

基于自编码器的网络异常检测研究综述

张国梁, 郭晓军

西藏民族大学信息工程学院 咸阳 中国 710200

摘要 网络入侵检测技术是指对危害计算机系统安全的行为进行检测的方法,它是计算机网络安全领域中的必不可少的防御机制。目前,基于有监督学习的网络异常入侵检测技术具有较高的效率和准确率,该类方法获得了广泛关注,取得了大量的研究成果。但是这类方法需要借助大量标注样本进行模型训练。为减少对标注样本依赖,基于无监督学习或半监督学习的网络入侵检测技术被提出,并逐渐成为该领域的研究热点。其中,基于自编码器的网络异常检测技术是这方面技术的典型代表。该文首先介绍了各类自编码器的基本原理、模型结构、损失函数和训练方法。然后在此基础上将其分为基于阈值和基于分类的方法。其中,基于阈值的方法又可分为基于重构误差和基于重构概率两类。合适的阈值对异常检测技术的成败至关重要,该文介绍了三种阈值的计算方法。接着对比分析了多个代表性研究工作的方法、性能及创新点,最后对该研究中存在的问题做了介绍,并对未来的研究方向做了展望。

关键词 网络安全; 入侵检测; 异常检测; 深度学习; 自编码器

中图分类号 TP393 DOI号 10.19363/J.cnki.cn10-1380/tn.2023.03.07

An Overview of Network Anomaly Detection Based on Autoencoders

ZHANG Guoliang, GUO Xiaojun

Department of Information Engineering, Xizang Minzu University, Xianyang 710200, China

Abstract Network intrusion detection technology refers to a method of detecting behaviors that endanger computer system security, such as collecting vulnerability information, denying access, and obtaining system control rights beyond the legal scope. It is an indispensable defense mechanism in the field of computer network security. It is widely recognized in academia and industry. At present, the network anomaly intrusion detection technology based on supervised learning has high processing efficiency and detection accuracy. However, such methods require a large number of labeled samples for model training, and the acquisition of these labeled samples is difficult and expensive. In order to reduce the dependence on labeled samples, network intrusion detection technology based on unsupervised learning or semi-supervised learning has been proposed, and has gradually become a research hotspot in this field. Among them, the network anomaly detection technology based on autoencoder is a typical representative of this technology. This paper sorts out and sums up the representative work of autoencoders in network anomaly detection, and reviews related literatures. Firstly, the basic principles, model structures, loss functions and training methods of various autoencoders are introduced. Secondly, it can be divided into threshold based and classification based methods on this basis. Among them, the threshold based method uses an autoencoder to calculate the reconstruction error or reconstruction probability, which can be divided into reconstruction error based and reconstruction probability based methods. Appropriate thresholds are critical to the success or failure of anomaly detection techniques. This paper introduces three calculation methods for thresholds. The classification based methods use an autoencoder for feature learning and dimensionality reduction, followed by a classifier for anomaly detection. Then, the method characteristics, performance evaluation and innovation points of several representative research works are compared and analyzed. Finally, the existing problems in the research are introduced, and the future research direction is prospected.

Key words network security; intrusion detection; anomaly detection; deep learning; autoencoder

通讯作者: 张国梁, 硕士, 讲师, zgl@xzmu.edu.cn.

本课题得到西藏自治区自然科学基金项目(No. XZ2019ZRG-36(Z))和西藏民族大学“涉藏网络信息内容与数据安全团队”项目(No. 324042000709)的资助。

收稿日期: 2021-12-16; 修改日期: 2022-04-07; 定稿日期: 2023-01-04

1 引言

网络入侵目的是破坏计算机和网络的安全性、机密性、完整性和可用性。入侵检测系统(Intrusion Detection System, IDS)是网络安全防御机制中一个重要的组成部分,扮演着数字空间“预警机”的角色^[1]。常用的网络入侵检查方法可分为两类:基于误用的入侵检测(Misuse-based Intrusion Detection System, MIDS)和基于异常的入侵检测(Anomaly-based Intrusion Detection System, AIDS)。MIDS 是建立恶意模式数据库,使用匹配算法对输入数据模式匹配,如果和已知的模式匹配,则发出警报^[2]。MIDS 是发展最成熟、应用最广泛的技术,其优点是误报率低。为了降低误报率,它需要不断更新恶意数据库。大部分商用的 IDS 主要采用误用入侵检测方法,并依靠不断提升规则库的完备性来提升检测的可靠性^[3]。MIDS 的缺点是不能检测出零日的攻击。由此,诞生另一类入侵检测技术——基于异常的入侵检测,这类检测方法并不依赖检测数据库,而是在网络流量中发现不符合预期正常行为的异常模式。异常入侵检测是从待检测的数据中检测出异常行为或异常数据的重要数据分析任务。它在统计和机器学习中得到了广泛的应用,也被称为离群点检测、新颖性检测、偏差检测或异常挖掘^[4]。目前,基于异常的网络入侵检测技术能够很好地实现检测已知和未知的攻击,已经成为入侵检测研究的热门研究方法。因此,本文主要研究基于异常的入侵检测技术。

最近几年来,深度学习取得很大发展,同时对许多应用领域产生了很大影响,已有了大量的工作将深度学习技术应用到网络入侵异常检测中。通常情况下,监督学习效率高、性能好,大多数研究专注于有监督的深度学习在网络异常检测方面中应用研究。例如在文献^[5-11]中都有很好的体现。

然而,基于监督学习的网络入侵检测技术,在训练过程需要大量的、有标注的训练样本数据,才可以准确地检测出仅在训练过程中出现过(或类似)的异常数据。从本质上讲,基于监督学习的网络异常检测方法不足以检测出新的攻击^[12]。更重要的事,监督学习需要大量的有标注的数据。然而这样的数据需要资深网络安全专家才能完成,因此数据的获得是非常困难和昂贵的^[13]。半监督学习是监督学习和非监督学习的结合,利用标注和未标记的数据进行学习。半监督学习仅需要少量有标记的样本数据^[14];而无监督学习,不需要有标注的数据。当前有越来越多的研究者尝试将无监督深度学习和半监督学习网络

入侵检测中,并取得了不少成果。

同时国内外学术界已有文献^[2-4,15-22]分别从不同角度对网络入侵检测技术进行了综述。这些工作都提供了很有价值的信息,但是都是粗粒度来综述网络入侵检测技术,概括性强,但没有深入展开。通过对国内有关半监督学习和无监督学习的网络入侵检测方面论文进行分析和梳理发现,在这一类研究中使用频率最多的深度学习模型是自编码器(Autoencoder, AE)。因此本文对自编码器在网络异常检测中代表性工作进行了梳理和总结。首先介绍各类自编码器的基本原理、模型结构、损失函数和训练方法。接着分别从基于阈值的方法和基于分类的方法描述了各类自编码器在网络异常检测技术中的应用。基于阈值的方法又可进一步分为重构误差(异常分数)和重构概率(正常分数),并对这两类阈值的计算方法进行对比分析。最后为方便快速且直观地了解当前的研究动态,归纳并总结文中的几类自编码器在网络入侵异常检测上的优劣,并对存在问题进行了梳理,对研究前景进行了展望。

本文结构如下:第 2 节介绍网络入侵检测基准数据集、预处理和评估测量方法;第 3 节对基于自编码器的网络异常检测技术进行全面分析;第 4 节总结当前存在的问题,展望未来研究;第 5 节结束语。

2 基准数据集及预处理和评估测量方法

2.1 常见网络入侵检测基准数据集级

有代表性的基准数据集在研究和评估网络入侵检测技术的性能方面起着重要作用。基准数据集是评估和比较不同网络入侵检测系统质量的良好基础。在数据集评估框架中,确定了构建可靠的基准数据集所需的 10 条标准,包括:完整的网络配置;完整的流量;完整的交互;标注的数据集;完全捕获;可用协议;攻击多样性;异构性;功能集;元数据^[22]。通过表 1 对四种常见的基准数据集进行对比说明。

表 1 四种常见基准数据集对比

Table 1 Comparison of four common benchmark datasets

参数	KDD CUP 99	NSL- KDD	UNSW- NB15	CICIDS 2017
创建年份	1998	2009	2015	2017
网络流量类型	仿真	仿真	仿真	仿真
数据量	5M	150K	2M	3.1M
持续时长	7 周	7 周	31 小时	5 天
特征数量	41	41	47	80
攻击特征数量	4	4	9	8

