网络首发时间：2025-01-24 13:39:21　网络首发地址：https://link.cnki.net/urlid/42.1671.TP.20250123.1559.012

基于深度聚类与对比学习的网络入侵检测

郭盈盈，张冬梅，李成龙

（山东建筑大学 计算机科学与技术学院，山东 济南 250101）

摘　要：互联网时代下，网络恶意入侵事件迅速增加，网络入侵检测的重要性愈发凸显。为了提升入侵检测的实时性，解决有标签的攻击样本不足地问题，提出一种基于深度聚类与对比学习的网络入侵检测模型。首先，利用修改的VGGNet与LSTM网络分别提取局部和整体特征，通过实例和对比聚类优化模型，基于聚类方法检测异常数据。其次，以网络流量原始数据包为输入，通过优化数据增强方法增强网络流量数据，以实现对比学习。在CIC-IDS-2017数据集的实验和比较实验表明，所提模型性能优于其他相近方法，修改后的VGGN更适合处理网络流量数据，引入LSTM与对比学习均能提升模型性能。

关键词：对比学习；深度聚类；卷积神经网络；网络入侵检测；异常检测

DOI：10. 11907/rjdk. 241257　开放科学（资源服务）标识码（OSID）：

中图分类号：TP18　**文献标识码：**A　**文章编号：**1672-7800（XXXX）0XX-0001-08

Network Intrusion Detection Based on Deep Clustering and Contrastive Learning

GUO Yingying, ZHANG Dongmei, LI Chenglong

(*School of Computer Science and Technology*, *Shandong Jianzhu University*, *Ji'nan* 250101, *China*)

Abstract: In the Internet era, network malicious intrusion events are increasing rapidly, and the importance of network intrusion detection is becoming increasingly prominent. In order to improve the real-time performance of intrusion detection and solve the problem of insufficient labeled attack samples, a network intrusion detection model based on deep clustering and contrastive learning is proposed. Firstly, the modified VGGNet and LSTM networks are used to extract local and global features, respectively. The clustering optimization model is optimized through examples and comparisons, and abnormal data is detected based on clustering methods. Secondly, taking the raw data packets of network traffic as input, the network traffic data is enhanced through optimized data augmentation methods to achieve contrastive learning. Experiments and comparative studies on the CIC-IDS-2017 dataset have shown that the proposed model outperforms other similar methods, and the modified VGGN is more suitable for processing network traffic data. The introduction of LSTM and contrastive learning can both improve model performance.

Key Words: contrastive learning; deep clustering; convolutional neural network; network intrusion detection; anomaly detection

0　引言

在互联网迅猛发展的背景下，网络恶意入侵事件数量迅速增加，因此维护网络安全，有效辨别和应对网络入侵已成为一项不容忽视的工作[1]。理想的网络入侵检测系统可实时监测网络流量，具备发现网络流量异常的能力，以便及时识别潜在的安全威胁和攻击行为。这些系统不仅需要识别已知的攻击模式，还能检测未知的威胁和攻击技术。

早期，网络入侵检测系统研究主要基于传统的机器学习方法。Goryunov等[2]以随机森林为主要方法研究其对单类攻击的识别效果。Bienzobas等[3]通过随机森林处理日志数据，以达到入侵检测的目的。Louk等[4]结合bagging

收稿日期：2024-03-15
扫描二维码阅读全文：
基金项目：国家自然科学基金项目（62102235）

作者简介：郭盈盈（1998-），女，CCF学生会员，山东建筑大学计算机科学与技术学院硕士研究生，研究方向为机器学习与数据挖掘；张冬梅（1970-），女，博士，CCF高级会员，山东建筑大学计算机科学与技术学院副教授、硕士生导师，研究方向为算法与复杂性、机器学习与数据挖掘；李成龙（1988-），男，博士，山东建筑大学计算机科学与技术学院副教授、硕士生导师，研究方向为计算机视觉、增强现实。本文通讯作者：张冬梅。

和梯度提升决策树，提出双集成入侵检测模型。Yang 等[5]使用基于树的 4 种机器学习方法构成的机器学习器进行入侵检测。Verkerken 等[6]将自动编码器与单类支持向量机、随机森林和神经网络相结合，提出多层入侵检测系统。然而，上述基于传统机器学习的方法大都为有监督或半监督形式，即模型训练全部或部分依赖有标记样本，因此既需要大量的标记样本，还依赖于人工提取的离线数据特征，有时还需要手动调整参数，并且部分方法仅能检测针单一攻击。在面对大规模和高维的网络流量数据时，基于传统机器学习的方法既难以获取足够的标签数据，又无法有效提取数据特征，此时检测准确率将大幅下降。

目前，深度学习在大规模和高维数据上表现优异，在入侵检测中得到了广泛应用。Wang 等[7]将人工神经网络应用于入侵检测领域，虽然使用的是公司内部数据集，网络也较为简单，但研究结果展现了深度学习在入侵检测领域的潜力。Belarbi 等[8]利用 DBN（Deep Belief Network）方法，通过两个阶段的训练来提升模型对罕见攻击的检测效果，但仍使用的是人工设计提取的特征集。Thakkar 等[9]提出一种通过深度神经网络进行入侵检测的特征选择方法，能有效提升检测模型的性能及效率。Yao 等[10]、Sun 等[11]结合自定义卷积网络与循环神经网络进行入侵检测，通过比较标签来计算损失。其中，前者通过卷积与循环神经网络并行提取特征；后者将数据包卷积后，以流为单位将卷积结果批量输入循环神经网络。类似的，文献[8-10]都基于离线人工提取的网络流特征，文献[11]还需要获取完整的数据流后才能进行检测，并且这些研究都需要大量含有标签的数据，但大量有标签的攻击样本很难获取，因此难以在线检测未知的异常情况。

