类的分类依据。

为了验证这一想法，以 ISCX 数据集为例，随机抽取 7 个类

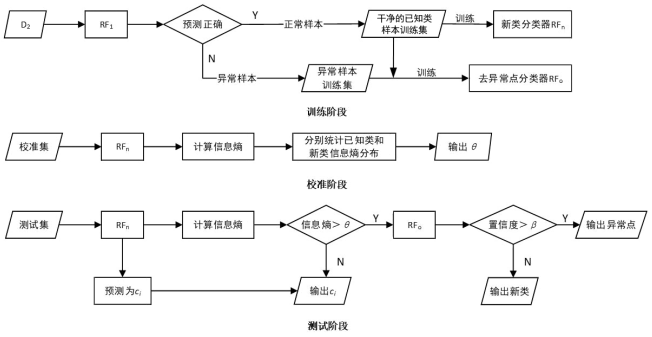

图 1 基于信息熵和 RF 的 NCD 方法的模型框架

Fig. 1 Model framework of NCD method based on information entropy and RF

作为已知类训练集和测试集，另外 3 个类作为新类测试集，分别测试并统计已知类和新类的信息熵分布[10]。

根据 RF 的投票结果计算样本信息熵的方法如下：首先将样本 d 判为已知类 c_i 的树的数目占树总数的比例作为样本 d 属于已知类 c_i 的概率，其次计算样本 d 被判为每个已知类的概率，并由此计算样本 d 的信息熵，计算已知类概率和信息熵的方法分别如公式（1）和公式（2）所示：

$$P(c_i \mid d) = \frac{1}{B}\sum_{b=1}^{B} I_b(c_i \mid d) \tag{1}$$

$$H_d = -\sum_{i=1}^{n} P(c_i \mid d)\log_2 P(c_i \mid d) \tag{2}$$

其中：$I_b(c_i \mid d) \in \{0,1\}$ 是第 b 棵树判断样本 d 是否为类 c_i 的结果，若判为 c_i，则设为 1，否则为 0；B 为 RF 中树的总数目，n 为已知类的类别数。

ISCX 数据集的信息熵分布统计结果如图 2 所示。已知类的信息熵值明显聚集于小于 1 的区域内，而新类的信息熵则普遍较大，这为基于信息熵的新类检测提供了可行性。

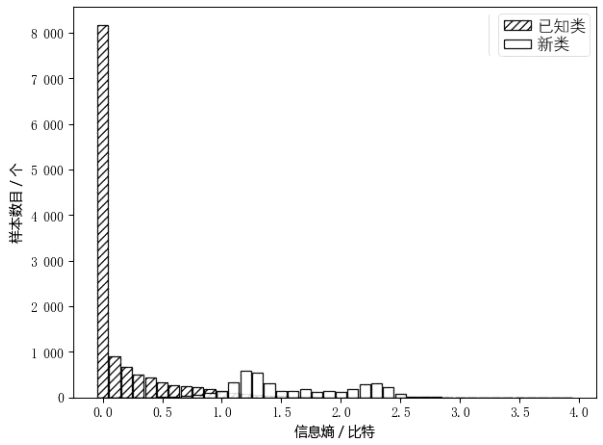

图 2 已知类和新类信息熵分布统计

Fig. 2 Information entropy distribution statistics for known and novel classes

2.2 去除异常流样本

在实际网络的流传输过程中会产生异常流样本，从而降低分类器学习的准确性。因此，训练前需筛选出训练集中的异常样本，具体步骤如表 1 中的算法 1 所示；得到干净的已知类样本训练集 D_t 和异常样本数据集 D_o，并用 D_t 训练新类分类器 RF_n。

表1　去除异常流样本算法

Tab.1 Algorithm for removing abnormal stream samples

算法 1:去除异常流样本

输入:
　D_t:已知类样本训练集
输出:
　D_t:干净的已知类样本训练集
　D_o:异常样本训练集
1:$D_o=\phi$
2:D_t 按 3∶7 随机平均分为 D_1、D_2
3:D_1 训练 RF 分类器 RF_1
4:for d in D_2 do
5:RF_1 预测 d 结果为 $\hat{y}$
6:d 的真实标签为 y_d
7:if $\hat{y}\neq y_d$
8:$D_t=D_t\setminus\{d\}$
9:$D_o=D_o\cup\{d\}$
10:end if
11:end for
12:return D_t、D_o

