Computer Science and Application 计算机科学与应用, 2021, 11(6), 1637-1648
Published Online June 2021 in Hans. http://www.hanspub.org/journal/csa
https://doi.org/10.12677/csa.2021.116169

Hans汉斯

基于改进条件变分自编码器的入侵检测研究

朱　琼[1]，袁永晖[2]，田春岐[2]

[1]中国航发上海商用航空发动机制造有限责任公司，上海
[2]同济大学电子与信息工程学院，上海
Email: lgy423@126.com

收稿日期：2021年5月10日；录用日期：2021年6月1日；发布日期：2021年6月8日

摘　要

现有的入侵检测方法大多集中于提高整体检测率，而应用于不平衡样本集上，传统方法往往在少数类攻击样本的识别上存在识别准确率低、误报率高的问题。因此，提出了一个结合入侵检测条件变分自编码器(Intrusion Detection Conditional Variational Auto Encoder, IDCVAE)和深度信念网络(Deep Belief Nets, DBN)的入侵检测方法。该方法首先利用IDCVAE学习数据的稀疏表示，然后使用其解码器部分扩充少数类样本，解决样本不均衡问题。最后利用DBN对平衡后的新数据集进行特征提取和分类。实验结果表明，本文的方法在保持整体检测率较高的同时，有效地提高了少数类攻击的检测率及误报率。

关键词

入侵检测，条件变分自编码器，生成网络，过采样，深度信念网络

Research on Intrusion Detection Based on Improved Conditional Variational Auto Encoder

Qiong Zhu[1], Yonghui Yuan[2], Chunqi Tian[2]

[1]AECC Shanghai Commercial Aircraft Engine Manufacturing Co., LTD., Shanghai
[2]College of Electronics and Information Engineering, Tongji University, Shanghai
Email: lgy423@126.com

Received: May 10th, 2021; accepted: Jun. 1st, 2021; published: Jun. 8th, 2021

Abstract

At present, most of the existing intrusion detection methods focus on improving the overall detection rate. However, traditional methods often perform poorly in detecting minority class samples.

文章引用: 朱琼, 袁永晖, 田春岐. 基于改进条件变分自编码器的入侵检测研究[J]. 计算机科学与应用, 2021, 11(6): 1637-1648. DOI: 10.12677/csa.2021.116169

Therefore, this paper proposed an intrusion detection method based on Intrusion Detection Conditional Variational Auto Encoder (IDCVAE) and Deep Belief Nets (DBN). IDCVAE can learn potential sparse representations in network data features and oversampling the minority class data. Deep belief network can effectively extract and classify the balanced new data set. Experimental results show that, the method in this paper effectively improves the detection rate of minority while keeping the high overall detection rate and low false alarm rate.

Keywords

Intrusion Detection, Conditional Variational Auto Encoder, Generative Network, Oversampling, Deep Belief Nets

1. 引言

近些年来，随着云计算、5G 通讯以及人工智能的迅猛发展，网络安全问题也日益严重。入侵检测系统(Intrusion Detection System, IDS) [1]作为防范网络攻击的重要手段，已经成为了许多大型组织和机构防御网络攻击的首选。相较于传统的被动防御策略，入侵检测系统可以主动检测并发掘网络内部和外部的攻击行为，在攻击生效前，就能拦截攻击并向管理人员发出安全预警。

国内外众多学者在这一领域提出了许多行之有效的方案，包括异常检测方法、聚类算法、集成学习、浅层学习、深度学习[2]等。文献[3]提出使用改进后的 K-means 算法对初始聚类中心进行优化，从而提升入侵检测模型的检测速度。文献[4]提出用遗传算法来提升 SVM 在入侵检测数据中的分类效果。文献[5]利用贝叶斯网络显著地提高了攻击数据的识别准确率。但其缺点在于需要大量的先验知识且对输入数据的表达形式非常敏感。文献[6]将一系列的集成方法应用到异常检测中，取得了不错的效果。随着近些年来深度网络的不断发展，神经网络在入侵检测中也得到了广泛的应用。文献[7]应用深度神经网络(deep neural network)与预处理后的数据，建立了深度学习模型。并在 KDD Cup 99 数据集上准确率、检测率都取得了较好的效果。文献[8]提出了一种多尺度卷积 CNN 模型。用多种不同的尺度卷积核对大量高维无标签原始数据进行了多层次特征提取，再采用正则化方法优化网络结构学习率，以获得原始数据的最优特征表示。相较于原始的 CNN 收敛更快，准确率更高。文献[9]提出了基于卷积神经网络(CNN)和长短时记忆网络(LSTM)的检测模型，首先利用 CNN 对输入的高维数据进行特征提取，然后将其输出作为 LSTM 的输入，最后利用全连接层得到分类结果。

