DOI:10.11918/202311091

堆叠循环卷积和头部注意力的工业互联网入侵检测

金志刚,刘 凯,武晓栋

(天津大学 电气自动化与信息工程学院,天津 300072)

摘 要:针对工业互联网中数据海量、有效信息与冗余信息混杂,且传统入侵检测方法忽略网络流量间复杂关联性的问题,为提升工业互联网场景下入侵检测系统的特征提取能力与检测准确性,提出一种堆叠循环卷积自编码器和头部注意力机制的工业互联网入侵检测方法。首先,利用卷积自编码器提取流量数据的空域特征,降低流量数据的冗余性;同时,引入二维卷积层和循环填充方式,通过对原始数据进行升维和空域拓展,扩大卷积核对数据的感受野,进而提高特征提取能力;其次,利用头部注意力机制提取流量间的相关性信息,并根据不同注意力头提取到的相关性信息的重要程度分配注意力头权重,在判断网络流量类型时考虑最重要的相关性信息;最后,在 UNSW-NB15 数据集上,与传统入侵检测方法及基于传统时序神经网络的入侵检测方法等进行二分类、多分类对比实验。实验结果表明,所提方法在二分类任务中准确率达 99.18%,在多分类任务中准确率达 87.36%,优于传统入侵检测方法。所提入侵检测方法在工业互联网场景中能有效对网络入侵进行检测。

关键词:入侵检测;深度学习;卷积自编码器;循环填充;头部注意力机制

中图分类号:TP393.08 **文献标志码**:A **文章编号**:0367-6234(2026)09-0104-11

Intrusion detection for industrial internet stacking circular convolution and head attention

Jin Zhigang, Liu Kai, Wu Xiaodong

(School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China)

Abstract: To address the problems of massive data, the mixture of effective and redundant information in the industrial internet, and the neglect of complex correlations among network traffic by traditional intrusion detection methods and to improve the feature extraction ability and detection accuracy of intrusion detection systems in industrial internet scenarios, this paper proposed an intrusion detection method for the industrial internet stacking circular convolution autoencoder and head attention mechanism. Firstly, a convolutional autoencoder was used to extract the spatial features of traffic data and reduce the redundancy of traffic data. Meanwhile, a two-dimensional convolutional layer and a circular padding method were introduced to expand the receptive field of the convolutional kernel on the data by increasing the dimension and expanding the spatial domain of the original data, thereby improving the feature extraction ability. Secondly, the head attention mechanism was utilized to extract the correlation information among the traffic, and attention head weights were assigned according to the importance of the correlation information extracted by different attention heads to consider the most important correlation information when determining the types of network traffic. Finally, binary classification and multi-classification comparative experiments were conducted with traditional intrusion detection methods and intrusion detection methods based on traditional time-series neural networks. Experimental results show that the intrusion detection method proposed in this paper achieves an accuracy of 99.18% in the binary classification task and 87.36% in the multi-classification task, which outperforms traditional intrusion detection methods. The proposed intrusion detection method can effectively detect network intrusions in industrial internet scenarios.

Keywords: intrusion detection; deep learning; convolutional autoencoder; circular padding; head attention mechanism

工业互联网深度融合5G、大数据等新一代信息技术,全面连接人类、设备、系统,已成为数字经济的新型基础设施。在国家高质量发展战略部署下,数以千万计的工业设备和传感器接入工业互联网。而

收稿日期:2023-11-30;**录用日期**:2024-03-02;**网络首发日期**:2024-06-24
网络首发地址:https://link.cnki.net/urlid/23.1235.t.20240622.1013.002
基金项目:国家自然科学基金(52171337)
作者简介:金志刚(1972—),男,教授,博士生导师
通信作者:金志刚,zgjin@tju.edu.cn

针对工业互联网的网络安全事件频发，如 2015 年乌克兰电力系统攻击事件、2020 年台湾 CPC 公司网络攻击事件，使得保护工业互联网的基础设施和网络安全成为一个关键问题[1]。作为一种旁路式部署的检测机制，网络入侵检测系统（intrusion detection system，IDS）能够有效监控和检测整个工业网络中的异常事件或网络攻击。互联网在工业经济中的扩张与发展使入侵检测成为研究人员关注的焦点[2]。