测试集中同样会存在异常已知类样本,因此分类器对其判定的不确定性会增大,使该样本的信息熵增大,容易被误判为新类。

为此,从 D_t 中抽取与 D_o 数量相等的样本集 D_p,D_o 和 D_p 分别作为正、负样本训练去异常点二分类器 RF_o。测试阶段通过级联 RF_o,对 RF_n 认定的新类样本进行再分类,删除其中的异常已知类样本。

2.3　确定新类判别阈值

依据校准集选取新类的判别阈值,校准数据集 D_v 中包括全部已知类和少量伪新类的样本;用 RF_n 进行预测,计算各个样本的信息熵,并以 0.1 为区间分别统计已知类和新类的信息熵分布,取两个分布的交点作为新类判别阈值 θ,具体过程表 2 中的算法 2 所示。

表2　确定分类阈值算法

Tab.2 Algorithm for determining the classification threshold

算法 2:确定分类阈值

输入:
　RF_n:由 D_t 训练的新类分类器
　D_v:包含已知类和少量伪新类的校准集
输出:
　θ:新类分类阈值
1:$K_{hi}=0$,$U_{hi}=0$
2:for each d in D_v do
3:用 RF_n 预测 d
4:$H_d=-\sum_i^n P(c_i|d)\log_2 P(c_i|d)$
5:$K_{hi}=K_{hi}+I(h_i,C_k|d)$
6:$U_{hi}=U_{hi}+I(h_i,C_u|d)$
7:end for
8:$h_\theta=\text{argmax}(\min(K_{h0},U_{h0}),\min(K_{h1},U_{h1}),\cdots)$
9:return θ

其中:h_i 表示$[i-0.05,i+0.05)$;K_{hi}、U_{hi}分别表示已知类和新类样本的信息熵在 h_i 区间内的样本数量;C_k、C_u分别表示已知类、新类;$I(h_i,C_k|d)\in\{0,1\}$表示若 $d\in C_k$ 且 $H_d\in h_i$,则 $I(h_i,C_k|d)$等于 1,否则为 0。

2.4　分类模型

如上文所述,测试集中异常样本的信息熵比正常样本高,导致误判为新类。因此,采用级联模式进行二次筛选。经过 RF_n 分类后,信息熵小于等于 θ 的样本被认定为已知类,并直接输出 RF_n 的分类结果;而信息熵大于 θ 的样本,称其为候选新类(包含新类和已知类中的异常样本)。

对于候选新类样本通过级联的去异常点二分类器 RF_o 进一步判断,并引入异常置信度 $ACon$,计算公式如下:

$$ACon(C_o\mid d)=\frac{1}{B}\sum_{b=1}^{B}I_b(C_o\mid d)\tag{3}$$

其中:C_o 表示异常类;$I_b(C_o|d)\in\{0,1\}$表示若第 b 棵树判断样本 $d\in C_o$,则 $I_b(C_o|d)$等于 1,否则为 0。

同时,引入异常置信度阈值 β 用于判断,对于异常置信度大于阈值 β 的样本,判为异常点,从候选新类中删除,反之则判为新类。本文方法完整的测试过程表 3 中的算法 3 所示。

表3　新类-异常样本检测算法

Tab.3 Algorithm for new class-abnormal sample detection

算法 3:新类-异常样本检测

输入:
　RF_n:由 D_t 训练的新类分类器
　RF_o:由 D_o 和 D_p 训练的去异常点二分类器
　D_s:包含全部已知类和新类的测试集
　θ:新类判别阈值
　β:异常置信度阈值
输出:
　$\hat{Y}$:测试样本分类结果
1:for each d in D_s do
2:用 RF_n 预测 d,预测结果为 $\hat{y}$
3:$H_d=-\sum_i^n P(c_i|d)\log_2 P(c_i|d)$
4:if $H_d\geqslant\theta$
5:用 RF_o 预测 d
6:$ACon(C_o|d)=\frac{1}{B}\sum_{b=1}^{B}I_b(C_o|d)$
7:if $ACon<\beta$:$\hat{y}=y_u$
8:else:$\hat{y}=y_o$
9:end if
10:end if
11:$\hat{Y}=\hat{Y}\cup\{\hat{y}\}$
12:end for
13:return $\hat{Y}$