在大量的网络数据中，攻击数据只占很小的一部分。这就导致我们所能够获取到的数据集往往是非常不平衡的。传统机器学习方法直接应用于不平衡数据上往往只能提高整体的检测精度，而对于少数类的分类效果则不尽如人意。如何解决入侵检测中不平衡数据的问题，许多专家学者进行了大量研究。文献[10]中作者引入了一个聚类一致性系数从而更好地找到了少数类的样本边界，进一步优化了 SMOTE (Synthetic Minority Oversampling Technique)算法的样本生成过程。其数据处理方法简单易于实现，分类器也较为通用，获得了较好的生成效果。文献[11]中，作者在原始的 SMOTE 基础上利用 K-means 算法对其改进，选取簇心的 k 个近邻进行插值操作，有效地避免了模糊样本边界的问题。在保证多数类样本信息前提下，提升了少数类分类精度，增强了入侵检测系统检测能力。文献[12]中使用变分自编码器(Variational

Auto Encoder, VAE)来作为入侵检测的半监督学习，但其缺点在于不能自由生成指定标签的数据。文献[13]在物联网领域的入侵检测中使用了条件变分自编码器对各类样本进行生成，最后通过比较生成样本和测试样本的距离来预测标签。文献[14]中提出使用 GAN (Generative Adversarial Networks)来生成少数类样本，GAN 在图片生成领域取得了巨大成功，应用于网络数据也显示出了优秀的性能。

针对以上情况，本文提出了一个基于改进条件变分自编码器的入侵检测方法。我们使用改进的条件变分自编码器(Intrusion Detection Conditional Variational Auto Encoder, IDCVAE)的编码器部分来学习原始复杂数据的分布，然后将具有高斯噪声的隐变量和需要生成类标签作为解码器的输入，生成少数类的攻击样本，解决了原始数据集的不平衡现象。之后将新的数据集作为深度信念网络(Deep Belief Nets，DBN)的输入，逐层对 DBN 进行训练，利用微调对参数进行全局调优，最终得到分类结果。该方法在保证了总体准确率较高的情况下，有效提高了少数类的检测准确率并降低了误报率，增强了入侵检测系统的检测性能。

2. 相关理论介绍

2.1. 变分自编码器

变分自编码器(Variational Auto Encoder, VAE)作为生成模型的重要一员，由 Diederik P. Kingma 和 Max Welling [15]于 2013 年提出。其结构如图 1 所示。它主要由一个编码器和一个解码器组成。编码器实现了从 X 到一组低维向量的映射，这些向量完全定义了关联的中间概率分布 $Q(Z|X)$ 的集合。之后再对这些中间分布进行采样，生成的样本构成一组隐变量作为解码器的输入，解码器则是和编码器进行相反的操作，将隐变量映射为一组新的参数，形成新的概率分布 $p(\hat{X}/Z)$，从中我们可以再次进行采样作为解码器的输出 $\hat{X}$ 。

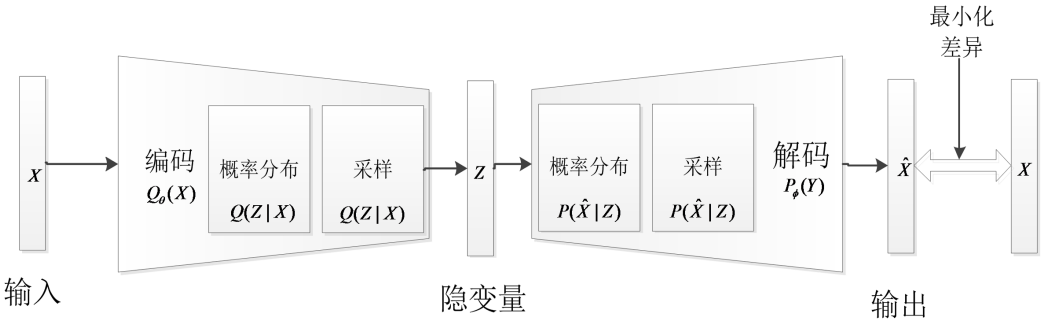

Figure 1. VAE architecture
图 1. VAE 结构图

VAE 的核心在于使用似然概率 $P(X)$ 在原始数据中进行采样，X 是一组随机变量。最后的目标在于尽可能的重建原始的输入数据。所以就必须最大化 $P(X)$ 的对数似然概率，公式如下：

$$\begin{aligned}\log P(X) &= \mathbb{E}\left[\log P(X|Z)\right] - D_{KL}\left[Q(Z|X)|P(Z)\right] + D_{KL}\left[Q(Z|X)\|P(Z|X)\right] \\ &\geq \mathbb{E}\left[\log P(X|Z)\right] - D_{KL}\left[Q(Z|X)\|P(Z)\right]\end{aligned} \tag{1}$$

