

赵峰, 董育宁, 邱晓晖. 一种增量更新模型的新类检测方法[J]. 智能计算机与应用, 2024, 14(4): 202-208. DOI: 10.20169/j.issn.2095-2163.240433

一种增量更新模型的新类检测方法

赵峰, 董育宁, 邱晓晖

(南京邮电大学通信与信息工程学院, 南京 210003)

摘要: 开集流识别网络流量分类是网络管理的重要组成部分。为了适应变化的网络环境, 已有许多研究瞄准开集流识别, 但这些方法不能以增量方式更新模型。针对这一问题, 本文提出了一种增量更新级联结构, 通过筛选的模拟新类, 利用置信度阈值进行新类检测, 采用分类器级联的方式逐步包含新出现的类; 当级联分类器个数达到设定值时, 重新训练多分类器, 以此减少分类时间。使用真实数据集对所提方法进行验证, 并与代表性文献方法进行对比。结果表明, 在分类性能上, 本文方法已知类 $F1$ 和综合性能指标 NA 均能达到 0.9 以上; 在时间性能上, 分类时间和平均模型更新时间显著减少, 均优于代表性文献方法, 有利于实现快速在线新类检测与分类。

关键词: 开集流识别; 新类检测; 模拟新类; 增量更新

中图分类号: TP391

文献标志码: A

文章编号: 2095-2163(2024)04-0202-07

A new class detection method for incremental model updating

ZHAO Feng, DONG Yuning, QIU Xiaohui

(School of Communications and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China)

Abstract: Open set flow identification network traffic classification is an important part of network management. In order to adapt to the changing network environment, there have been many researches aiming at open-set flow recognition, but these methods cannot update the model in an incremental way. To address this problem, this paper proposes an incremental update cascade structure, which reduces the classification time by simulating new classes through screening, using confidence thresholds for new class detection, and using a cascade of classifiers to progressively include emerging classes; and re-training multiple classifiers when the number of cascaded classifiers reaches a set value. The proposed method is validated using real datasets and compared with representative literature methods. The results show that, in terms of classification performance, the known class $F1$ and the comprehensive performance index NA of this method can reach more than 0.9; in terms of time performance, the classification time and the average model update time are significantly reduced, which are both better than the representative literature methods, and are conducive to the realization of fast online new class detection and classification.

Key words: open set flow recognition; new class detection; simulate new class; incremental update

0 引言

网络流量分类 (Network Traffic Classification, NTC) 在服务质量 (Quality of Service, QoS) 保证、网络安全、流量趋势分析等方面发挥着关键作用^[1-2]。随着机器学习 (Machine Learning, ML) 和深度学习 (Deep Learning, DL) 技术的不断进步, 网络流量分类取得了较高的分类精度^[3-4]。但在现实 NTC 问题

中, 随着环境的变化, 可能出现新类, 如果使用训练好的分类器对新类样本进行分类, 预测精度会严重下降^[5]。这种在测试过程中存在新类的问题, 即开集流识别 (Open Set Flow Recognition, OSFR) 问题^[6]。

对于 OSFR, 极限值理论 (Extreme Value Theory, EVT) 被用来对决策边界处的正训练样本进行建模, 如果已知类数据被准确建模就可以拒绝新类^[7]。文献[8]提出了 Vote-based EVT (V-EVT) 方法, 通

作者简介: 赵峰 (1998-), 男, 硕士研究生, 主要研究方向: 网络流量分类; 邱晓晖 (1968-), 女, 博士, 教授, 主要研究方向: 智能信号处理、图像处理与模式识别。

通讯作者: 董育宁 (1955-), 男, 博士, 教授, 主要研究方向: 无线通信网络, 图像处理。Email: 19900011@njupt.edu.cn

收稿日期: 2023-03-21

过卷积神经网络 (Convolutional Neural Networks, CNN) 和随机森林模型 (Random Forest Classifier, RFC), 将投票模式与 EVT 相结合, 用于拟合威布尔分布 (Weibull Distribution), 通过阈值来拒绝新类样本, 但在实验测试中发现正样本数据并不能完全拟合 Weibull, 已知类只有 78% 左右的召回率。

