基于聚类过采样和自动编码器的网络入侵检测方法

蹇诗婕[1,2], 刘 岳[1,2], 姜 波[1], 卢志刚[1,2], 刘玉岭[1,2], 刘宝旭[1,2]

[1]中国科学院信息工程研究所, 北京 中国 100093
[2]中国科学院大学网络空间安全学院, 北京 中国 100049

摘要 近年来, 随着互联网技术的不断发展, 入侵检测在维护网络空间安全方面发挥着越来越重要的作用。但是, 由于网络入侵行为的数据稀疏性, 已有的检测方法对于海量流量数据的检测效果较差, 模型准确率、F-measure 等指标数值较低, 并且高维数据处理的成本过高。为了解决这些问题, 本文提出了一种基于稀疏异常样本数据场景下的新型深度神经网络入侵检测方法, 该方法能够有效地识别不平衡数据集中的异常行为。本文首先使用 k 均值综合少数过采样方法来处理不平衡的流量数据, 解决网络流量数据类别分布不平衡问题, 平衡网络流量数据分布。再采用自动编码器来处理海量高维数据并训练检测模型, 来提升海量高维流量中异常行为的检测精度, 并在两个真实典型的入侵检测数据集上进行了大量的实验。实验结果表明, 本文所提出的方法在两个真实典型数据集上的检测准确率分别为 99.06％和 99.16%, F-measure 分别为 99.15%和 98.22%。相比于常用的欠采样和过采样方法, k 均值综合少数过采样技术能够有效地解决网络流量数据类别分布不平衡的问题, 提升模型对低频攻击行为的检测效果。同时, 与已有的网络入侵检测方法相比, 本文所提出的方法在准确率、F-measure 和检测性能上均有明显提升, 证明了本文所提出的方法对于海量网络流量数据的检测具有较高的检测精度和良好的应用前景。

关键词 入侵检测; 海量流量数据; 类别不平衡; 自动编码器; k 均值综合少数过采样技术
中图法分类号 TP393.08 DOI 号 10.19363/J.cnki.cn10-1380/tn.2023.11.10

Network Intrusion Detection Using Cluster Oversampling and Auto-Encoder

JIAN Shijie[1,2], LIU Yue[1,2], JIANG Bo[1], LU Zhigang[1,2], LIU Yuling[1,2], LIU Baoxu[1,2]

[1] Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093, China
[2] School of Cyber Security, University of Chinese Academy of Sciences, Beijing 100049, China

Abstract With the continuous development of Internet technology, intrusion detection is becoming more and more important to safeguard the security of cyberspace in these years. However, existing detection methods work poorly on massive traffic data due to the data sparsity of the network intrusion behaviors. The accuracy rate, F-measure and other indicators are relatively low. In addition, the cost of high-dimensional data processing is too high. To address these issues, we propose a novel deep neural network intrusion detection method based on sparse abnormal sample data scenarios, which is called K-means Sparse Anomaly Intrusion Detection System (KSAIDS). It can be used to effectively identify the abnormal behaviors in imbalanced datasets. In particular, we first use k-means Synthetic Minority Over-sampling Technique method to deal with the imbalanced traffic data, which can effectively solve the problem of unbalanced distribution of network traffic data categories and balance the distribution of network traffic data. The proposed model then employs Auto-Encoder to process the massive high-dimensional data and train detection model so as to improve the detection accuracy of abnormal behaviors in massive high-dimensional traffic. And extensive experiments are carried out on two real-world typical intrusion detection datasets. Experimental analysis results demonstrate that the detection accuracy of the proposed method on two real-world typical datasets is 99.06% and 99.16%, and the F-measure is 99.15% and 98.22%, respectively. Compared with the commonly used under-sampling and over-sampling methods, the k-means Synthetic Minority Over-sampling Technique method can effectively solve the problem of unbalanced distribution of network traffic data categories and improve the model's detection effect on low-frequency attack behavior. At the same time, compared with the state-of-the-art models of intrusion detection, the detection accuracy rate, F-Measure and detection performance of the KSAIDS method are significantly improved, which proves that the KSAIDS method has high detection accuracy and great application prospects for the detection of large-scale network traffic data.

通讯作者: 姜波, 博士, 副研究员, Email: jiangbo@iie.ac.cn。

本论文得到国家重点研发计划(No. 2019QY1303, No. 2019QY1302, No. 2021YFC3300401)、中国科学院战略性先导 C 类(No. XDC02040100)、中国科学院青年创新促进会(No. 2021156)的资助。这项工作也得到了中国科学院网络评估技术重点实验室和北京市网络安全与保护技术重点实验室的部分支持。
收稿日期: 2020-04-14; 修改日期: 2020-06-08; 定稿日期: 2022-12-20

Key words intrusion detection; massive traffic data; class imbalanced; auto-encoder; k-means synthetic minority over-sampling technique

1 引言

随着信息化时代的飞速发展, 互联网已经成为人们生活中不可或缺的一部分。然而, 网络中攻击行为的频率和规模仍然呈现不断增长的趋势。这些攻击行为不仅会造成巨大的经济损失, 而且严重威胁着社会稳定和国家安全。维护网络空间安全已经成为当前亟待解决的问题。为了更好地维护网络空间安全并防御各种攻击, 入侵检测技术作为一种主动防御方法已经成为当前研究的热点。入侵检测系统(Intrusion Detection System, IDS)[1]是一种主动的安全防护技术, 能够监视网络中数据的传输行为, 并在发现可疑传输行为或异常中断时发出警报。

James Anderson 在 1980 年首次提出入侵检测的概念[1], 即使用系统来监视入侵行为。目前已经有许多关于检查入侵行为的研究工作, 这些工作可以分为基于误用的入侵检测系统和基于异常的入侵检测系统。基于误用的入侵检测系统, 也称为基于签名的入侵检测系统, 它是一种基于知识的入侵检测系统, 能够基于已有的签名知识库来检测攻击行为[2]。尽管基于误用的入侵检测系统具有较高的准确率和较低的误报率, 但是它难以检测不在签名知识库中的未知攻击。不同于基于误用的入侵检测系统, 基于异常的入侵检测系统通过对正常行为与异常行为进行比较, 能够检测出未知攻击[3]。因此, 基于异常的入侵检测系统是目前研究的热点内容。其中, 最常用的一种方法是基于特征的传统机器学习方法, 例如决策树(Decision Tree, DT), 随机森林(Random Forests, RF), 朴素贝叶斯, Adaboost, 极端梯度增强(Extreme Gradient Boosting, XGBoost)等。然而, 基于传统机器学习方法的入侵检测通常侧重于特征工程, 属于浅层的学习方法。随着网络中海量高维数据的增加、网络带宽的增加、数据的复杂性和特征的多样性也在增加, 使用浅层学习方法难以达到分析和预测的目的。

近年来, 深度神经网络技术已经在图像识别, 自然语言处理和语音识别等方面均取得了巨大的成功。深度神经网络[4]是一种学习数据表征信息的方法, 通过构造由多个隐藏层组成的非线性网络结构, 能够学习数据的内部信息, 并能适应高维度学习和预测的要求。目前, 基于深度学习的入侵检测方法包括自动编码器, 循环神经网络(Recurrent Neural Network, RNN), 长短期记忆网络(Long Short Term Memory Network, LSTM), 门控循环单元(Gated Recurrent Unit, GRU), 和卷积神经网络(Convolutional Neural Networks, CNN)等, 并且这些方法已经在一些研究中取得了成功。但是, 这些用于入侵检测的深度学习方法仍然存在一些问题。例如, 网络流量数据存在类别分布不平衡的问题, 而许多研究都没有很好地解决这一问题。由于很多算法的基本假设是数据均匀分布[5], 因此, 当数据存在类别不平衡问题时, 预测算法和评价指标的选取会对实际预测结果造成较大影响, 通常难以取得理想的预测结果, 且训练效率较低。除此以外, 研究人员通常更关注少数类别数据的预测情况, 由于少数类别数据数量远远少于多数类别数据数量, 存在较多稀疏样本数据, 导致模型难以有效地对这些稀疏样本数据进行分类, 最终学习到的分类边界偏移, 更偏向多数类别数据, 模型分类性能较差[6]。例如在网络安全领域, 学者们更关心攻击行为的预测结果, 如果不考虑网络流量数据的整体分布, 就会导致决策函数偏向多数类别的样本数据, 低频攻击样本被视为噪声而被忽略, 使得模型难以捕获有效特征, 并且很难检测到低频攻击[7]。因此, 改善网络流量数据的整体分布, 能够有效提升入侵检测模型的预测能力, 使得模型能够更好地对未知流量数据进行异常检测[8]。另一方面, 部分研究在将符号数据转换为数值数据时, 并未对高维数据进行处理, 导致模型训练效率较低、检测性能较差, 并且消耗了大量存储空间[9]。因此, 对网络流量数据进行降维处理能够更好地提升入侵检测的效率和检测性能。