$P(X)$ 的对数下界即变分下界目标如下所示：

$$\mathcal{L}(\theta,\phi;X) = \mathbb{E}\left[\log P(X|Z)\right] - D_{KL}\left[Q(Z|X)\|P(Z)\right] \tag{2}$$

$\mathcal{L}(\theta,\phi;X)$ 就是变分下界同时也是 VAE 的目标函数。公式的第二项就是利用 KL (Kullback-Leibler)散度来最小化编码器的分布 $Q(Z|X)$ 和隐变量 Z 的先验分布 $P(Z)$ 之间的分布距离，也就是说将学习到的

$Q(Z|X)$尽可能接近于先验分布$P(Z)$。因此，VAE 的训练目标就是最大化数据生成概率$\log P(X|Z)$的同时最小化编码器的分布$Q(Z|X)$和先验概率$P(Z)$之间的分布距离。

2.2. 入侵检测条件变分自编码器

基于 VAE 模型，条件变分自编码器(Conditional Variational Auto Encoder, CVAE)和它有着类似的思想。但不同于 VAE 只将随机变量 X 作为输入，同时也将标签 Y 也同时作为编码器和解码器的输入条件。故称之为条件变分自编码器。正如图 2 所示，编码器$Q(Z|X,Y)$和解码器$P(Z|X,Y)$输入条件有两个分别为 X 和 Y。

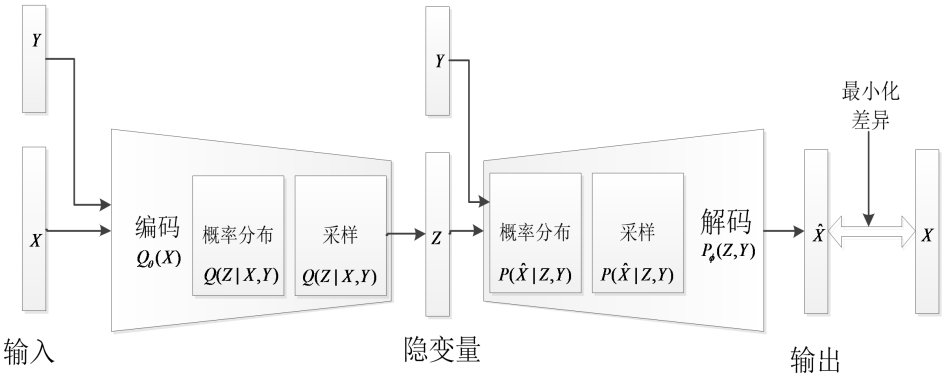

Figure 2. CVAE architecture
图 2. CVAE 结构图

CVAE 变分下界目标如下

$$\begin{aligned}&\log P(X|Y)-D_{KL}\left[Q(Z|X,Y)\|P(Z|X,Y)\right]\\&=\mathbb{E}\left[\log P(X|Z,Y)\right]-D_{KL}\left[Q(Z|X)\|P(Z|Y)\right]\end{aligned}\tag{3}$$

标准 CVAE 的编码器和解码器的条件概率分布都和标签 Y 相关，而在入侵检测任务中编码器部分无需要将特征与标签建立联系，故我们改进 CVAE 称之为入侵检测条件变分自编码器(Intrusion Detection Conditional Variational Auto Encoder, IDCVAE)。我们只在解码器阶段添加标签 Y 的信息，这样仅仅解码器的条件概率与 Y 相关，编码器部分则没有变化，更有利于编码阶段的特征提取。其结构如图 3 所示：

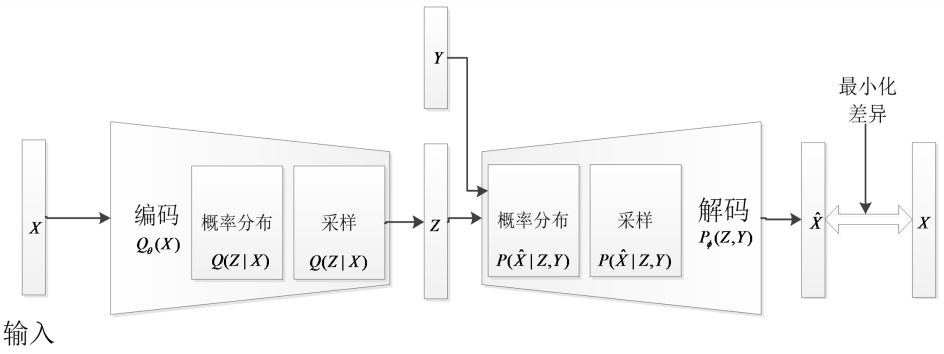

Figure 3. IDCVAE architecture
图 3. IDCVAE 结构图

IDCVAE 中编码器网络由$Q(Z|X)$和解码器网络$P(X|Z,Y)$组成，解码过程中隐变量 Z 和类标签 Y

会被合并在一起作为解码器的输入，所以指定标签的新少数类样本会被生成出来。IDCVAE 的变分下界如下：

$$\begin{aligned}&\log P(X\mid Y)-D_{KL}\left[Q(Z\mid X)\|P(Z\mid X,Y)\right]\\&=\mathbb{E}\left[\log P(X\mid Z,Y)\right]-D_{KL}\left[Q(Z\mid X)\|P(Z\mid Y)\right]\end{aligned}\tag{4}$$