目前, 已有大量方法被探索用于 OSFR, 这些方法可以分成基于判别式模型方法和基于生成式模型方法^[9-10]。虽然这些工作可以识别未知类, 但不能对随后出现的不同新类实例快速更新模型, 主要难点是:

(1) 在检测新类样本的同时, 如何将已知类样本准确的细分类;

(2) 如何在不重新训练模型的情况下, 采用增量方式逐步应对不同的新类实例, 从而减少模型更新时间。

本文提出一种增量更新级联方法 (Incremental Update Cascade Structure Model, IUCS)。在训练阶段, 根据已知类置信度分布筛选模拟新类; 测试阶段, 构建 RFC 级联模型, 在线分类并增量更新模型。实验表明, 该方法的分类精度和时间性能都优于代表性文献方法。

1 相关工作

支持向量机 (Support Vector Machine, SVM) 通过控制最优分类超平面两侧间隔来控制分类器性能, 文献[11]提出 1-vs-Set 方法, 在建模中加入了开放空间风险项, 以应对超出已知类支持范围的空间。但是一些已知类所占空间是无限的, 开放空间风险仍然存在; 文献[12]将非线性纳入解决方案, 制定了一个紧凑衰减概率 (Compact Abating Probability) 模型, 进一步限制开放空间风险。在此基础上, 威布尔校准支持向量机方法 (Weibull-calibrated SVM, W-SVM) 和基于概率的 SVM 方法 (P-SVM) 通过阈值的分类方案有效限制了开放空间风险^[12-13]。除此以外还有一些传统机器学习方法应用在 OSFR 问题上, 例如基于稀疏表示的开放集识别模型 (Sparse Representation-based Open Set Recognition, SROSR), 开放集版本的最近邻分类器 (Open Set Version of Nearest Neighbor Classifier, OSNN) 等^[14-15]。

然而上述方法并不能随着类数量增加而扩展。类增量学习 (Class-Incremental Learning, C-IL) 是增量学习的一个分支, 加强训练过的分类器来处理

新类^[16]。文献[17]提出了一种集成学习方法, 通过动态地调节分类器的权重来实现类别增量学习; 文献[18]提出利用未标记数据学习新类 (Learning with Augmented Class with Unlabeled data, LACU) 框架的 LACU-SVM 半监督学习算法。一些 C-IL 算法使用聚类方法来检测新类, 但无监督学习方法的新类检测性能和分类精确度并不理想, 文献[19]基于孤立森林 (Isolation Forest, IForest) 提出基于完全随机树的无监督学习算法 SENCForest (Classification under Streaming Emerging New Class)。

上述方法在 OSFR 问题上取得了成效, 但在模型更新问题上存在不足。本文利用模拟新类和分类置信度阈值有效检测新类, 并通过级联二分类器实现模型增量更新, 能在一定程度上解决上述问题。

2 本文方法

2.1 模型框架

本文方法框架如图 1 所示。主要的设计思路如下: 为了提高分类性能, 先做特征提取 (Feature Extraction, FE), 再进行特征选择 (Feature Selection, FS)。模型训练阶段, 利用无标签数据集, 根据样本置信度筛选模拟新类样本, 用于二分类器的训练。在线上分类模块使用多分类器完成已知类细分类。同时, 利用时变二分类器级联结构逐次应对顺序出现的不同新类, 实现模型增量更新。

2.2 特征提取与特征选择

为了做到在线流分类, 将采集到的流数据按照 1 s 划分为流段, 使用每条流 (段) 的前 10 个数据包计算特征。在传统网络流统计特征的基础上, 加入条件频度和速率序列特征可以提高分类的总体准确度^[20]。

(1) 特征提取

将流数据转换成包大小序列、包达到时间序列、时间戳序列、包差值序列、上行速率序列和下行速率序列组成的五元组, 再用五元组计算出 102 个统计特征, 并且计算出 25 个下行条件频率特征和 4 个上行条件频率特征, 共提取 131 个流特征。

(2) 特征选择

首先, 计算特征与标签之间的皮尔森相关系数 (Pearson Correlation Coefficient, PCC), 赋予每个特征一个重要性分数, 然后计算所有特征两两之间的 PCC, 删除 PCC 大于 0.9 的特征对中与标签相关性小的特征, 通过极限树对剩余特征进行重要性排序, 最后通过 RFC 观察逐个增加特征时分类准确率的变化, 寻找性能拐点, 获得最优特征子集。