依据检测方法的不同，IDS 分为误用入侵检测和异常入侵检测。误用入侵检测的检测原理决定了其无法识别新型攻击，且具有规则库更新维护成本高、滞后性强等问题，在现代网络环境下具有明显的局限性，因此目前研究人员的研究重点为异常入侵检测。Alazab 等[3]使用飞蛾-火焰优化算法作为搜索算法，为 IDS 生成有效的特征子集，使得作为分类器的决策树算法在该特征子集上获得最佳性能。丁红卫等[4]以降低数据维度和消除冗余信息为目的，提出了一种基于深度自编码器网络和误差反向传播算法的入侵检测模型，通过逐层学习网络数据的分布规则，获得数据的低维特征，并通过使用正则化修正和添加噪声的方式避免模型过拟合。Imrana 等[5]针对传统 IDS 对网络攻击尤其是 U2R 攻击和 R2L 攻击的误报率较高的问题，提出了一种基于双向长短期记忆网络（long short-term memory network，LSTM）的 IDS，双向学习网络流量的时序信息。实验结果表明，该系统优于传统基于 LSTM 的 IDS。Mushtaq 等[6]提出了一种基于自编码器（autoencoder，AE）和 LSTM 的入侵检测模型，利用 AE 和 LSTM 提取网络流量的最佳特征，进而对网络流量进行分类。Zhao 等[7]提出了一种基于轻量级深度神经网络的物联网 IDS，在数据预处理阶段，为避免原始流量特征高维导致模型复杂度高，采用主成分分析（principal component analysis，PCA）算法实现特征降维。在分类模块中使用扩展压缩结构、逆残差结构等方式实现有效的特征提取。郭志民等[8]提出了一种基于特征降维的多头注意力机制 Transformer 模型，并应用于网络入侵检测，该模型考虑了入侵行为的时间相关性，实验结果表明，该模型在多个数据集上表现良好。Basset 等[9]提出了一种多尺度残差时间卷积网络提取网络流量的时空信息，并引入改进的注意力机制计算特征重要性得分，帮助模型在学习过程中专注于重要信息。

尽管上述研究取得了一定成果，但在工业互联网场景下应用仍存在局限性。一方面，工业互联网通常由大量设备、传感器和控制系统组成，产生了海量数据，且数据中混杂大量与表征网络行为类型相关性较弱的冗余特征[10]。而工业领域许多关键系统需要实时响应和决策。冗余信息在干扰网络行为类型有效表征的同时增加了 IDS 数据推理时间成本。部分研究未充分考虑冗余信息的影响，特征提取针对性不足。另一方面，网络行为通常具有时序相关性[11]，部分研究未考虑流量数据的相关性，丢失时序信息。尽管部分研究使用循环神经网络（recurrent neural network，RNN）及其变体网络学习时序特征，但在具有大规模数据的工业互联网场景下可能会出现梯度消失和爆炸[12]的问题。虽然注意力机制能有效避免这一问题，但针对 IDS 注意力机制的应用研究大多集中于提取流量数据全局信息，缺乏对流量数据间时序相关性的进一步分析。

综上，为应对工业互联网中存在的网络安全问题，针对 IDS 应用所面临的挑战，本文提出堆叠循环卷积自编码器和头部注意力机制（circular convolutional autoencoder and head attention mechanism，CCAE-HAM）的入侵检测方法。该方法利用循环卷积自编码器（circular convolutional autoencoder，CCAE）提取网络流量空域信息，通过头部注意力机制（head attention mechanism，HAM）融合时域信息，从时空角度全面分析网络流量，实现网络攻击的精准检测。在卷积自编码器（convolutional autoencoder，CAE）中引入二维卷积层以扩大对原始数据的感受野，这一升维处理能够更有效地提取网络数据空域特征。同时，在二维卷积层引入循环填充方式，在空域上对原始数据进行拼接延展，使卷积核能够跨越数据边界提取信息，进一步提升空域特征提取能力，避免冗余信息进入后续处理模块，影响最终检测性能。之后，提出一种 HAM，以实现网络流量间相关性的特异性提取。该机制考虑到各类流量在同一特征提取角度下的相关性差异，综合流量特征和注意力分数矩阵，为每个注意力头赋予不同的权重，以获得时序相关性对于流量类型判别的贡献大小，进而提升对流量数据之间相关性差异的捕获能力。