其中:θ 为算法 2 中获取的新类判别阈值,β 为异常置信度阈值,可以灵活调节以平衡分类的准确率和覆盖率;H_d 为根据多分类器分类结果计算的信息熵;$ACon(C_o|d)$为根据 RF_o 得到的异常置信度;y_u 和 y_o 分别表示预测标签为新类和异常点。

3　实验(Experiment)

3.1　实验环境

实验使用惠普笔记本电脑,硬件和软件的配置如下:CPU 为 AMD R5-4600H@3.00 GHz,GPU 为 NVIDIA GTX 1650 Ti-4G,16 GB 运存,操作系统为 64 位 Windows 10;在 Python 编程语言环境下运行。

分类器均采用 RF,树的数目设置为 100 棵,叶节点最小样本数设置为 1 个,所有实验采用五折交叉验证。

3.2 评估指标

3.2.1 新类分类指标

采用分类准确率 A_o 作为分类准确性指标，定义如下：

$$A_o=\frac{\sum_{i=1}^{n}(TP_i+TN_i)+TU}{\sum_{i=1}^{n}(TP_i+TN_i+FP_i+FN_i)+TU+FU} \tag{4}$$

其中：TP_i、TN_i、FP_i、FN_i 分别代表已知类的真阳性、真阴性、假阳性、假阴性，TU、FU 分别代表新类的正确判断和错误判断，n 为已知类类别数目。

采用 F_1 值作为评估指标，由精确率 P 和召回率 R 计算得出，计算公式如下：

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

需要注意，计算 F_1 时未将新类作为一个额外的样本类加入计算，因为在分类器中，没有新类的训练样本，所以将新类作为一个真阳性分类没有意义。但是，在计算已知类的 P 和 R 时，FP 和 FN 中也会包含错误分类的新类样本。

3.2.2 滤除异常样本指标

本文方法包含从候选新类样本中过滤异常样本的模块，准确率仍然使用 A_o，但是样本总数减少。因此，定义覆盖率指标 *Coverage* 如下所示：

$$Coverage=1-\frac{N_n}{N} \tag{6}$$

其中：N 表示预测样本总数，N_n 表示判为异常样本的数目。

定义 *ORR*(Outlier Removal Rate)表示已知类异常样本的滤除率、*FDR*(False Deletion Rate)表示新类样本被判为异常点的比例。

3.2.3 时间性能指标

分别用 T_t 和 T_c 表示训练时间和分类时间，单位为 ms/样本，分别表示逐样本的平均训练时间和分类时间。

3.3 数据集

使用南邮数据集(NJUPT Dataset，NDset)、ISCX 数据集进行方法验证。NDset 是通过 WireShark 于 2020 年在南京邮电大学校园网环境下采集的[11]。NDset 和 ISCX 数据集的具体类别和样本数如表 4 和表 5 所示。

表4 南邮数据集

Tab.4 NDset

类别	样本数/个
480P_bilibili	3 120
480P_douyu	3 149
480P_huya	3 098
720P_huya	3 129
720P_tencent	3 120
1080P_bilibli	3 128
1080P_douyu	3 102
1080P_huya	3 134
1080P_tencent	3 126
kougou_music	3 111
QQ_music	3 101

表5 ISCX 数据集

Tab.5 ISCX Dataset

类别	样本数/个
BitTorrent	3 178
Facebook_audio	3 016
Facebook_chat	3 065
Ftp	3 016
Hangouts_audio	3 175
Skype_audio	3 065
Skype_file	3 165
Skype_video	3 177
VoipBuster	3 075
YouTube	3 061