2.1.1 KDD CUP 99 数据集

KDD CUP 99 是基于 DARPA 数据集, 是网络入侵检测中使用最广泛的数据集。该数据集中的一个网络连接被定义为某个时间内从开始到结束的 TCP 数据包序列。每个网络连接被标注为正常(normal)或异常(attack), 其中, 异常类型又被细分为 39 种类型, 其中 22 种攻击类型出现在训练集中, 另有 17 种未知攻击类型出现在测试集中。在训练集中包含了 1 种正常的标识类型和 22 种训练攻击类型。异常分为四个主要类别: 1)DOS 拒绝服务; 2)R2L: 来自远程计算机的未经授权的访问, 例如猜测密码; 3)U2R: 未经授权访问本地超级用户(root)特权, 例如缓冲区溢出攻击; 4)探测: 监视和其他探测, 例如端口扫描^[23]。该数据集的出现时间较早, 但它为基于机器学习的网络入侵检测研究奠定了良好的基础。

2.1.2 NSL-KDD 数据集

KDD CUP 99 数据集中最重要的缺陷之一是在训练集和测试集中分别有 78%和 75%的重复记录, 这导致学习算法偏向于频繁记录, 从而阻止学习诸如 U2R 之类威胁更大的不频繁记录^[22]。NSL-KDD 数据集是 KDD CUP 99 数据集的改进, 去除了大量冗余数据。NSL-KDD 数据集的优点是: 在训练集中没有冗余记录, 所以分类器不会产生任何有偏的结果; 测试集中没有重复的记录, 这些记录具有更好的还原率; 样本的数量与特定攻击类别的难度成反比^[24]。相比 KDD99 而言, NSL-KDD 数据集中的训练和测试的记录数量是合理的; 2)R2L: 来自远程计算机的未经授权的访问, 例如猜测密码; 3)U2R: 未经授权访问本地超级用户(root)特权, 例如缓冲区溢出攻击; 4)探测: 监视和其他探测, 例如端口扫描^[23]。该数据集的出现时间较早, 但它为基于机器学习的网络入侵检测研究奠定了良好的基础。

2.1.3 UNSW-NB15 数据集

UNSW-NB15 数据集的原始网络数据包是在澳大利亚网络安全中心(ACCS)的实验室网络中创建的^[17]。使用 IXIA 完美风暴工具用于创建现代网络场景和九类攻击流量的混合。由于它包含了 CVE(公共已知信息安全漏洞字典)中不断更新的新攻击类型的信息, 因此该数据集捕捉了现代网络入侵特征。进行了两种模拟场景, 一种是 16 小时, 1 次攻击/秒, 另一种是 15 小时, 10 次攻击/秒。通过 IXIA 报告对数据集进行标注。从网络轨迹中提取了 43 个特征, 包括基本特征(14 个)、内容特征(8 个)、时间特征(9 个)、连接特征(7 个)和附加特征(5)^[15]。它包含九种不同的攻击类型, 即: 模糊器、分析、后门、DoS、漏洞利用、

泛型、侦察、Shellcode 和蠕虫。使用了 Argus, Bro-IDS 工具, 并开发了 12 种算法以生成带有类标签的 49 个特征^[25]。数据集也以基于流的格式提供, 并带有其他属性。UNSW-NB15 是一个复杂的数据集, 它代表了现代网络和攻击流量, 可以用于可靠的评估。

2.1.4 CICIDS2017 数据集

CICIDS 2017 是在一个模拟环境中创建的, 历时 5 天, 包含基于包和双向流格式的网络流量。对于每个流, 作者提取了 80 多个属性, 包含正常网络数据和最新的常见攻击。它还包括使用 CICFlowMeter 进行网络流量分析的结果, 并基于时间戳、源和目标 IP、源和目标端口, 协议和攻击标注流量。也可以使用提取的特征定义、协议和攻击的标注流量^[22]。基于 HTTP、HTTPS、FTP、SSH 和电子邮件协议构建了 25 个用户的抽象行为。主要攻击包括: 蛮力 FTP、蛮力 SSH、DoS、Heartbleed、Web 攻击、渗透、僵尸网络和 DDoS 等攻击。对于每个流提供了有关 IP 地址和攻击的元数据^[26]。

文献^[6-7,10,23-24]使用 KDD CUP 99 和 NSL-KDD 两个数据集来验证模型的有效性。文献^[25-35]中只使用了 NSL-KDD 数据集。文献^[5,36]使用 UNSW-NB15, 文献^[6,37]使用 CICIDS 2017。在这四个数据集中 NSL-KDD 是使用频率最高的, 是因为该数据集中, 测试数据中部分类别没有在训练集中出现, 从而增加了模型预测的难度。更多的工作选择该数据集是为了能够更好地验证模型的泛化性能。UNSW-NB15 和 CICIDS 2017 是相对比较新的数据集, 因此当前使用相对较少。

在选择入侵检测数据集, 尽量选择符合上面提到的 10 项准则, 同时还要尽量选择较新的数据集。因为使用较新的数据集将会更具说服力, 更能有效地检测入侵行为。在这四种数据集中只有 CICIDS2017 能基本满足上面提到 10 条标准^[22]。

2.2 数据预处理

数据预处理是机器学习理论的重要步骤, 因为从网络流量中提取的数据往往控制松散, 导致数据值不相关或重复。它通过去除冗余、噪声或不相关的信息来过滤网络数据, 从而提高检测攻击行为的性能。网络数据的数据预处理包括特性的创建、缩减、转换和规范化^[19]。本节主要介绍特征的数字化和标准化。

1) 特征的数字化: 数据集中一般会有一些符号型的特征, 不能直接进行数值计算, 必须对这些特征进行数字化。KDD CUP 99 的 41 个特征中有 3 个非数字特征。这 3 个特征分别是: Protocol(协议)、

Service(网络服务)和 Flag(连接状态)。UNSW NB15 数据集有 41 个数字特征和 3 个非数字特征^[19]。这些非数字化特征可以根据分布情况指定具体数值。为了能够更好体现这些特征取值差异,也可以使用 One-Hot 编码等方法进行转换。如果协议是 TCP 表示为(0, 0, 1), 如果协议是 UDP 表示为(0, 1, 0), 如果协议是 ICMP 表示为(1, 0, 0)。服务和标志以相同的方式编码。有三种不同的协议, 70 种不同的服务, 11 种不同的标志和其他 38 个数字功能。因此, 有总共 122 个功能^[3]。

2) 对数据集归一化,数据集中各属性取值范围差异很大:例如在 KDD CUP 99 中大部分属性取值 0 到 9 的个位数,但还有一些是属性的取值范围是几千或几万范围。这样不利于机器学习算法的计算。因此在数据预处理阶段需要对数据集进行归一化处理。常见的归一化函数是线性变换和 z-score, 如式(1)和式(2)所示^[19]。

$$X_n = \frac{(X - \min(X))}{((\max(X) - \min(X)))} \quad (1)$$

$$Z = \frac{(X - \mu)}{\sigma} \quad (2)$$

在(2)式中 X 表示特性值, μ 是特征值的均值, σ 是特征值的标准偏差。

2.3 评估测量方法

在网络入侵检测领域,通过统计方法评估模型的性能。将数据集中的攻击连接标注为“正样本”记录,正常连接标注为“负样本”记录,形成二分类问题^[38]。分别使用以下术语确定分类模型的质量:

1) 真阳性(True Positive TP): 网络连接被正确分类为攻击的数量。

2) 真阴性(True Negative TN): 网络连接被正确分类为正常的数量。

3) 假阳性(False Positive FP): 正常网络连接被错误分类为攻击的数量。

4) 假阴性(False Negative FN): 攻击网络连接被错误分类为正常的数量。

分类结果的“混淆矩阵”^[39]。如表 2 所示。

表 2 分类结果混淆矩阵

Table 2 Classification result confusion matrix

真实情况	正例(预测)	反例(预测)
正例	真阳性(TP)	假阴性(FN)
反例	假阳性(FP)	真阴性(TN)

基于上述术语,考虑以下最常用的评估指标。基于上述术语,考虑以下最常用的评估指标^[20]。

1) 准确性(Accuracy ACC): 用于计算正确识别的

连接记录与整个测试数据集的比率。如果准确性较高,则机器学习模型会更好(准确性 $\in [0, 1]$)。正确性定义如下:

$$ACC = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

2) 精度(Precision PR): 用于计算正确识别的攻击连接记录与所有已识别攻击连接记录的数量之比。如果精度较高,则机器学习模型会更好。精度定义如下:

$$PR = \frac{TP}{TP+FP} \quad (4)$$

3) 查全率(RE, Recall): 也称为召回率。它计算正确分类的攻击连接记录占攻击连接记录总数的比率。如果 RE 较高,则机器学习模型会更好。定义如下:

$$RE = \frac{TP}{TP+FN} \quad (5)$$

4) F1-分数(F1-Score): F1-Score 也称为 F1-Measure。它是精确度和查全率的调和平均值^[22]。如果 F1-分数越高,精确度和检测率都高机器学习模型越好。定义如下:

$$F1-Score = 2 \times \left(\frac{PR \times RE}{PR + RE} \right) \quad (6)$$

5) 误报率(False Positive Rate, FPR): 用于计算标注为“攻击”的正常连接记录占总连接记录的比率。如果 FPR 较低,则机器学习模型会更好。如果 FPR 较高,安全人员会用较多的时间用来处理假的入侵报警,而真正的报警可能被埋没。FPR 定义如下:

$$FPR = \frac{FP}{FP+TN} \quad (7)$$

6) ROC(Receiver Operating Characteristic)曲线: 根据不同阈值在 y 轴上的真正例率(True Positive Rate, TPR), 与 x 轴上的假正例率(FPR)之间的权衡取舍,绘制了 ROC。ROC 曲线下的面积(AUC)是 ROC 曲线下的面积的大小,与 ROC 一起用作机器学习模型的比较指标。如果 AUC 较高,则机器学习模型会更好^[38]。

