

doi:10.14132/j.cnki.1673-5439.2022.02.011

基于代价敏感度的改进型 K 近邻异常流量检测算法

李泽一, 王攀

(南京邮电大学 现代邮政学院, 江苏 南京 210003)

摘要:随着互联网的快速发展,网络安全越来越受到人们的重视。传统的异常流量检测模型虽然具有较好的识别率,但需要大量有标记的数据进行训练。因此,基于无监督学习的网络异常流量检测方法被广泛采用。近年来,随着深度学习算法在异常检测中的运用,无监督深度学习模型也不同程度地提升了检测算法的性能。然而,无监督深度学习方法往往无法避免异常检测阈值选择的问题。因此,针对现有数据标记困难和阈值选择的问题,文中提出了一种基于代价敏感度改进的 K 近邻算法结合阈值选择方法的异常流量检测系统。该系统不但可以准确识别恶意流量,也无需有标记数据集,极大减少了人工标注数据的工作量。实验使用 UNSW-NB15、NSL-KDD 和 CICIDS2017 数据集来验证模型的适用性,并分别与经典的机器学习算法 One Class SVM 以及深度学习方法 AutoEncoder 进行了对比。实验结果表明,在 3 类数据集上,与深度学习算法和传统的无监督机器学习算法相比,该算法有效提升了网络异常流量检测的性能。

关键词:异常检测;无监督学习;K 近邻算法;入侵检测系统

中图分类号:TP393.08;TN915.05 文献标志码:A 文章编号:1673-5439(2022)02-0085-08

Unsupervised network abnormal traffic detection method based on improved KNN

LI Zeyi, WANG Pan

(School of Modern Posts, Nanjing University of Posts and Telecommunications, Nanjing 210003, China)

Abstract: With the rapid development of the Internet, people pay an increasing attention to network security. Although the traditional abnormal traffic detection model has a reasonable recognition rate, it needs considerable labeled data for training. Therefore, an abnormal network traffic detection method based on unsupervised learning has been widely used. In recent years, with the application of deep learning algorithms in anomaly detection, unsupervised deep learning models have also improved the performance of detection algorithms to varying degrees. However, unsupervised deep learning methods cannot avoid the problem of threshold selection for anomaly detection. Therefore, given the difficulty of data labeling and threshold selection, this paper proposes an abnormal traffic detection system based on a cost-sensitive improved K-nearest neighbor (KNN) algorithm combined with a threshold selection method. As a result, the system can accurately identify malicious traffic and does not require a labeled data set, which dramatically reduces the workload of manually labeling data. The experiment uses three data sets of UNSW-NB15, NSL-KDD, and CICIDS2017 to verify the model's applicability by comparing the proposed method with the classic machine learning algorithm One-Class SVM and the deep learning method AutoEncoder. The results show that compared with the deep learning algorithms and traditional

收稿日期:2021-10-19;修回日期:2021-11-04 本刊网址:<http://nyzr.njupt.edu.cn>

基金项目:国家重点研发计划(2020YFB1804701)和国家自然科学基金(61972211)资助项目

作者简介:李泽一,男,硕士研究生;王攀(通信作者),男,博士,研究员, wangpan@njupt.edu.cn

引用本文:李泽一,王攀.基于代价敏感度的改进型 K 近邻异常流量检测算法[J].南京邮电大学学报(自然科学版),2022,42(2):85-92.

unsupervised machine learning algorithms on the three types of data sets, the proposed algorithm effectively improves the performance of abnormal network traffic detection.

Keywords: anomaly detection; unsupervised learning; K-nearest neighbor (KNN) algorithm; intrusion detection system

移动互联网时代,网络流量呈现爆发式增长,随之而来的是网络攻击事件频发,攻击形式呈现出复杂多变的特点。于是入侵检测系统不再局限于基于传统的 DPI 技术,开始采用有监督式的机器学习方法检测流量,例如卷积神经网络^[1]、支持向量机等。尽管这些方法都表现出了良好的示范性能,但随着网络数据流量规模的急速扩大,使得有监督的机器学习方法识别的准确度下降,出现误报、漏报的情况。同时有监督的机器学习方法,存在数据标签困难、恶意样本少和无法检测未知恶意攻击的问题。而无监督的深度学习方法可以训练不带标签的数据集,并能成功检测未知的攻击。因此,使用无监督的机器学习方法检测恶意攻击流量已成为网络安全领域的一个热门研究方向。现阶段异常检测模型在图像领域取得令人瞩目的成就。但是基于图像的异常检测算法运用在网络流量上效果未可知。