为了实现实时入侵检测，解决标记样本不足的问题，本文直接对网络流原始数据包进行处理，提出了基于深度聚类与对比学习的入侵检测模型。首先在网络入侵检测中引入对比学习方法，利用大规模未标记的数据进行训练；其次通过 VGGNet 与长短时记忆网络（Long Short-Term Memory Network，LSTM）网络处理高维的网络流数据，独立提取数据中的局部与全局特征；最后利用对比学习计算损失来优化网络，基于聚类检测异常数据。

1 相关工作

1.1 基于聚类的异常检测

异常检测方法包含基于统计、基于距离、基于密度、基于集成等。其中，基于聚类的异常检测是异常检测中常用的方法之一。聚类是无监督学习的一种形式，无需预先标记训练数据即可发现数据中的潜在结构和模式，适用于不同类型的数据，包括数值型、图像型、文本型等，具有较好的灵活性。

基于聚类的异常检测与聚类的目的不同。聚类的目标是将数据划分为若干簇，异常检测的目的是找出数据中的异常点。基于聚类的异常检测是在聚簇的基础上，比较每个数据点与其所属簇的相似度来检测异常，这种方法有利于发现数据中的未知异常。基于深度聚类的异常检测则使用神经网络来学习数据表示，具有更强大的表示学习能力，能适应复杂的数据结构，在处理大规模的高维数据时更高效。然而，深度学习的模型更加复杂，既需要更多计算资源和大量标记数据计算损失以训练网络，还需要根据数据和应用特点选择合适的网络结构。

1.2 对比学习

对比学习是自监督学习中的一种方法，利用数据本身的特征来学习，基本思想是通过比较数据中不同样本间的相似性和差异性来学习表示，在图像识别、自然语言处理、推荐系统等领域中应用广泛，特别是在无监督或半监督学习情境下更加普遍[13]。

无监督对比学习应用方便、可解释性强、无需标记数据，具体实现方法为对对比样本进行数据增强操作。数据增强是扩充数据集的常用方法，通过对原始样本进行一系列变换或扰动来生成新样本，使模型更好地学习样本特征。目前，与图像有关的数据增强方法最成熟，增强操作包括翻转、裁剪、缩放等，但大多数非图像数据并没有通用的数据增强方法，需要针对特定数据类型和应用领域设计特定方法。为此，针对不同类型的数据，各种不同的数据增强方法被提出，例如在文本类数据中替换同义词、随机插入词语，改变音频类数据播放速度、添加噪声等，或对随时间变化的单值数据波形图进行拉伸、延展、位移等操作。Li 等[14]利用无标签数据，通过对图像进行数据增强后进行对比学习，实现了图像数据识别工作。

本文模型使用对比学习的原因为无需标签数据，可有效利用大规模未标记的数据进行训练，以学习出更具泛化能力的表示。然而，在网络入侵检测中应用对比学习时，通常难以直接使用图像的数据增强方法，所以本文提出了一种新的针对网络原始数据包的数据增强方法，以更好的使用对比学习。

2 本文方法

2.1 总体结构

在特征提取时，通常需要使用合适的编码器网络将数据映射到潜在表示。理想情况下，潜在空间既能适合后续的对比学习，还要保留数据的重要特征。考虑到网络流量数据的特点，本文结合卷积与循环神经网络提取特征，但该方法仍需要大量有标记的样本训练网络，还需要考虑如何发挥各自优势。

为此，本文首先引入对比学习解决模型缺乏大量标记样本的问题，直接使用未经标记的网络流量原始数据作为输入；其次令卷积网络与循环神经网络各自提取特征；最

后结合卷积网络提取的局部特征和循环神经网络获取的全局特征，以在一定程度上避免单一类型特征的局限性，实现不同特征间的优势互补。综上，本文模型使用对比学习与深度聚类实现全程无监督，利用对比学习拉近同类数据的特性，令模型中基于深度聚类的异常检测结果更准确。在特征提取网络中分别构建 VGGNet 与 LSTM 网络，两者独立处理网络数据包，再融合输出特征，使提取的数据特征具有多重特性，特征空间更完善。

图 1 为本文模型主要结构，关键步骤包括特征提取和样本对比学习。首先，通过 VGG-LSTM 网络提取数据特征，在实例对比与聚类对比中通过对比学习计算损失、训练提取的特征，并根据聚类对比结果识别入侵流量。

同时，为了达到无监督的效果，对比学习需要对原始样本进行数据增强操作。首先，将原始数据与其对应的增强数据输入网络，通过网络提取样本的特征表示；其次将提取的表示进行实例级与聚类级对比学习，通过对比学习计算损失优化模型；最后利用训练好的模型直接检测网络入侵数据。

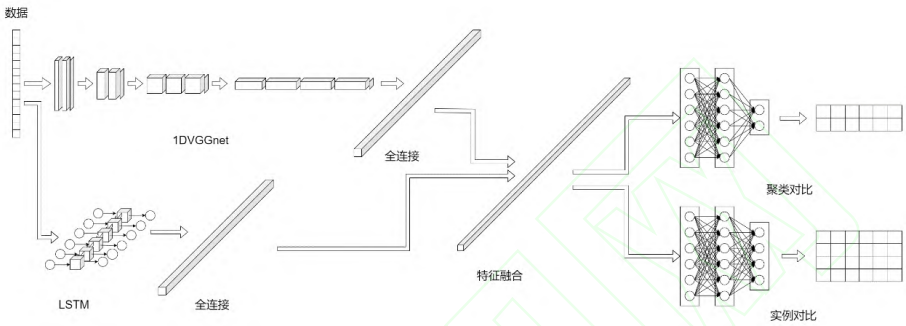

Fig. 1　Main structure of the model

图 1　模型主要结构

2.2　特征提取

2.2.1　VGGNet 提取局部特征

由于网络流量数据包的高维特点，导致传统机器学习方法提取的特征不理想，因此需要选用卷积神经网络 AlexNet、GoogleNet 和 VGGNet 等高维数据提取方法。其中，卷积神经网络包含多层卷积层，卷积层通过卷积核从数据中提取局部特征，每一层对应一个权重和一个偏差系数。在卷积层的卷积核 i 运行时，假设权重系数为 $\boldsymbol{w}_i$，偏差量为 $\boldsymbol{b}_i$，卷积层 i 的输入为 $\boldsymbol{X}_{i-1}$，卷积过程如式(1)所示。