3 基于自编码器的网络异常检测技术

3.1 常用自编码器模型

自编码器是一种用于无监督方式学习的人工神经网络。该网络由两部分组成: 一个由函数表示的编码器和一个生成重构的解码器, 分为为表示: $h = f(x)$ 和 $\hat{x} = g(h)$ 。编码器可以学习输入的数据表示(编码); 解码器通过学习从编码中还原或重构输入数据。自编码器将 x 和 h 看作是随机变量, 将编码器和解码器概念推广, 将确定函数扩展为随机映射 $p_{encoder}(h|x)$ 和 $p_{decoder}(x|h)$ ^[40]。图 1 展示了自编码

器各个变量之间的关系图。

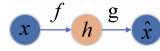


图1 AE中各变量之间关系图^[40]

Figure 1 Relationship diagram between variables in AE^[40]

自编码器模型存在多种变体。通过约束隐藏层 h 的维数形成欠完备自编码器(Undercomplete Autoencoder, UAE)和过完备自编码器(Overcomplete Autoencoder)。通过在损失函数中加入正则化方法形成: 稀疏自编码器(Sparse AutoEncoder, SAE)、去噪自编码器(Denoising Autoencoder, DAE)、收缩自编码器(Contractive Autoencoder, CAE)。将多个自编码器堆叠形成堆叠自编码(Stacked Autoencoders, StackAE)。将自编码器与隐变量模型理论结合形成的变分自编码器(Variational Autoencoder VAE)^[40]。常用自编码器分类如图2所示。

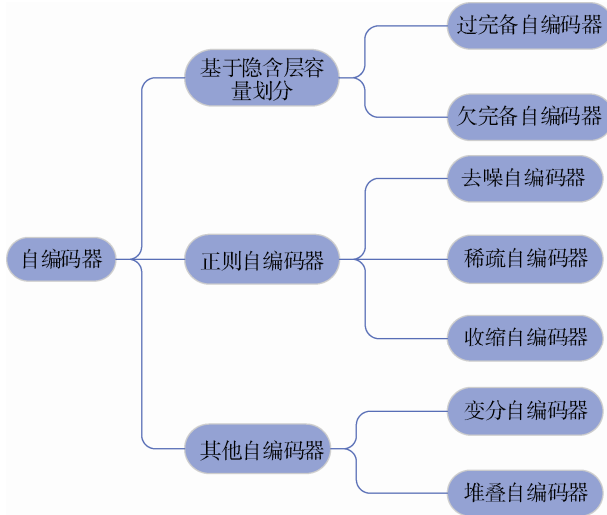


图2 常用自编码器分类图

Figure 2 Autoencoder classification diagram

3.1.1 欠完备自编码器和过完备自编码器

在设计自编码器的结构时, 隐藏层的维数是关键性的参数。依据该参数和输入层维数的大小关系, 将自编码器分为欠完备自编码器和过完备自编码器。

在通常情况下, 希望自编码器能够学习到训练数据的有用信息。但是如果自编码器的容量太大, 那训练出执行复制任务的自编码器可能无法学习到数据集的任何有用信息。因此在自编码器中设计隐藏层 h 维度小于输入层 x 的维度, 将这类自编码器称为欠完备自编码器, 反之称为过完备自编码器。欠完备自

编码器的编码器部分实现了对输入数据的压缩或降维, 因此可以捕捉输入数据中显著的特征^[40]。其结构如图3所示。

对于过完备自编码器来说, 不具这样的功能。但是可以通过在损失函数加入一定的正则项来鼓励模型学习某些特性。这两类自编码器的损失函数的一般形式为:

$$\theta, \phi = \arg \min_{\theta, \phi} L(x, g_{\phi}(f_{\theta}(x))) \quad (8)$$

其中 $h = f_{\theta}(x)$ 为编码器, $g_{\phi}(f_{\theta}(x))$ 为解码器^[40]。损失函数 L 表示输入数据的重构误差。通过最小化重构误差计算 θ, ϕ , 得到最优模型。

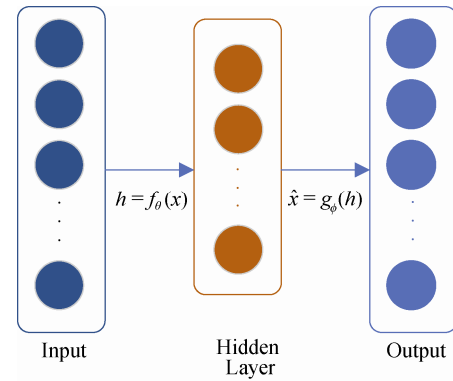


图3 欠完备自编码器结构

Figure 3 Architecture of Undercomplete Autoencoder

3.1.2 正则自编码器

正则自编码器通过在损失函数中加正则项, 对于优化结果进行约束。强制模型优先重建输入数据的某些部分, 因此这样可能学习到数据的有用特性。

(1) 稀疏自编码器

稀疏自编码器在损失函数中加入了编码层的稀疏惩罚项 $\Omega(h)$ 。损失函数一般形式:

$$\theta, \phi = \arg \min_{\theta, \phi} L(x, g_{\phi}(f_{\theta}(x))) + \alpha \Omega(h) \quad (9)$$

其中 $h = f_{\theta}(x)$ 为编码器, $g_{\phi}(f_{\theta}(x))$ 为解码器, $\Omega(h)$ 是稀疏正则项, α 惩罚系数。损失函数 L 表示输入数据的重构误差。通过最小化重构误差计算 θ, ϕ , 得到最优模型。同时通过足够大的 α 实现稀疏化, 使得隐藏层 h 中的一些参数为 0, 实现稀疏化。因此, 该方法可以用于特征选择^[40]。

(2) 去噪自编码器

去噪自编码器在训练过程中, 输入的数据 \tilde{x} 是被某种噪声“损坏”的 x 的副本, 通过训练能够预测原始未被损坏数据。能够恢复出原始信号的模型是并非最好, 但是能够对“损坏”的原始数据编码、解码, 然后还能恢复真正的原始数据, 这样模型将具

有较强的鲁棒性, 从而增强模型的泛化能力。添加噪声的另一个好处是减少了生成模型的过拟合^[13]。定义两个随机变量 x 和 \tilde{x} , x 表示数据集中样本数据, \tilde{x} 表示样本数据被损坏后的数据。引入条件分布 $C(\tilde{x}|x)$ 破坏 x , 将其变为 \tilde{x} 。其结构如图 4 所示。去噪自编码器的其损失函数一般形式:

$$\theta, \phi = \arg \min_{\theta, \phi} L(x, g(f(\tilde{x}))) \quad (10)$$

其中 $h = f(\tilde{x})$ 为编码器, $g(f(\tilde{x}))$ 为解码器。损失函数 L 表示输入数据的重构误差。通过最小化重构误差计算 θ, ϕ , 得到最优模型。

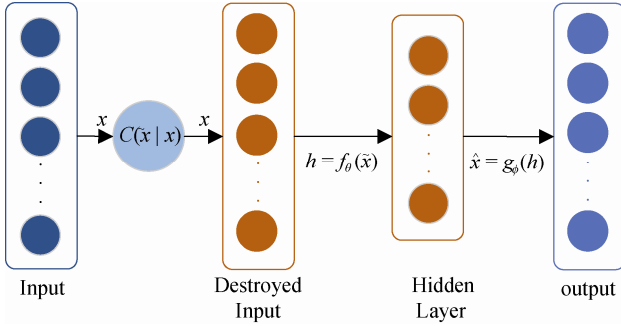


图 4 去噪自编码器结构^[13]

Figure 4 Denoising Autoencoder Architecture^[13]