为了解决上述问题, 本文提出一种基于聚类过采样和自动编码器的网络入侵检测方法, 称为 k 均值稀疏异常入侵检测方法(k-means Sparse Anomaly Intrusion Detection System, KSAIDS), 它能够有效地识别类别不平衡数据集中的异常行为。具体而言, KSAIDS 使用 k 均值综合少数过采样技术(k-means Synthetic minority over-sampling technique, k-means SMOTE)来处理数据集中不平衡的正常流量数据和异常流量数据, 再结合自动编码器方法, 从而能够有效地从大规模流量数据中提取非线性结构信息。本文在两个真实典型的网络数据 UNSW-NB15 和 CICIDS2017 数据集上进行实验, 来验证所提出方法的有效性。实验结果表明, 本文所提出的 KSAIDS 优

于最新的入侵检测方法。本文的主要贡献如下:

1) 本文使用 k-means SMOTE 过采样方法来处理不平衡的正常流量数据和异常流量数据, 通过减少样本的偏差来提升低频攻击行为的检测率。

2) 本文提出了一种用于入侵检测的自动编码器方法, 能够减少数据维数, 去除噪声数据, 并捕获更有效的特征信息。

3) 本文通过在 UNSW-NB15 和 CICIDS2017 数据集上进行实验, 来验证所提出方法的有效性。实验结果一致表明, 所提出的方法是有效且具有竞争力的。

本文的其余部分安排如下。第 2 节介绍了入侵检测的研究现状; 第 3 节介绍了所提出的 KSAIDS 的框架; 第 4 节说明了所提出的方法在 UNSW-NB15 和 CICIDS2017 数据集上的实验和实验结果; 第 5 节总结结论。

2 相关工作

这一部分将详细阐述基于误用的入侵检测系统和基于异常的入侵检测系统的研究现状。其中, 基于异常的入侵检测系统分为基于传统机器学习的入侵检测和基于深度学习的入侵检测两部分。

2.1 基于误用的入侵检测系统

基于误用的入侵检测系统通过将网络流量与已有签名数据进行匹配来检测异常行为。

Wutyi 等人[10]提出了一种通过合并攻击签名来手动制定启发式规则的入侵检测方法。然而, 该方法需要人为手工设置多种规则。Wang 等人[11]提出了一种隐私保护框架, 用于基于雾设备的分布式网络中的基于签名的入侵检测。但是, 签名匹配是一项高成本的操作, 需要大量的工作量。

基于签名的入侵检测系统有一个共同的缺点, 即它们过于依赖已有的签名规则, 只能基于已有规则库来检测攻击行为, 难以检测未知攻击行为。因此, 基于异常的入侵检测是当前学者们研究的重点。

2.2 基于异常的入侵检测系统

2.1.1 基于机器学习的入侵检测系统

传统的基于机器学习的检测方法可以有效地检测未知攻击行为。Sahu 等人[12]使用 J48 决策树来检测异常。Farnaaz 等人[13]提出了一种基于随机森林算法的入侵检测系统。Yu 等人[14]结合主成分分析法提出了基于朴素贝叶斯的网络入侵最优路由检测的选择方法。Hu 等人[15]提出了两种基于 Adaboost 的在线分布式入侵检测算法。Dhaliwal 等人[16]通过用 XGBoost 方法学习完整的数据信息, 来提升预测准确性。但是, 传统的基于机器学习的入侵检测系统强调特征选择和特征提取, 属于浅层的学习方法。浅层学习方法难以应对网络中的大规模高维数据。并且许多研究都没有考虑对不平衡数据进行处理。

2.1.2 基于深度学习的入侵检测系统

基于深度学习的入侵检测方法能够更好地检测大规模网络流量数据。Yin 等人[17]提出了一种基于循环神经网络的入侵检测方法。Agarap 等人[18]用支持向量机替换了门控循环单元模型输出层的 Softmax 函数, 以实现入侵检测。Ding 等人[19]提出了一种基于卷积神经网络的入侵检测方法。但是, 这些方法均没有对高维数据进行预处理, 并且在测试数据集上的检测准确率较低。Kim 等人[20]提出了一种利用 CNN-LSTM 神经网络对用户角色和权限进行分类的异常数据库入侵检测方法。Qureshi 等人[21]使用深度稀疏自编码和自学习方法来构建入侵检测模型。Abolhasanzadeh 等人[22]使用自动编码器提取神经网络的瓶颈特征来进行入侵检测, 使用了具有七个隐藏层的自动编码器进行实验。但是, 这些工作均没有考虑不平衡流量数据集的整体分布情况, 并且检测能力仍有待提升。现有的基于深度学习的入侵检测技术对海量数据中类别不平衡问题的研究仍处于初期阶段, 对低频异常样本的检测效果仍然不理想。因此, 本文提出了一种基于 k 均值综合少数过采样和自动编码器的入侵检测方法来解决这些问题。

3 k 均值稀疏异常入侵检测方法

在这一部分, 本文将对所提出的 k 均值稀疏异常入侵检测方法的执行步骤进行详细阐述。如图 1 所示, 通过使用 k-means SMOTE 过采样方法和自动编码器方法来实现 KSAIDS, 从而检测异常攻击行为。其中, KSAIDS 的具体实施步骤如下:

1) **数据预处理**。首先对数据集进行预处理, 将符号数据使用独热编码转换为数值型数据, 并对数值型数据进行归一化处理, 从而获得标准化的数据。

2) **不平衡数据处理**。由于网络流量数据集中的正负例样本数据数量高度不平衡, 因此本文使用 k-means SMOTE 过采样方法对异常流量数据进行处理, 得到平衡数据集。注意, 在这里, 本文只对训练数据集进行不平衡数据处理, 并未改变测试数据集。

3) **模型训练**。本文使用经过处理得到的平衡流量数据集作为模型的输入, 并使用自动编码器来构建入侵检测模型。

4) **异常行为检测**。本文使用训练好的模型来检测测试数据集中的稀疏异常行为, 并使用评估指标来衡量所提出模型的检测能力。

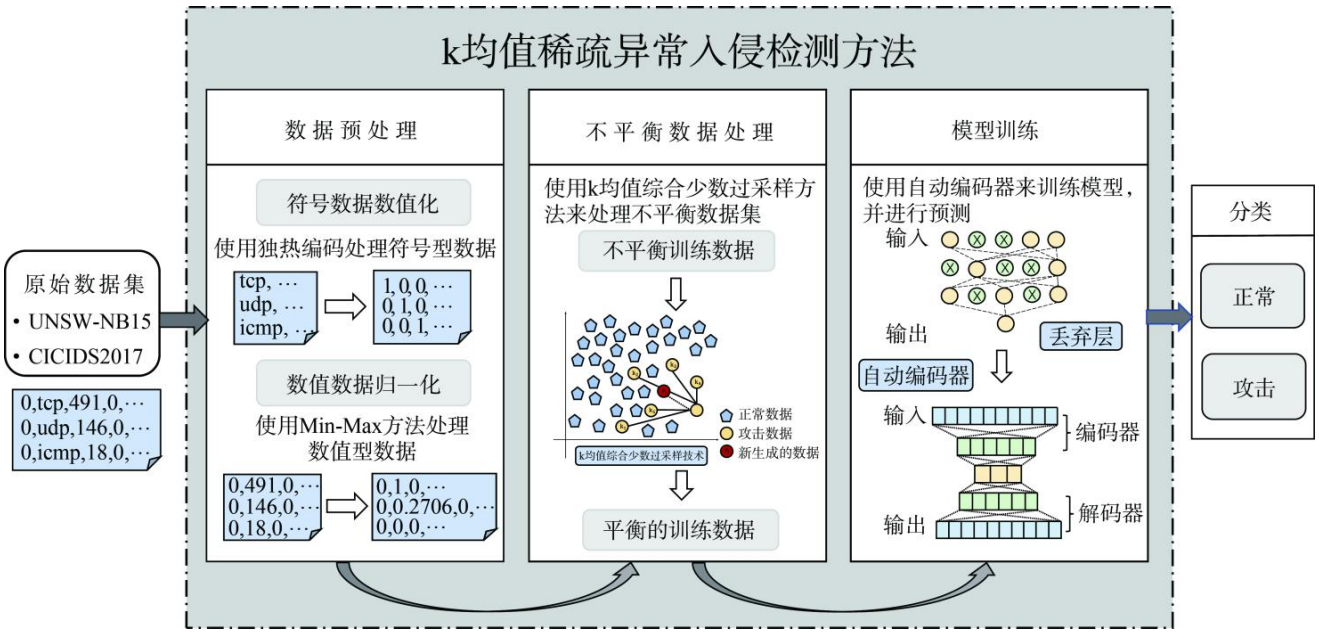

图 1　KSAIDS 入侵检测方法框架结构

Figure 1　Proposed KSAIDS framework structure

3.1　数据预处理

通常网络流量数据中包含两种类型的数据, 分别是符号特征类型数据和数值特征类型数据, 需要在预处理阶段进行处理。因此, 数据预处理包括两个步骤: 1)符号特征数值化和 2)数值特征归一化。