其目标函数为：

$$\mathcal{L}(\theta,\phi;X,Y)=\mathbb{E}\left[\log P(X\mid Z,Y)\right]-D_{KL}\left[Q(Z\mid X)\|P(Z\mid Y)\right]\tag{5}$$

公式主要包含两个部分：一个对数重建似然 $\mathbb{E}\left[\log P(X\mid Z,Y)\right]$ 和一个 KL 散度 $D_{KL}\left[Q(Z\mid X)\|P(Z\mid Y)\right]$。第一项利用条件概率分布 $P(X\mid Z,Y)$ 重建原始输入 X，第二项利用 KL 散度来度量编码器分布 $Q(Z\mid X)$ 和先验概率分布 $P(Z\mid Y)$ 的距离。在模型中，我们使用 NSL-KDD 的类标签作为条件变量 Y，因此根据需要生成指定标签的攻击类样本。

2.3. 深度信念网络

传统神经网络往往不能有太多隐层，因为隐层的堆叠会导致模型参数数量迅速增长，同时也会导致训练时间过长。2006 年，Hinton 提出了深度信念网络 DBN，也被称为深度置信网络，一举解决了深层次神经网络的训练问题。其结构主要由多层的无监督的受限玻尔兹曼机(Restricted Boltzmann Machine)和一层有监督的反向传播(Back Propagation)网络组成。RBM 结构如图 4 所示，它是一个由可见层(Visible Layer)和隐层(Hidden Layer)组成的双层网络。可见层一般用作数据的输入，隐层作为特征提取层。层与层的节点全连接，层内节点无连接。其结构也可以看作是一个有向无环图。RBM 是一个基于能量模型的结构，其能量函数如公式 6 所示：

$$E(v,h\mid\theta)=-\sum_{i=1}^{m}a_iv_i-\sum_{j=1}^{n}b_jh_j-\sum_{i=1}^{m}\sum_{j=1}^{n}v_ih_jw_{ij}\tag{6}$$

上式中，a_i 和 b_j 分别表示可见层节点 v_i 和隐层节点 h_j 的偏置值，w_{ij} 表示可见层和隐层的关联权重。关联权重在最小化 RBM 的能量函数过程中，进行更新和优化，最后达到最优。RBM 的目标就是使 $E(v,h\mid\theta)$ 最小，此时网络最稳定，网络最优。

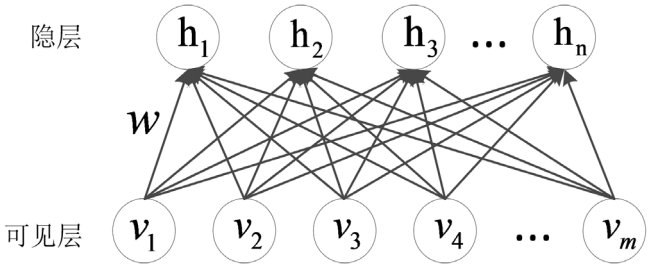

Figure 4. RBM architecture
图 4. RBM 结构图

DBN 由多个 RBM 堆叠而成，也就说想要获得全局的最优参数是非常困难的。因此，DBN 采用了逐层训练的方法。每次只训练相邻的两层，依次训练，直到所有 RBM 训练完成。再通过 BP 神经网络逐层进行反向传播，最终获得全局的最优参数。也就是说包含了一个正向训练和微调(fine-tuning)的过程。这个方法也是由 Hinton 提出，也称为对比散度(Contrastive Divergence, CD)训练方法。具体的过程如图 5 所示：