3.1.3 其他自编码器

(1) 变分自编码器

变分自编码器是一种概率生成模型, 它模拟两个随机变量 x 和 z 之间的关系。用 x 表示观测变量, z 表示隐变量。将数据集 $X = \{x^{(i)}\}_{i=1}^N$ 看作是随机变量 x 相互独立的样本。假设该数据是由隐变量 z 确定的随机过程生成^[41]。数据 x 的生成包括两个步骤:(1)由先验分布 $p_\theta(z)$ 生成 $z^{(i)}$;(2)再由条件概率分布 $p_\phi(x|z)$ 生成 $x^{(i)}$ 。VAE 的核心是如何确定隐变量 z 的概率分布和参数 θ 。

在 VAE 中数据 x 的生成概率可以用公式(11)表示^[42]:

$$p(x) = \int p(x|z; \theta) p(z) dz \quad (11)$$

VAE 的目标使训练数据中每一个样本的生成概率最大化。通过最大化 $p(x)$ 来求解隐变量 z 概率分布和参数 θ 。直接求解 $p(x) = \int p(x|z; \theta) p(z) dz$ 和后验概率 $p_\theta(z|x)$ 是不可行的。引入一个容易求解的 $q_\phi(z|x)$ 近似逼近真实后验概率 $p_\theta(z|x)$ 。 $q_\phi(z|x)$ 为近似后验概率分布, 现在转为求参数 ϕ 和 θ , 可以通过变分推断来求解。图 5 表述了观测变量和隐变量 z 关系。

由公式(12)^[42]结合贝叶斯公式和对数似然函数可以推导出公式(12):

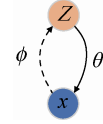


图 5 VAE 观测变量 x 和隐变量 z 关系图

Figure 5 The relationship between the observed variable x and the latent variable z of the VAE

$$\log p_\theta(x^{(i)}) = E_{q_\phi(z|x^{(i)})} [\log p_\theta(x^{(i)}, z) - \log q_\phi(z|x^{(i)})] + D_{KL}(q_\phi(z|x^{(i)}) || p_\theta(z|x^{(i)})) \quad (12)$$

第二项 D_{KL} 表示近似后验分布和真实后验分布的差异度, 是非负的值。从公式(12)推导出公式(13)。

$$\log p_\theta(x^{(i)}) \geq \mathcal{L}(\theta, \phi; x^{(i)}) = E_{q_\phi(z|x^{(i)})} [\log p_\theta(x^{(i)}, z) - \log q_\phi(z|x^{(i)})] \quad (13)$$

将 $\mathcal{L}(\theta, \phi; x^{(i)})$ 记作 $ELBO$ (evidence lower bound), 称为变分下届。当 $q_\phi(z|x^{(i)})$ 和 $p_\theta(z|x^{(i)})$ 相等时 D_{KL} 项为零, 公式(13)的等号成立。因此问题转为最大化 $ELBO$ 。 $ELBO$ 可以变形为:

$$\mathcal{L}(\theta, \phi; x^{(i)}) = E_{q_\phi(z|x^{(i)})} [\log p_\theta(x^{(i)}|z)] - D_{KL}(q_\phi(z|x^{(i)}) || p_\theta(z)) \quad (14)$$

该式中右边第一项为条概率分布 $\log p_\theta(x|z)$ 关于近似后验 $q_\phi(z|x)$ 的期望, 可以看作负的重构误差项。第二项为近似后验和隐变量 z 的先验的交叉熵, 可以看作正则化项。在 VAE 中 $q_\phi(z|x)$ 看作是编码器, 而 $p_\theta(x|z)$ 看作解码器。

从理论来讲, 可以使用随机梯度估计的方法最大化目标函数 $\mathcal{L}(\theta, \phi; x^{(i)})$, 求出参数 ϕ 和 θ 。但是直接使用随机梯度估计方法会有很大的方差, 在实际计算中是不可行。通常可以使用重参数化技巧来克服这个问题。

重参数化技巧方法的原理: 将 $z \sim q_\phi(z|x)$ 表示为: $z = g_\phi(\epsilon, x)$, $\epsilon \sim p(\epsilon)$ 。其中 g_ϕ 为可微的确定性转换函数, ϵ 是服从简单分布的随机变量(通常选择 ϵ 服从标准高斯分布)^[42]。通过这种变换, g_ϕ 和 ϵ 有了确定形式。可以用随机梯度估计的方法来最大化目标函数 $\mathcal{L}(\theta, \phi; x^{(i)})$, 求出参数 θ, ϕ 。公式(14)变换为(15)形式^[42]。

$$\mathcal{L}(\theta, \phi; x^{(i)}) \simeq \frac{1}{L} \sum_l \left(\log p_\theta(x^{(i)} | z^{(i,l)}) \right) - D_{KL}(q_\phi(z|x^{(i)}) || p_\theta(z)) \quad (15)$$

其中 $z^{(i,l)} = g_\phi(\epsilon^{(i,l)}, x^{(i)})$ 同时 $\epsilon^{(i,l)} \sim p(\epsilon)$ 。

(2) 堆叠自编码器

堆叠自编码器是指将多个自编码器堆叠形成多层次的自编码器。模型结构如图 6 所示。采用贪心

逐层方式来训练模型中的每一个自编码器。训练具体步骤: 第一个 AE 通过最小化原始输入数据的重构误差来训练隐藏层, 训练完成后获得隐藏层 h^1 , 再将第一个 AE 的隐藏层的输出作为第二个 AE 的输入; 第二步对再用相同方法训练第二个 AE, 获取 h^2 。这样, 一层一层地训练, 直到获得最后一个 AE。将在

训练好的隐藏层逐层接在一起, 就可以完成逐层特征提取的任务^[43]。这样得到的特征更有代表性, 且网络维度较小, 降低运算复杂度。

由于堆叠自编码的隐藏层包括多层, 因此也称为深度自编码器。可以将欠完备自编码、稀疏自编码器和去噪自编码器等均可构建为堆叠自编码器。

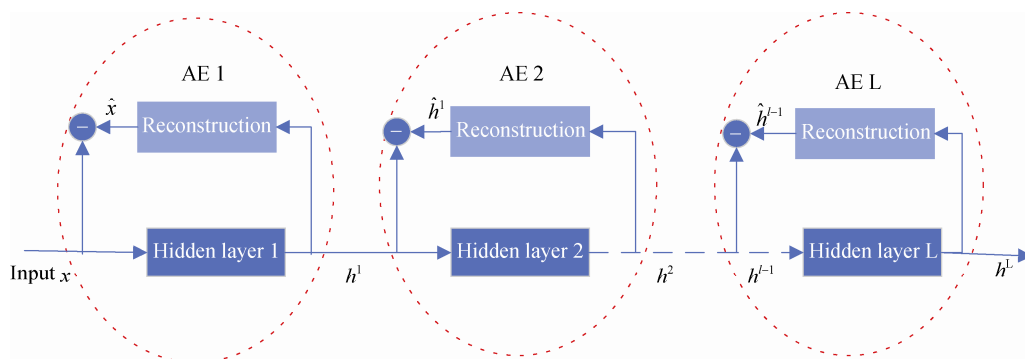


图 6 堆叠自编码器结构^[43]

Figure 6 Stacked Autoencoder Architecture^[43]

3.2 基于自编码器的网络异常检测技术

本文对基于自编码器的网络异常检测技术进行了分类, 如图 7 所示。首先根据网络异常检测方法分为基于阈值的方法和基于分类的方法。前一种方法进一步分为重构误差和重构概率两种。后者一种方法自编码器用于对数据进行特征学习和降维, 用分类器进行异常检测。

3.2.1 基于阈值的网络异常检测技术

基于阈值的检测方法基本原理是, 当数据通过检测系统时都会给出相应的分数(异常分数或正常分数), 再根据预设阈值就可以判断数据是否异常。因此这类方法可以进一步分为重构误差和重构概率两种方法。前者将数据的重构误差作为异常分数, 当异常分数大于预设的阈值时为异常, 反之为正常。后者