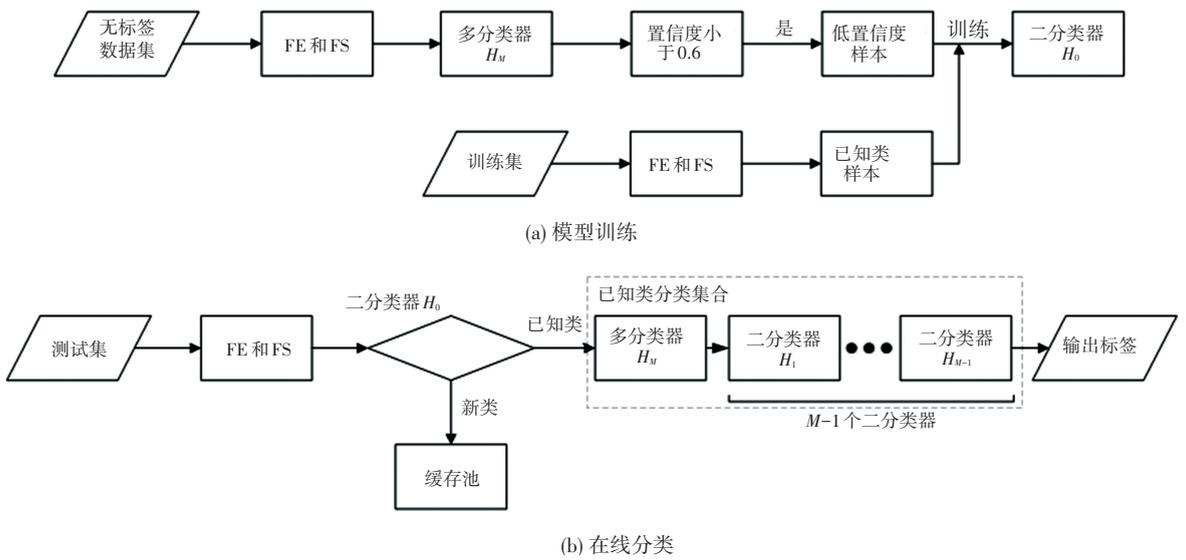


图1 本文方法框架

Fig. 1 Method framework of this paper

2.3 模拟新类筛选

在RFC中,将样本 x 判为类别 c_i 的概率定义为投票给 c_i 类树的数目占总数目的比例,即式(1):

$$P(c_i | x_j, \theta_r) = \frac{1}{B} \sum_{b=1}^B I_b(c_i | x_j, \theta_r) \quad (1)$$

其中, $I_b(c_i | x_j, \theta_r)$ 为第 b 棵树判断样本 x 是否属于 c_i 类,1 表示属于,0 相反; θ_r 为RFC参数; B 为RFC中树的总数。

一个样本的置信度 (Confidence Measure, CM) 可以定义为式(2):

$$CM = \text{Max}_{i=1, \dots, n} P(c_i | x_j, \theta_r) \quad (2)$$

也就是说,对一个数据样本,如果投票给 c_i 类的树的数目最大,就判为 c_i 类。

因为RFC只用已知类进行训练,如果一个样本不属于任何已知类,那么CM值较小,本文使用恶意流量数据集来验证这一想法。选出5类作为已知类,5类作为新类,每个类200个样本,置信度分布直方图如图2所示。从图2可以看出,接近1000个已知类样本在0.9~1.0的置信度区间,新类样本的CM明显较小。

每类样本的置信度均值如图3所示。由于RFC是在1-5类已知类上训练的,投票给1-5类的平均CM非常大,几乎接近1;相比之下投票给新类6-10的CM平均接近0.6,由此可以采用这样的模拟新类样本在训练时用。

