基于深度森林的 BGP 异常检测方法

赵智男[1,3] 张健毅[2,3] 池亚平[2,3]

[1](西安电子科技大学通信工程学院 陕西 西安 710701)

[2](北京电子科技学院 北京 100070)

[3](中国科学院信息工程研究所中国科学院网络测评技术重点实验室 北京 100093)

摘 要 一直以来,边界网关协议(Border Gateway Protocol, BGP)异常事件严重影响着互联网的稳定与安全,因此 BGP 异常检测算法的研究显得尤为重要。针对已应用于 BGP 异常检测的机器学习算法准确率不高且实验数据集异常种类单一的问题,为了提高准确率并提高方法普适性,引入基于深度森林的异常分类算法。实验采用多个异常事件数据集,根据皮尔森相关系数来剔除冗余无关特征,用于对 BGP 异常分类,分别采用深度森林算法和其他机器学习算法对数据分类。实验结果表明,深度森林的性能是优于其他算法的。

关键词 边界网关协议 异常检测 深度森林 机器学习

中图分类号 TP3 **文献标志码** A DOI:10.3969/j.issn.1000-386x.2024.10.054

BGP ANOMALY DETECTION METHOD BASED ON DEEP FOREEST

Zhao Zhinan[1,3] Zhang Jianyi[2,3] Chi Yaping[2,3]

[1](*School of Telecommunications Engineering, Xidian University, Xi'an 710701, Shaanxi, China*)

[2](*Beijing Electronic Science and Technology Institute, Beijing 100070, China*)

[3](*Key Laboratory of Network Assessment Technology, Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093, China*)

Abstract For a long time, BGP abnormal events have seriously affected the stability and security of the Internet, so the research on BGP anomaly detection algorithms is particularly important. In view of the low accuracy of the machine learning algorithm that has been applied to BGP anomaly detection and the single type of anomaly in the experimental data set, in order to improve the accuracy and universality of the algorithm, an anomaly classification algorithm based on deep forest is introduced. The experiment used multiple abnormal event data sets, and eliminated redundant and irrelevant features based on Pearson correlation coefficients, which were used to classify BGP anomalies. The deep forest algorithm and other machine learning algorithms were used to classify the data. Comprehensive experimental results show that the performance of deep forest is better than other algorithms.

Keywords Border gateway protocol Anomaly detection Deep forest Machine learning

0 引 言

BGP 作为最复杂且应用最广的路由协议,在它被设计之初默认 BGP 邻居之间相互信任,未采用任何措施来确保 BGP 邻居间的通信不被攻击[1]。因 BGP 的安全脆弱性,互联网路由基础设施易受到攻击,破坏了互联网的网络可达性[2]。频繁发生的 BGP 异常事件对广大互联网用户造成了非常恶劣的影响,例如 2008 年巴基斯坦将 YouTube 列入路由黑洞,但是封杀范围无意中扩大到其他国家,使得 YouTube 停运两小时;2014 年,印尼运营商 Indosat 发生大规模路由劫持使得印尼、泰国和美国的网络连接受阻三小时;2015 年,印度运营商 BHARTI Airtel 发生了路由泄露事件导致两千多个自治域网络故障。这些事件极大地影响了互联网的安全和稳定,因此 BGP 异常检测引起了研究者们

收稿日期:2021-05-06。**赵智男**,硕士生,主研领域:BGP 异常检测。**张健毅**,讲师。**池亚平**,教授。

的关注。具体的 BGP 异常事件可分为直接异常、间接异常和链接异常[3]。直接异常可以分为直接预期异常和直接意外异常,直接预期异常主要包括 BGP 劫持,例如前缀劫持和子前缀劫持。当攻击者声称拥有属于另一个自治域(AS)的前缀或子前缀,从而导致从此 AS 到攻击者的路由重定向时,就会发生劫持。直接意外异常是指 BGP 路由器运营商对 BGP 的错误配置,可能导致宣布使用错误的前缀,宣布合法的错误的前缀导致的后果和前缀劫持相似,宣布不合法的前缀则会导致路由泄露或者路由黑洞。间接异常是指针对 Internet 组件(例如 Web 服务器)的恶意活动,例如 Slammer 蠕虫攻击。链接异常是指连接链路或者某个核心 AS 内部发生故障引起的异常,例如 2005 年的莫斯科大停电事件。由于上述 BGP 异常事件发生都伴随着 BGP 更新报文数量的激增这一特点,因此可以将异常检测问题转化为一个二分类问题,即将从 BGP 更新报文中提取的数据分为正常数据和异常数据。