$$X_i = f\left(w_i \otimes X_{i-1} + b_i\right) \tag{1}$$

然而，网络流量数据并没有图像数据那么高的维度，所以本文模型选择 VGGNet 提取网络数据特征。VGGNet 是由牛津大学视觉几何小组提出的深层卷积网络，通过多个 3×3 卷积核代替 AlexNet 中的 7×7 卷积核，由于多个小型的卷积核相较于一个大型的卷积核计算参数更少，从而解决了神经网络过深而出现的参数和梯度爆炸问题[15]。VGGNet 结构简单、可拓展性强，因其小卷积核的特点，可通过不断堆叠相似的网络结构加深网络，在图像识别等领域应用广泛。例如，Ding 等[16]提出新的 VGGNet 模型，通过结构重新参数化的方法将训练与推理解耦，以均衡模型精度和速度。

本文选用 VGGNet 的原因为结构简单、层数加深或减少时都能有效提取高维数据特征。由于除图像数据外的数据大都为一维数据，且非图像数据之间不存在图像数据间的位置关系，局部之间的联系不够紧密，因此本文针对非图像数据构建了适用于处理一维数据的 VGGNet，通过一维卷积计算代替二维卷积计算。同时，考虑到实时入侵检测需要很高的时效性，还适当缩减原始的 VGGNet 结构以提升运算效率，达到实时检测的目的。

传统的 VGGNet 中全部使用 3×3 大小的卷积核，包含 5 个卷积层、5 个最大值池化层和 3 个全连接层。其中，卷积层中前 3 层依次由 64 通道的两层卷积，128 通道的两层卷积，256 通道的 3 层卷积组成；第 4、5 层都是由 512 通道的 3 层卷积组成，中间以池化层分隔开来。出于对网络流量数据特性和模型深度对时效的考虑，本文模型的 VGGNet 使用一维卷积网络层与池化层代替二维卷积和池化层，将卷积核大小设置为 1×3，第 4 层修改为包含 4 层 1×3 卷积核、通道数为 512 的卷积层，并删去第 5 层卷积层。

此外，将 ReLU 作为卷积层的激活函数，如式(2)所示。该函数相较于 *sigmoid*、*tanh* 等饱和激活函数而言，求导更便捷，既能有效提升模型速度，又能解决梯度消失问题。修改后的 VGGNet 整体结构如图 2 所示。

$$\mathrm{ReLU}\left(X_i\right) = \begin{cases} X_i & \left(X_i > 0\right) \\ 0 & \left(X_i \leqslant 0\right) \end{cases} \tag{2}$$

2.2.2　LSTM 提取全局特征

为了在获取网络流量数据特征的同时，获得局部与全局特征，还使用 LSTM 提取数据特征[17]。LSTM 网络使用记忆单元代替循环神经网络中的隐藏单元，以解决传统网络的梯度爆炸或梯度消失问题。由于 LSTM 能有效捕捉长距离依赖性，处理变长序列输入，允许在一定程度上进行并行计算，因此在大规模数据集中的效率较高，从而在需要分析长距离依赖或学习时序特征的研究中得到了广泛应用。本文模型选用 LSTM 网络的是原因为：①可同时保存长期与短期信息，以有效表达、分析较长数据的整体信息；②可获取序列属性数据的时序特征和结构；③网络结

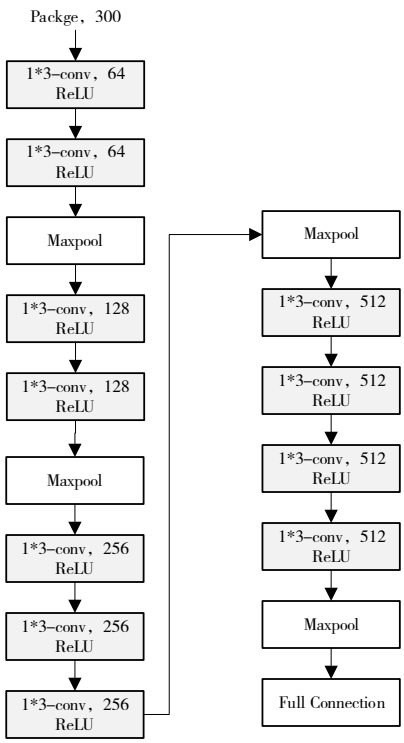

Fig. 2 Structure of modified VGGNet

图2 修改后的VGGNet结构

构相对简单,训练或在实际应用中消耗的时间较短。

LSTM网络由特定网络的循环调用组成,包含记忆单元和输入门、遗忘门和输出门。其中,输入门将新的信息选择性地记录到记忆单元中;遗忘门选择性遗忘上一时刻记忆单元中的信息;记忆单元负责存储目前为止网络学到的信息;输出门决定从记忆单元中提取多少信息用于输出,并生成新的输出。具体结构如图3所示。