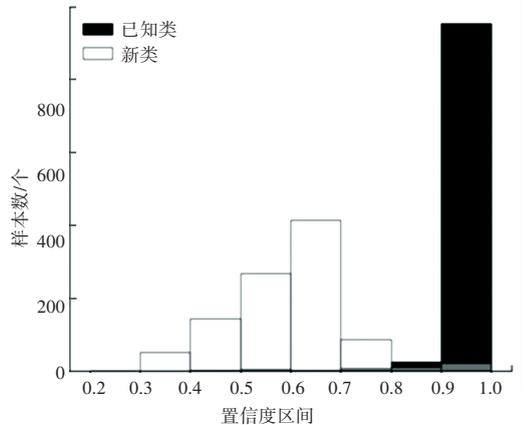


图2 置信度分布直方图

Fig. 2 Histogram of confidence measure distribution

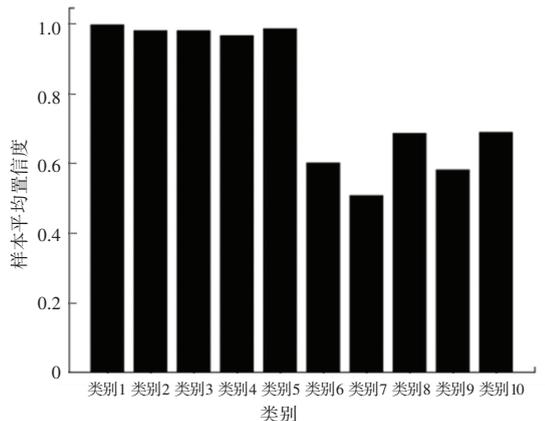


图3 样本置信度均值

Fig. 3 The mean of the sample confidence

据此得出模拟新类样本筛选的主要思路:首先,训练已知类多分类器 H_M ;其次,利用无标签数据,从中计算样本 CM 值,筛选出 CM 小于 0.6 的作为模拟新类。用筛选出的模拟新类样本和已知类样本训练新类-已知类二分类器 H_0 。

2.4 分类模型

线上分类模块采用级联模型,先利用 H_0 检测出新类,再对剩下的已知类进行细分类。 H_0 计算每个样本的置信度,将低 CM 样本和判为模拟新类的高 CM 认为是新类,并放入缓存池中。已知类细分类过程如算法 1 所示。

算法 1 已知类分类过程

Input

$rfcs$ - 已知类分类器集合
 x - 测试集样本
 s - 缓存池 β 大小
 th - 分类阈值

Output

y -测试集样本预测标签

```

1  for each  $x$  do:
2    for classifier in  $rfcs$ :
3      flag  $\leftarrow$  false
4      calculate  $CM$  of the sample  $x$ 
5      if  $CM > th$ :
6         $y \leftarrow$  classifier ( $x$ )
7        flag  $\leftarrow$  true
8        break
9    end for
10   if flag == false:
11      $\beta \leftarrow \beta \cup \{x\}$ 
12     if  $|\beta| > s$ :
13       updateMode ( $\beta$ )
14        $\beta \leftarrow$  null
15   end for
16    $y \in \{C_1, C_2, C_3, \dots, C_n, newclass\}$ 

```

2.5 模型更新

当缓存池快满时,进行模型更新。在不重新训练模型的情况下,采用增量方式逐步包含顺序出现的类,从而减少模型更新时间。模型更新主要包括两个过程:

- (1)重新训练新类-已知类二分类器 H_0 ;
- (2)训练并级联新的二分类器 H_1 等。

随着级联二分类器数量增多,分类时间也随之增加。当级联二分类器个数达到设定值 (M) 时,用

重新训练的多分类器 H_M 代替级联的多个二分类器。具体过程如算法 2 所示,模型更新过程如图 4 所示。

算法 2 模型更新过程

Input

$rfcs$ - 已知类分类器集合
 $set1$ -已知类训练集
 $buffer \beta$ -缓存池
 $set2$ -已知类标签集
 H_0 -新类-已知类二分类器
 nth - 模型更新阈值

```

1  generate a new label  $l$ 
2  label the data in buffer  $\beta$  with label  $l$ 
3   $set2 \leftarrow set2 \cup \{l\}$ 
4  size(current known class)  $n$  is size( $set2$ )
5  if size( $rfcs$ ) <  $nth$ :
6    train a binary classifier  $rfc$  with data in
7    buffer  $\beta$  and some samples in  $set1$ 
8     $rfcs \leftarrow rfcs \cup \{rfc\}$ 
9     $set1 \leftarrow set1 \cup \beta$ 
10   retrain  $H_0$  with samples in  $set1$ 
11  else
12    $rfcs \leftarrow$  null
13    $set1 \leftarrow set1 \cup \beta$ 
14   train a  $n$  Classifier  $rfc$  with  $set1$ 
15    $rfcs \leftarrow rfcs \cup \{rfc\}$ 
16   retrain  $rfc1$  with samples in  $set1$ 
17    $\beta \leftarrow$  null