目前已有的基于机器学习的 BGP 异常检测方法主要是针对间接异常的检测,研究使用数据集为 Slammer、Nimda 和 Code-red-I 三个典型异常事件的数据组成。检测算法包括支持向量机(Support Vector Machine,SVM)算法、隐马尔可夫(Hidden Markov Models,HMMs)模型、朴素贝叶斯算法(Naive Bayesian,NB)、J48 决策树和长短期记忆模型(Long Short-Term Memory,LSTM),其中 SVM 算法为目前性能最好的 BGP 异常检测算法。但是由于已有研究中只针对了间接异常,所以无法保证相关算法模型可以有效检测到其他种类的异常。

经过本文实验,在采用多种 BGP 异常数据组合的数据集训练模型后,文献[4-7]中算法对各个测试集的检测性能都有了明显的下降,证明其都无法有效地学习到不同种类 BGP 异常之间的共同特征,从而影响检测准确率。而本文所提的基于深度森林的 BGP 异常检测方法显示出其深度集成学习的优越性,通过逐层的特征变换和堆叠结构实现了特征逐层的融合,挖掘出深层融合特征,从而能够更加准确地检测出异常数据。

本文所提模型相比以往机器学习 BGP 异常检测方法可以达到更高的准确率,且能更好地挖掘出不同种类 BGP 异常之间的深层融合特征,适应多种类型的 BGP 异常,拥有更好的普适性。

1　相关工作

当前已有的 BGP 异常检测方法主要有时间序列分析、统计模式识别、机器学习、基于历史 BGP 数据验证和可达性检查[3]。基于时间序列分析的方法可以根据 BGP 更新报文数量激增检测少数大规模异常,例如 Mai 等[8]使用小波变换和聚类思想对 BGP 异常进行检测,但是按照一定的时间窗口统计报文数量对更短的时间内的异常不敏感,由于大部分小规模 BGP 异常通常在 10 到 20 分钟内收敛恢复正常,因此基于时间序列分析的方法不具有普适性。基于历史 BGP 数据验证的方法和基于可达性检查的方法仅能检测直接预期异常且需要的存储资源非常大,前者利用历史数据根据长期趋势对 BGP 异常路由进行检测,例如 Karlin 等[9]基于前缀变化利用十天的历史 BGP 更新报文对异常进行分析检测,后者根据路由表中的路径信息验证结果进行异常检测,例如 Zheng 等[10]最早设计的根据源 IP 到某个 IP 的网络距离(源与目标之间的跳数)的显著变化作为 BGP 劫持的指标从而判断异常的方法。基于统计模式识别的方法采用统计概率论进行模式识别,根据模式之间的距离函数来判定异常,能够同时检测到多种异常,例如 Huang 等[11]采用的基于 PCA 的 BGP 异常检测方法,但该方法正确估计高维数据分布困难、检测速度慢。基于机器学习的方法,可以根据更新报文数量激增检测到异常,将异常检测问题抽象为分类异常和正常数据的二分类问题,虽然目前检测准确率低于其他方法,但是适合各种维度大小的数据而且检测速度快。

当前研究使用的基于机器学习的 BGP 异常检测方法主要包括 SVM 算法、隐马尔可夫模型、朴素贝叶斯算法、J48 决策树和长短期记忆模型,文献[4]中使用了 SVM 和 HMMs 算法进行 BGP 异常检测实验,得出 SVM 模型的检测效果强于 HMMs 模型。文献[5]比较了 NB、SVM 以及 J48 决策树算法应用在 BGP 异常检测中的效果,得出基于高斯核(RBF)的 SVM 算法检测效果最好的结论。文献[6]中比较了 SVM 和 LSTM 在 BGP 异常检测中的应用效果,同样得出 SVM 检测效果较优的结论。文献[7]中研究了 SVM 算法基于不同的核函数时,针对 BGP 异常检测的效果,得出结论,在选取最相关特征之后的基于线性核(Linear)的 SVM 算法检测效果最优。但是这些文献中使用的训练数据集都是用 Slammer、Nimda 和 Code-red-I 三个典型异常事件的数据组成,异常数据较少且异常种类都是间接异常,这会造成如果测试对象中包含其他种类异常时检测准确率过低的问题。针对这一问题将采用更多异常事件的数据作为训练集,包含更多种类的异常,这样更有利于训练更全面有效的 BGP 异常检测模型。原有的单一机器学习检测算法在扩增种类后的数据集上显现出准确率不高和运算速度较慢等弊端,