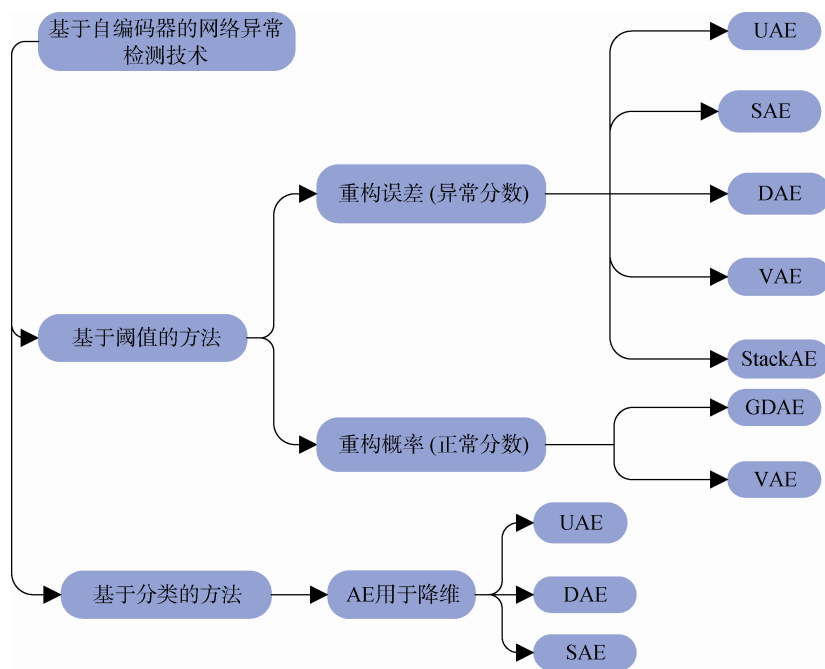


图 7 基于自编码器的网络异常检测分类图

Figure 7 Classification diagram of network anomaly detection based on autoencoder

将数据的重构概率作为正常分数, 当正常分数小于预设的阈值时为异常, 反之为正常。

(1) 基于重构误差(异常分数)的方法

基重构误差的网络异常检测技术是指使用训练样本的重构误差来训练自编码器。基重构误差的网络异常检测技术过程如图 8 所示。首先对数据集进行数字化、归一化。接着利用重采样将数据集分为训练数据集、验证数据集和测试数据集。在模型训练阶段, 损失函数中都是以样本数据重构误差作为基本项, 在此基

础可以加上各类正则项。模型训练过程仅使用正常样本数据。验证数据集用于计算检测阈值。异常检测时, 当数据通过检测系统时都会生成相应的异常分数, 再根据预设阈值就可以判断数据是否异常。这种方法的基本假设是测试数据与训练数据相似的数据具有低重构误差。而与训练数据不同的异常数据将具有较高的重构误差。在测试阶段, 当重构误差大于给定阈值时, 确定为异常, 反之亦然。下面具体介绍常用的自编码器进行异常检测的具体过程。

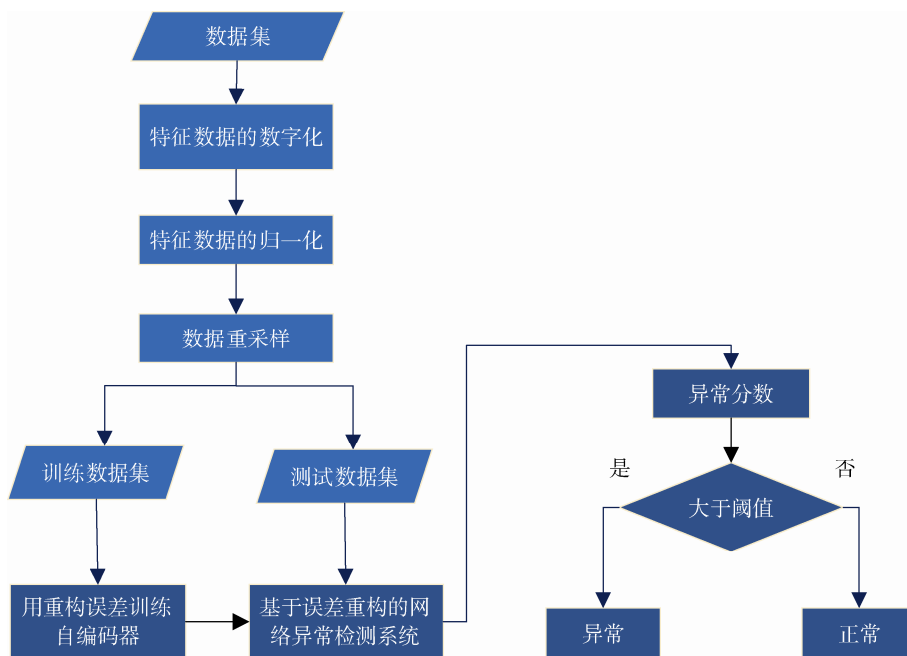


图 8 基于重构误差的网络异常检测技术过程图

Figure 8 Process diagram of network anomaly detection technology based on reconstruction error

文献^[12,27,29,44-45]使用了欠完备自编码器进行网络异常检测。模型训练时仅选用正常流量数据, 不需要标注数据, 自编码器可以学习到正常样本数据信息, 并进行重构。公式(8)损失函数具体形式可以选择误差平方和(Sum of squared errors, SSE)或交叉熵(Cross-Entropy), 如公式(16)和(17):

$$L_{res}(x - \hat{x}) = \sum (x_i - \hat{x}_i)^2 \quad (16)$$

$$L_{rec}(x - \hat{x}) = \sum [x_i \log \hat{x}_i + (1 - x_i) \log(1 - \hat{x}_i)] \quad (17)$$

当测试样本通过训练好的自编码器时, 将产生重构误差(异常分数), 通过该值来进行异常检测。令重构误差为 RE , 预先设定的异常检测阈值为 τ 。当 $RE < \tau$ 时为正常, 反之为异常。为了提高检测效率, 可以将多个自编码器集成一起使用。Mirsky Y 等人^[46]提出了一种基于自编码器的小型网络入侵检测系统, 数据不需要标注, 以无监督学习方式训练, 并以一种高效的在线方式检测本地网络的攻击。核心算法将多个自编码器集成在一起, 组成一个功能强大网络

入侵检测系统。该系统中, 在特征数据提取之后加入了特征映射层, 将特征数据根据相关性进行划分、并映射为 k 个向量。 k 个向量分别对应 k 个自编码器。该系统由两级组成: 第一级为集成层, 由 k 个自编码器组成; 第二级为输出层, 包含一个自编码器。使用均方根误差(RMSE)进行模型训练。将 k 个自编码器输出的 RMSE 归一化后组成一个向量, 作为输出层自编码器的输入。最后输出层输出误差平方和作为异常分数, 通过预设的阈值判断输入是否为异常。由于对数据特征进行了划分, 使得每一个自编码器的网络规模变得较小, 从而降低系统复杂度, 同时提高性能。根据评估表明, 该检测系统可以检测出各种攻击, 其性能可与离线异常检测器媲美。

稀疏自编码器除了可以将重构误差作为异常分数之外, 还可以将稀疏率作为异常分数。稀疏自编码器在损失函数中加入了可以稀疏隐含层的正则项。公式(18)损失函数的一般形式:

$$J_{sparse} = \|X - \hat{X}\|_2^2 + \beta \sum_{j=0}^m KL(\rho \| \hat{\rho}) \quad (18)$$

在 Gharib 等人^[12]提出通过串联两个自编码器构成检测器 D1 和 D2 检测网络异常。D1 为稀疏自编码, D2 为欠完备自编码器。先利用 D1 对流量进行初步检测, 再用 D2 进行精细检测。通过串联两个自编码器不仅可以提高检测效率, 同时还能提高检测准确率。在 D1 中计算隐含层单元激活率 $\hat{\rho}$, 并预先设定一个目标稀疏因子 ρ (超参数)。将 ρ 和 $\hat{\rho}$ 的 Cross-Entropy 作为稀疏的正则项加入损失函数中, 具体形式如公式(18)。D1 可以从正常流量中学习稀疏表示。当面对异常样本时, 隐藏层单元的激活率会变大。因此可以将激活率作为异常分数来检测是否异常。

(2) 基于重构概率(正常分数)的方法

对于概率生成模型来说, 可以计算出样本数据的重构概率。在利用概率生成模型进行网络异常检测时, 就可以将数据的重构概率值作为正常分数来判断是否异常的依据。这类方法具体过程如图 9 所示。首先对数据集进行预处理、并将其分为训练数据(仅含正常流量数据)和测试数据, 以无监督方式训练模型, 接着使用训练好的模型计算测试数据的重构概率, 将该值作为正常分数。如果正常分数小于预设的阈值, 则判定为异常。反之为正常。这种方法的基本原理是: 训练模型时仅用正常流量数据, 因此正常流量数据将有很高重构概率, 而异常数据方差较大, 重构概率较低。在整个过程中不需要生成样本