```

3 实验结果与讨论

将本文方法 IUCS 与文献方法 V-EVT^[8] 进行性能对比。

3.1 数据集

实验在 ISCX 数据集(简称 ISCX)上进行。ISCX 数据集是一个公共数据集,共 12 类数据,具体信息见表 1。

表 1 ISCX 具体信息

Table 1 Specific information about ISCX

流类型	应用	样本数量
文件传输	bitTorrent,ftp,skype	2 000 * 3
语音通话	facebook,hangouts,skype	2 000 * 3
视频	skype,youtube,netflix	2 000 * 3
IP 语音	voipbuster	2 000
聊天	facebook	2 000
邮件	email	2 000

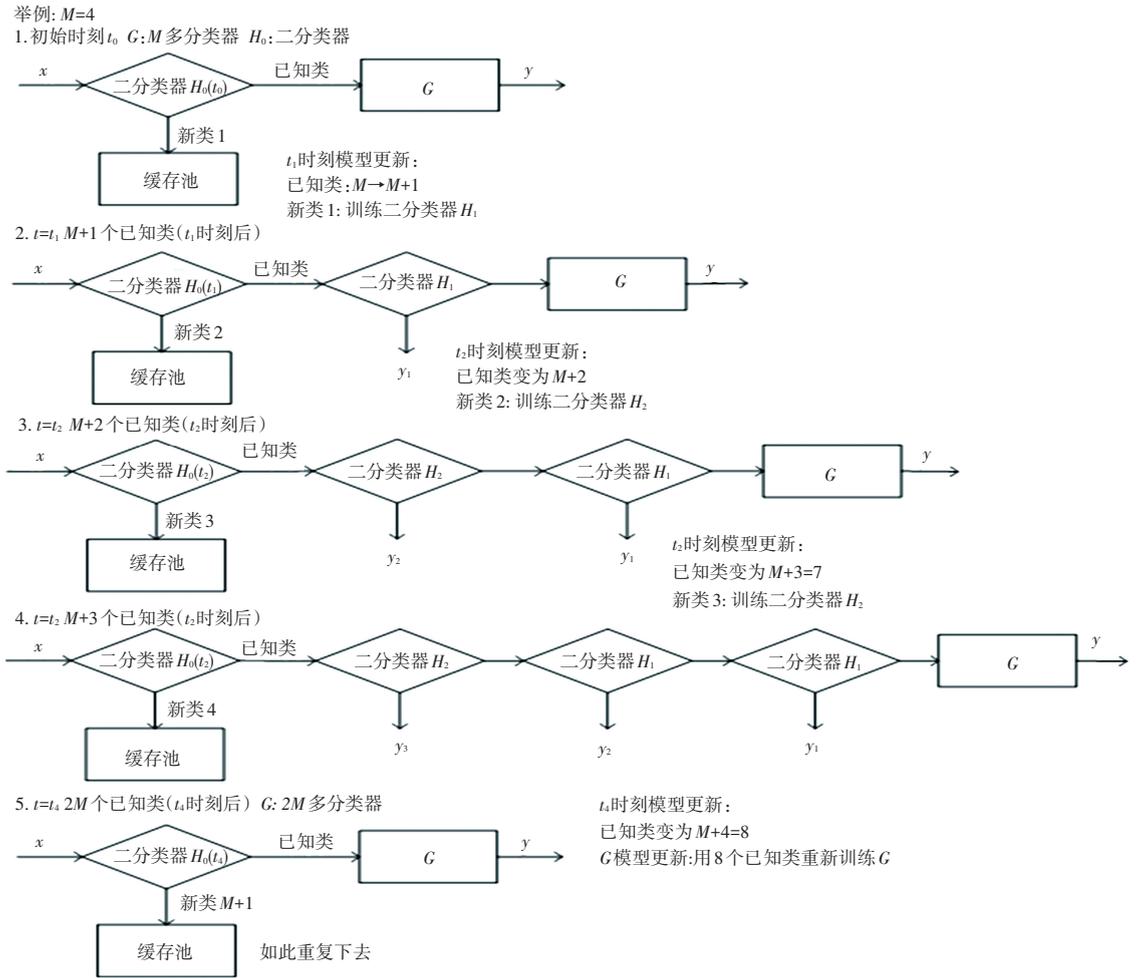


图4 模型更新示意图

Fig. 4 Model updating diagram

3.2 评价指标

在评估过程中,既要考虑分类性能,也要对时间效率进行衡量。分类性能采用 4 种评价指标,分别是查准率(P)、查全率(R)、 $F1$ 分数($F1_score$) 和归化正确率(NA)。

P 是预测为正样本的结果中,正确预测的比例,公式(3):

$$P = \frac{TP}{TP + FP} \quad (3)$$

R 是所有正样本中被找出的比例,公式(4):

$$R = \frac{TP}{TP + FN} \quad (4)$$

$F1_score$ 是 P 和 R 的一种调和平均,公式(5):

$$F1_score = \frac{2 \times P \times R}{P + R} \quad (5)$$

其中, TP 和 TN 分别是真例和假例被正确分类的样本数, FP 和 FN 分别是真例和假例被错误分类

的样本数。

NA 是 AKS 和 AUS 的加权平均, AKS 表示已知类准确率, AUS 表示新类查全率,具体计算如式(6) ~ (8) 所示:

$$NA = \lambda AKS + (1 - \lambda) AUS \quad (6)$$

$$AKS = \frac{\sum (TP + TN)}{\sum (TP + FP + FN + TN)} \quad (7)$$