相反地,集成学习更适合应用于较复杂的环境。深度森林已经在多种分类问题中表现出优越的性能[12],因此提出基于深度森林的机器学习算法进行 BGP 异常检测。

2 算法介绍

本文提出的 BGP 异常检测模型引入了深度森林的思想,利用级联堆叠的结构对原始特征和变换特征进行多次融合,经过逐层的特征变换来挖掘数据的深层信息,最后利用融合后的特征对异常数据进行检测。

2.1 模型介绍

BGP 异常检测模型基于深度森林(gcForest)算法[13],算法所采用的级联结构为 BGP 异常检测方法提供了重要的思路,模型将使用级联结构并通过逐层特征变换融合使得模型能够挖掘更深层的特征。图 1 为文本所提模型结构。

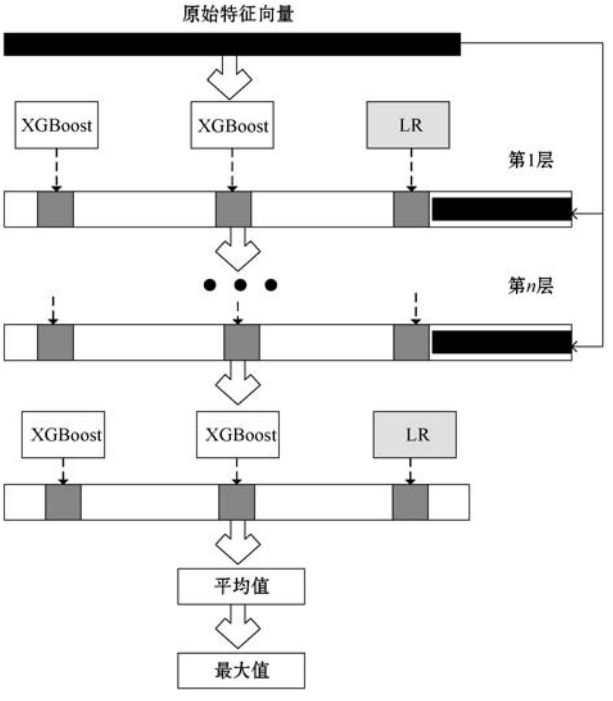

图 1 BGP 异常检测模型结构图

本文模型结构每层由两个极端梯度提升分类器(XGBoost)和一个逻辑回归分类器(LR)组成,首先将 BGP 异常检测问题抽象为二分类问题,那模型的一层中每个分类器都会生成一个长度为 2 的概率向量,然后综合此集成分类器中所有分类器的预测结果取均值从而生成每个集成分类器的决策结果,本层所有集成分类器得到的结果和原始输入连接作为下一层的输入。训练过程中,每一层都会使用三折交叉验证计算本层的分类准确率,层数会不断增加直到触发迭代终止条件,终止条件为当前层的 3 层之内模型准确率没有提升或者达到设置的最大层数,则停止并删除多余层。取最后一层各分类器决策结果的平均值,并从此两个类别中选出概率最大的类作为分类结果。具体的模型训练流程如下。

输入:原始 BGP 更新特征向量。

输出:分类准确率。

1) 将两个 XGBoost 和一个 LR 组成模型第 1 层;将特征向量输入当前层各分类器,使用 3 折交叉验证根据训练样本和标签训练当前层中的分类器得到每个分类器的预测输出并拼接成类别向量。

2) 将两个 XGBoost 和一个 LR 组成模型第 2 层;将特征向量与前一层生成的类别向量拼接作为新的特征向量输入当前层,使用 3 折交叉验证训练当前层各个分类器得到预测输出并拼接成类别向量,计算当前层对训练样本的分类准确率,定义为 Ac_2,记当前最大准确率并初始化为 $Ac_{\max}=Ac_2$。