在上述背景下本文使用无监督方法来检测网络异常流量。首先 K 近邻^[2](KNN)算法作为传统的机器学习算法,在异常检测方面发展已较为成熟。该算法以数据样本与其邻居的距离为基准,对每个样本到其邻居进行距离排序,并将此排序中的前 k 个点声明为异常值。而自动编码器(AE)作为深度学习异常检测的最典型代表,最初起源于图像^[3],现在被迁移到网络流量领域中。其算法旨在学习一组基向量,可以通过组合特征进行压缩,之后解压成一个新的向量。其可以生成接近原始输入的特征输出,反映相似数据的相似信息。在异常检测的应用中,AE 通过计算样本的最大重构误差来判断样本是否异常。但是网络流量与图像的数据结构有着本质的差别,因此使用该模型会有一些的局限性。

本文为了识别异常网络流量,提出了一种基于改进 KNN 的无监督异常检测方法。本文的主要贡献:

(1) 本文提出了一种基于改进型 KNN 的异常流量检测方法。该方法解决了恶意流量样本少、难以采集和标注的问题。

(2) 文章提出了一种代价敏感性的指标体系,用于衡量异常检测模型的性能。基于此指标,改进的模型表现出色。

(3) 实验使用了 3 种规模大小相异的数据集和两种经典的机器学习和深度学习的方法,以印证本文提出的算法的性能。实验结果表明,本文提出的改进 KNN 异常检测算法相比深度学习和其他传统的机器学习方法,更加适合识别异常网络流量。

1 相关工作

1.1 异常检测研究

随着机器学习的不断发展,许多学者将机器学习应用到异常检测领域。Zhao 等^[4]提出一个 Pyod 开源框架,对各种异常检测算法进行归纳总结,供各位同行使用。深度学习由于其可以自动提取高级抽象特征,故利于处理大规模复杂数据。因此深度异常检测首先在图像领域发展起来,而且已在医学图像领域逐步发展成熟。Schlegl 等^[5]提出了一种深度卷积生成对抗网络(fast-ANOGAN)的无监督学习方法,以检测视网膜的解剖是否变异。其模型是根据特征空间中观察位置的局部密度确定异常值。Akçay 等^[6]提出一种较为新颖的异常检测模型(GANomaly),通过使用卷积生成对抗网络学习高维图像空间的生成和潜在空间的推理。当模型学习出数据分布的较大距离度量时,表明该分布存在异常值。2020 年,Wolleb 等^[7]提出基于胸膜腔变化的异常检测模型。通过深度学习模型提取疾病高度特异性的特征,学者们能更详细地检测结构变化。在图像异常检测上,深度学习已经发展相当成熟。接下来,本文探讨异常检测模型在网络流量检测领域中的应用。

1.2 网络流量异常检测

应用传统机器学习和深度学习的方法,网络异常检测目前也得到了长足的发展。目前已有学者将机器学习应用在网络流量检测中。Ramaswamy 等^[8]提出了基于 KNN 的进化推理系统(kENFIS)。该系统主要用于检测计算机蠕虫。Amer 等^[9]用数据集训练 OCSVM,然后考虑数据点与确定的决策边界的归一化距离,对每个数据点进行分类。Falcão 等^[10]将 6 种不同种类的异常检测方法在网络流量上进行综合对比实验。同样他们也针对各类数据集进行分析和整理,但并未对深度学习的异常检测进行对比实

验,无法了解深度学习在网络流量上的效果。Chen 等^[11]提出使用卷积自动编码器(CAE)检测网络异常流量,实验在 NSL-KDD 数据集上表现较好,但是仅仅在单一数据集上不具有说服力,不能说明 CAE 就是适用网络流量的最佳选择。文献[12]提出了一种基于自动编码器的入侵检测系统,将模型收敛后的损失值作为阈值,形成一套端到端的入侵检测系统,但如果模型收敛性不够好,就不能保证检测效果。Zavrak 等^[13]使用无监督深度学习方法和半监督学习方法检测异常网络流量。具体地说,使用自动编码器和变分自动编码器方法来识别基于流特征的未知攻击。实验结果表明,VAE 在很大程度上优于 AE 和 One-Class SVM。但是他们只是使用 ROC 曲线和 AUC 值展示模型的好坏,并没有展示其具体的异常检测效果。文献[14]提出了一种基于改进重建概率的异常检测方法。其中作者重点凸显 VAE 模型的重建概率是一种概率度量,能考虑变量分布的可变性。生成对抗网络(GAN)能够对现实世界数据的复杂高维分布进行建模,这表明它们可以有效地进行异常检测。因此 Zenati 等^[15]使用 GAN 模型进行异常检测,通过在网络入侵数据集上的测试,表现出模型较好的性能。但是文献[14-15]实验的数据集使用 KDD cup^[16],该数据集较小。因此这两种模型,在该数据上的实验效果不具有说服力。