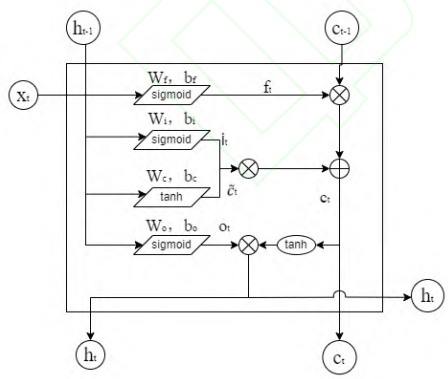

Fig. 3 Network structure of LSTM

图3 LSTM网络结构

由图3可见,一方面输入的当前时刻数据x_t与上一时刻数据h_{t-1}经过遗忘门的sigmoid函数计算,为上一步记忆单元c_{t-1}里的每个数据输出一个0~1间的值f_t,从而判断输入数据中应该丢弃哪些信息。计算公式如式(3)所示。

$$f_t = \sigma\left(W_f \cdot \left[h_{t-1}, x_t\right] + b_f\right) \tag{3}$$

另一方面,x_t与h_{t-1}同时经过输入门的sigmoid函数计算得到i_t,以决定本次要更新哪个值;同时,通过tanh函数创建一个更新的候选值向量$\tilde{c}_t$来提取和保存信息,该向量将会被增加到记忆单元中,计算公式如式(4)、式(5)所示。

$$i_t = \sigma\left(W_i \cdot \left[h_{t-1}, x_t\right] + b_i\right) \tag{4}$$

$$\tilde{c}_t = tanh\left(W_c \cdot \left[h_{t-1}, x_t\right] + b_c\right) \tag{5}$$

其次,根据遗忘门与输入门的计算结果,对记忆单元进行更新并保存操作,具体计算公式如式(6)所示。

$$c_t = f_t * c_{t-1} + i_t * \tilde{c}_t \tag{6}$$

接下来,将记忆单元中的数据c_t输入输出门,以决定从记忆单元中提取多少信息用于输出。最后,通过函数sigmoid的计算结果o_t,决定输出该次记忆单元的哪些部分,并将记忆单元通过函数tanh得到-1~1之间的值,再与o_t相乘得到最终的隐藏层输出h_t,计算公式如式(7)、式(8)所示。循环上述过程,直至数据包中所有数据处理完毕。

$$o_t = \sigma\left(W_o \cdot \left[h_{t-1}, x_t\right] + b_o\right) \tag{7}$$

$$h_t = o_t * \tanh\left(c_t\right) \tag{8}$$

2.3 实例对比与聚类对比

2.3.1 实例对比

实例对比学习可通过拉近正样本,并将他们远离负样本来完成操作。由于训练是无监督学习,没有先验标签作为标准,因此将每个样本视为一个单独的类,同一样本的数据增强认为是正样本,而一同输入的其余样本均视为负样本。针对网络流量问题,Yao等[10]将人工特征集数据转换为图像,通过卷积网络进行特征提取。本文考虑到图像数据中值与值的位置关系,对网络数据使用图像增强的方法可能可解释性不够强。目前,有研究采用生成对抗方法,使用神经网络学习特征来生成同类数据,例如王正文等[18]应用生成对抗方法预测类别,将网络数据转化为灰度图像进行处理。Hore等[19]利用生成对抗网络对攻击样本的原始数据进行扰动,以生成恶意攻击的网络数据包,但如果在模型中加入生成对抗网络会增加模型复杂度。

为此,本文考虑到同一数据流的目的相同,其数据包必然属于同一类别,因此在训练时将同属于一个数据流的数据包视为该数据的数据增强,作为正样本进行对比学习。对于数据传输过程中存在一个数据流只成功发送一个数据包的特殊情况,通过以下方法进行数据增强:保持数据包头部数据不变,随机调换其余数据的位置,以不影响所属类别的方式生成增强数据。具体而言,对*batchsize*为N的一批样本集合$\boldsymbol{X}=\{x_1, x_2, \cdots\cdots, x_N\}$,当样本$x$输入模型时,首先与数据增强的结果$x^a$组成样本对$(x, x^a)$,其次将其视为该批次唯一的正样本对,最后将其余一同输入的样本及其增强样本视为样本x的负样本。

为了获得更好的效果,模型没有直接对输入数据提取的特征表示$\boldsymbol{h}$应用对比学习,而是定义了多层感知器f_l。f_l定义为两层全连接网络,使用ReLu函数进行激活。首先,将特征通过感知器映射到另一个子空间,以获得$z_i=f_l(h_i)(i \in [1, N])$;其次最小化损失函数,即该批次中所有正样本

对损失的平均值，如式(9)所示。

$$Loss_I = \frac{\sum_{i=1}^{N}\left(loss_i + loss_i^a\right)}{2N} \tag{9}$$

$$loss_i = -\log \frac{\exp\left(\frac{\mathrm{sim}\left(z_i, z_i^a\right)}{\tau_I}\right)}{\sum_{k=1}^{N}\exp\left(\frac{\mathrm{sim}\left(z_i, z_k\right)}{\tau_I}\right) + \sum_{k=1}^{N}\exp\left(\frac{\mathrm{sim}\left(z_i, z_k^a\right)}{\tau_I}\right)} \tag{10}$$

$$loss_i^a = -\log \frac{\exp\left(\frac{\mathrm{sim}\left(z_i^a, z_i\right)}{\tau_I}\right)}{\sum_{k=1}^{N}\exp\left(\frac{\mathrm{sim}\left(z_i^a, z_k^a\right)}{\tau_I}\right) + \sum_{k=1}^{N}\exp\left(\frac{\mathrm{sim}\left(z_i^a, z_k\right)}{\tau_I}\right)} \tag{11}$$

$$\mathrm{sim}\left(z_i, z_j\right) = \frac{\left(z_i\right)^T\left(z_j\right)}{\left\|z_i\right\|\left\|z_j\right\|} \tag{12}$$

式中：$loss_i$ 为 x_i 的损失函数；$loss_i^a$ 为 x_i^a 的损失函数；τ_I 是定义样本之间吸引和排斥程度的温度参数；sim 函数用于计算余弦相似度。