3) 初始化迭代次数 $k=1$,定义连续没有性能提升的层数 $m=0$,开始迭代。

4) 将两个 XGBoost 和一个 LR 组成模型第 $2k+1$ 层;将特征向量与前一层生成的类别向量拼接作为新的特征向量输入当前层,使用三折交叉验证训练当前层各个分类器得到预测输出并拼接成类别向量,计算当前层对训练样本的分类准确率 Ac_{2k+1}。

5) 比较当前层分类准确率 Ac_{2k+1} 和 $Ac_{\max}$,若 $Ac_{2k+1}>Ac_{\max}$,令 $m=0$,$Ac_{\max}=Ac_{2k+1}$ 并继续执行;否则判断 $++m\geqslant 3$,若为真,则删除最新的 3 层并执行 7),若为假则继续执行。

6) 令 $k=k+1$,并返回 4)。

7) 得到最终层各分类器预测结果,分别计算预测为正常和异常的平均概率,概率大的类别为最终分类结果,输出 $Ac_{\max}$。

本文所提基于深度森林的 BGP 异常检测模型,每一层都通过 XGBoost 和 LR 将当前输入的特征向量映射到更加抽象的类别空间,挖掘了更深层的信息,再与原始特征向量进行拼接作为当前层输出,如此逐层训练实现更加充分的特征信息学习从而得到好的检测模型。

2.2 特征描述

异常事件发生时,都会伴随着 BGP 更新报文数量会巨幅增长这一特点,因此当前用于 BGP 异常检测研究的数据集都是根据异常事件相关的 BGP 更新报文生成的。具体数据特征及描述如表 1 所示。

表 1 数据特征及其描述

特征属性	描述
announcements	宣告数量
withdrawals	撤回数量
dups	重复宣告数量
wd_dups	重复撤回数量
flaps	撤回并再次宣告数量

续表 1

特征属性	描述
nadas	撤回之后新的宣告数量
news	普通新宣告数量
imp_wd	隐式撤回数量
imp_wd_dpath	相同路径隐式撤回数量
imp_wd_spath	不同路径隐式撤回数量
origin_0/origin_1/origin_2	IGP/EGP/INCOMPLETE 消息数量
origin_changes	ORIGIN 变化数量
nlri_ann	宣告前缀数量
ann_to_longer/ann_to_shorter	路径变长/短数量
as_path_max/as_path_avg	最长/平均 AS 路径
unique_as_path_avg	唯一 AS 平均路径
unique_as_path_max	最大路径长度
edit_distance_max	最大编辑距离
edit_distance_avg	平均编辑距离
edit_distance_dict_k	编辑距离为 k 的数量(k 取 0,1,…,10)
edit_distance_unique_dict_k	无环编辑距离为 k 的数量(k 取 0,1,…,10)
number_rare_ases	在 95% 的 AS 路径从未出现过的 AS 数量
rare_ases_avg	稀有 AS 出现平均次数
rare_ases_max	稀有 AS 出现最大次数

目前用于 BGP 异常检测研究的数据集都是从原始更新报文中提取相关特征生成的。

本文所用数据集来自文献[14],文中提出了一种 BGP 异常事件数据集生成的方法,从原始更新报文中提取如表 1 中的特征,以一分钟为时间间隔,将每一分钟收集到的报文中所包含的具体特征值组合生成特征向量。根据异常事件的发生时间,从异常事件发生前一段时间开始提取数据,直到异常事件结束后一段时间生成的特征向量集合作为一个异常事件数据集。

BGP 异常事件发生时,不同种类的异常影响偏重的特征也有所不同。例如链接异常发生时,宣告、撤回和宣告振动的数量会较以往明显减小,稀有 AS 数量会有所增加,因为链接异常时更新报文都会走备用路径从而产生新的 AS 声明;直接异常发生时,新宣告的数量会明显增加,因为这类异常通常为劫持或错误配置会产生新的伪造路由路径;间接异常发生时,隐式撤回数量会明显减少。可以见得,不同异常种类对特征的影响偏重不同,所以深度挖掘各类异常所影响的特征之间的信息对统一的 BGP 异常检测具有积极意义,也体现出本文所提模型的优越性。