3.1.1　符号特征数值化

对于网络流量中的符号数据, 本文使用独热编码的方式将符号数据转化为数值数据。举例而言, 本文在 UNSW-NB15 和 CICIDS2017 数据集上进行了实验。

UNSW-NB15 数据集包含了三种类型的符号特征数据, 分别是: proto 特征, 服务特征和状态特征。在该步骤中, 使用独热编码方法将符号数据转换为数值数据, 映射为二进制向量, 特征数量从 45 变化为 196, 对于类别标签, 将数据集中的正常流量数据标记为 0, 异常流量数据标记为 1。CICIDS2017 数据集不包含符号类型特征, 因此 CICIDS2017 数据集不需要处理符号类型数据。同样地, 对于类别标签, 将 CICIDS2017 数据集中的正常流量数据标记为 0, 异常流量数据标记为 1。

3.1.2　数值特征归一化

数据归一化可以解决不同特征数据之间维数差异较大的问题, 因此被广泛用于数据预处理步骤中。为了确保检测结果的可靠性, 需要对数据集中的数值特征进行归一化处理。归一化是指将所有特征数据缩小到[0, 1]的范围内, 常用的归一化方法有 z-score 归一化和最小最大归一化方法。本文使用最小最大归一化方法来处理数据。其转换公式如公式(1)所示:

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{1}$$

其中, x 代表某一种特征的属性值, $x_{\max}$ 代表这种特征属性的最大值, $x_{\min}$ 代表这种特征属性的最小值, x' 代表对 x 进行归一化处理后的结果。

3.2　不平衡数据处理

3.2.1　综合少数过采样技术

在网络流量数据中, 本文进行了大量统计工作, 并观察到正常和异常流量数据存在着高度不平衡的现象, 如果不考虑网络流量数据的整体分布, 会导致决策函数偏向多数类别的样本数据, 低频攻击样本被视为噪声, 难以检测低频攻击。为了提升模型的检测能力, 需要对不平衡数据集进行处理。常见的数据不平衡问题的解决方案包括欠采样方法和过采样方法。

欠采样通过减少多数类别的数据量来平衡类别分布, 但是这样的处理方式可能会丢失重要数据信息。另一方面, 过采样方法通常比欠采样方法处理效果更好, 因此使用频率更高。最著名的一种过采样方法是综合少数过采样技术(Synthetic Minority Over-sampling Technique, SMOTE)[23], SMOTE 方法的优点是它减少了采样过程中的局限性, 并且 SMOTE 方法采用线性插值的理论, 能够有效地减少过拟合现象。通过 SMOTE 方法来生成合成数据的公式如公式(2)

所示:

$$y_{\text{new}} = y_i + (y_i - y_j) \times \delta \tag{2}$$

其中, y_{new} 是新生成的综合数据, y_i是来自少数类别的样例数据, y_j代表从y_i的 k 个最近邻居中随机选择的邻居, δ代表 0～1 之间的随机数。

但是 SMOTE 算法在处理不平衡数据方面存在一些缺陷。一方面, SMOTE 使用均匀概率对少数类别数据进行过采样, 忽略了类内不平衡的问题, 且该方法容易产生噪声数据。具体而言, 当存在噪声的情况下, SMOTE 方法可能会在多数类别数据的区域生成少数类别的样本, 加剧了类内不平衡问题。另一方面, SMOTE 方法放大了数据中存在的噪声, 因为它无法将重叠的类别区域和安全区域划分开。如图 2 所示。

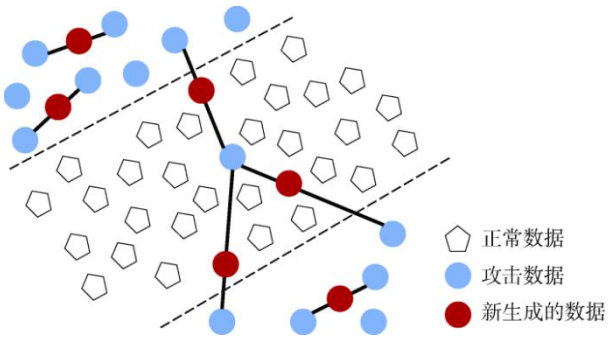

图 2　SMOTE 方法的缺陷

Figure 2　The disadvantage of the SMOTE method

3.2.2　k 均值综合少数过采样技术

k 均值综合少数过采样技术[24]通过仅在安全区域中进行过采样来避免产生噪声, 并有效地克服了类间不平衡和类内不平衡问题。样本数据分布依据于簇密度, 在稀疏的少数类别数据区域生成更多的样本数据, 从而消除了类内不平衡问题, 如图 3 所示。考虑到异常流量数据的稀疏性, 本文采用的是 k-means SMOTE 方法来解决数据类别不平衡的问题, 从而提升模型的检测性能。

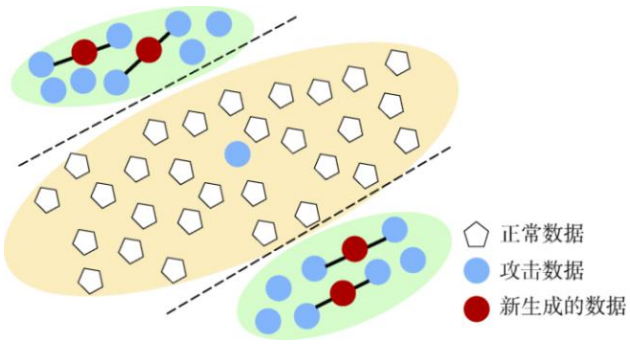

图 3　k-means SMOTE 方法在安全区域进行过采样

Figure 3　Oversampling in safe areas using the k-means SMOTE method

具体而言, k-means SMOTE 包括了三个步骤, 分别是聚类、过滤和过采样。1)在聚类步骤中, 使用 k-means 方法将数据划分为 k 个簇。2)在过滤步骤中, 选择要进行过采样的簇, 保留少数类别样本比例较高的簇。然后, 分配要生成的合成样本的数量, 将更多样本分配给稀疏少数类别样本的簇。3)在过采样步骤中, 每个选定的集群使用 SMOTE 方法, 从而实现少数类别和多数类别实例的目标比例。其中, k-means 算法的聚类准则函数公式如公式(3)所示:

$$S = \sum_{i=1}^{k} \sum_{p \in C_j} |p - m_j|^2 \tag{3}$$

其中, S 代表所有数据的误差平方和, k 是指定聚类簇数量, C_j是第 j 类簇, p 是簇C_j中的数据, m_j是簇C_j中数据的平均值。当聚类准则函数收敛时, 或达到指定迭代次数时, 聚类迭代过程终止。k-means SMOTE 具体算法过程[25]如算法 1 所述:

算法 1. k-means SMOTE 算法.

输入:

X_train: 训练数据集

Threshold: 不平衡比率阈值

ex: 计算密度的指数, 默认值为 *X_train* 的特征数

n: 需要生成的样本数量

knn: 过采样使用的最近邻居的数目

输出:

Y: 生成的新数据

方法:

//第一步: 对输入数据进行聚类, 并筛选出少数类别数据多于多数类别数据的簇。

// 对输入数据进行 kmeans 聚类

Clusters ← *kmeans*(X_train);

FilterClusters = NULL;

FOR each $t \in Clusters$ do

$Ratio = \frac{CountMajor(t)+1}{CountMinor(t)+1}$;

IF *Ratio*< *Threshold* THEN

FilterClusters←{*t*}∪*FilterClusters*;

END IF

END FOR

//第二步: 对于每个过滤后的簇, 根据其少数类别数据的密度计算采样权值。

FOR each $m \in FilterClusters$ do

DT = *euclideanDistances(m)*;

meanMinorDistance(m) ← mean(DT);

$density(m) = \frac{CountMinor(m)}{meanMinorDistance(m)^{ex}}$;

$sparsity(m) = \frac{1}{density(m)}$;

END FOR

$Sumofsparsity \leftarrow \sum_{m\epsilon FilterClusters} sparsity(m)$

$Sampleweight(m) \leftarrow \frac{sparsity(m)}{Sumofsparsity}$

//第三步: 使用 SMOTE 方法对每个过滤后的簇进行处理, 使用抽样权值计算得到生成的样本数量。

Y=NULL;

FOR each $m \in$ FilterClusters do

$samples \leftarrow ||n * Sampleweight(m)||$;

$Y \leftarrow \{SMOTE(m, samples, knn)\} \cup Y$;