3.4 不同置信度异常阈值对比

为了验证级联式去除异常样本模块的有效性，以 NDset 为例，新类类别选取为[1080P_douyu，1080P_huya，720P_tencent，QQ_music]共 4 类，校准数据集 D_v 选取的伪新类为 1080P_huya。通过修改阈值 β 对比去除异常点前后的各项评估指标的变化，结果如表 6 所示，$\beta=1$ 表示未做去除异常点处理。

表6 不同 β 的性能对比

Tab.6 Performance comparison of different β

β	A_o	F_1	*Coverage*	*R*	*P*	*ORR*	*FDR*
1	0.891	0.981	1	0.986	0.976	0	0
0.8	0.924	0.981	0.948	0.986	0.976	0.371	0.360
0.6	0.943	0.981	0.899	0.986	0.976	0.606	0.940
0.5	0.95	0.981	0.854	0.986	0.976	0.663	0.185
0.4	0.967	0.981	0.722	0.986	0.976	0.852	0.445

在未进行去除异常点的情况下，6 330 个已知类测试样本中有 1 133 个被新类识别模块判为候选新类，约占所有已知类测试样本的 17.9%，而 4 910 个新类测试样本被判断为候选新类的个数为 4 856 个，约占比 98.9%。级联去异常点模块后，β 使用 0.5 时，会有 66.3%的已知类异常样本被删除，而新类中有 18.5%的样本被当作异常样本被误删。表 6 中的数据表明，去异常点模块能从候选新类样本中删除大部分的已知类异常样本，并且保留大多数新类样本，进一步提高新类样本的纯度，并且可以根据需要自行调节阈值。需要注意，F_1 没有跟随阈值变化是因为 R 和 P 的计算中未包含判为候选新类的样本。

3.5 不同新类分类阈值对比

根据本文提出的算法 2，计算得到一个新类分类的阈值，会对于分类的最终性能有着较强的影响，因此设置实验通过修改 θ 值进行对比，验证其有效性。新类和校准集选取同本文“3.4”小节，根据算法 2 得到阈值 θ 为 0.9，阈值 β 统一设置为 0.5，不同 θ 的性能对比结果如表 7 所示。

表7 不同 θ 的性能对比

Tab.7 Performance comparison of different θ

θ	A_o	F_1	*Coverage*	*R*	*P*	*ORR*	*FDR*
1.1	0.940	0.957	0.878	0.971	0.943	0.696	0.177
0.9	0.950	0.981	0.865	0.986	0.976	0.663	0.157
0.7	0.952	0.987	0.842	0.987	0.985	0.662	0.178

当θ取0.9时,覆盖率比θ取1.1时小1.3%,但准确率高1%,F_1值也高2.4%;而相比于θ取0.7时,准确率几乎一样,但覆盖率高2.3%,只有F_1值低0.6%且θ取0.9时,对新类样本的误删率最低。因此,由算法2计算的阈值θ的分类性能较好。

3.6 不同方法的性能对比

将本文方法EntC-NCD与文献方法V-EVT分别在NDset和ISCX两个数据集上进行实验对比,采用本文所提方法进行去异常点处理时,阈值β分别设置为0.5、0.8,结果如表8和表9所示,EntC-NCD-1表示未做去异常点处理。

表8　不同分类方法在NDset上的对比结果

Tab.8 Comparison of different classification methods on NDset

方法	A_o	F_1	Coverage	R	P	ORR	FDR
EntC-NCD-1	0.891	0.981	1	0.986	0.976	—	—
EntC-NCD	0.950	0.981	0.854	0.986	0.976	0.663	0.185
V-EVT	0.876	0.968	1	0.990	0.948	—	—

表9　不同分类方法在ISCX数据集上的对比结果

Tab.9 Comparison of different classification methods on ISCX Dataset

方法	A_o	F_1	Coverage	R	P	ORR	FDR
EntC-NCD-1	0.925	0.967	1	0.976	0.959	—	—
EntC-NCD	0.946	0.967	0.844	0.976	0.959	0.512	0.251
V-EVT	0.899	0.968	1	0.988	0.949	—	—