2 模型与数据选择

2.1 算法建立

K 近邻算法是基于邻居节点计算最近距离的算法(KNN),主要用于识别异常值^[17]。该算法对一个新样本进行分类时,必须计算它与集合中每个样本点的距离。对于每个数据点,检查整个数据集,以提取具有最相似特征值的数据集,即最接近的邻居。本文提出的改进 KNN 算法使用欧几里得距离作为距离度量,其涉及的计算并不复杂,具体见式(1)。 p 表示一个数据样本, p_i 则代表该样本的第 i 个特征, $[p_1, p_2, \dots, p_s]$ 表示 s 维特征的数据样本,同时用 $\text{dist}(p, q)$ 表示两个流样本 p 和 q 之间的距离。

$$\text{dist}(p_i, q_i) = \left(\sum_{l=1}^n |p_i^{(l)} - q_i^{(l)}|^2 \right)^{\frac{1}{2}} \quad (1)$$

基于近邻的异常检测算法的核心思想是对每个点计算它的 k 近邻距离,然后在测试集中按照每个 k 近邻距离降序排序。前 n 个点即可认为是离群点。执行步骤见算法 1。

算法 1 基于改进 KNN 的异常检测

1. 输入:训练数据集为 train_data,测试数据集为 test_data。
2. 算法:
3. Begin:
4. for p in train_data:
5. 根据式(1)计算样本 p 的 k 近邻距离;
6. 对于 p 赋予异常得分;
7. 归一化 p 的异常得分;
8. end
9. 预定阈值;
10. for p in test_data:
11. if p 归一化得分 > 阈值:
12. p 为异常数据;
13. else:
14. p 为正常数据;
15. end

图 1 展示了异常检测的流程,原始的网络流量特征先预处理,处理完成的数据进入 KNN 算法。在整个过程中,异常检测如何判断离群的样本,变成了问题的关键。目前的异常分数是根据不同的度量方法计算得出的,为了一致性,异常值被分配了更大的异常分数。本次实验不讨论异常分数的计算,重点讨论阈值选择的方法。由于目前阈值的设定还没有较好的选择。基于此,本文提出改进阈值选择方法,即把异常得分做最值归一化,将异常得分映射到 $[0, 1]$ 的区间内。利用验证集获取阈值,达到阈值修正的目的。同时为了更好地展现异常检测能力,传统的评价指标已不再适用。为了凸显模型异常检测性能,引入代价敏感度评价指标,即赋予良性精确率和恶意召回率更多的权重。

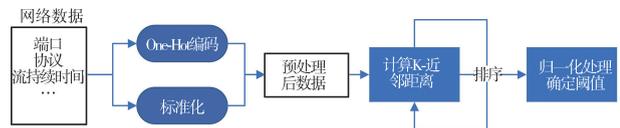


图 1 算法流程图

2.2 数据集选择

本实验使用 3 种类型的数据集,分别是 CICIDS2017^[18]、UNSW-NB15^[19] 和 NSL-KDD^[20] 的 3 个数据集。

(1) CICIDS2017: CICIDS2017 是一个包含入侵检测和入侵防御数据的公共数据集。它还包括使用 CICFlowMeter^[21] 的网络流量分析结果,使用基于时间戳、源 IP、目标 IP、源端口、目标端口和协议等特征组成的标签流。具体的攻击类型如表 1 所示。图 2 展现出 t-SNE 可视化中数据集的乱序性。

表 1 CICIDS2017 不同流量占比

序号	流量分类名称	百分比/%
0	正常流量	80.30
1	Bot	0.07
2	DDOS	4.52
3	DoS GoldenEye	0.36
4	DoS Hulk	8.16
5	DoS Slowhttptest	0.19
6	DoS slowloris	0.20
7	FTP-Patator	0.28
8	PortScan	5.61
9	SSH-Patator	0.21
10	Web Attack	0.07

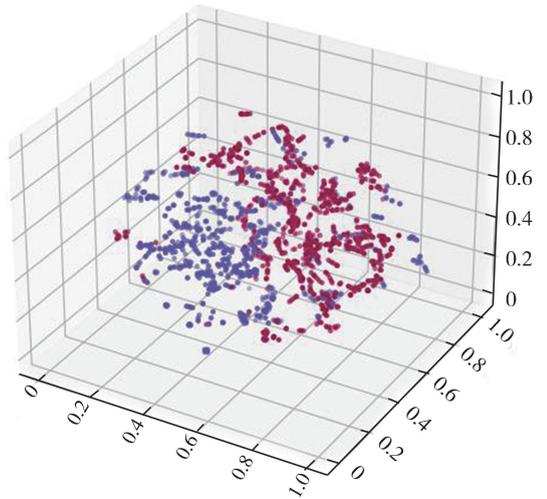


图 3 NSL-KDD t-SNE 可视化

(3) UNSW-NB15: UNSW-NB15 是澳大利亚网络安全中心(ACCS)2015年在 LABS 收集的正常网络活动和综合攻击活动的混合体。具体的攻击类型如表 3 所示。从图 4 可以清晰地看出该数据集的乱序性。