2.3　特征处理

在监督学习中冗余和无关特征常常会影响分类的准确性和效率,适当地删除冗余和无关特征可以在一定程度上提高模型的性能。本文根据训练集计算各个特征之间的皮尔森相关系数,并将与类别标签相关系数最小的十个特征删除,然后依据剩余特征进行分类研究,对删除冗余特征前后的数据集进行实验测试。图 2 为根据训练集得到的各个特征之间皮尔森相关系数热力图,每个单元中是对应行特征和列特征的相关系数值,值的绝对值越大,相关性越大。

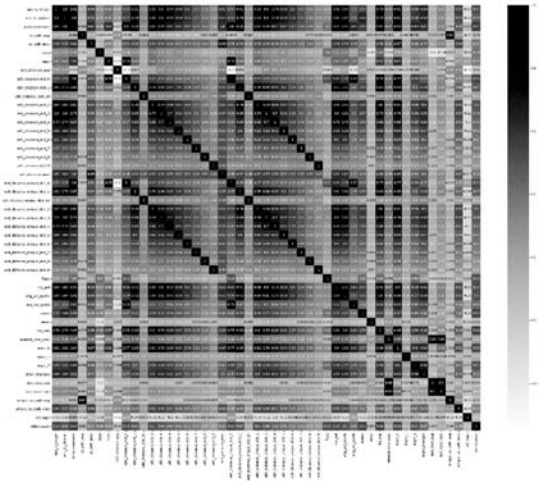

图 2　特征皮尔森相关系数热力图

如上热力图主要关注计算标签和其余所有特征之间的皮尔森相关系数,对相关系数的绝对值进行排序。如果某一特征与分类标签的相关性小,那么此特征对于异常分类的贡献就小;如果某特征与分类标签的相关性大,那么此特征对于异常分类的贡献就大。根据皮尔森相关系数计算出对 BGP 异常分类贡献最小的十个特征作为冗余特征,并且剔除,以此来更为高效准确地得到分类结果。

由计算结果可得到与分类标签的相关性最小的十个特征如表 2 所示。

表 2　冗余特征

冗余特征
as_path_avg
as_path_max
edit_distance_avg
edit_distance_dict_10
edit_distance_dict_9
edit_distance_unique_dict_10
nlri_ann

续表 2

冗余特征
rare_ases_max
unique_as_path_avg
unique_as_path_max

为了证明剔除冗余特征之后,能够提高 BGP 异常检测模型的性能,本文使用 Slammer 事件数据集对剔除冗余特征前后的深度森林训练模型进行性能比较,实验结果如表 3 所示。

表 3 全特征和提出冗余特征之后性能对比

测试集		准确率/%	F1-Score/%
Slammer513[14]	全特征	83.56	59.8
	剔除冗余	88.71	61.1
Slammer6893[14]	全特征	83.83	58.8
	剔除冗余	86.37	62.5

由表 3 可以看出在剔除冗余特征之后,在 Slammer 事件的两个数据集上进行测试,分类的准确率和 F1-Score 都有了提升,准确率提高 3 百分点左右,F1-Score 平均提高了 2 百分点。因此可以得知剔除冗余特征之后的模型性能更好。

3 实验结果和分析

本文所使用的数据集来自文献[14]提供的开源数据集,基于以往 BGP 异常检测算法研究所使用的数据集异常数据少异常种类单一的问题,将集合若干数据集中标签为异常的特征向量与从这些数据集中随机选取的标签为正常的特征向量组合成为一个平衡的数据集作为训练集训练模型,如表 4 所示,AS9121RTL、Aws-leak 和 AS3561-filtering 异常为直接异常,Code-red-I 和 Malasian-telecom 为间接异常,Moscow Blackout 和 Japan-earthquake 为链接异常。以 Nimda、Slammer 和 As-path-error 事件数据集作为测试集,具体如表 5 所示。

表 4 训练集

异常事件名称	异常数据	正常数据
AS9121RTL	547	547
Moscow Blackout	680	680
Aws-leak	526	526
Code-red-I	1 892	1 892
Japan-earthquake	849	849
Malasian-telecom	623	623
AS3561-filtering	391	391
合计	5 508	5 508

表 5 测试集

事件	异常数据	正常数据
Nimda513	1 178	6 802
Nimda6893	1 178	6 802
Slammer513	1 129	6 071
Slammer6893	1 129	6 071
As-path-error	341	6 432