2.3.2　聚类对比

样本在进行实例对比学习的同时，也会进入聚类对比学习网络进行聚类对比学习。如果将聚类对比的输出表示为矩阵 $\boldsymbol{Y}$，将聚类网络的输出宽度设置为簇类数量，则在 $\boldsymbol{Y}$ 中 y_{ij} 对应的是样本 i 属于第 j 个簇的概率。

与实例对比类似，聚类对比集中将正样本远离负样本，试图将正样本聚类到同一个簇，避免负样本与正样本聚在同一个簇中。具体为，将输入数据 x 及其增强 x^a 视为同一批数据中的正样本，其余输入数据及其增强共 $2N-2$ 个数据被视为负样本。同理，聚类对比使用多层感知器，通过 ReLU 函数激活的两层全连接网络 f_C，在输出前使用 softmax 函数计算样本属于某个簇的概率。

本文将感知器的输出宽度设置为簇类数量 M，首先通过 f_C 获得大小为 $N\times M$ 的原样本的输出 $\boldsymbol{Y}$，以及增强样本的输出 $\boldsymbol{Y^a}$；其次最小化聚类对比损失函数，如式(13)所示。$Loss'_C$ 为该批次所有簇类的平均损失，如式(14)所示。

$$Loss_C = Loss'_C + H\left(Y, Y^a\right) \tag{13}$$

$$Loss'_C = \frac{\sum_{j=1}^{M}\left(\widehat{loss}_j + \widehat{loss}_j^{\,a}\right)}{2M} \tag{14}$$

$$\widehat{loss}_j = -\log \frac{\exp\left(\frac{\mathrm{sim}\left(y_j, y_j^a\right)}{\tau_C}\right)}{\sum_{k=1}^{M}\exp\left(\frac{\mathrm{sim}\left(y_j, y_k\right)}{\tau_C}\right) + \sum_{k=1}^{M}\exp\left(\frac{\mathrm{sim}\left(y_j, y_k^a\right)}{\tau_C}\right)} \tag{15}$$

$$\widehat{loss}_j^{\,a} = -\log \frac{\exp\left(\frac{\mathrm{sim}\left(y_j^a, y_j\right)}{\tau_C}\right)}{\sum_{k=1}^{M}\exp\left(\frac{\mathrm{sim}\left(y_j^a, y_k^a\right)}{\tau_C}\right) + \sum_{k=1}^{M}\exp\left(\frac{\mathrm{sim}\left(y_j^a, y_k\right)}{\tau_C}\right)} \tag{16}$$

$$H\left(Y, Y^a\right) = -\sum_{j=1}^{M}\left[P\left(y_j\right)\log P\left(y_j\right) + P\left(y_j^a\right)\log P\left(y_j^a\right)\right] \tag{17}$$

$$P\left(y_j\right) = \frac{\sum_{i=1}^{N} Y_{ij}}{N} \tag{18}$$

$$P\left(y_j^a\right) = \frac{\sum_{i=1}^{N} Y_{ij}^a}{N} \tag{19}$$

式中：$\widehat{loss}_j$ 为 y_j 的损失函数；$\widehat{loss}_j^{\,a}$ 为 y_j^a 的损失函数；y_j 为 $\boldsymbol{Y}$ 的第 j 列；$H\left(Y, Y^a\right)$ 为聚类分配概率的熵，引入该项是为了避免聚类将大多数数据点映射到同一簇中。

2.4　模型优化

模型总损失由实例对比损失与聚类对比损失组成，总损失如式(20)所示。

$$Loss = Loss_I + Loss_C \tag{20}$$

为进一步提升模型的聚类效果，训练后在预训练模型的实例对比阶段加入超参数 γ 作为阈值，即当非正样本对的其他样本对的相似度时大于或等于 γ，则认为这些样本相似。由此将式(10)、式(11)改为式(21)、式(22)。

$$loss'^{\,a}_i = -\log \frac{\sum_{j=1}^{N} 1_{\left\{sim\left(z_i^a, z_j\right)\geqslant\gamma\right\}}\exp\left(sim\left(z_i^a, z_j\right)\right) + \sum_{j=1}^{N} 1_{\left\{sim\left(z_i^a, z_j^a\right)\geqslant\gamma\right\}}\exp\left(sim\left(z_i^a, z_j^a\right)\right)}{\sum_{k=1}^{N}\exp\left(\frac{sim\left(z_i^a, z_k^a\right)}{\tau_I}\right) + \sum_{k=1}^{N}\exp\left(\frac{sim\left(z_i^a, z_k\right)}{\tau_I}\right)} \tag{21}$$

$$loss'_i = -\log \frac{\sum_{j=1}^{N} 1_{\left\{sim\left(z_i, z_j\right)\geqslant\gamma\right\}}\exp\left(\mathrm{sim}\left(z_i, z_j\right)\right) + \sum_{j=1}^{N} 1_{\left\{sim\left(z_i, z_j^a\right)\geqslant\gamma\right\}}\exp\left(sim\left(z_i, z_j^a\right)\right)}{\sum_{k=1}^{N}\exp\left(\frac{\mathrm{sim}\left(z_i, z_k\right)}{\tau_I}\right) + \sum_{k=1}^{N}\exp\left(\frac{\mathrm{sim}\left(z_i, z_k^a\right)}{\tau_I}\right)} \tag{22}$$

3　实验结果与分析

3.1　实验环境与数据集

本文代码使用Python语言编程，基于开源机器学习库Pytorch进行实验，具体实验环境如表1所示。目前，常用的网络入侵公共数据集包括KDD99、NSL-KDD[20]、UNSW-NB15[21]、CIC-IDS-2017和CSE-CIC-IDS-2018[22]等。其中，KDD99、NSL-KDD、UNSW-NB15数据集只有人工提取的数据特征集，包括与主机相关的统计特征、流的持续