END FOR

RETURN *Y*

对使用 k-means SMOTE 方法生成的异常数据进行分析, 例如对于 UNSW-NB15 数据集而言, sload 特征指的是每小时传输的源比特数量, dload 特征指的是每小时传输的目的比特数量。对于正常类别数据, sload 值通常较小, 例如值为 279、284 等, dload 值也较小, 通常为 0。而对于 Worms 攻击, 攻击者通过复制自身, 并使用计算机网络进行广泛传播, 从而对目标计算机进行攻击。因此 Worms 攻击 sload 值通常较大, 例如值为 7321、13862、28424 等。Worms 攻击的 dload 值通常也较大, 例如值为 1421、2640、5455 等。查看使用 k-means SMOTE 方法生成的异常流量数据, 其 sload 值为 11824、13485、24068 等, 其 dload 值为 2256、2569、4601 等, 这样的生成数据仍符合攻击数据的基本特征, 通过传输大量数据而产生攻击行为。因此, 使用 k-means SMOTE 方法生成的异常数据仍符合攻击数据的基本特征, 可以用于实验。并且, 通过改善网络流量数据的整体分布, 能够更好地训练模型, 使得决策函数不再偏向多数类别的样本数据, 低频攻击样本不会被视为噪声数据而被忽略, 从而提升模型对低频攻击行为的检测能力, 提升模型整体的预测能力。

3.3 基于自动编码器的模型训练方法

目前, 基于深度学习的入侵检测相关研究使用的方法有循环神经网络、长短期记忆网络、门控循环单元、卷积神经网络以及卷积神经网络和长短期记忆网络的结合方法 CNN-LSTM。但是这些研究仍存在一些问题。循环神经网络、长短期记忆网络、门控循环单元都是基于时序的方法, 更适合用于处理时间序列特征较强的问题, 例如自然语言处理等领域相关问题。而在入侵检测领域中, 时间序列特征并不明显, 因此使用这些基于时序的方法来检测异常流量, 效果较差。卷积神经网络主要通过卷积层和池化层来提取有效特征信息, 更适合用于图像处理等相关领域。而在入侵检测领域中, 使用卷积神经网络来进行异常检测, 需要将流量数据构造为类似图像的输入形式, 在这个过程中会损失部分重要信息, 导致检测结果较差。CNN-LSTM 方法结合了卷积神经网络对图像提取特征的优势和长短期记忆网络对序列数据处理的优势, 更适合用于图像处理等相关领域。并且这些相关研究并未对高维数据进行处理, 导致模型训练效率低, 检测能力较弱。

由于数据的维数过高可能导致训练效率较低, 降低数据的维度能够减少所需要的存储空间, 加快计算速度, 消除冗余特征并且更好地表达数据。传统的线性降维方法, 例如主成分分析方法, 难以捕获数据中的非线性信息。而基于核函数的主成分分析方法等非线性降维方法计算复杂度较高, 并且难以应用于大规模数据集。自动编码器(Auto-Encoder, AE)[26]是一种包含了输入层、隐藏层(编码层)和解码层的三层神经网络, 是一种由编码器和解码器组成的深度学习算法, 如图 4 所示。自动编码器通过对数据进行压缩, 能够自动从样本数据中学习非线性有效特征信息; 通过重构输入, 能够有效去除噪声数据; 并且通过对海量数据进行降维处理, 能够减少存储耗费的资源、提升模型的检测效率。使用自动编码器方法来进行入侵检测, 不需要提取时间序列相关特征, 也不需要将输入数据构造为类似图像数据的输入形式。作为深度学习中的降维方法, 自动编码器能够有效地从海量数据集中提取非线性结构信息、去除噪声信息并获得更高级的特征, 能够有效地提升模型的分类能力。因此, 本文在模型训练过程中, 使用自动编码器来构建入侵检测模型, 从而提升模型的检测能力。

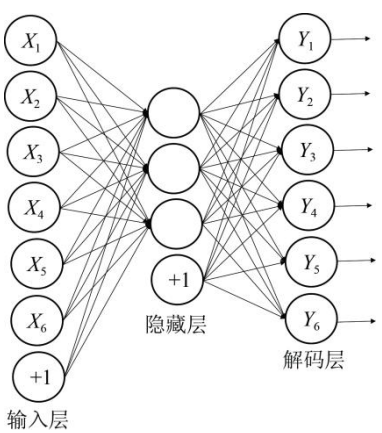

图 4 自动编码器的框架结构图

Figure 4 The frame structure diagram of Auto-Encoder.

常用的深度学习降维方法除了自动编码器方法, 还有稀疏自动编码器(Sparse Auto-Encoder, SAE) [27]方法。其中, 稀疏自动编码器在自动编码器的基础上增加正则化限制, 使得每一层大部分节点为零。但是, 当给隐藏神经元加入了稀疏性限制后, 会丢失部分有效信息, 因此, 本文选择了自动编码器来构建模型。

经过数据预处理和不平衡数据处理后, 本文使用自动编码器来减少训练数据的维数并训练模型。具体而言, 为避免过拟合现象, 本文添加了丢弃层来处理通过 k-means SMOTE 方法获得的平衡数据集。然后, 将丢弃层处理的数据作为自动编码器的输入。对于自动编码器而言, 编码器将输入数据映射为潜在变量, 而解码器使用映射得到的潜在变量重新构建输入数据。通过重构输入, 隐藏层能够学习到输入数据的非线性特征信息, 从而有效地减小数据集的特征维度, 并保证特征信息的完整性。因此, 自动编码器非常适合用于降维和特征学习任务。其中, 丢弃层的公式如公式(4)所示, 其中 r 是一个独立且均匀分布的向量, 满足概率为 p 的伯努利分布, 它的形状和$\lambda(n)$一致。*表示元素的乘积, $\lambda(n)$是当前层的输出, $\lambda(n)'$是使用丢弃层后的输出。其中, 丢弃层的框架结构图如图 5 所示。

$$\begin{aligned} r &\sim Bernoulli(p) \\ \lambda(n)' &= r * \lambda(n) \end{aligned} \tag{4}$$

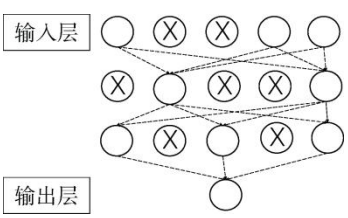

图 5　丢弃层的框架结构图

Figure 5　The frame structure diagram of the Dropout layer

3.4　异常行为检测

在对模型进行训练后, 本文使用训练后的模型对测试数据进行分类, 并获得相应的预测结果和评估结果。最终, 本文选择了 Adam 优化器, 并使用了 sigmoid 激活函数来预测类别概率, 预测概率值大于 0.5 则预测为异常类别数据, 反之则预测为正常类别数据。sigmoid 函数的计算公式如公式(5)所示, 其中, t 为输入, $\sigma(t)$为对应输出的概率值。

$$\sigma(t) = \frac{1}{1+e^{-t}} \tag{5}$$

本文同时使用了交叉熵来计算损失值。交叉熵的计算公式如公式(6)所示, 其中, n 代表类别数目, x_i代表第 i 个样本, $p(x_i)$代表第 i 个样本的真实值, $q(x_i)$ 代表第 i 个样本的预测值, $J(p,q)$代表交叉熵的计算值。

$$J(p,q) = -\sum_{i=1}^{n} p(x_i)\log\left(q(x_i)\right) \tag{6}$$

4　实验和结果

4.1　数据集

UNSW-NB15 和 CICIDS2017 数据集是近几年新创建的入侵检测数据集, 包含了较全面的攻击行为, 是入侵检测领域研究的通用数据集, 被广泛用于入侵检测领域。目前众多国内外学者[28-31]均使用 UNSW-NB15 和 CICIDS2017 数据集来验证所提出的入侵检测方法检测能力。

4.1.1　UNSW-NB15 数据集

UNSW-NB15 数据集[32]是在澳大利亚网络安全中心的网络范围实验室中创建的, 收集了 2015 年 1 月 22 日的 16 个小时数据和 2015 年 2 月 17 日的 15 个小时数据。该数据集是当代真实世界的网络流量, 包含了较全面的综合攻击活动。UNSW-NB15 数据集包含 10 种数据类型, 分别是正常流量, 模糊攻击(Fuzzers), 分析攻击(Analysis), 后门攻击(Backdoor), 拒绝服务攻击(DoS), 漏洞利用攻击(Exploits), 通用攻击(Generic), 侦察攻击(Reconnaissance), Shellcode 和蠕虫攻击(Worms)。

UNSW-NB15 数据集中的一个分区被配置为训练数据集和测试数据集。其中, 训练数据集 UNSW_NB15_training-set.csv 中有 175341 条流量数据, 测试集 UNSW_NB15_testing-set.csv 中有 82332 条流量数据。本文使用了整个 UNSW-NB15 的训练数据和测试数据来评估所提出方法的检测能力。