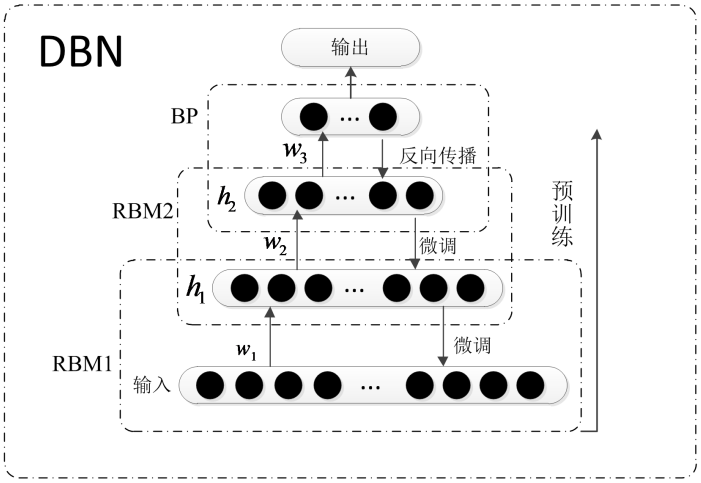

Figure 5. DBN architecture
图 5. DBN 结构图

3. 本文入侵检测算法

本文所提出的 IDCVAE-DBN 的结构图如图 6 所示。整个实验过程主要包含了三个阶段：1) 训练 IDCVAE。我们使用训练集中的少数类入侵样本作为编码器的输入，再解码器阶段将样本的类标签作为输入。训练的最终目的是保证重建后的样本与真实训练集中的样本差距尽可能的小。2) 生成新的攻击样本。我们将需要生成的样本类标签作为已经训练好的解码器的输入生成新的样本。将生成的所有样本与原始数据集合并，形成新的平衡数据集。3) 未知攻击检测。将新的数据集作为 DBN 的输入进行训练以及微调。训练好的 DBN 分类器用来对测试集中的攻击进行分类。算法具体步骤如下：

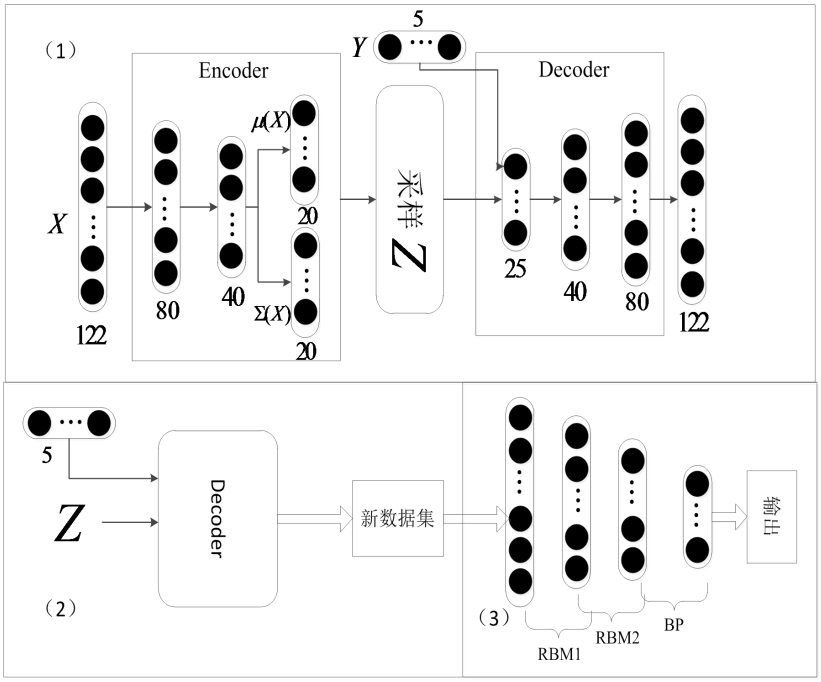

Figure 6. IDCVAE-DBN architecture
图 6. IDCVAE-DBN 模型图

a) 数据预处理。首先利用 one-hot 编码对非数值数据进行数值化，然后对数据进行标准化。数据预处理能够解决原始数据表达形式不一致，满足计算机处理要求。标准化的特征也有利于算法的学习。

b) 模型权重初始化为截断的正太分布中输出的随机值。偏移量均被初始化为 0。IDCVAE 输入数据维度为 122，隐变量 Z 的维度为 20，输出数据维度为 122。

c) 训练集 X 作为编码器的输入，对编码器进行训练。对训练得到的均值和方差进行采样，得到隐变量。不断迭代获取最优的模型参数。

d) 将训练得到的最优的隐变量 Z 以及需要生成的样本标签 $\bar{Y}$ 作为解码器的输入，得到新的样本，并将其并入原始数据集，得到新的平衡数据集 $\bar{X}$ 。

e) 新的训练集 $\bar{X}$ 作为 DBN 的输入，通过 CD 算法对 DBN 中包含的多层 RBM 进行逐层预训练，获得初步参数。之后再利用反向传播算法对参数进行微调，最终获得最优的 DBN 模型参数。

用已经训练好的 DBN 对测试集中的攻击进行分类，获得最终的分类结果。

3.1. IDCVAE 模型细节

本文使用了 NSL-KDD 数据集作为训练集。训练集中包含了 38 个数值特征和 3 个符号特征，总计 41 维特征。其中，数值特征可直接提取。符号特征需要用 one-hot 编码成数值特征。因为模型的输入必须为向量组，所以所有的符号特征需要经过 one-hot 编码，最后得到每个输入向量的维度为 122。之后对所有数据进行最值归一化，保证每个特征在同一个维度下。IDCVAE 的损失主要由一个重建损失和 KL 损失组成。KL 损失用来衡量隐变量的分布和单位高斯分布的差异。重建损失使用平均平方误差来度量数据的重构误差。