时间和总字节数、流的平均用时等特征。CIC-IDS-2017[12]和CSE-CIC-IDS-2018数据集除了包含人工提取的特征集之外，还包含更多更新的攻击测试数据，例如流的平均时间和长度、反馈时间、流中各种类型数据包的数量等特征，并提供网络流量原始数据包的数据，以十六进制表示。

Table 1 Experimental environment and configuration

表1 实验环境与配置

OS	Windows 10 家庭版
GPU	8G/16G NVIDIA GeForce RTX 2070
PyCharm	212.5086.64
Pytorch	2.0.0
Python	3.6.5

由于在人工提取的数据特征集中，类似流持续时间、正反向数据包数量及大小、流平均长度及流活动时间等特征，需要分析完整的数据流才能得到，因此会大幅影响在线检测的时效。为此，本文直接将传输过程中的原始数据包作为实验数据。同时，为了更好地在实验中复现真实网络场景，本文选择CIC-IDS-2017与CSE-CIC-IDS-2018数据集中的原始数据包作为研究数据，但受限于实验设备和算力，只能在数据量相对更小的CIC-IDS-2017的原始数据包上进行实验。

CIC-IDS-2017数据集提供的是存放原始十六进制表示数据的PCAP文件，攻击流量包括FTP-Patator、SSH-Patator、DoS(Denial of Service)、Heartbleed、Web Attack、渗透、Botnet和DDoS(Distributed Denial of Service)8种类型，包含2017年中5天内模拟的网络流量，周二、周三、周四、周五的上午和下午存在不同的攻击行为，周一全天、周二至周五的非高峰时段均产生正常非攻击流量。

本文呢受限于算力因素，选择数据集中周二至周四的网络流量作为实验数据，包含5种类型流量，分别为正常和FTP-Patator、SSH-Patator、DoS及渗透攻击等4种攻击类型，即可将所有数据分为正常和异常(攻击)两类，其中80%作为训练数据，20%作为测试使用。

3.2 数据预处理

CIC-IDS-2017数据集提供的原始数据包数据为捕获网络流量时常用的数据报存储格式——PCAP文件。该类型文件需要按特定格式解读，并且一个PCAP文件包含一天所有流量，不同数据流的数据包交错混杂，无法直接输入模型进行训练和检测。因此，本文在数据预处理阶段，将PCAP文件中的数据，以实时检测时的数据包为单位进行区分，并标记其所属数据流类型。具体为，首先从PCAP文件中按数据包提取原始十六进制数据；其次提取每个数据包的五元组数据(源IP，目的IP，源端口，目的端口和协议号)，将其作为数据包id，并与十六进制表示的数据包数据一起输入网络；最后以数据包为单位进行实验，即一个数据包就是一个实例x。为了便于实验，本文选择每个数据包中前300个字节作为测试数据。

由于网络流量的特殊性，直接应用图像增强方法显然不够，数据增强是对比学习中实现无监督的重要手段，目的是在原数据的基础上生成与原数据相同类型的测试数据。由于相同五元组的数据包可视为同一个流的数据包，为此寻找到与数据x相同五元组的其他数据包x^a，将其视为x的增强数据。对于同一个数据流而言，属于该流的数据包必定属于同一种类型，无论是攻击流量还是正常流量。

3.3 实验步骤

本文实验可分为数据预处理、预训练模型、网络入侵检测与比较3个阶段，通过训练40个epoch即可得到训练好的模型，如图4所示。在预训练过程中，首先通过VGG-LSTM网络提取样本及其增强组成的数据对(x,x^a)的特征，其次利用最大池化操作筛选组合局部与全局特征，最后在实例对比和聚类对比中通过对比学习最小化损失函数优化特征提取、实例对比和聚类对比网络。

由图4可见，该模型如果用于实际网络入侵检测，应该以实时捕获的数据包为输入。由于缺乏实验条件，本文在实验中将数据集提供的网络流量数据作为输入，经过特征提取和聚类对比，根据聚类结果检测该数据是否为异常数据。

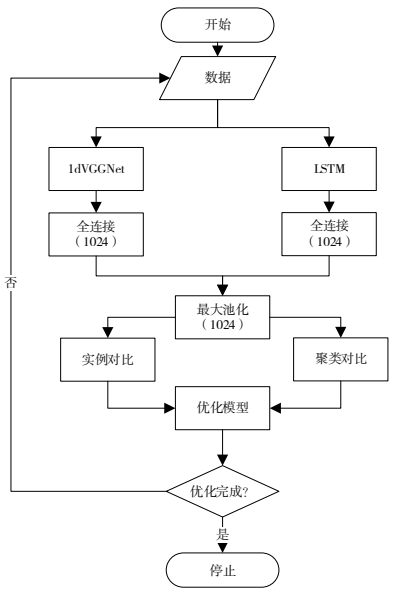

Fig. 4 Pre-training process

图4 预训练流程

图5为流量数据的检测过程。由此可见，本文在检测过程中使用了基于深度聚类的异常检测方法，能在一定程度上检测出某些未知类型的攻击流量。

3.4 实验结果

在入侵检测过程中，通常使用准确率(Accuracy)、精确率(Precision)、召回率(Recall)、F1-分数(F1-score)等指标评价模型性能[23]。其中，准确率是指正确分类的样本数与样本总数的比值；精确率为正确识别的异常样本数与真实异常样本数之比，表示攻击检测的查准率；召回率定义为