4.1.2　CICIDS2017 数据集

CICIDS2017 数据集[33]是 Sharafaldin 等人创建的基于真实网络场景的可靠的入侵检测数据集, 收集了 2017 年 7 月 3 日(周一)到 7 月 7 日(周五)共 5 天的数据。该数据集包含了全面的新型攻击类型, 例如端口扫描攻击(PortScan)、分布式拒绝服务攻击(DDoS)、僵尸网络攻击(Botnet)等。本文使用了周一、周五两天的数据集来评估所提出的方法的检测能力。其中, 周一、周五两天的数据集共包含 1233163 条流量数据, 本文按照训练集：测试集=8：2 的比例来对整体流量数据进行划分, 划分后训练集包含 986530 条流量数据, 测试集包含 246633 条流量数据。

本文使用了 UNSW-NB15 和 CICIDS2017 数据集来验证所提出方法的检测能力。两个数据集中都存在数据类别不平衡问题。例如, UNSW-NB15 数据集中异常流量数据(例如 Analysis, Shellcode, Backdoor 和 Worms)数量远小于正常流量数据数量。如果不对这些不平衡数据集进行处理, 异常流量样本很可能会被视为噪声数据而被忽略, 这种不平衡数据集难以较好地训练模型。本文使用 k-means SMOTE 过采样方法, 并参考了 Seo 等人[34]的过采样处理比例, 对不平衡数据集进行处理, 来改善对少数异常类别数据的检测能力。UNSW-NB15 和 CICIDS2017 的原始训练数据集的分布以及通过 k-means SMOTE 方法处理的训练数据集的分布分别如图 6、图 7 所示。

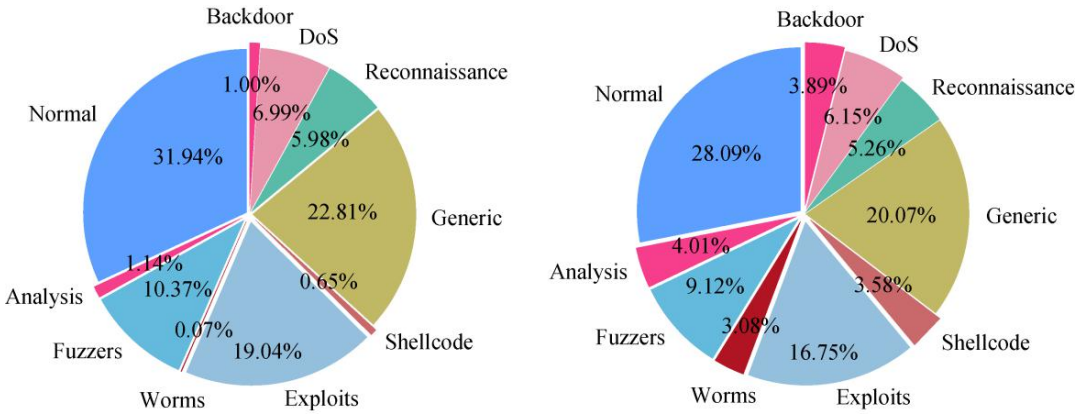

图 6　UNSW-NB15 数据集分布。左边图像是原始训练数据集, 右边图像是经过 k-means SMOTE 方法处理的数据集

Figure 6　The distribution of UNSW-NB15 dataset. The left image is the original training dataset, and the right image is the dataset processed by the k-means SMOTE method.

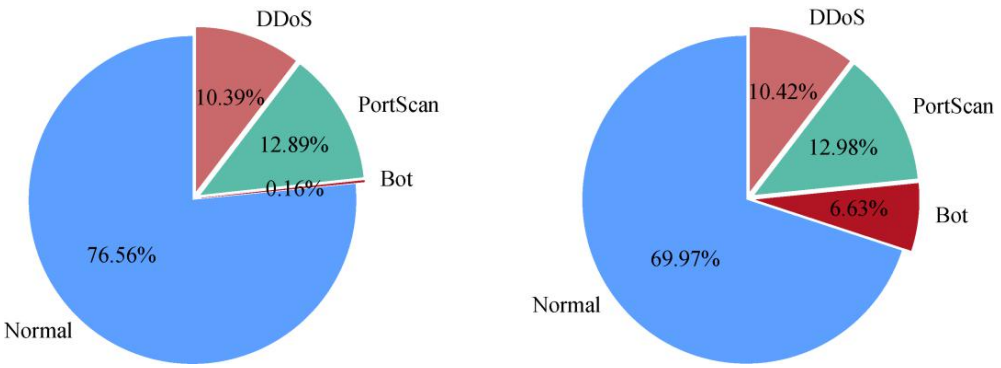

图 7　CICIDS2017 数据集分布。左边图像是原始训练数据集, 右边图像是经过 k-means SMOTE 方法处理的数据集

Figure 7　The distribution of CICIDS2017 dataset. The left image is the original training dataset, and the right image is the dataset processed by the k-means SMOTE method.

4.2　实验设置

本文使用自动编码器对数据进行降维处理。具体而言, 为了更好地表示原始数据集, 使得自动编码器能够最大程度地保留原始数据的特征。本文测试了多种自动编码器的结构, 对不同的自动编码器神经网络的隐藏层神经元数量进行实验, 通过使用交叉熵来计算损失值。测试了从 1 个隐藏神经元到 100 个隐藏神经元的训练损失值。在 UNSW-NB15 和 CICIDS2017 数据集上自动编码器的最佳隐藏层神经元数量分别为 24 和 35。在后续的实验中, 本文使用了最佳隐藏层数量来训练模型。

4.3　评价指标

为了评估所提出的方法, 本文使用了准确率, 精准率, 召回率, F1 值 4 个指标。这 4 个评价指标依赖于混淆矩阵的 4 个参数, 分别是真阳性、真阴性、假阳性、假阴性。

- 真阳性(*True Positive*, *TP*): 正确分类为攻击类别的攻击样本数量。
- 真阴性(*True Negative*, *TN*): 正确分类为正常类别的正常样本数量。

- 假阳性(*False Positive*, *FP*): 错误分类为攻击类别的正常样本数量。
- 假阴性(*False Negative*, *FN*): 错误分类为正常类别的攻击样本数量。

4 个指标的计算公式如下所示:

- 准确率(*Accuracy*): 正确分类的样本数占总样本数的百分比, 是一个用于评估入侵检测方法整体性能的指标:

$$ACC = \frac{TP + TN}{TP + FP + TN + FN} \times 100\% \tag{7}$$

- 精准率(*Precision*): 正确分类为攻击的攻击样本数占分类为攻击样本总数的百分比:

$$Precision = \frac{TP}{TP + FP} \times 100\% \tag{8}$$

- 召回率(*Recall*): 正确分类为攻击的攻击样本数占攻击样本总数的百分比:

$$Recall = \frac{TP}{TP + FN} \times 100\% \tag{9}$$

- F1 值(*F-measure*): *Precision* 和 *Recall* 的综合评价指标:

$$F\text{-}measure = 2 \times \frac{Precision \times Recall}{Precision + Recall} \times 100\% \tag{10}$$

对于入侵检测方法而言, 准确率, 精准率, 召回率和 F1 值越高, 则所提出的入侵检测方法的检测性能越好。

4.4 基线对比算法

为了证明 KSAIDS 的有效性, 本文对机器学习方法、深度学习方法、及不同的采样方法进行了对比实验。使用了六种常见的机器学习方法进行对比实验, 分别是决策树、随机森林、高斯朴素贝叶斯(Gaussian Naive Bayes, GNB)、伯努利朴素贝叶斯(Bernoulli Naive Bayes, BNB)、AdaBoost 和 XGBoost, 并且与 Jing 等人[28]和 Ahmim 等人[29]在 2019 年提出的基于机器学习的入侵检测方法的检测结果进行对比。本文还使用了六种常见的深度学习方法进行对比实验, 分别为 RNN、GRU、CNN、CNN-LSTM、SAE、AE, 并且与 Roy 等人[30]和 Andresini 等人[31]在近两年提出的基于深度学习的入侵检测方法的检测结果进行对比。除此之外, 本文使用了七种常用的处理不平衡数据的方法作为对照实验, 包含了三种欠采样方法和四种过采样方法。其中, 3 种欠采样方法, 分别为随机欠采样、NearMiss、CondensedNearestNeighbour; 4 种过采样方法, 分别为随机过采样、SMOTE、ADASYN、BorderlineSMOTE。

4.5 结果分析

4.5.1 机器学习方法的性能比较

表 1 显示了本文所提出的 KSAIDS 与传统机器学习方法在 UNSW-NB15 和 CICIDS2017 数据集上的检测对比结果及与 Jing 等人[28]和 Ahmim 等人[29]研究工作的对比。从表 1 可以看出, 对于传统的机器学习方法, 基于树的方法的检测结果通常优于基于概率的方法, 原因是朴素贝叶斯方法难以处理具有相关性的特征。但是, 基于概率的方法比基于树的方法耗费更少的执行时间, 其中, 执行时间为训练过程和测试过程所花费的总时间。集成学习方法的检测效果通常优于单一方法, 因为单一方法难以捕获到更多的数据信息, 然而, 单一方法比集成方法消耗的时间更少。