对于编码器网络 $Q(Z|X)$ 的分布我们使用的是多元高斯分布，而对于解码器 $P(X|Z,Y)$，我们使用多元伯努利分布进行拟合。隐变量 Z 的先验概率分布为标准正态分布 $N(0,I)$。其中多元高斯分布的均值为 $\mu(X)$，方差为 $\Sigma(X)\to\sigma_i^2(X)$，隐变量 Z 是对多元高斯分布进行采样而得。但由于从 $N(\mu,\sigma^2)$ 中直接采样是不可导的，所以利用了一个重参数的技巧。先在 $N(0,I)$ 中采样一个 ε，然后另 $Z=\mu+\varepsilon\times\sigma$，这样整个过程就变的可导了，从而使得整个模型可训练了。

3.2. DBN 模型细节

DBN 的输入为经过阶段二形成的平衡数据集。输出向量的维度为 5。DBN 由两个 RBM 以及一个 BP 层组成。隐层的激活函数为 Sigmoid，输出层的激活函数为 Softmax，DBN 和 RBM 的学习率均为 1×10^{-4}。利用 CD 算法对每个 RBM 分别进行无监督的训练。预训练过程中 RBM 的权值和偏置值都是通过随机初始化获得。优化算法使用的是 Adam。预训练完成后，利用 BP 层进行全局的参数调优。

4. 实验及结果分析

4.1. 实验环境及数据集介绍

实验使用的是 Windows10 操作系统、CPU 2.9 GHz、内存 16.0 GB 的 PC 机。编程语言使用 Python，版本为 3.7.4。基于 tensorflow、sklearn 等开源库实现。数据集选用 NSL-KDD 数据集。此数据集在 KDD-CUP1999 数据集的基础上去除了大量冗余和重复的数据，并且包含了一些 KDD-CUP1999 中所没有的新的攻击类型。NSL-KDD 数据集训练集和测试集样本数量适中，因此可以避免由于样本数过大、采样方法各不相同所造成的各类检测方法无法进行比较的问题。NSL-KDD 每条样本包含了 41 个特征属性标签，分别由 38 个数值型特征和 3 个非数值型特征组成。

数据集中主要包含 5 种类别的数据包括正常数据和另外四种攻击数据。四种攻击分别为：端口监视或扫描(Probe)、拒绝服务攻击(denial of service, DoS)、未授权的本地超级用户特权访问(user to root, U2R)、来自远程主的未授权访问(remote to local, R2L)。四类攻击类型又可详细划分为 39 类具体攻击。数据集详细情况如表 1 所示。

我们使用 KDDTrain+_20Percent.txt (包含了全部训练数据的 20%)作为训练集，KDDTest+.txt 作为测试集。训练集数据集本身极度不平衡，部分攻击类的样本远远少于正常数据。其中正常数据 13,449 条，Probe 类 2486 条，DoS 类 9044 条，U2R 类 24 条，R2L 类 189 条。本文攻击样本生成过程中，总计生成了 42,053 条数据。其中正常样本 0 条，Probe 类 10,963 条，DoS 类 4405 条，U2R 类 13,425，R2L 类 13,260 条。使得所有类别样本数目相同。数据集详细情况如表 1 所示：

Table 1. Sample category statistics
表 1. 样本类别统计

类别	攻击子类型	KDDTrain+_20Percent	KDDTest+
Normal	normal	13,449	9711
Probe	ipsweep	710	141
	nmap	301	73
	portsweep	587	157
	saint	0	319
	satan	691	735
Dos	apache2	0	737
	back	196	359
	land	1	7
	mailbomb	0	293
	neptune	8282	4657
	pod	38	41
	processtable	0	685
	smurf	529	665
	teardrop	188	12
	udpstorm	0	2
U2R	buffer_overflow	6	20
	httptunnel	0	133
	loadmodule	1	2
	perl	0	2
	ps	0	15
	rootkit	4	13
	sqlatack	0	2
	xterm	0	13

Continued

	ftp_write	1	3
	guess_passwd	10	1231
	imap	5	1
	multihop	2	18
	named	0	17
	phf	2	2
	sendmail	0	14
R2L	snmpgetattack	0	178
	snmpguess	0	331
	spy	1	0
	warezclient	181	0
	warezmaster	7	944
	worm	0	2
	xlock	0	9
	xsnoop	0	4
数据集样本总数		25,192	22,544

4.2. 评价标准

我们使用混淆矩阵常见的 5 种常见的度量指标，分别为：准确率(accuracy)、精准率(precision)、召回率(recall)、FPR (false positive rate)以及 F1-score。二分类混淆矩阵如表 2 所示，将样本分为 true positive (TP)，true negative (TN)，false positive (FP)，and false negative (FN)。TP 表示正类样本被正确分类、TN 表示负类样本被正确分类、FP 表示正类样本被错误的分为负类、FN 表示负类样本被错误的分为正类。