$$AUS = \frac{TU}{TU + FU} \quad (8)$$

其中, TU 和 FU 分别代表新类样本正确和错误分类的样本数, λ 是样本中已知类所占比例, $0 < \lambda < 1$ 。

时间性能分为 3 个部分(均为每样本平均时间): 模型训练时间, 模型更新时间, 在线分类时间。

3.3 实验环境

实验在硬件配置为 Intel (R) Core (TM) i7-

10750H CPU @ 2.60 GHz, 16 G 内存的联想笔记本电脑完成。分类器均采用 RFC, 分类阈值为 0.9, 采用五折交叉验证, 训练集和测试集的比例为 4 : 1。设多分类器树个数为 100, 二分类器树个数为 20, 缓存池大小为 400 样本。

3.4 实验结果

将实验分成 5 个时段, 逐次增加新类, 完成新类检测并更新模型, 实验时序见表 2。

表 2 实验时序步骤
Table 2 Experiment sequence step

时序	已知类	新类
时段 1 ($t_0 - t_1$)	$c1, c2, c3, c4$	$c5$
时段 2 ($t_1 - t_2$)	$c1, c2, c3, c4, c5$	$c6$
时段 3 ($t_2 - t_3$)	$c1, c2, c3, c4, c5, c6$	$c7$
时段 4 ($t_3 - t_4$)	$c1, c2, c3, c4, c5, c6, c7$	$c8$
时段 5 (t_4 之后)	$c1, c2, c3, c4, c5, c6, c7, c8$	

3.5 不同方法对比

5 个时段上已知类和新类各个评价指标以及归化正确率见表 3。IUCS 前 4 个时段新类识别 P 在 90% 左右, R 高于 90%, 但是已知类的 R 较低; 因为新类有较高的 R , 所以 NA 较高。经过 4 次模型更新, 时段 5 中 $c1 - c8$ 为被识别为已知类, 取得了较好的分类性能。

表 3 5 个时段上已知类和新类各个评价指标以及归化正确率
Table 3 Evaluation indicators and normalization accuracy of known and new classes in 5 time periods