在两个数据集上,EntC-NCD-1比V-EVT的A_o高1.5%~2.6%;F_1则是在ISCX数据集上两者相似,在NDset上是本文所提方法较优;EntC-NCD通过去除异常点处理,进一步提高了分类准确率,其A_o高于V-EVT方法4.7%~7.4%。V-EVT是通过RF投票数分布拟合每个已知类的Weibull分布,再通过计算测试样本的累积概率判断是否属于该类;若不属于所有已知类,则判为新类。但是,实际的拟合结果并不完全贴合实际投票的分布,导致V-EVT的分类性能不如本文所提方法。

在不同数据集上的时间性能对比结果如表10所示。EntC-NCD只需要进行一次多分类并计算一次信息熵,预测时间较短,在NDset上,即使加上去异常点处理,平均一个样本也仅需0.079 ms;V-EVT虽然只需要进行一次分类器分类,但是需要分别计算每一个已知类的Weibull分布值进行判断,所以需要0.592 ms,分类时间仍较本文所提方法高一个数量级。

在训练时间上,EntC-NCD需要多训练一个去异常点分类器,V-EVT则是需要拟合每一个已知类的Weibull分布,训练耗时相差不大。

表10　不同分类方法的时间性能对比结果

Tab.10 Comparison of temporal performance of different classification methods

方法	T_t/(ms·样本$^{-1}$)		T_c/(ms·样本$^{-1}$)	
	NDset	ISCX	NDset	ISCX
EntC-NCD	0.355	0.371	0.079	0.078
V-EVT	0.458	0.509	0.592	0.456

综上所述,相比于V-EVT,本文方法在不同的数据集上均有更好的表现,同时具有一定的普适性。

4 结论(Conclusion)

本文提出了一种基于信息熵的级联式新类识别和去异常点模型,并针对新类分类阈值的选取给出了优选方法。此外,本文还讨论了不同新类判别阈值、异常置信度阈值对分类性能的影响,在两个真实的网络数据集上对本文所提方法进行验证,并与文献方法进行对比。实验数据表明,本文所提方法的识别准确率均可达到约95%,单个样本的识别时间仅需0.079 ms,在分类精度和时间性能上均优于对比方法且有一定的普适性,更加适用于不同需求的新类分类场景。

参考文献(References)

[1] GENG C X, HUANG S J, CHEN S C. Recent advances in open set recognition: a survey[J]. IEEE Transactions on Pattern Analys. is and Machine Intelligence, 2021, 43(10): 3614-3631.

[2] 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.

[3] BALASUBRAMANIAN L, KRUBER F, BOTSCH M, et al. Open-set recognition based on the combination of deep learning and ensemble method for detecting unknown traffic scenarios[C]//IEEE. Proceeding of the 2021 IEEE Intelligent Vehicles Symposium. New York: IEEE, 2021: 674-681.

[4] MU X, TING K M, ZHOU Z H. Classification under streaming emerging new classes: a solution using completely-random trees[J]. IEEE Transactions on Knowledge and Data Engineering, 2017, 29(8): 1605-1618.

[5] LIU F T, TING K M, ZHOU Z H. Isolation-based anomaly detection[J]. ACM Transactions on Knowledge Discovery from Data, 2012, 6(1): 3.

[6] 武炜杰,张景祥. 有新类的动态数据流分类算法研究[J]. 计算机科学与探索,2021,15(1):132-140.

[7] 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.

[8] SCHEIRER W J, JAIN L P, BOULT T E. Probability models for open set recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 36(11): 2317-2324.

[9] JAIN L P, SCHEIRER W J, BOULT T E. Multi-class open set recognition using probability of inclusion[C]// Springer. Proceeding of the Computer Vision-ECCV 2014: 13th European Conference. Zurich: Springer, 2014: 393-409.

[10] DRAPER-GIL G, LASHKARI A H, MAMUN M S I, et al. Characterization of encrypted and vpn traffic using time-related[C]//Springer. Proceedings of the 2nd international conference on information systems security and privacy (ICISSP). Bucharest: Spring, 2016: 407-414.

[11] 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.

作者简介:

曾文玺(1998-),男,硕士生。研究领域:网络流分类与识别。

董育宁(1955-),男,博士,教授。研究领域:网络流分类与识别,图像和视频信息处理。本文通信作者。