BGP 异常事件往往不只影响一个 AS,所以同一个异常事件可能从不同的 AS 的不同采集点采集多个异常数据集。训练集是将每个事件从每个采集点生成的异常数据集合起来,测试集将每个采集点生成的数据集分别作为一个单独的测试集,测试集表中 Nimda 和 Slammer 事件都分为从 AS513 和 AS6893 两个采集点生成的数据集,加上 As-path-error 事件的一个数据集合计五个数据集作为测试集。

原始 BGP 更新报文解析后的具体格式为:BGP 版本|采集时间|UPDATE 类型|采集点 IP|采集点 AS 号|IP 前缀|AS 路径|ORIGIN。从原始更新报文中,以一分钟为单位,根据 2.2 节所述特征进行提取,将每分钟收集到的报文中所包含的具体特征值组合生成一条实验数据。根据各个异常事件的具体发生时间生成各自的事件数据集。然后根据 2.3 节中计算的冗余特征,对数据集进行剔除冗余特征的处理,得到训练集和测试集。

训练集中统计异常数据量为 5 508 条,随机从各个事件数据集中选取与异常数据数量相等的正常数据 5 508 条,合计 11 016 条,来构造一个平衡的训练集。

测试集共有 5 个,每个测试集中的异常数据量和正常数据量分布具体展示在表 5。

本文采用基于深度森林的机器学习算法进行 BGP 异常检测方法的设计,并与文献[4-7]中性能最优的算法进行性能对比,再和极端梯度提升算法(XGBoost)、逻辑回归(LR)算法对比,来评估本文深度森林模型和单独应用的基学习器的优劣。

实验中,首先对数据集进行特征处理,根据计算皮尔森相关系数剔除冗余无关特征。然后使用训练集分别对深度森林、基于 RBF 核的 SVM[4-5]、基于线性核的 SVM[6-7]、XGBoost 和 LR 模型进行训练,最终使用测试集对各个模型进行性能评估。实验基于 TensorFlow 2.4.1 下的 Keras 2.4.3,使用 Python 3.8 完成。

实验首先使用准确率和 F1-Score 对分类模型的性能进行评估。以下为评估参数的定义解释。异常数据分类为异常的数量定义为 T_P,正常数据分类为异常的数量定义为 F_P,异常数据分类为正常的数量定义为

F_N,正常数据分类为正常的数量定义为 T_N。异常数据的查全率 R 和查准率 P 计算式如下:

$$R = \frac{T_P}{T_P + F_N} \tag{1}$$

$$P = \frac{T_P}{T_P + F_P} \tag{2}$$

准确率和 F1-Score 的计算公式如下:

$$A_{ccuracy} = \frac{T_P + T_N}{T_P + T_N + F_P + F_N} \tag{3}$$

$$F_{1\text{-Score}} = \frac{2 \times P \times R}{P + R} \tag{4}$$

实验首先使用训练集训练各个模型,然后利用分别五个测试集进行测试实验,得到每个模型的分类准确率和 F1-Score 值,具体数据如表 6 – 表 10 所示。

表 6　测试集 Nimda-513 各模型性能对比

模型	准确率/%	F1-Score/%
gcForest(本文)	70.13	61.4
SVM(rbf 核)[4-5]	65.93	5.3
SVM(线性核)[6-7]	66.15	37.0
XGBoost	61.96	61.1
LogisticRegression	67.17	54.9

表 7　测试集 Nimda-6893 各模型性能对比

模型	准确率/%	F1-Score/%
gcForest(本文)	70.67	65.7
SVM(rbf 核)[4-5]	67.52	13.8
SVM(线性核)[6-7]	66.80	54.8
XGBoost	65.61	61.1
LogisticRegression	68.35	39.4

表 8　测试集 Slammer-513 各模型性能对比

模型	准确率/%	F1-Score/%
gcForest(本文)	88.71	61.1
SVM(rbf 核)[4-5]	87.10	35.7
SVM(线性核)[6-7]	75.85	45.6
XGBoost	66.74	46.3
LogisticRegression	82.22	56.5

表 9　测试集 Slammer-6893 各模型性能对比

模型	准确率/%	F1-Score/%
gcForest(本文)	86.37	62.5
SVM(rbf 核)[4-5]	85.67	21.8
SVM(线性核)[6-7]	80.69	52.6
XGBoost	77.26	54.8
LogisticRegression	82.24	55.2