表 3 UNSW-NB15 不同流量占比

序号	流量分类名称	百分比/%
0	正常流量	31.9
1	Generic	22.8
2	DOS	7.0
3	Exploits	19.0
4	Shellcode	0.6
5	Fuzzers	10.4
6	Analysis	1.0
7	Backdoor	1.1
8	Worms	0.1
9	Reconnaissance	6.0

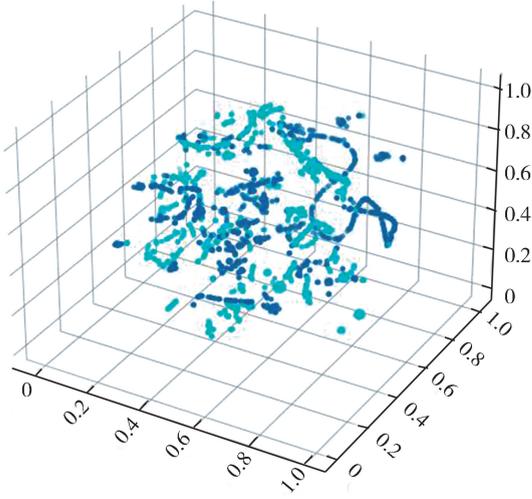


图 2 CICIDS2017 t-SNE 可视化

(2) NSL-KDD: NSL-KDD 是对 KDD99 数据集的改进。NSL-KDD 数据集的训练集不包含冗余记录,所以模型分类器不会受到冗余记录的影响。具体的攻击类型如表 2 所示。图 3 的可视化展现了该数据集未分离时的乱序性。同时 NSL-KDD 数据集的测试集中没有重复记录,使得检测率更加准确。

表 2 NSL-KDD 不同流量占比

序号	流量分类名称	百分比/%
0	正常流量	16.69
1	Probe	0.83
2	DOS	79.24
3	R2L	0.23
4	U2R	0.01

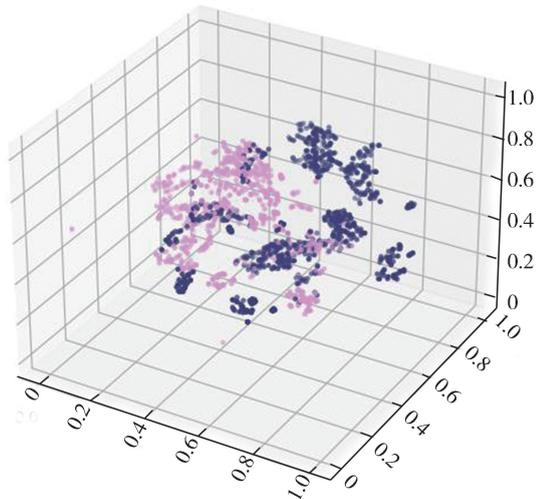


图 4 UNSW-NB15 t-SNE 可视化

为了模拟真实的网络环境,本次实验分别选取正常流量和恶意流量构成上述各数据集的测试集,比例为 100 : 1。实验将剩余的正常流组合成一个训练集进行模型训练。每个数据集中都有符号特征和数字特征。本文对于文字特征,使用 one-hot 编码格式进行处理,对于数字符号,使用最值归一化公式(2)进行处理。

$$X_{\text{norm}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \quad (2)$$

3 实验结果与分析

3.1 实验环境

本文以 Python3 作为主要编程语言。具体参数如表 4 所示。本文使用以下 3 种算法对网络入侵检测算法进行对比实验,包括 KNN、AE 和 OCSVM。同时本次算法 KNN 的近邻参数默认选择为 5。文章重点讨论模型结构对于网络流量的适用性,参数不过多讨论。

表 4 设备参数

开发工具	名称和版本
GPU	NVIDIA RTX1660
CPU	AMD2700X
CUDA 版本	7.5
CuDNN 版本	10.5

3.2 性能评价指标

性能评价指标包括接收者操作特征曲线 (ROC) 曲线下与坐标轴围成的面积 (AUC)、准确率 (Accuracy)、精确率 (Precision)、召回率 (Recall)、混淆矩阵 (Confusion Matrix) 等,这些指标来源于 4 个参数指标:预测正确正样本 (True Positive, TP) 的数量、预测错误的正样本 (False Positive, FP) 数量、预测正确的负样本 (True Negative, TN) 数量、预测错误的负样本 (False Negative, FN) 数量。下面对文章的绩效评价指标进行详细说明:

准确率 (Accuracy) 指的是正确预测的样本数占总预测样本数的比值,反映的是模型算法整体性能。其计算公式为:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

精确率 (Precision) 指的是正确预测的正样本数占所有预测为正样本的数量的比值,反映的重点是

正样本的比重。其计算公式为:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

召回率 (Recall) 指的是正确预测的正样本数量占正样本数量总数的比值,反映的重点是预测正确的正样本的比重。其计算公式为:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

曲线下面积 (AUC)。ROC 曲线表示二元分类器在区分阈值变化时的性能的图形图:ROC 曲线下方面积值衡量识别算法对目标数据集的适合度。在异常检测算法的研究中,学者们着重评价全局的效果。因此在本文实验中,使用 AUC 作为衡量模型的重要因素。

3.3 代价敏感性处理

在异常检测领域,模型会存在将良性样本判断为恶意样本的问题。但是,出现模型将恶意样本判断为良性样本的问题更为严重。因此,从表 5 中可以看出,伪阳性值越小越好。根据表 6 发现,良性的精确率和恶意的召回率是由 FP 计算出。所以,在计算时存在 FP 的公式需要考虑更大的权重因素。本实验通过加权平均来评估精确率。由于良性样本的数量较多,加权平均的精度率会偏向良性流量。同时,本文选择宏平均计算召回率,这可以确保在计算召回率时更少的恶意流量样本,仍然获得更多的权重。

表 5 评价系数 1

参数指标	预测阳性	预测阴性
真实阳性	真阳性 (TP)	伪阴性 (FN)
真实阴性	伪阳性 (FP)	真阴性 (TN)

表 6 评价系数 2

参数指标	精确率	召回率
阳性	$\frac{\text{TP}}{\text{TP} + \text{FP}}$	$\frac{\text{TP}}{\text{TP} + \text{TN}}$
阴性	$\frac{\text{FN}}{\text{FN} + \text{TN}}$	$\frac{\text{FN}}{\text{FN} + \text{FP}}$

因此,根据式(3)使用加权平均精确率 (WP) 和宏平均召回率 (MR) 来计算 F1。

$$\text{F1} = \frac{2 * \text{WP} * \text{MR}}{\text{WP} + \text{MR}} \quad (3)$$

3.4 阈值选择

以往 KNN 算法,使用异常得分的后 10%作为离群的异常值,即阈值设置为 10%的置信区间的横坐标点。本文采用验证集设置阈值。根据图 5(a~c),本次实验在 CICIDS2017、NSL-KDD 和 UNSW-NB15 这 3 个数据集上,分别设置的阈值是 0.025、

0.05 和 0.19。根据表 7 可以发现,使用改进后的阈值效果更加出色。尤其是在 CICIDS2017 数据集上,原有选取阈值的方法不能检测恶意流量。通过采用改进阈值的方法,KNN 达到一个很好的性能。从图 5(b)来看,恶意流量除了在左侧有一个峰度难以区分以外,其他都可以很好地识别出来。