Table 2. Confusion matrix
表 2. 混淆矩阵

	预测正类	Recall
实际正类	TP	FN
实际负类	FP	TN

在这些度量之中，召回率是覆盖面的度量，度量有多个正例被分为正例。FPR 表示预测为正例但真实情况为反例的，占所有真实情况中反例的比率。F1-score 是精确率和召回率的调和平均数。精准率与召回率是衡量分类性能最基本的两个评价标准，F1-score 则兼顾了两者。对于非平衡数据集来说 F1-score 用来评估模型效果好坏是更加准确的。这些指标定义如下：

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{7}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{8}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{9}$$

$$\text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}} \tag{10}$$

$$\text{F1-score} = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}} \tag{11}$$

4.3. 实验结果与分析

本文实验主要分为三个部分。第一部分比较了进行过采样和不进行过采样的性能，第二部分比较不同过采样方法的性能，第三部分进行本文入侵检测方法与其他相关方法的对比实验。

由表 3 可知，本文所使用的方法相较于传统的机器学习方法，在各项指标上均获得了更好的结果。在同样使用 DBN 分类的情况下，在经过 IDCVAE 进行过采样后，准确率和提升了接近 10%，同时保证了 FPR 的下降。表 4 中，SVM 和 GBDT 虽然能在 Normal 等大类上取得较好的分类效果，然而其在 R2L 这种少数类上标签上的表现却差强人意，尤其是在 U2R 上两个方法的准确率都为 0。直接采用 DBN 进行分类后，对少数类检测准确率提升也非常有限。仅仅只有不到 3%。而应用了 IDCVAE 进行数据增强后，少数类的检测准确率相较于 DBN 得到了不小的提升。其中 R2L 提升了 24.17%，U2R 提升了 12.4%。由此可见使用 IDCVAE 进行过采样后，对总体以及少数类检测率的提升是非常有效的。

Table 3. Classification results of oversampling methods and normal methods (%)
表 3. 采用过采样和不采用过采样方法分类效果(%)

分类方法	Accuracy	Recall	Precision	F1-Score	FPR
SVM	70.33	54.75	88.52	67.66	7.46
GBDT	69.01	46.32	91.38	61.48	6.22
DBN	73.82	59.19	95.89	73.20	3.14
IDCVAE-DBN	82.36	75.34	97.21	84.89	2.81

Table 4. Subclass detection rate of oversampling methods and normal methods (%)
表 4. 采用过采样和不采用过采样方法子类检测率对比(%)

分类方法	Normal	Probe	DoS	R2L	U2R
SVM	92.34	75.45	43.77	6.87	0.00
GBDT	89.32	72.10	47.36	8.25	0.00
DBN	90.81	72.42	56.78	11.46	2.10
IDCVAE-DBN	96.82	86.36	72.53	35.63	14.50

如表 5 和表 6 所示，在同样使用 DBN 作为分类器时，通过三种过采样方法使得识别的整体准确率都得到了一定程度的提升。充分证明了过采样方法对少数类检测率提升的有效性。其中 IDCVAE 相较于其他两种机器学习过采样方法对于各项指标的提升是最为显著的。其中相比于 SMOTE，R2L 的检测准确率提高了 26.3%。相比于 ADASYN，U2R 的检测准确率提高了 6.43%。

我们也与其他表现较好的方法进行了比较。结果如表 7 所示。实验结果表明在准确率、召回率以及 FPR 上 IDCVAE-DBN 均优于其他方法。由此可见，IDCVAE-DBN 在不平衡数据的入侵检测问题中，整体效果符合预期，有着更好的检测性能。

Table 5. Classification results of different oversampling methods (%)
表 5. 不同过采样方法分类效果(%)

分类方法	Accuracy	Recall	Precision	F1-Score	FPR
SMOTE-DBN	80.15	65.13	94.50	77.11	3.55
ADASYN-DBN	80.33	64.27	95.12	76.71	3.79
IDCVAE-DBN	82.36	75.34	97.21	84.89	2.81

Table 6. Results compared with other methods (%)
表 6. 其他方法对比结果(%)

分类方法	Normal	Probe	DoS	R2L	U2R
SMOTE-DBN	92.63	66.99	58.15	9.33	8.00
ADASYN-DBN	91.42	59.21	56.23	8.77	8.07
IDCVAE-DBN	96.82	86.36	72.53	35.63	14.50

Table 7. Results compared with other methods (%)
表 7. 其他方法对比结果(%)

分类方法	Accuracy	Recall	FPR
1 改进 SMOTE + DBN + GBDT [16]	73.29	53.65	10.23
OCL [17]	481.25	/	/
SCDNN [18]	72.64	57.48	/
IDCVAE-DBN	82.36	75.34	2.81