表 10　测试集 As-path-error 各模型性能对比

模型	准确率/%	F1-Score/%
gcForest(本文)	97.61	69.5
SVM(rbf 核)[4-5]	74.10	43.8
SVM(线性核)[6-7]	88.63	57.1
XGBoost	96.18	55.4
LogisticRegression	75.40	38.2

可以看出,深度森林算法模型的准确率和 F1-Score 在五个测试集中都高于其他四种算法。SVM 算法的分类准确率并未显示出明显的劣势,在除了 AS-path-error 的其他测试集上都只略低于深度森林模型,但是 F1-Score 明显低于其他算法,尤其是基于 RBF 核的 SVM 算法,在 Nimda-513 和 Nimda-6893 上甚至低于 0.2,由此可知基于 RBF 核的 SVM 算法在使用包含多种类型 BGP 异常的训练集时,模型效果不佳。还可以看出,单独使用 XGBoost 和 LR 模型的准确率和 F1-Score 比基于 XGBoost 和逻辑回归的深度森林集成算法低。分类器所识别出的异常数据占所有异常数据的比例 TPR 定义为:

$$R_{TP} = \frac{T_P}{T_P + F_N} \tag{5}$$

分类器错认为是异常数据的正常数据占所有正常数据的比例 FPR 定义为:

$$R_{FP} = \frac{F_P}{F_P + T_N} \tag{6}$$

以 FPR 为横轴,TPR 为纵轴得到 ROC 曲线图。图 3 – 图 5 为在 Slammer、Nimda、AS-path-error 三个异常事件上五个算法模型的 ROC 曲线对比图。