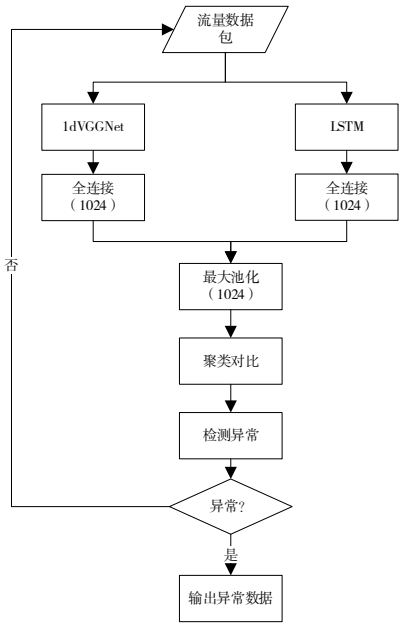

Fig. 5　Anomaly detection process

图5　异常检测流程

正确识别的异常样本数与预测的异常样本数之比；Recall反映模型识别攻击的能力，是入侵检测系统中的重要指标；F1-分数为精确率与召回率的调和平均数。具体计算如式(23)—式(26)所示。实验结果如图6所示。

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \tag{23}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{24}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{25}$$

$$\text{F1-score} = \frac{2*\text{Precision}*\text{Recall}}{\text{Precision} + \text{Recall}} \tag{26}$$

式中：TP为正确分类为该类型的样本数；TN为正确分类为非该类型的样本数；FP为错误分类为该类型的样本数；FN为错误分类为非该类型的样本数。

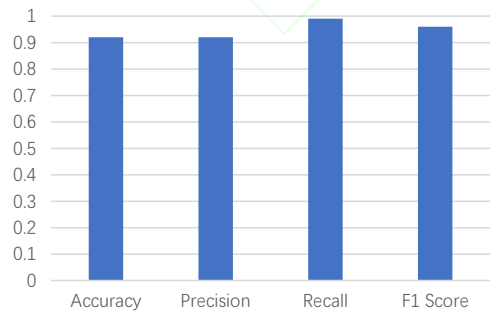

Fig. 6　Model experiment results

图6　模型实验结果

本文模型借鉴了图像分类方法，首先使用卷积网络VGGNet提取特征，根据网络流数据的特性修改VGGNet；其次在模型中加入LSTM网络来增加特征空间的多样性，与卷积网络互相独立获取全局特征；最后引入对比学习优化模型实现无监督训练。考虑到[10,11]等文献同样使用了卷积神经网络与LSTM网络进行入侵检测的研究，故本文在验证模型性能的同时进行以下比较实验。

(1)为了验证LSTM对数据检测性能的影响，将通过对比学习实现的VGGNet入侵检测模型与本文模型进行比较，实验结果如表2所示。对照模型A代表只使用本文模型相同的VGGNet提取数据特征，并使用对比学习计算损失来优化模型。由此可知，本文模型相较于对照模型A入侵检测性能更强。

Table 2　Impact of LSTM on model performance

表2　LSTM对模型性能的影响

实验模型	Accuracy	Precision	Recall	F1 Score
对照模型A	0.80	0.80	0.99	0.88
本文模型	**0.92**	**0.92**	**0.99**	**0.96**

(2)为了验证加入对比学习对网络流量数据检测性能的影响，将特征提取网络相近但使用交叉熵计算损失的模型与本文模型进行比较，实验结果如表3所示。由于缺乏文献[10]的卷积网络细节，在对照者B、C中均使用二维VGGNet，将数据处理为20×20的灰度图像实现卷积操作，并与LSTM提取的特征相互结合。其中，对照模型B为使用文献[10]中有监督的交叉熵函数计算损失；对照模型C使用本文的无监督对比学习计算损失。

Table 3　Results of contrastive learning loss vs. cross entropy loss (2D convolution)

表3　对比学习与交叉熵计算损失对比的结果(二维卷积)

实验模型	Accuracy	Precision	Recall	F1 Score
对照模型B	0.81	0.99	0.81	0.89
对照模型C	**0.88**	**0.99**	**0.87**	**0.92**

为了验证修改后一维VGGNet对模型的影响，本文参考文献[10]的思路设计对照模型D，即使用本文修改后的一维VGGNet与LSTM分别提取特征，将交叉熵函数作为损失函数优化模型优化。实验结果如表4所示。

Table 4　Results of contrastive learning loss vs. cross entropy loss (1D convolution)

表4　对比损失与交叉熵损失函数对比的结果(一维卷积)

实验模型	Accuracy	Precision	Recall	F1 Score
对照模型D	0.82	0.82	0.99	0.90
本文模型	**0.92**	**0.92**	**0.99**	**0.96**

由表3、表4可知，无论是采用二维卷积还是一维卷积提取特征，加入对比学习的模型性能均优于对照模型。同时，使用一维卷积模型的3个指标都优于二维卷积，仅在Precision方面存在不足，原因可能为二维卷积能在更大的区域提取局部特征。然而，二维卷积的可解释性不够强、模型参数更为复杂，从综合指标F1-Score而言一维卷积模型性能更优，为此本文模型选择一维卷积方法。

(3)为了验证使用卷积神经网络与LSTM提取特征时，不同组合方式对模型性能的影响，将两者顺序连接的模型与本文模型进行比较，实验结果如表5所示。文献[11]的方法将CNN输出的结果组合后输入LSTM中，以顺序连接方式相结合，并将数据以包、流为单位分别进行处理，以获

得更全面的特征。同时,该方法使用的是自定义卷积网络,利用交叉熵函数计算损失。由于缺乏相应的模型细节,本文仅参考文献[11]的思路设计对照模型E,通过本文修改后的VGGNet与LSTM进行顺序连接,并使用交叉熵函数计算网络损失。

Table 5 Comparison of VGGNet and LSTM combination approaches

表5 VGGNet和LSTM组合方式对比

实验模型	Accuracy	Precision	Recall	F1 Score
对照模型E	0.80	0.81	0.99	0.89
本文模型	**0.92**	**0.92**	**0.99**	**0.96**