数据, 只需要计算出重构概率即可。这样可以减小计算量, 提高检测效率。

An 等人^[45]中提出了一种使用 VAE 重构样本数据概率进行异常检测的方法。在使用 VAE 的具体实践中, 隐变量 z 的先验分布 $p_\theta(z)$ 和近似后验 $q_\phi(z|x)$ 通常选择各向同性高斯分布, 似然的分布 $p_\theta(x|z)$ 选择多元高斯分布。令 z 的先验分布为 $p_\theta(z) = \mathcal{N}(0, I)$, 近似后验概率分布为 $q_\phi(z|x) = \mathcal{N}(z; \mu_z, \sigma_z)$, 似然概率分布 $p_\theta(x|z) = \mathcal{N}(z; \mu_x, \sigma_x)$ 。通过优化目标函数公式(7)训练数据来确定参数 $\phi\{\mu_z, \sigma_z\}$ 和 $\theta\{\mu_x, \sigma_x\}$ 。概率重构输入变量分布的参数, 而不是输入变量本身。Zavrak S 等人^[39]中提出利用这一方法进行网络异常检测。

(3) 阈值的计算方法

在基于阈值的网络异常检测方法中, 确定合适的阈值是异常检测成败的关键。如果阈值不合适, 则可能无法执行有效的异常检测。异常检测阈值的计算方法有三种: 朴素方法、随机方法和上分位数法。值计算方法分类如图 10 所示。

朴素方法: 如果将重构误差值作为阈值。最朴素的做法是, 在模型训练阶段, 仅用正常流量数据, 得到训练好的模型后, 对验证数据再进行重构, 并记录重构误差值, 选取最大值作为阈值^[36]。

随机方法: Aygun 等人^[47]中假设正常流量数据重构误差和异常数据重构误差分别服从不同高斯分布,

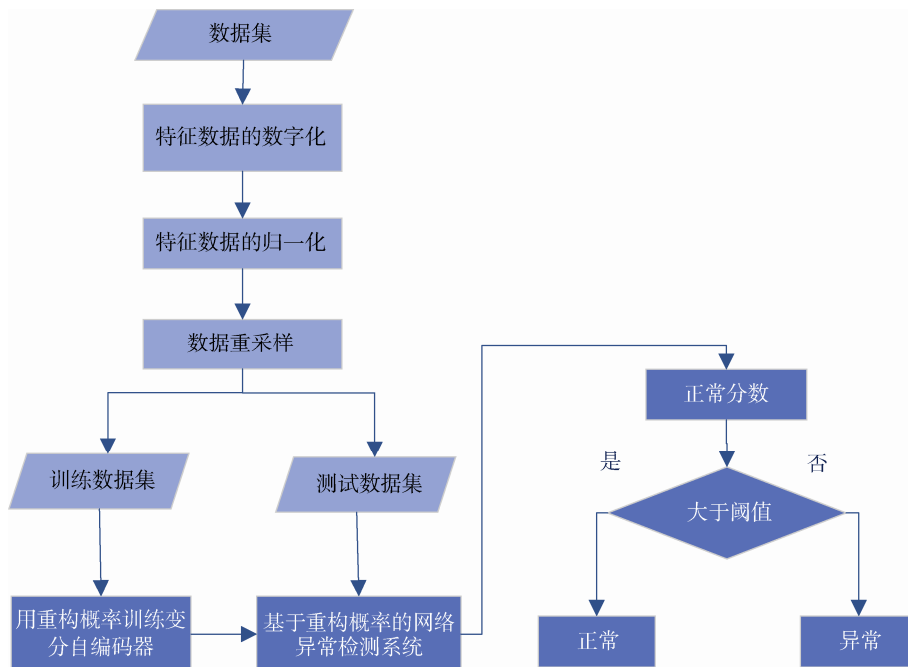


图 9 基于重构概率的异常检测技术过程图

Figure 9 Process diagram of anomaly detection technology based on reconstruction probability

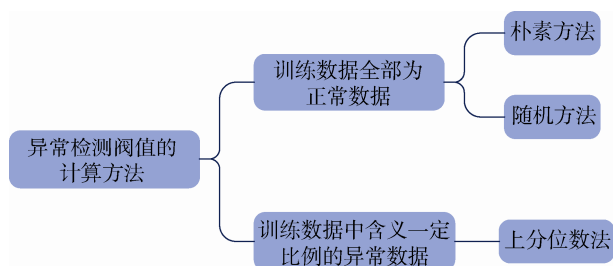


图 10 异常检测阈值计算方法分类图

Figure 10 Classification diagram of anomaly detection threshold calculation methods

正常流量数据重构误差均值为 μ , 方差为 σ 。检测阈值用 τ 表示, τ 将位于 $[\mu, \mu + \lambda\sigma]$ 区间内。其中 λ 为大于等于 1 整数。经过多次训练, 在此区间选择一个能使检测精确度最高的值作为检测阈值。

上分位数法: Choi 等人^[13]中提出了一种基于异常流量数据在训练数据中所占百分比为条件的检测阈值设置方法。将训练数据中异常数据的百分比 α 作为先验, 假设重建误差服从正态分布 $N(\mu, \sigma)$ 。 z_α 为标准正态分布的上 α 分位点。检测阈值的计算公式为: $\tau = \mu + z_\alpha * \sigma$ 。该方法是假定正常流量数据远大

于异常数据。在训练模型时训练数据中含有一定比例的异常数据。

3.2.2 基于分类的网络异常检测技术

自编码器最成功应用是以无监督的方式进行特征提取和降维任务^[40]。Gurung 等人^[27]提出了构建了两级稀疏自编码器进行特征选择, 将数据从 115 维降到 10 维, 再利用 Logistic 回归网络对所学特征进行分类。Al-Qatf 等人^[28]提出的方法是利用基于稀疏自编码器进行特征学习和降维, 再使用支持向量机(SVM)进行分类。Naseer 等人^[29]提出的方法是利用卷积自编码器进行特征学习和降维, 使用全连接层和 Sigmoid 单元类的方法组成输出层进行分类。Shone 等人^[48]。堆叠自编码器的深度学习能力和浅层学习分类器相结合。浅层学习分类器使用了随机森林。实验结果表明, 该方法可以提高浅层网络的分类精度, 并降低了训练和测试时间。这类方法的具体过程如图 11 所示。首先对数据集进行预处理、并分为训练数据和测试数据。接着使用重构误差以无监督方式训练自编码器, 再用训练好的自编码器对数据进行降维, 然后使用浅层网络对降维后的数据进行分类。

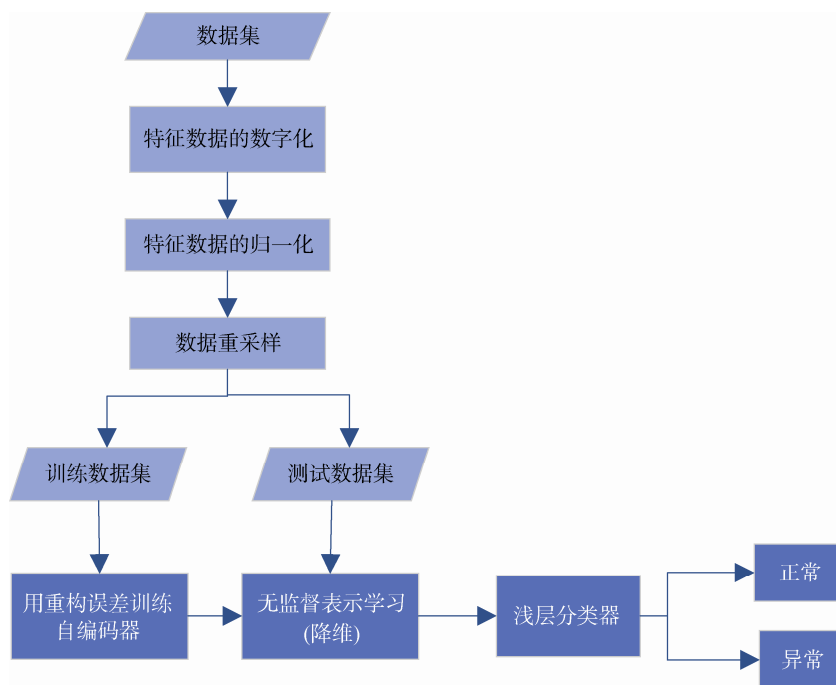


图 11 基于分类的网络异常检测技术过程图

Figure 11 Process diagram of classification-based network anomaly detection technology

3.3 小结

表 2 总结了基于自编码器的网络异常检测研究代表性的工作。分别从所用自编码器的类型, 异常检测方法、阈值计算、数据集和性能评价等方面进行了对比。由该表可知: 1)在异常检测方法中, 基于重

构误差的方法最常用; 2)在评估模型时, 使用最多是 NSL-KDD 数据集; 3)在基于阈值的异常检测中, 设置阈值是至关重要的, 但是仅有三个文献有具体的计算阈值方法, 而其需要计算阈值的文献中没有体现这一部分内容; 4)通过对比发现文献^[12-13,27,29,45]中