从表 1 还可以看出, KSAIDS 在 UNSW-NB15 和 CICIDS2017 数据集上的检测准确率分别为 99.06％和 99.16%, F-measure 分别为 99.15%和 98.22%, 优于传统的机器学习方法, 证明了该方法的有效性。其中, 召回率也称为检测率, KSAIDS 在两个数据集上的召回率分别为 100％和 98.28%, 相比于 7 种机器学习方法, KSAIDS 的召回率仍较高。同时, KSAIDS 花费的时间多于决策树等单一模型方法, 但是少于随机森林等集成学习方法。这些现象的原因在于, KSAIDS 对不平衡流量数据进行处理, 避免了由于样本数量过多而使得模型偏向正常类别的情况。并且随着数据数量和复杂度的增加, 浅层学习方法的学习能力会受到一定的限制。

本文还与 Jing 等人[28]和 Ahmim 等人[29]的工作进行了比较。如表 1 所示, Jing 等人[28]提出了一种结合了非线性缩放方法的支持向量机入侵检测方法, 该方法在 UNSW-NB15 数据集上的检测准确率为 85.99%, 召回率为 86%。Ahmim 等人[29]提出了一种基于决策树和规则的入侵检测方法, 该方法在 CICIDS2017 数据集上的检测准确率为 96.66%, 召回率为 94.47%。由表可知, KSAIDS 的检测能力优于 Jing 等人[28]和 Ahmim 等人[29]提出的方法, 证明了本文所提方法的有效性。原因是他们的工作均使用浅层的方法来学习海量数据中的信息, 并且未考虑到网络流量数据的整体分布情况。

图 8 展示了本文所提出的 KSAIDS 与传统机器学习方法在 UNSW-NB15 和 CICIDS2017 数据集上的检测准确率、精准率、召回率、F-measure 的对比结果, 由图 8 可知, 本文所提出的 KSAIDS 的检测能力优于其他 6 种机器学习方法, 能够有效地检测网络异常流量。

表 1　机器学习方法在 UNSW-NB15 数据集和 CICIDS2017 数据集上的检测结果比较

Table 1　Detection results comparison of machine learning methods on the UNSW-NB15 and the CICIDS2017 dataset

方法	UNSW-NB15 数据集				
	准确率(Accuary)	精准率(Precision)	召回率(Recall)	F-measure	执行时间(s)
决策树[12]	0.8858	0.9858	0.8014	0.8841	1.4201
随机森林[13]	0.9289	0.9874	0.8805	0.9309	56.0396
高斯朴素贝叶斯[14]	0.6381	0.9792	0.3415	0.5064	1.2242
伯努利朴素贝叶斯[14]	0.7455	0.8384	0.6588	0.7378	0.6970
AdaBoost[15]	0.9288	0.9873	0.8803	0.9307	33.3993
XGBoost[16]	0.9182	0.9878	0.8602	0.9196	445.6249
Jing 等人[28]	0.8599	\	0.8600	\	\
KSAIDS	**0.9906**	**0.9832**	**1.0**	**0.9915**	**31.9911**
方法	CICIDS2017 数据集				
	准确率(Accuary)	精准率(Precision)	召回率(Recall)	F-measure	执行时间(s)
决策树[12]	0.8891	0.9610	0.5483	0.6983	8.5593
随机森林[13]	0.9451	0.9292	0.8287	0.8761	162.2207
高斯朴素贝叶斯[14]	0.8283	0.5770	0.9961	0.7308	1.5116
伯努利朴素贝叶斯[14]	0.8993	0.7009	0.9933	0.8219	1.5668
AdaBoost[15]	0.9364	0.8923	0.8282	0.8590	305.7122
XGBoost[16]	0.9325	0.8799	0.8240	0.8510	1327.1984
Ahmim 等人[29]	0.9666	\	0.9447	\	167.11
KSAIDS	**0.9916**	**0.9816**	**0.9828**	**0.9822**	**149.7567**

(注: “\”表示该方案无法参与该评分项。)

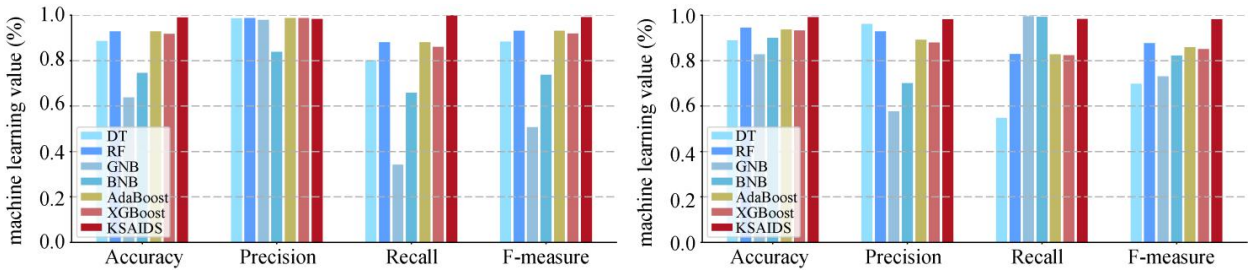

图 8　机器学习方法的检测结果比较。左边为 UNSW-NB15 数据集检测结果, 右边为 CICIDS2017 数据集检测结果

Figure 8　Detection results comparison of machine learning methods. The left results are on the UNSW-NB15 dataset, and the right results are on the CICIDS2017 dataset.

4.5.2　深度学习方法的性能比较

由于深度学习方法可以学习数据的内在规律, 因此能够更好地对海量网络流量数据进行拟合。为了验证 KSAIDS 具有较好的检测能力, 本文使用了入侵检测领域常用的六种深度神经网络方法作为对照实验, 在 UNSW-NB15 和 CICIDS2017 数据集上的检测对比结果及与 Roy 等人[30]和 Andresini 等人[31]研究工作的对比如表 2 所示。

从表 2 可以看出, 在 UNSW-NB15 数据集上, CNN-LSTM 的检测结果最差, 准确率为 49.67%, F-measure 为 48.85%, 且耗时最长, 为 1287.8755s。同时, 在 CICIDS2017 数据集上, CNN-LSTM 的检测准确率为 86.85%, F-measure 为 65.91%, 耗时仍最长, 为 3055.3014s, 检测结果仍较差。实验结果证明网络流量数据并未较好地发挥 CNN-LSTM 方法的优势, 导致检测结果不尽人意。在两个数据集上, GRU 检测结果均优于 RNN 且耗费的时间也更长, 因为 GRU 方法改善了 RNN 的梯度消失的问题, 能够建模长时间依赖, 因此检测效果更优。CNN 的检测结果也优于 RNN, 且耗费时间更短, 因为 CNN 通过提取特征图和池化操作, 能够减少模型参数, 提升训练速度。在两个数据集上, AE 的检测准确率和 F-measure 均高于 SAE, 因为 SAE 加入了正则化限制, 丢失了部分有效信息。

表 2 深度学习方法在 UNSW-NB15 数据集和 CICIDS2017 数据集上的检测结果比较

Table 2 Detection results comparison of deep learning methods on the UNSW-NB15 and the CICIDS2017 dataset

方法	UNSW-NB15 数据集				
	准确率(Accuary)	精准率(Precision)	召回率(Recall)	F-measure	执行时间(s)
RNN[17]	0.7346	0.7122	0.8692	0.7829	707.7427
GRU[18]	0.8879	0.8689	0.9379	0.9021	888.7156
CNN[19]	0.7939	0.7350	0.9783	0.8394	219.7034
CNN-LSTM[20]	0.4967	0.5546	0.4364	0.4885	1287.8755
SAE[21]	0.9201	0.8733	1.0	0.9324	56.7223
AE[22]	0.9415	0.9040	1.0	0.9496	55.6398
Roy 等人[30]	0.9571	1.00	0.9600	0.9800	\
KSAIDS	**0.9906**	**0.9832**	**1.0**	**0.9915**	**31.9911**
方法	CICIDS2017 数据集				
	准确率(Accuary)	精准率(Precision)	召回率(Recall)	F-measure	执行时间(s)
RNN[17]	0.8347	0.6833	0.5468	0.6075	1558.3316
GRU[18]	0.8581	0.7256	0.6332	0.6762	1883.8156
CNN[19]	0.8937	0.7770	0.7653	0.7711	444.6815
CNN-LSTM[20]	0.8685	0.8372	0.5435	0.6591	3055.3014
SAE[21]	0.9495	0.9469	0.8307	0.8850	166.7451
AE[22]	0.9536	0.9682	0.8289	0.8932	161.6302
Andresini 等人[31]	0.9790	\	\	0.9493	\
KSAIDS	**0.9916**	**0.9816**	**0.9828**	**0.9822**	**149.7567**

(注: “\”表示该方案无法参与该评分项。)