由表5可知,本文模型在3个指标上有所提升,说明通过VGGNet与LSTM独立提取特征后,进行组合筛选的特征更有效。同时,分析对照模型D与对照模型E的结构可知,通过VGGNet与LSTM网络独立提取特征的结构相较于顺序连接性能更优。

4 结语

本文提出一种基于深度聚类与对比学习的入侵检测模型,并在网络入侵检测数据集CIC-IDS-2017上进行实验验证。首先通过创新的数据增强方法对网络流量进行数据增强;其次通过修改的VGGNet与LSTM网络提取数据的局部、全局特征;最后将二者筛选后结合,通过实例对比和聚类对比计算损失函数来优化模型。总而言之,模型的整个训练过程实现了无监督学习,无需使用有标签的数据,同时基于深度聚类的异常检测方法能使模型在一定程度上检测出未知的攻击。

实验表明,修改后的VGGNet更适合处理网络流量数据,并且加入对比学习与多重特征后能有效提升模型性能。然而,实验中也发现网络流量数据集中存在的数据类别不平衡问题,在一定程度上影响了模型性能,未来可考虑在损失函数中添加类别权重来解决该问题,还可进一步从理论方面完善模型结构,研究更有普适性的模型。

参考文献:

[1] THAKKAR A, LOHIYA R. A survey on intrusion detection system: feature selection, model, performance measures, application perspective, challenges, and future research directions[J]. Artificial Intelligence Review, 2022, 55(1): 453-563.

[2] GORYUNOV M N, MATSKEVICH A G, RYBOLOVLEV D A. Synthesis of a machine learning model for detecting computer attacks based on the cicids2017 dataset[J]. Proceedings of the Institute for System Programming of the RAS, 2020, 32(5): 81-94.

[3] BIENZOBAS Á C, SÁNCHEZ-MACIÁN A. Threat trekker: an approach to cyber threat hunting[DB/OL]. https://arxiv. org/abs/2310. 04197.

[4] LOUK M H L, TAMA B A. Dual-IDS: a bagging-based gradient boosting decision tree model for network anomaly intrusion detection system[J]. Expert Systems with Applications, 2023, 213: 119030.

[5] YANG L, MOUBAYED A, SHAMI A. MTH-IDS: a multitiered hybrid intrusion detection system for internet of vehicles[J]. IEEE Internet of Things Journal, 2021, 9(1): 616-632.

[6] VERKERKEN M, D'HOOGE L, SUDYANA D, et al. A novel multi-stage approach for hierarchical intrusion detection[J]. IEEE Transactions on Network and Service Management, 2023,20(3):3915-3929.

[7] WANG Z. The applications of deep learning on traffic identification[J]. BlackHat USA, 2015, 24(11): 1-10.

[8] BELARBI O, KHAN A, CARNELLI P, et al. An intrusion detection system based on deep belief networks[C]// International Conference on Science of Cyber Security, 2022: 377-392.

[9] THAKKAR A, LOHIYA R. Fusion of statistical importance for feature selection in deep neural network-based intrusion detection system[J]. Information Fusion, 2023, 90: 353-363.

[10] YAO R, WANG N, LIU Z, et al. Intrusion detection system in the advanced metering infrastructure: a cross-layer feature-fusion CNN-LSTM-based approach[J]. Sensors, 2021, 21(2): 626.

[11] SUN P, LIU P, LI Q, et al. DL-IDS: extracting features using CNN-LSTM hybrid network for intrusion detection system[J]. Security and Communication Networks, 2020, 2020: 1-11.

[12] SHARAFALDIN I, LASHKARI A H, GHORBANI A A. Toward generating a new intrusion detection dataset and intrusion traffic characterization[C]// International Conference on Information Systems Security and Privacy, 2018: 108-116.

[13] HADSELL R, CHOPRA S, LECUN Y. Dimensionality reduction by learning an invariant mapping[C]// IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2006: 1735-1742.

[14] LI Y, HU P, LIU Z, et al. Contrastive clustering[C]// Proceedings of the AAAI Conference on Artificial Intelligence, 2021: 8547-8555.

[15] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [DB/OL]. https://arxiv. org/abs/1409. 1556.

[16] DING X, ZHANG X, MA N, et al. RepVGG: making VGG-style convnets great again[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2021: 13733-13742.

[17] JAVAID A, NIYAZ Q, SUN W, et al. A deep learning approach for network intrusion detection system[C]// Proceedings of the 9th EAI International Conference on Bio-inspired Information and Communications Technologies, 2016: 21-26.

[18] WANG Z W, WANG J F. Unknown malicious traffic classification method based on generative zero-shot learning[J]. Journal of Sichuan University(Natural Science Edition), 2023, 60(4): 67-74.
王正文, 王俊峰. 基于生成式零样本学习的未知恶意流量分类方法[J]. 四川大学学报(自然科学版),2023, 60(4): 67-74.

[19] HORE S, GHADERMAZI J, PAUDEL D, et al. Deep packgen: a deep reinforcement learning framework for adversarial network packet generation[DB/OL]. https://arxiv. org/abs/2305. 11039.

[20] RAMOS V, ABRAHAM A. ANTIDS: self organized ant-based clustering model for intrusion detection system[C]// Proceedings of the Fourth IEEE International Workshop, 2005: 977-986.

[21] 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.

[22] SHARAFALDIN I, LASHKARI A H, GHORBANI A A. Toward generating a new intrusion detection dataset and intrusion traffic characterization[C]// International Conference on Information Systems Security and Privacy, 2018: 108-116.

[23] XU C, SHEN J, DU X, et al. An intrusion detection system using a deep neural network with gated recurrent units[J]. IEEE Access, 2018, 6: 48697-48707.

(责任编辑:刘嘉文)