检测性能较好。对于基于阈值方法, 自编码器模型在训练时仅使用正常流量数据, 或者含使用包含很少量异常流数据, 数据不要需要标注, 通过阈值来异常检测, 因此属于半监督或无监督的异常检测方法。而基于分类的方法中, 自编码器用于为预训练阶段, 无监督的特征提出, 而分类阶段还需要有标注的数据。因此从本质上讲这类方法主体上还是监督学习方法。想要减少于对标注数据的使用或者不完全不依赖于标准数据, 应选择基于阈值的检测方法。通过进一步分析和对比可知, 要想提高网络异常检测的性能, 应该从以下三个方面考虑: 1)采用多级检测方法; 2)将选择将多个 AE 集成或堆叠使用;3)对于基于阈值的异常检测方法来说, 能够计算出更加精确的阈值。

4 问题与挑战

虽然自编码器在网络异常检测任务具有不错的效果, 但是其中还存在不少问题与挑战。

1) 阈值的确定问题

在基于自编码器的无监督或半监督的网络异常检测方法中, 自编码器模型可以采用深度神经网络,

可以使得重构误差很小。但是异常检测的阈值确定成为成败的关键。阈值大小直接影响检测的性能。本文的 3.2 节总结了三种计算阈值的方法。第一类朴素方法最简单直接, 在正常流量中一些特殊数据使得最大重构误差值很大, 不能直接作为阈值使用, 需要手工调整。第二类随机方法, 需要在一个较大重构误差区间内找到使得检测性能最好的阈值, 非常困难的。因此在基于自编码器的无监督或半监督的网络异常检测方法中, 最优阈值的确定是需要解决的难点问题。

2) 数据不平衡的问题

通常异常流量在总的流量中分布较少。尤其一下特殊的攻击流量分布更为罕见, 这严重影响了网络入侵检测系统的正确率。当前的主要方式是从数据分布和机器学习算法两方面进行解决。比较常用的方法是使用过采样和采用的方法来减弱数据平衡的影响。深度生成模型进行模拟生成异常流量。还有一些方法在算法中对分布较少流量设置较大权重方法来解决数据不平衡的影响。但是对于分布极少的流量检测效果很差, 是需要解决的难点问题。

表 3 基于自编码器的网络异常检测技术代表性的工作

Table 3 Representative work on network anomaly detection technology based on Autoencoder

文献	模型	异常检测方法	自编码器的用法	阈值的计算	数据集	性能评价	创新点
[12]	SAE, UAE	基于重构误差(异常分数)	两级检测, 通过稀疏率和重构误差检测	未体现	NSL-KDD	ACC:96.45%, RE: 97.43%, 通过两级检测提高性能 FS: 96.49%	
[44]	SAE	基于分类	特征提取和降维	不需要	NSL-KDD	RE: 84.6%, PR: 92.8%	使用 SAE 进行特征提取
[46]	UAE	基于重构误差(异常分数)	集成多个自编码器, 通过重构误差检测	朴素方法	在线检测	AUC:92.22%	集成多个自编码器提高检测效率
[48]	DAE	基于重构误差(异常分数)	通过重构误差检测	随机方法	KDDTrain+	ACC:94.35%, PR: 95.56% RE: 94.35%, FS: 94.80%	随机方法计算阈值
[13]	StackAE	基于重构误差(异常分数)	通过重构误差检测	上分位数法	NSL-KDD	ACC:87.66%, PR: 97.84% RE: 97.04%, FS: 85.59%	使用上分位数法计算阈值
[13]	VAE	基于重构误差(异常分数)	通过重构误差检测	上分位数法	NSL-KDD	ACC:87.82%, PR: 95.27% RE: 96.04%, FS: 86.59%	使用上分位数法计算阈值
[47]	VAE	基于重构概率(正常分数)	通过重构概率检测	未体现	NSL-KDD	AUC: 97.6%	使用 VAE 进行概率重构
[53]	VAE	基于重构概率(正常分数)	通过重构概率检测	未体现	CICDS2017	AUC:75.96%	使用 VAE 基于流异常检测
[29]	ConvAE	基于分类	特征提取和降维	不需要	KDDTrain+	PR: 97%, AUC:89%	使用 ConvAE 进行特征提取
[48]	StackAE	基于分类	特征提取和降维	不需要	KDDCup99	ACC:97.85%, PR: 99.56%	使用 StackAE 进行特征提取

3) 异常检测方法

基于自编码器的无监督或半监督的网络异常检测方法有一定局限性。未来应该尝试将其他的无监督学习、半监督学习方法。迁移学习和小样本学习网络入侵检测技术方面。

5 结束语

随着信息技术的不断发展, 当今的社会已对计算机网络高度依赖, 网络入侵检测系统是网络安全必不可少的防御机制, 因此在学术界和工业界获得了广泛的研究。基于半监督学习和无监督学习的网络入侵检测技术是当前研究的热点之一。

自编码器是典型的无监督的生成模型。本文首先对常用自编码器的基本原理、模型结构、损失函数和训练方法进行了详细的介绍。重点从阈值方法和分类方法两个方面, 描述了自编码器在网络异常检测技术中具体用法, 并对异常检查中阈值的计算方法进行了总结。最后还对这一方面代表性工作进行比较。本文旨在对基于自编码器的网络异常检测研究进行综述, 并为相关研究人员提供研究现状、面临的挑战以及有待改进的方向。

参考文献

- [1] Qing S H, Jiang J C, Ma H T, et al. Research on Intrusion Detection Techniques: A Survey[J]. *Journal of China Institute of Communications*, 2004, 25(7): 19-29.
(卿斯汉, 蒋建春, 马恒太, 等. 入侵检测技术研究综述[J]. *通信学报*, 2004, 25(7): 19-29.)
- [2] Bhuyan M H, Bhattacharyya D K, Kalita J K. Network Anomaly Detection: Methods, Systems and Tools[J]. *IEEE Communications Surveys & Tutorials*, 2014, 16(1): 303-336.
- [3] Zhang Y D, Chen S Y, Peng Y H, et al. A Survey of Deep Learning Based Network Intrusion Detection[J]. *Journal of Guangzhou University (Natural Science Edition)*, 2019, 18(3): 17-26.
(张勇东, 陈思洋, 彭雨荷, 等. 基于深度学习的网络入侵检测研究综述[J]. *广州大学学报(自然科学版)*, 2019, 18(3): 17-26.)
- [4] Ahmed M, Naser Mahmood A, Hu J K. A Survey of Network Anomaly Detection Techniques[J]. *Journal of Network and Computer Applications*, 2016, 60: 19-31.
- [5] Azizjon M, Jumabek A, Kim W. 1D CNN Based Network Intrusion Detection with Normalization on Imbalanced Data[C]. *2020 International Conference on Artificial Intelligence in Information and Communication*, 2020: 218-224.
- [6] Vinayakumar R, Alazab M, Soman K P, et al. Deep Learning Approach for Intelligent Intrusion Detection System[J]. *IEEE Access*, 2019, 7: 41525-41550.
- [7] Vinayakumar R, Soman K P, Poornachandran P. Evaluating Effectiveness of Shallow and Deep Networks to Intrusion Detection System[C]. *2017 International Conference on Advances in Computing, Communications and Informatics*, 2017: 1282-1289.
- [8] Lian H F, Zhang H, Guo W Z. Netflow Anomaly Detection Based on Data Enhancement and Hybrid Neural Network[J]. *Journal of Chinese Computer Systems*, 2020, 41(4): 786-793.
(连鸿飞, 张浩, 郭文忠. 一种数据增强与混合神经网络的异常流量检测[J]. *小型微型计算机系统*, 2020, 41(4): 786-793.)
- [9] Radford B J, Apolonio L M, Trias A J, et al. Network Traffic Anomaly Detection Using Recurrent Neural Networks[EB/OL]. 2018: arXiv: 1803.10769. <https://arxiv.org/abs/1803.10769>.
- [10] Hsu C M, Hsieh H Y, Prakosa S W, et al. Using Long-Short-Term Memory Based Convolutional Neural Networks for Network Intrusion Detection[M]. *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*. Cham: Springer International Publishing, 2019: 86-94.
- [11] Vinayakumar R, Soman K P, Poornachandran P. Applying Convolutional Neural Network for Network Intrusion Detection[C]. *2017 International Conference on Advances in Computing, Communications and Informatics*, 2017: 1222-1228.
- [12] Gharib M, Mohammadi B, Dastgerdi S H, et al. AutoIDS: Auto-Encoder Based Method for Intrusion Detection System[EB/OL]. 2019: arXiv: 1911.03306. <https://arxiv.org/abs/1911.03306>.
- [13] Choi H, Kim M, Lee G, et al. Unsupervised Learning Approach for Network Intrusion Detection System Using Autoencoders[J]. *The Journal of Supercomputing*, 2019, 75(9): 5597-5621.
- [14] Zhou Z H. Disagreement-Based Semi-Supervised Learning[J]. *Acta Automatica Sinica*, 2013, 39(11): 1871-1878.
(周志华. 基于分歧的半监督学习[J]. *自动化学报*, 2013, 39(11): 1871-1878.)
- [15] Lee C H, Su Y Y, Lin Y C, et al. Machine Learning Based Network Intrusion Detection[C]. *2017 2nd IEEE International Conference on Computational Intelligence and Applications*, 2017: 79-83.
- [16] Hindy H, Brosset D, Bayne E, et al. A taxonomy and survey of intrusion detection system design techniques, network threats and datasets[J]. arXiv preprint arXiv:1806.03517, 2018.
- [17] Weller-Fahy D J, Borghetti B J, Sodemann A A. A Survey of Distance and Similarity Measures Used within Network Intrusion Anomaly Detection[J]. *IEEE Communications Surveys & Tutorials*, 2014, 17(1): 70-91.
- [18] Buczak A L, Guven E. A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection[J]. *IEEE Communications Surveys & Tutorials*, 2015, 18(2): 1153-1176.
- [19] Moustafa N, Hu J K, Slay J. A Holistic Review of Network Anomaly Detection Systems: A Comprehensive Survey[J]. *Journal of Network and Computer Applications*, 2019, 128: 33-55.
- [20] Jian S J, Lu Z G, Du D, et al. Overview of Network Intrusion Detection Technology[J]. *Journal of Cyber Security*, 2020, 5(4): 96-122.
(蹇诗婕, 卢志刚, 牡丹, 等. 网络入侵检测技术综述[J]. *信息安全学报*, 2020, 5(4): 96-122.)