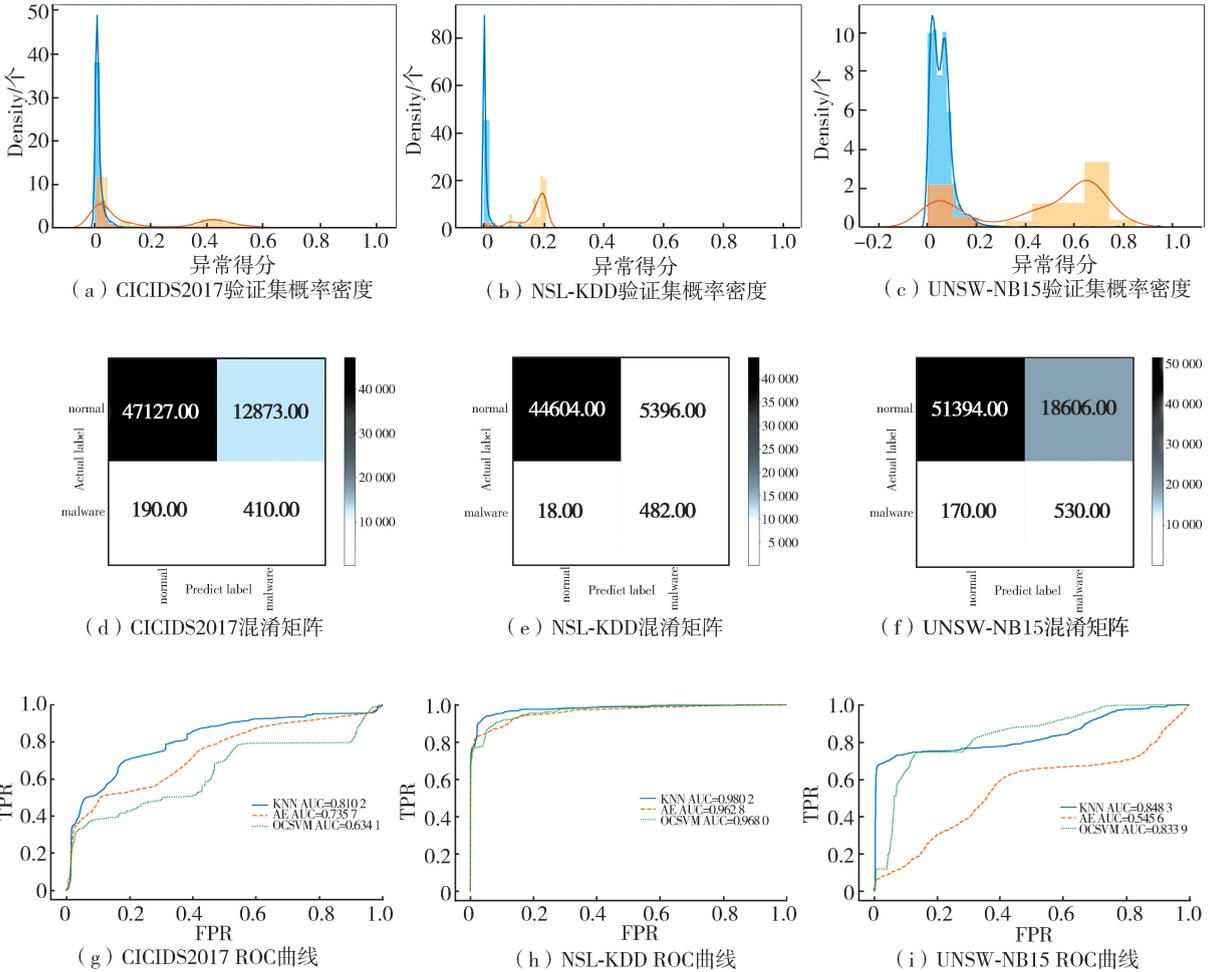


图 5 各类数据集的效果

表 7 阈值表现评价

数据集	参数	WP	MR	F1
UNSW-NB15	原始	0.987 0	0.742 8	0.847 7
	改进	0.989 2	0.826 7	0.900 7
NSL-KDD	原始	0.990 3	0.921 0	0.954 4
	改进	0.991 1	0.935 0	0.962 2
CICIDS2017	原始	0.000 1	0.500 0	0.000 2
	改进	0.986 4	0.734 4	0.841 9

最后通过混淆矩阵图 5(d~f)可以看出经过改进的 KNN 算法能够很好地检测异常流量,为系统提供报警服务。在 UNSW-NB15 数据集上,模型能够

识别 530 条恶意流量,170 条恶意流量被忽略;CICIDS2017 数据集上的 410 条恶意流量被模型识别,但仍有 190 条被遗漏;在 NSL-KDD 数据集中有 482 条流量被模型识别,效果最为显著。从表 7 可以看出,NSL-KDD 数据集流量较为简单,并不复杂。所以模型在原始和改进阈值下,两者效果并没有特别大的差异。另外两个数据集经过改进阈值选择方法后,效果得到显著提升。最后本次实验使用 NSL-KDD 测试集画出箱线图,如图 6 所示。在左列中正正常流量的异常得分较低。右侧则展现出恶意流量异常得分的分位线比左侧正常流量高。由此可以看