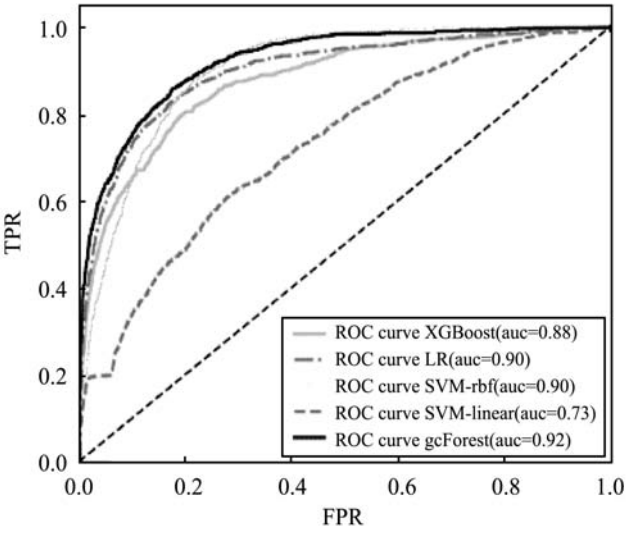

图 3　Slammer 事件各模型 ROC 曲线

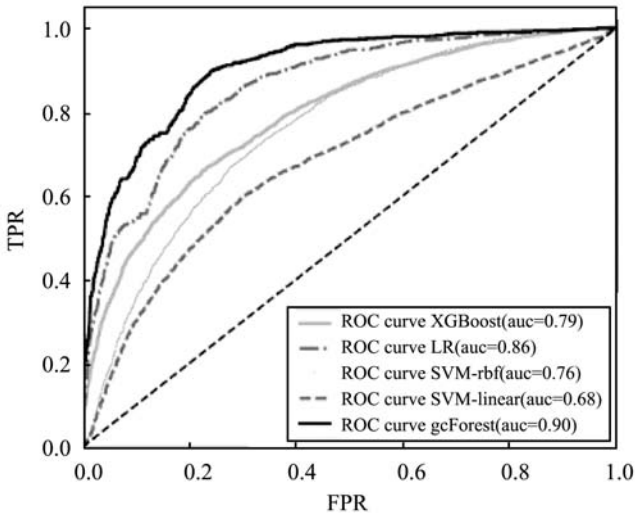

图 4 Nimda 事件各模型 ROC 曲线

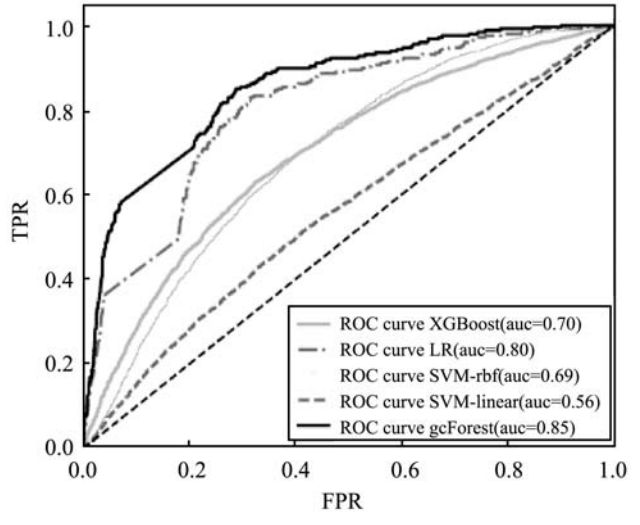

图 5 AS-path-error 事件各模型 ROC 曲线

若算法为完全随机分类则其 ROC 曲线为斜直虚线,分类算法的 ROC 曲线距离直虚线越远表示算法分类效果越好,AUC 值为 ROC 曲线包围的面积,AUC 值越大也表明算法分类效果越好。由图 3 – 图 5 可以看出,三个异常事件的 ROC 曲线图中深度森林模型 ROC 曲线距离虚线最远且 AUC 值最大,表明此模型分类效果强于其他四种模型。

4 结 语

为了提高 BGP 异常检测准确率并提高算法普适性,本文提出基于深度森林的 BGP 异常检测模型,在由直接、间接和链接异常数据集组成的训练集上训练模型,并用三个异常事件共五个测试集对模型进行了测试对比实验。由综合准确率、F1-Score 以及五种模型的 ROC 曲线对比可以得出结论,基于深度森林的 BGP 异常事件检测算法模型的性能优于 SVM-rbf、SVM-linear、XGBoost 和 LR 算法。

参 考 文 献

[1] 兰迪·张,米卡·巴特尔. BGP 设计与实现[M]. 黄博,葛建立,译. 北京:人民邮电出版社,2012.

[2] 王娜,杜学绘,王文娟,等. 边界网关协议安全研究综述[J]. 计算机学报,2017,40(7):1626 – 1648.

[3] Al-Musawi B, Branch P, Armitage G. BGP anomaly detection techniques: A survey[J]. IEEE Communications Surveys & Tutorials,2017,19(1):377 – 396.

[4] Al-Rousan N M, Trajkovi ć L. Machine learning models for classification of BGP anomalies[C]//13th International Conference on High Performance Switching and Routing,2012:103 – 108.

[5] Ćosović M, Obradovi ć S, Trajkovi ć L. Performance evaluation of BGP anomaly classifiers[C]//3rd International Conference on Digital Information, Networking, and Wireless Communications,2015:115 – 120.

[6] Ding Q Y, Li Z D, Batta P, et al. Detecting BGP anomalies using machine learning techniques[C]//IEEE International Conference on Systems, Man, and Cybernetics,2016:3352 – 3355.

[7] Dai X B, Wang N, Wang W J. Application of machine learning in BGP anomaly detection[C]//Journal of Physics: Conference Series,2019,10:1742 – 1796.

[8] Mai J N, Yuan L H, Chuah N. Detecting BGP anomalies with wavelet[C]//IEEE Network Operations and Management Symposium,2008:465 – 472.

[9] Karlin J, Forrest S, Rexford J. Pretty good BGP: Improving BGP by cautiously adopting routes[C]//14th IEEE International Conference on Network Protocols,2006:290 – 299.

[10] Zheng C X, Ji L S, Pei D, et al. A light-weight distributed scheme for detecting IP prefix hijacks in real-time[J]. ACM Sigcomm Computer Communication Review, 2007, 37(4): 277 – 288.

[11] Huang Y, Feamster N, Lakhina A, et al. Diagnosing network disruptions with network-wide analysis[C]//ACM SIGMETRICS Performance Evaluation Review,2007:61 – 72.

[12] 戴瑾,王天宇,王少尉. 基于深度森林的网络流量分类方法[J]. 国防科技大学学报,2020,42(4):30 – 34.

[13] Zhou Z H, Feng J. Deep forest[J]. National Science Review,2019,6(1):74 – 86.

[14] Fonseca P, Mota E S, Bennesby R, et al. BGP dataset generation and feature extraction for anomaly detection[C]//IEEE Symposium on Computers and Communications,2019: 1 – 6.