- [21] Li J L, Zhang H. Survey on Semi-Supervised Anomaly Traffic Detection[J]. *Journal of Chinese Computer Systems*, 2020, 41(11): 2371-2379.
(李杰铃, 张浩. 半监督异常流量检测研究综述[J]. *小型微型计算机系统*, 2020, 41(11): 2371-2379.)
- [22] Ring M, Wunderlich S, Scheuring D, et al. A Survey of Network-Based Intrusion Detection Data Sets[J]. *Computers & Security*, 2019, 86: 147-167.
- [23] Poornachandran P, Vinayakumar R, Kp S. A Comparative Analysis of Deep Learning Approaches for Network Intrusion Detection Systems (N-IDSs): Deep Learning for N-IDSs[J]. *International Journal of Digital Crime and Forensics*, 2019, 11(3): 65-89.
- [24] Vinayakumar R, Soman K P, Poornachandran P. Evaluation of Recurrent Neural Network and Its Variants for Intrusion Detection System IDS[J]. *International Journal of Information System Modeling and Design*, 2017, 8(3): 43-63.
- [25] Moustafa N, Slay J. UNSW-NB15: A Comprehensive Data Set for Network Intrusion Detection Systems (UNSW-NB15 Network Data Set)[C]. *2015 Military Communications and Information Systems Conference*, 2015: 1-6.
- [26] Yulianto A, Sukarno P, Suwastika N A. Improving AdaBoost-Based Intrusion Detection System (IDS) Performance on CIC IDS 2017 Dataset[J]. *Journal of Physics: Conference Series*, 2019, 1192: 012018.
- [27] Gurung S, Kanti Ghose M, Subedi A. Deep Learning Approach on Network Intrusion Detection System Using NSL-KDD Dataset[J]. *International Journal of Computer Network and Information Security*, 2019, 11(3): 8-14.
- [28] Al-Qatf M, Yu L S, Al-Habib M, et al. Deep Learning Approach Combining Sparse Autoencoder with SVM for Network Intrusion Detection[J]. *IEEE Access*, 2018, 6: 52843-52856.
- [29] Naseer S, Saleem Y, Khalid S, et al. Enhanced Network Anomaly Detection Based on Deep Neural Networks[J]. *IEEE Access*, 2018, 6: 48231-48246.
- [30] Devan P, Khare N. An Efficient XGBoost-DNN-Based Classification Model for Network Intrusion Detection System[J]. *Neural Computing and Applications*, 2020, 32(16): 12499-12514.
- [31] Le T T H, Kim Y, Kim H. Network Intrusion Detection Based on Novel Feature Selection Model and Various Recurrent Neural Networks[J]. *Applied Sciences*, 2019, 9(7): 1392.
- [32] Kevric J, Jukic S, Subasi A. An Effective Combining Classifier Approach Using Tree Algorithms for Network Intrusion Detection[J]. *Neural Computing and Applications*, 2017, 28(1): 1051-1058.
- [33] Caminero G, Lopez-Martin M, Carro B. Adversarial Environment Reinforcement Learning Algorithm for Intrusion Detection[J]. *Computer Networks*, 2019, 159: 96-109.
- [34] Lopez-Martin M, Carro B, Sanchez-Esguevillas A. Application of Deep Reinforcement Learning to Intrusion Detection for Supervised Problems[J]. *Expert Systems With Applications*, 2020, 141: 112963.
- [35] Lai X F, Liang X W, Xie Z C, et al. Intrusion Detection Method Based on Entity Embedding and Long Short-Term Memory Networks[J]. *Journal of University of Chinese Academy of Sciences*, 2020, 37(4): 553-561.
(赖训飞, 梁旭文, 谢卓辰, 等. 基于实体嵌入和长短时记忆网络的入侵检测方法[J]. *中国科学院大学学报*, 2020, 37(4): 553-561.)
- [36] Baig M M, Awais M M, El-Alfy E S M, et al. A Multiclass Cascade of Artificial Neural Network for Network Intrusion Detection[J]. *Journal of Intelligent & Fuzzy Systems: Applications in Engineering and Technology*, 2017, 32(4): 2875-2883.
- [37] Zavrak S, İskefiyeli M. Anomaly-Based Intrusion Detection from Network Flow Features Using Variational Autoencoder[J]. *IEEE Access*, 2020, 8: 108346-108358.
- [38] Zhou Z H. Machine learning[M]. Beijing: Tsinghua University Press, 2016: 30.
(周志华. 机器学习[M]. 北京: 清华大学出版社, 2016: 30.)
- [39] Xu C Y. Research of network intrusion detection method based on deep learning[D]. Hangzhou: Zhejiang University, 2019.
(许聪源. 基于深度学习的网络入侵检测方法研究[D]. 杭州: 浙江大学, 2019.)
- [40] Goodfellow I, Bengio Y, Courville A. Deep learning[M]. Cambridge, Massachusetts: The MIT Press, 2016.
- [41] Kingma D P, Welling M. Auto-Encoding Variational Bayes [EB/OL]. 2013: arXiv: 1312.6114. <https://arxiv.org/abs/1312.6114>.
- [42] Doersch C. Tutorial on Variational Autoencoders[EB/OL]. 2016: arXiv: 1606.05908. <https://arxiv.org/abs/1606.05908>.
- [43] Yuan X F, Huang B, Wang Y L, et al. Deep Learning-Based Feature Representation and Its Application for Soft Sensor Modeling with Variable-Wise Weighted SAE[J]. *IEEE Transactions on Industrial Informatics*, 2018, 14(7): 3235-3243.
- [44] Van N T, Tinh T N, Sach L T. An Anomaly-Based Network Intrusion Detection System Using Deep Learning[C]. *2017 International Conference on System Science and Engineering*, 2017: 210-214.
- [45] An J, Cho S. Variational autoencoder based anomaly detection using reconstruction probability[J]. *Special Lecture on IE*, 2015, 2(1): 1-18.
- [46] Mirsky Y, Doitshman T, Elovici Y, et al. Kitsune: An Ensemble of Autoencoders for Online Network Intrusion Detection[EB/OL]. 2018: arXiv: 1802.09089. <https://arxiv.org/abs/1802.09089>.
- [47] Aygun R C, Yavuz A G. Network Anomaly Detection with Stochastically Improved Autoencoder Based Models[C]. *2017 IEEE 4th International Conference on Cyber Security and Cloud Computing*, 2017: 193-198.
- [48] Shone N, Ngoc T N, Dinh Phai V, et al. A Deep Learning Approach to Network Intrusion Detection[J]. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 2018, 2(1): 41-50.



张国梁 于 2009 年在西安石油大学计算机应用技术专业获得硕士学位。现任西藏民族大学信息工程学院讲师。研究领域为网络安全、深度学习。研究兴趣包括: 网络异常检测、自然语言处理和深度学习等。Email: zgl@xzmu.edu.cn



郭晓军 于 2017 年在东南大学计算机应用技术专业获得博士学位。现任西藏民族大学信息工程学院副教授。研究领域为网络安全、网络测量。研究兴趣包括: Web 安全、网络异常检测。Email: aikt@xzmu.edu.cn