5. 结束语

本文提出了一种基于改进条件变分自编码器的入侵检测方法，为入侵检测中少数类样本的检测问题提出了一个新思路。改进的条件变分自编码器可以学习样本的内在稀疏表示，其解码器可以有效地生成少数类样本，改善了数据集不平衡的情况下分类器的分类准确性。深度信念网络也有效地提取了数据的特征。实验结果表明，本文所提出的方法，对于不平衡数据集上的入侵检测问题有着良好的处理能力。在保证多数类样本准确率不下降的同时，大大提高了少数类的检测准确率。下一步，我们打算研究如何进一步提高样本的子类检测率以及大型数据集下的模型适应能力。

基金项目

上海市工业互联网资助项目(2018-GYHLW-02043)；国家自然科学基金资助项目(61771346, 61772372)，上海市信息化发展专项资金(新一代信息基础设施建设)项目(201901010)。

参考文献

[1] 李威, 杨忠明. 入侵检测系统的研究综述[J]. 吉林大学学报(信息科学版), 2016, 34(5): 657-662.

[2] 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.

[3] 于立婷, 谭小波, 解羽. 基于改进人工蜂群优化 K-means 的入侵检测模型[J]. 沈阳理工大学学报, 2019, 38(6): 8-14+27

[4] 柯钢. 改进粒子群算法优化支持向量机的入侵检测方法[J]. 合肥工业大学学报(自然科学版), 2019, 42(10):

1341-1345.

[5] 王洋, 吴建英, 黄金垒, 等. 基于贝叶斯攻击图的网络入侵意图识别方法[J]. 计算机工程与应用, 2019, 55(22): 73-79.

[6] Cabrera, J.B.D., Gutiérrez, C. and Mehra, R.K. (2008) Ensemble Methods for Anomaly Detection and Distributed Intrusion Detection in Mobile Ad-Hoc Networks. *Information Fusion*, **9**, 96-119. https://doi.org/10.1016/j.inffus.2007.03.001

[7] Jin, K., Nara, S., Jo, S.Y., *et al*. (2017) Method of Intrusion Detection Using Deep Neural Network. 2017 *IEEE International Conference on Big Data and Smart Computing* (*BigComp*), Jeju Island, 13-16 February 2017, 313-316. https://doi.org/10.1109/BIGCOMP.2017.7881684

[8] 刘月峰, 王成, 张亚斌, 等. 用于网络入侵检测的多尺度卷积 CNN 模型[J]. 计算机工程与应用, 2019, 55(3): 90-95+153.

[9] 刘月峰, 蔡爽, 杨涵晰, 等. 融合 CNN 与 BiLSTM 的网络入侵检测方法[J]. 计算机工程, 2019, 45(12): 127-133.

[10] Lou, X. (2013) Clustering Boundary Over-Sampling Classification Method for Imbalanced Data Sets. *Journal of Zhejiang University* (*Engineering Science*), **47**, 944-950.

[11] 沈学利, 覃淑娟. 基于 SMOTE 和深度信念网络的异常检测[J]. 计算机应用, 2018, 38(7): 1941-1945.

[12] 曹卫东, 许志香, 王静. 基于深度生成模型的半监督入侵检测算法[J]. 计算机科学, 2019, 46(3): 197-201.

[13] Lopez-Martin, M., Carro, B., Sanchez-Esguevillas, A., *et al*. (2017) Conditional Variational Autoencoder for Prediction and Feature Recovery Applied to Intrusion Detection in IoT. *Sensors*, **17**, 1967. https://doi.org/10.3390/s17091967

[14] Lee, J. and Park, K. (2021) GAN-Based Imbalanced Data Intrusion Detection System. *Personal and Ubiquitous Computing*, **25**, 121-128. https://doi.org/10.1007/s00779-019-01332-y

[15] Kingma, D.P. and Welling, M. (2014) Auto-Encoding Variational Bayes. http://arxiv.org/abs/1312.6114

[16] 陈虹, 肖越, 肖成龙, 等. 融合最大相异系数密度的 SMOTE 算法的入侵检测方法[J]. 信息网络安全, 2019(3): 61-71.

[17] Su, T. Sun, H. and Wang, S. (2019) Intrusion Detection Using Convolutional Recurrent Neural Network. In: *Proceedings of the* 2019 8*th International Conference on Computing and Pattern Recognition*, ACM, Beijing, 413-419. https://doi.org/10.1145/3373509.3373539

[18] Ma, T., Wang, F., Cheng, J., *et al*. (2016) A Hybrid Spectral Clustering and Deep Neural Network Ensemble Algorithm for Intrusion Detection in Sensor Networks. *Sensors*, **16**, 1701. https://doi.org/10.3390/s16101701