出, KNN 能显著检测异常流量。

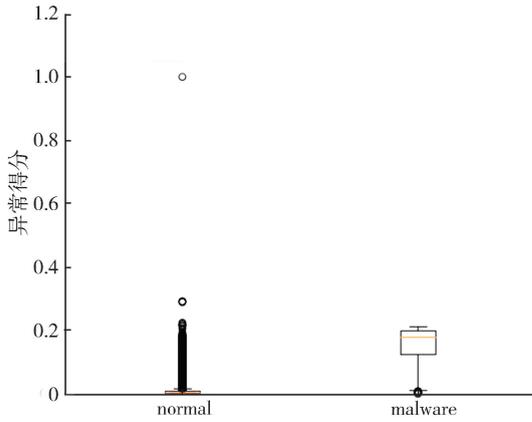


图 6 NSL-KDD 测试集异常得分箱线图

3.5 测试集结果对比分析

通过 3 种不同类型的异常检测模型在 NSL-KDD、UNSW-NB15 和 CICIDS2017 数据集上的测试对比,由图 5(g~i)可以看出, KNN 模型在这 3 类基准数据集上都有很好的效果。在 NSL-KDD 数据集上, KNN 算法相比 AE 和 OCSVM, 其 AUC 值仅仅高出约 0.01, 两者差距不大, 并不能凸显 KNN 算法的优越性。由于 NSL-KDD 数据集构建时间较早, 当时的网络流量较为简单, 该数据集的流量不具有当前网络的时效性。因此, 每个模型都能学习到其特征, 可以达到较好的效果。但是在 UNSW-NB15 和 CICIDS2017 这两个数据集上, 本文算法则展现出强大的性能。

UNSW-NB15 数据集是一个相对较新数据集。其中的流量种类也会更加丰富。在该数据集下, KNN 和 OCSVM 都取得比较好的效果, 但是 AE 表现却差强人意。AE 从模型上来说仅仅是维度的压缩和解压缩, 因此遇到未知的流量效果不一定好。

CICIDS2017 数据集是 3 个数据集中最新的数据集, 同时也是规模最大的数据集。在此数据集上, 充分显现出不同模型的特点。AE 作为深度学习模型在数据量特别大的情况下, 依然还有一定成效。AE 对比 KNN 算法的 AUC 值降低 0.074 5。OCSVM 效果最糟糕, 说明该算法不适用较大规模的数据。

从表 8 可以看出 KNN 在 CICIDS2017 数据集上有着很好的准确性。尽管在 F1 指标上, 与 AE 模型相比, 少了 0.022 8, 但是在 3 类数据集上都保持较高的水准。由于 AE 的神经网络结构是由图像迁移到网络流量中, 从 UNSW-NB15 的数据集中可以看

出该模型的不稳定性, 因此未必适合网络流量。反观 OCSVM 算法, 在 NSL-KDD 和 UNSW-NB15 这两个数据集上效果不错, 而运用到较大规模的数据集上, 该算法的性能就会被限制, 因此该算法也不适用于网络流量异常检测。实验结果表明, KNN 算法在检测异常网络流量方面效果更优。

表 8 3 种算法在 3 类数据集集中的表现评价

数据集	模型	F1	准确率
CICIDS2017	KNN	0.841 9	0.976 1
	OCSVM	0.727 4	0.554 5
	AE	0.806 1	0.838 5
NSL-KDD	KNN	0.962 2	0.892 7
	OCSVM	0.945 5	0.905 9
	AE	0.908 5	0.695 2
UNSW-NB15	KNN	0.900 7	0.734 5
	OCSVM	0.748 2	0.773 0
	AE	0.615 4	0.196 9

3.6 算法效率对比

随着基于人工智能的网络安全技术广泛运用, 模型的轻便性也受到人们的关注。本文在考虑算法效果的同时, 也对每一类算法在不同的数据集上做了效率评估。对每个数据集, 都采用同样大小的数据让不同的模型进行训练, 以此来检测不同模型所需要的时间。从表 9 来看, KNN 在每个数据集中, 相较于其他模型用时最少, 效率最高, 更加适合网络流量。

表 9 模型时间参数

模型	NSL-KDD	CICIDS2017	UNSW-NB15
KNN	14.31	45.86	14.32
AE	83.31	99.90	86.07
OCSVM	132.27	377.37	131.90

4 结束语

本文提出了一种改进的 KNN 算法来解决在网络流量结构复杂的环境下未知流量检测的问题。由于现网环境中捕获标记恶意流数据集具有挑战性, 因此本文提出的无监督算法不需要数据标记, 直接使用数据集进行训练, 降低标注成本。通过实验证明, 该算法在公共入侵检测数据集 NSL-KDD、UNSW-NB15 和 CICIDS2017 上表现出比 AE 和 OCSVM 更高的性能。

然而,基于流的无监督检测算法研究还很少且不成熟。在本文中,无监督算法虽然取得了一些成绩,但仍有很大的改进空间。未来,在改进入侵检测模型方面,实验可以增加其他真实网络的流量数据集,以印证算法的泛化能力。例如,模型放入工业环境中,通过工业内复杂的网络环境鉴别其可靠性。同时,在精确率方面,尝试寻找更适合网络流量的深度学习模型,可以让其更好地学习正常流量的多维分布,以达到提高检测异常网络流量精确率的目的。

参考文献:

- [1] 陈雪娇,王攀,俞家辉. 基于卷积神经网络的加密流量识别方法[J]. 南京邮电大学学报(自然科学版), 2018, 38(6): 36-41.
CHEN Xuejiao, WANG Pan, YU Jiahui. CNN based encrypted traffic identification method[J]. Journal of Nanjing University of Posts and Telecommunications (Natural Science Edition), 2018, 38(6): 36-41. (in Chinese)
- [2] SHUBAIR A, RAMADASS S, ALTYEB A A. KENFIS: KNN-based evolving neuro-fuzzy inference system for computer worms detection[J]. Journal of Intelligent and Fuzzy Systems, 2014, 26(4): 1893-1908.
- [3] GONG D, LIU L Q, LE V, et al. Memorizing normality to detect anomaly: memory-augmented deep autoencoder for unsupervised anomaly detection[C]//IEEE/CVF International Conference on Computer Vision (ICCV). 2019: 1705-1714.
- [4] ZHAO Y, NASRULLAH Z, LI Z. PyOD: a python toolbox for scalable outlier detection [J]. Journal of Machine Learning Research, 2019, 20(96): 1-7.
- [5] SCHLEGL T, SEEBÖCK P, WALDSTEIN S M, et al. F-AnoGAN: fast unsupervised anomaly detection with generative adversarial networks [J]. Medical Image Analysis, 2019, 54: 30-44.
- [6] AKCAY S, ATAPOUR-ABARGHOU EI A, BRECKON T P. GANomaly: semi-supervised anomaly detection via adversarial training[C]//Computer Vision (ACCV). 2019.
- [7] WOLLEB J, SANDKÜHLER R, CATTIN P C. DeScarGAN: disease-specific anomaly detection with weak supervision[C]//Medical Image Computing and Computer Assisted Intervention (MICCAI). 2020.
- [8] RAMASWAMY S, RASTOGI R, SHIM K. Efficient algorithms for mining outliers from large data sets[C]//Proceedings of the ACM SIGMOD International Conference on Management of Data. 2000: 427-438.
- [9] AMER M, GOLDSTEIN M, ABDENNADHER S. Enhancing one-class support vector machines for unsupervised anomaly detection[C]//Proceedings of the ACM SIGKDD Workshop on Outlier Detection and Description. 2013: 8-15.
- [10] FALCÃO F, ZOPPI T, SILVA C B V, et al. Quantitative comparison of unsupervised anomaly detection algorithms for intrusion detection [C] // Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing. 2019: 318-327.
- [11] CHEN Z M, YEO C K, LEE B S, et al. Autoencoder-based network anomaly detection[C]//Wireless Telecommunications Symposium (WTS). 2018: 1-5.
- [12] MOHAMED S, EJBALI R, ZAIED M. Denoising autoencoder with dropout based network anomaly detection [C]//ICSEA.2019: 110.
- [13] ZAVRAK S, İSKEFIYELI M. Anomaly-based intrusion detection from network flow features using variational autoencoder[J]. IEEE Access, 2020, 8: 108346-108358.
- [14] AN J, CHO S. Variational autoencoder based anomaly detection using reconstruction probability [J]. Special Lecture on IE, 2015, 2(1): 1-18.
- [15] ZENATI H, FOO C S, LECOAT B, et al. Efficient GAN-based anomaly detection[EB/OL].[2021-09-20]. <https://arxiv.org/abs/1802.06222>.
- [16] TAVALLAEE M, BAGHERI E, LU W, et al. A detailed analysis of the KDD CUP 99 data set[C]//IEEE Symposium on Computational Intelligence for Security and Defense Applications. 2009: 1-6.
- [17] SU M Y. Real-time anomaly detection systems for denial-of-service attacks by weighted k-nearest-neighbor classifiers[J]. Expert Systems With Applications, 2011, 38(4): 3492-3498.
- [18] PANIGRAHI R, BORAH S. A detailed analysis of CIC-IDS2017 dataset for designing intrusion detection systems [J]. International Journal of Engineering & Technology, 2018, 7(3): 479-482.
- [19] MOUSTAFA N, SLAY J. UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set) [C]//Military Communications and Information Systems Conference (MilCIS). 2015: 1-6.
- [20] REVATHI S, MALATHI A. A detailed analysis on NSL-KDD dataset using various machine learning techniques for intrusion detection[J]. International Journal of Engineering Research & Technology, 2013, 2(12): 1848-1853.
- [21] HABIBI LASHKARI A, DRAPER GIL G, MAMUN M S I, et al. Characterization of tor traffic using time based features[C]//Proceedings of the 3rd International Conference on Information Systems Security and Privacy. 2017: 253-262.