本文还与 Roy 等人[30]和 Andresini 等人[31]的工作进行了比较。Roy 等人[30]提出了一种基于双向长短期记忆网络的入侵检测方法, 在 UNSW-NB15 数据集上的检测准确率为 95.71%, F-measure 为 98.00%。Andresini 等人[31]提出了一种基于卷积神经网络的多通道网络流表示的入侵检测方法, 该方法在 CICIDS2017 数据集上的检测准确率为 97.90%, F-measure 为 94.93%。如表 2 所示, KSAIDS 的检测能力优于 Roy 等人[30]和 Andresini 等人[31]提出的方法, 原因是 Roy 等人仅使用了五种特征, 并且他们的工作都没有对不平衡流量数据进行处理, 也证明了 KSAIDS 的有效性。

图 9 展示了本文所提出的 KSAIDS 与常用的深度学习方法在 UNSW-NB15 和 CICIDS2017 数据集上的检测准确率、精准率、召回率、F-measure 的对比结果, 从图 9 中可以看出, 本文所提出的 KSAIDS 的检测能力优于其他六种深度学习方法, 能够有效地检测网络异常流量。并且 KSAIDS 检测耗费时间远少于其他深度学习方法, 在 UNSW-NB15 和 CICIDS2017 数据集上的检测时间分别为 31.9911s 和 149.7567s。由于 CICIDS2017 测试数据集数量多于 UNSW-NB15 测试数据集数量, 因此, 在 CICIDS2017 数据集上检测结果的耗费时间普遍多于在 UNSW-NB15 数据集上检测结果的耗费时间。

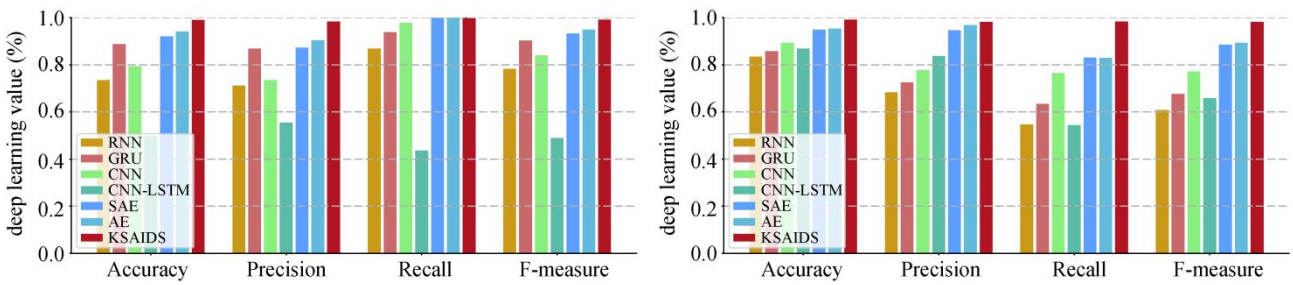

图 9 深度学习方法的检测结果比较。左边为 UNSW-NB15 数据集检测结果, 右边为 CICIDS2017 数据集检测结果

Figure 9 Detection results comparison of deep learning methods. The left results are on the UNSW-NB15 dataset, and the right results are on the CICIDS2017 dataset.

4.5.3 处理不平衡数据方法的性能比较

本文使用七种常用的处理不平衡数据的方法结合 AE 作为对照实验, 与本文所提出的 KSAIDS 的检测对比结果如表 3 所示。

三种欠采样对比方法的采样原理分别为: 随机欠采样方法即随机选取多数类别样本数据, 并将其删除; NearMiss 方法是一种启发式算法, 通过删除多数类别样本中与多个最近邻/最远邻少数类别样本平均距离最短的样本, 来平衡数据分布; CondensedNearestNeighbour 方法通过使用 1 近邻原则迭代地决定是否删除一个样本, 该方法的缺点是对于噪声数据非常敏感。由表 3 可知, 相比于基线实验 AE, 使用欠采样方法来处理不平衡数据后, 在两个数据集上的检测准确率、F-measure 等指标均大幅下降, 这是因为欠采样的方式删除了大部分正常流量数据, 丢失了很多重要的数据信息, 虽然减少了检测时间, 但是检测结果也随之大幅降低。其中, 随机欠采样方法在两个数据集上的检测准确率、F-measure 均高于另外两种欠采样方法。

四种过采样对比方法的采样原理分别为: 随机过采样即随机选取少数类别样本数据, 进行复制, 该方法存在过拟合问题; SMOTE 方法通过线性插值来增加少数类别数据, 避免过拟合, 但是存在加剧类内不平衡问题和放大噪声的问题; ADASYN 方法采用自适应合成抽样, 对不同的少数类别样本赋予不同的权重, 从而生成不同数量的样本; BorderlineSMOTE 方法仅使用边界上的少数类样本来合成新样本, 从而改善样本的类别分布。

由表 3 还可以发现, 相比于基线实验 AE, 使用过采样方法对不平衡流量数据进行处理后, 检测准确率、F-measure 等指标均有提升, 其中, 检测效果最好的就是本文提出的使用 k-means SMOTE 过采样方法的 KSAIDS, 这是因为 k-means SMOTE 方法在 SMOTE 方法的基础上, 通过结合 k-means 聚类方法, 仅在安全区域中进行过采样, 解决了类间不平衡和类内不平衡问题, 及 SMOTE 方法放大噪声的问题。并且在两个数据集上, 使用 SMOTE、ADASYN、BorderlineSMOTE、k-means SMOTE 四种过采样方法处理不平衡数据后, 检测结果均优于简单的随机过采样方法, 因为简单的随机过采样方法存在较严重的过拟合问题。这一实验结果也证明了通过改善不平衡数据的整体分布, 能够有效地提升模型的检测能力。

表 3 处理不平衡数据方法的检测结果比较

Table 3 Comparison of detection results of methods for processing unbalanced data

类别	方法	UNSW-NB15 数据集				
		准确率(Accuary)	精准率(Precision)	召回率(Recall)	F-measure	执行时间(s)
基线	AE	0.9415	0.9040	1.0	0.9496	55.6398
欠采样	随机欠采样+AE	0.6689	0.6245	1.0	0.7688	2.7115
	NearMiss+AE	0.5550	0.5530	1.0	0.7122	2.6846
	CondensedNearestNeighbour+AE	0.5506	0.5506	1.0	0.7101	2.9052
过采样	随机过采样+AE	0.9552	0.9248	1.0	0.9609	31.3458
	SMOTE +AE	0.9697	0.9479	1.0	0.9732	32.8164
	ADASYN+AE	0.9678	0.9448	1.0	0.9716	32.9510
	BorderlineSMOTE +AE	0.9728	0.9530	1.0	0.9759	33.0405
	k-means SMOTE+AE(KSAIDS)	**0.9906**	**0.9832**	**1.0**	**0.9915**	**31.9911**
类别	方法	CICIDS2017 数据集				
		准确率(Accuary)	精准率(Precision)	召回率(Recall)	F-measure	执行时间(s)
基线	AE	0.9536	0.9682	0.8289	0.8932	161.6302
欠采样	随机欠采样+AE	0.7864	0.5228	0.9970	0.6860	3.2128
	NearMiss+AE	0.3737	0.2345	0.7409	0.3563	3.2241
	CondensedNearestNeighbour+AE	0.5038	0.2882	0.7631	0.4184	2.7317
过采样	随机过采样+AE	0.9806	0.9737	0.9428	0.9580	150.9925
	SMOTE +AE	0.9842	0.9834	0.9484	0.9656	152.7613
	ADASYN+AE	0.9850	0.9787	0.9567	0.9676	159.6545
	BorderlineSMOTE +AE	0.9849	0.9829	0.9520	0.9672	152.7245
	k-means SMOTE+AE(KSAIDS)	**0.9916**	**0.9816**	**0.9828**	**0.9822**	**149.7567**

5 结论

在入侵检测研究中, 大规模网络流量数据存在严重的稀疏性问题。本文提出了一种新的入侵检测方法, 称为 k 均值稀疏异常入侵检测方法, 通过使用 k-means SMOTE 过采样方法来处理不平衡的正负样本实例, 并结合自动编码器方法从大规模数据集中提取有效的非线性结构信息, 能够检测低频异常行为。为了验证所提出方法的有效性, 本文使用了 UNSW-NB15 和 CICIDS2017 数据集进行了广泛的实验。实验结果表明, 与最新的入侵检测算法相比, KSAIDS 显着地提升了入侵检测的检测结果与性能, 且通用性、泛化性较强。

KSAIDS 需要使用带标签的训练数据集, 通过改善网络流量数据的分布, 来训练检测能力更强的模型, 从而能够更好地对未知流量数据进行预测分类。通过改善网络流量数据的分布, 以及对海量数据进行降维, 有效地提升了入侵检测模型的检测能力, 缺点在于 KSAIDS 仍然需要带标签的训练数据来训练模型, 在真实的网络环境下, 需要人工进行类别标注。在未来的研究工作中, 我们会使用 KSAIDS 对现实中真实的网络流量数据进行检测, 来验证该方法的有效性。

致 谢 感谢中国科学院网络测评技术重点实验室的各位老师和同学提出的有益建议。感谢审稿专家和编辑部老师对本文提出的有益建议及指导。

参考文献

[1] J.P. Anderson, Computer security threat monitoring and surveillance[R]. Technical report, James P.Anderson Company, Fort Washington, Pennsylvania, 1980.