时序	类别	P	R	$F1$	NA
时段 1	已知类 ($c1 - c4$)	0.989 5	0.885 0	0.934 3	0.989 40
	新类 ($c5$)	0.895 0	0.994 4	0.942 1	
时段 2	已知类 ($c1 - c5$)	0.978 3	0.945 3	0.961 5	0.988 30
	新类 ($c6$)	0.905 0	0.983 7	0.942 7	
时段 3	已知类 ($c1 - c6$)	0.923 7	0.893 3	0.908 3	0.934 60
	新类 ($c7$)	0.912 5	0.885 9	0.899 0	
时段 4	已知类 ($c1 - c7$)	0.937 3	0.931 6	0.934 4	0.995 60
	新类 ($c8$)	0.917 5	1	0.957 0	
时段 5	已知类 ($c1 - c8$)	0.964 2	0.883 7	0.922 2	0.981 25

注: $c1$: facebook_audio, $c2$: ftp, $c3$: skype_video, $c4$: voipbuster, $c5$: bitorrent, $c6$: hangouts_audio, $c7$: skype_file, $c8$: youtube

本文方法 IUCS 与 V-EVT 分类性能对比见表 4。与 V-EVT 相比, 本文方法在已知类 R 和新类 P 上提高 5~10 个百分点, $F1$ 提高 3~5 个百分点, 整体 NA 提高 1~2 个百分点。V-EVT 的 $F1$ 表现不

好, 可能是其更多关注新类的 R 而略微忽略了新类 P 。同时, RFC 的投票分布不能很好地拟合威布尔模型, 也可能是其 $F1$ 不高的原因之一。因为其新类 R 较高, 所以 NA 表现不错。

不同方法的时间性能见表 5。在训练阶段, V-EVT 除了训练 RFC, 还需要拟合各个类的威布尔模型, 增加了模型训练时间; 相比之下, IUCS 只需要训练 H_0 和 H_M 。V-EVT (除了计算 RFC 投票树) 要计算每个威布尔模型的累积密度函数值。随着级联二分类器数量增多, IUCS 分类时间也随之增加, 如图 5 所示。 t_4 时刻重新训练多分类器 (暂时取代级联的二分类器), 时段 5 的分类时间基本回归到时段 1 水平。总体来说, IUCS 的平均分类时间应为时段 1~时段 4 的分类耗时平均值。

表 4 不同方法新类检测性能对比

Table 4 Comparison of new class detection performance of different methods

方法	已知类			新类			NA
	P	R	$F1$	P	R	$F1$	
V-EVT	0.897 0	0.805 3	0.848 7	0.817 8	0.870 7	0.843 4	0.919 0
IUCS	0.925 0	0.878 0	0.900 9	0.884 8	0.915 2	0.899 7	0.947 0

IUCS 不同时刻模型更新时间如图 5 所示, 模型更新需要重新训练, 本文方法的平均模型更新时间显著减少, 有利于实现快速在线分类。

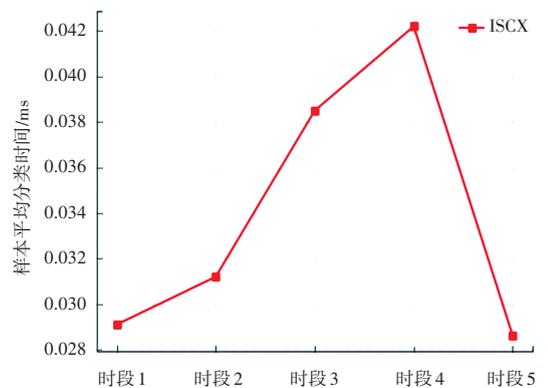


图 5 不同时段 IUCS 分类时间

Fig. 5 IUCS classification time in different periods

表 5 不同方法时间性能 (ms/样本)

Table 5 Time performance of different methods (ms/sample)