[2] AlYousef M Y, Abdelmajeed N T. Dynamically Detecting Security Threats and Updating a Signature-Based Intrusion Detection System's Database[J]. *Procedia Computer Science*, 2019, 159: 1507-1516.

[3] Nikolova E, Jecheva V. Some Similarity Coefficients and Application of Data Mining Techniques to the Anomaly-Based IDS[J]. *Telecommunication Systems*, 2012, 50(2): 127-135.

[4] LeCun Y, Bengio Y, Hinton G. Deep Learning[J]. *Nature*, 2015, 521(7553): 436-444.

[5] Roy D D, Shin D, Bioengineering, et al. Network intrusion detection in smart grids for imbalanced attack types using machine learning models[C]. *2019 International Conference on Information and Communication Technology Convergence*, 2019: 576-581.

[6] Azizjon M, Jumabek A, Kim W, et al. 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.

[7] Yan B H, Han G D. LA-GRU: Building Combined Intrusion Detection Model Based on Imbalanced Learning and Gated Recurrent Unit Neural Network[J]. *Security and Communication Networks*, 2018, 2018: 1-13.

[8] Yang J, Li T, Liang G, et al. A Simple Recurrent Unit Model Based Intrusion Detection System with DCGAN[J]. *IEEE Access*, 7: 83286-83296.

[9] Abdulhammed R, Faezipour M, Musafer H, et al. Efficient network intrusion detection using PCA-based dimensionality reduction of features[C]. *2019 International Symposium on Networks, Computers and Communications*, 2019: 1-6.

[10] Wutyi K S, Thwin M M S. Heuristic Rules for Attack Detection Charged by NSL KDD Dataset[M]. Advances in Intelligent Systems and Computing. Cham: Springer International Publishing, 2015: 137-153.

[11] Wang Y, Meng W Z, Li W J, et al. A Fog-Based Privacy-Preserving Approach for Distributed Signature-Based Intrusion Detection[J]. *Journal of Parallel and Distributed Computing*, 2018, 122: 26-35.

[12] Sahu S, Mehtre B M, Communication N A B T, et al. Network intrusion detection system using J48 Decision Tree[C]. *2015 International Conference on Advances in Computing, Communications and Informatics*, 2015: 2023-2026.

[13] Farnaaz N, Jabbar M A. Random Forest Modeling for Network Intrusion Detection System[J]. *Procedia Computer Science*, 2016, 89: 213-217.

[14] Yu N. A Novel Selection Method of Network Intrusion Optimal Route Detection Based on Naive Bayesian[J]. *International Journal of Applied Decision Sciences*, 2018, 11(1): 1.

[15] Hu W M, Gao J, Wang Y G, et al. Online Adaboost-Based Parameterized Methods for Dynamic Distributed Network Intrusion Detection[J]. *IEEE Transactions on Cybernetics*, 2014, 44(1): 66-82.

[16] Dhaliwal S, Nahid A A, Abbas R. Effective Intrusion Detection System Using XGBoost[J]. *Information*, 2018, 9(7): 149.

[17] Yin C L, Zhu Y F, Fei J L, et al. A Deep Learning Approach for Intrusion Detection Using Recurrent Neural Networks[J]. *IEEE Access*, 5: 21954-21961.

[18] Agarap A F M. A Neural Network Architecture Combining Gated Recurrent Unit (GRU) and Support Vector Machine (SVM) for Intrusion Detection in Network Traffic Data[C]. *The 2018 10th International Conference on Machine Learning and Computing*, 2018: 26-30.

[19] Ding Y L, Zhai Y Q. Intrusion Detection System for NSL-KDD Dataset Using Convolutional Neural Networks[C]. *The 2018 2nd International Conference on Computer Science and Artificial Intelligence*, 2018: 81-85.

[20] Kim T Y, Cho S B. CNN-LSTM Neural Networks for Anomalous Database Intrusion Detection in RBAC-Administered Model[M]. Communications in Computer and Information Science. Cham: Springer International Publishing, 2019: 131-139.

[21] Qureshi A S, Khan A, Shamim N, et al. Intrusion Detection Using Deep Sparse Auto-Encoder and Self-Taught Learning[J]. *Neural*

Computing and Applications, 2020, 32(8): 3135-3147.

[22] Abolhasanzadeh B, Communication N A B T, Processing C A, et al. Nonlinear dimensionality reduction for intrusion detection using auto-encoder bottleneck features[C]. *2015 7th Conference on Information and Knowledge Technology*, 2015: 1-5.

[23] Bajer D, Zonć B, Dudjak M, et al. Performance analysis of SMOTE-based oversampling techniques when dealing with data imbalance[C]. *2019 International Conference on Systems, Signals and Image Processing*, 2019: 265-271.

[24] Douzas G, Bacao F, Last F. Improving Imbalanced Learning through a Heuristic Oversampling Method Based on K-Means and SMOTE[J]. *Information Sciences*, 2018, 465: 1-20.

[25] Last F, Douzas G, Bacao F. Oversampling for Imbalanced Learning Based on K-Means and SMOTE[EB/OL]. 2017: arXiv: 1711.00837. https://arxiv.org/abs/1711.00837.

[26] Chicco D, Sadowski P, Baldi P. Deep Autoencoder Neural Networks for Gene Ontology Annotation Predictions[C]. *The 5th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics*, 2014: 533-540.

[27] Varastehpour S, Sharifzadeh H, Ardekani I, et al. Vein pattern visualisation and feature extraction using sparse auto-encoder for forensic purposes[C]. *2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance*, 2019: 1-8.

[28] Jing D S, Chen H B, Aerospace, et al. SVM based network intrusion detection for the UNSW-NB15 dataset[C]. *2019 IEEE 13th International Conference on ASIC*, 2020: 1-4.

[29] Ahmim A, Maglaras L, Ferrag M A, et al. A novel hierarchical intrusion detection system based on decision tree and rules-based models[C]. *2019 15th International Conference on Distributed Computing in Sensor Systems*, 2019: 228-233.

[30] Roy B, Cheung H, Communication N A B T, et al. A deep learning approach for intrusion detection in Internet of Things using Bi-directional long short-term memory recurrent neural network[C]. *2018 28th International Telecommunication Networks and Applications Conference*, 2019: 1-6.

[31] Andresini G, Appice A, Mauro N D, et al. Multi-Channel Deep Feature Learning for Intrusion Detection[J]. *IEEE Access*, 8: 53346.

[32] Moustafa N, Slay J, Communication N A B T, et al. 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.

[33] Sharafaldin I, Habibi Lashkari A, Ghorbani A A. Toward generating a new intrusion detection dataset and intrusion traffic characterization[C]. *The 4th International Conference on Information Systems Security and Privacy*, 2018: 108-116.

[34] Seo J H, Kim Y H. Machine-Learning Approach to Optimize SMOTE Ratio in Class Imbalance Dataset for Intrusion Detection[J]. *Computational Intelligence and Neuroscience*, 2018, 2018: 9704672.

蹇诗婕 于 2018 年在北京科技大学信息安全专业获得学士学位。现在中国科学院信息工程研究所第六研究室攻读硕士学位。研究领域为网络安全态势感知、入侵检测、流量异常检测等。Email: jianshijie@iie.ac.cn

刘岳 于 2018 年在重庆大学信息安全专业获得学士学位。现在中国科学院信息工程研究所第六研究室攻读硕士学位。研究领域为恶意代码检测。Email: liuyue0909@iie.ac.cn

姜波 于 2016 年在中国科学院大学计算机系统结构专业获得博士学位。现任中国科学院信息工程研究所副研究员。研究领域为网络安全态势感知、知识图谱、数据挖掘等。Email: jiangbo@iie.ac.cn

卢志刚 于 2010 年在中国科学院研究生院获得博士学位。现任中国科学院信息工程研究所高级工程师, 中国科学院网络空间安全学院副教授。研究领域为网络安全态势感知、网络攻击检测、移动终端安全等。Email: luzhigang@iie.ac.cn

刘玉岭 于 2013 年在中国科学院软件研究所获得博士学位。现任中国科学院信息工程研究所高级工程师。研究领域为网络安全态势感知、网安大数据分析、安全测评认证等。Email: liuyuling@iie.ac.cn

刘宝旭 于 2002 年在中国科学院研究生院获得博士学位。现任中国科学院信息工程研究所研究员, 第六研究室主任。研究领域为网络安全攻防对抗、网络安全测评技术等。Email: liubaoxu@iie.ac.cn