方法	训练时间	分类时间
V-EVT	0.745 7	0.407 5
IUCS	0.119 7	0.035 3

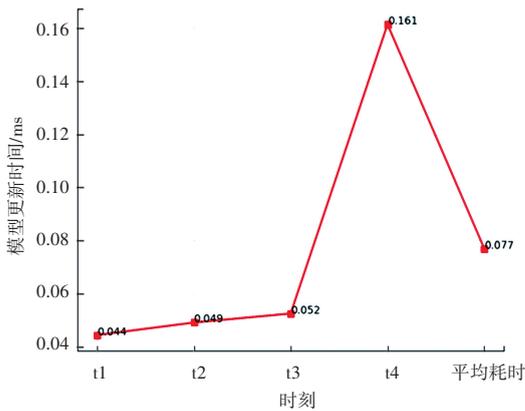


图6 不同时刻 IUCS 模型更新时间

Fig. 6 IUCS model update time at different times

4 结束语

本文提出了一种增量更新级联结构新类检测方法。首先,利用无标签数据集筛选模拟新类样本与已知类样本结合训练新类-已知类二分类器,利用分类置信度阈值有效检测新类;其次,利用时变二分类器级联结构逐次应对顺序出现的不同新类。在真实网络数据集上进行方法验证的结果表明,新类检测率高于90%,并在模型更新后分类准确率大于90%。与文献方法相比,本文在分类准确率和时间性能上具有明显优势。

本文还存在以下局限性:

- (1)同一时段对一个新类进行模型更新,无法区分同一时段内不同新类;
- (2)较高的CMT以降低R为代价获得更高的P,导致缓存池的新类样本中掺杂一些已知类样本。

参考文献

- [1] WANG W, ZHU M, WANG J, et al. End-to-end encrypted traffic classification with one-dimensional convolution neural networks [C]//Proceedings of 2017 IEEE International Conference on Intelligence and Security Informatics (ISI). IEEE, 2017: 43-48.
- [2] SUSANTO H, KIM B G. Congestion control and user utility function for real-time traffic [C]// Proceedings of 2013 IEEE Globecom Workshops (GC Wkshps). IEEE, 2013: 974-979.
- [3] TRINH H D, GAMBIN A, GIUPPONI L, et al. Mobile traffic classification through physical control channel fingerprinting: a deep learning approach [J]. IEEE Transactions on Network and Service Management, 2021, 18(2): 1946-1961.
- [4] KATTDIGE C, CHOI K N, WIJESINGHE A, et al. SETA++: real-time scalable encrypted traffic analytics in multi-Gbps networks [J]. IEEE Transactions on Network and Service Management, 2021, 18(3): 3244-3259.
- [5] OZA P, PATEL V M. Deepcnn-based multi-task learning for open-set recognition [J]. arXiv preprint arXiv:1903.03161, 2019.
- [6] GENG C, HUANG S, CHEN S. Recent advances in open set recognition: A survey [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 43(10): 3614-3631.
- [7] KOTZ S, NADARAJAH S. Extreme value distributions. Theory and applications [M]. London: Imperial College Press, 2020: 61-63.
- [8] 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]// Proceedings of 2021 IEEE Intelligent Vehicles Symposium (IV). IEEE, 2021: 674-681.
- [9] JOSEPH K J, KHAN S, KHAN F S, et al. Towards open world object detection [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, 2021: 5830-5840.
- [10] MENDES JÚNIOR P R, DE SOUZA R M, WERNECK R O, et al. Nearest neighbors distance ratio open-set classifier [J]. Machine Learning, 2017, 106(3): 359-386.
- [11] SCHEIRER W J, ANDERSON D R R, SAPKOTA A, et al. Toward open set recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(7): 1757-1772.
- [12] 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.
- [13] JAIN L P, SCHEIRER W J, BOULT T E. Multi-class open set recognition using probability of inclusion [C]//Proceedings of ECCV 2014. Cham: Springer, 2014. DOI: 10.1007/978-3-319-10578-9_26
- [14] ZHANG H, PATEL V M. Sparse representation-based open set recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 39(8): 1690-1696.
- [15] BENDALE A, BOULT T. Towards open world recognition [C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2015: 1893-1902.
- [16] HU C, CHEN Y, HU L, et al. A novel random forests based class incremental learning method for activity recognition [J]. Pattern Recognition, 2018, 78: 277-290.
- [17] MUHLBAIER M D, TOPALIS A, POLIKAR R. Learn++. NC: Combining ensemble of classifiers with dynamically weighted consult-and-vote for efficient incremental learning of new classes [J]. IEEE Transactions on Neural Networks, 2008, 20(1): 152-168.
- [18] DA Q, YU Y, ZHOU Z H. Learning with augmented class by exploiting unlabeled data [C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2014: 1760-1766.
- [19] XIN M, KAI M T, 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.
- [20] 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.