1　空域特征提取

特征提取旨在去除原始数据中的噪声和冗余部分，保留关键信息。有效的特征提取能够加快模型的训练和推理速度，提高模型的泛化能力。

目前常用的特征提取方法很多，如 PCA、支持向量机、自编码器等。然而，这些方法应用于工业互联网入侵检测存在局限性。PCA 算法难以应对特征复杂且非线性相关的工业互联网数据[13]。支持向量机在处理大规模非线性数据时则面临计算复杂度过高[14]的问题。而自编码器作为一种自监督学

习算法,通过训练学习数据的关键特征,可以充分拟合非线性相关的工业互联网数据[15]。然而,传统的自编码器可能无法充分捕捉到网络流量的空域局部特征,导致空间信息的损失。

CAE 凭借卷积操作对局部信息的感知实现对网络流量空域特征的提取,包含编码和解码两个模块,式(1)为编码过程,式(2)为解码过程。

$$\boldsymbol{h}=f(E(x))=\boldsymbol{W}\cdot E(x)+b \tag{1}$$

$$y=g(D(\boldsymbol{h}))=\boldsymbol{W}^{\mathrm{T}}\cdot D(\boldsymbol{h})+c \tag{2}$$

式中:$f(\cdot)$、$g(\cdot)$分别为卷积层和反卷积层,$\boldsymbol{W}$、$\boldsymbol{W}^{\mathrm{T}}$分别为卷积层和反卷积层中卷积核的权重矩阵,$b$、$c$分别为卷积层和反卷积层中的偏置项,$E(\cdot)$、$D(\cdot)$分别为编码模块和解码模块中除卷积层和反卷积层之外的网络层,x 为输入样本,$\boldsymbol{h}$ 为 CAE 提取到的网络流量空域特征,y 为 CAE 根据样本特征解码后所得的还原样本。

为提升 CAE 对网络流量空域特征的提取能力,本文采用二维卷积层替代原始的一维卷积层。现有基于深度学习的 IDS 大多将网络流量处理成一维特征向量的形式(通常包括协议类型、服务类型、失败登录次数等特征),并直接进行网络流量类型判定。目前的公开数据集,如 UNSW-NB15[16] 等也为一维特征向量的形式。然而这种一维数据形式限制了 CAE 对空域信息特征的提取。具体而言,在相同的卷积核尺寸下,一维卷积核对数据特征的感受野远远小于二维卷积核,这意味着经过同样的卷积操作后,二维卷积核可以比一维卷积核提取到更多的局部信息,如图 1 所示,其中 X 为流量中的特征,D 表示维度。尽管可以通过堆叠多层一维卷积层来扩大信息提取广度,但会导致网络复杂度增加。另一方面,网络流量的数据特性也使 CAE 对其进行二维卷积操作具有可行性。与语音信号、图像信号等连续信号不同,网络流量的数据特征具有语义独立性,在语音和图像数据中,每个数据(像素)通常没有单独的含义,需要结合上下文或组合其他数据来表达具体含义。而在网络流量中,每个特征(如目的端口、协议类型等)都具有具体含义。这种语义独立性使得改变网络流量的数据顺序并不会破坏特征之间的相关性。

为进一步提高卷积核对空域信息的提取度,本文采用循环填充方式[17]替换传统卷积层中的复制填充方式。循环填充通过在特征图边缘填充远距离边缘的数据,实现了特征图的拼接,从而使卷积核在进行卷积操作时能捕捉到特征图边缘特征之间的相关性信息,避免了复制填充中因卷积核作用于相同数据而带来的计算冗余问题。循环填充的计算函数为:

$$\boldsymbol{A}^{n\times n}\rightarrow\begin{bmatrix}\boldsymbol{A}_{n+1-\frac{p}{2}:n,n+1-\frac{p}{2}:n} & \boldsymbol{A}_{n+1-\frac{p}{2}:n,:} & \boldsymbol{A}_{n+1-\frac{p}{2}:n,1:\frac{p}{2}}\\ \boldsymbol{A}_{:,n+1-\frac{p}{2}:n} & \boldsymbol{A}_{:,:}^{n\times n} & \boldsymbol{A}_{:,1:\frac{p}{2}}\\ \boldsymbol{A}_{1:\frac{p}{2},n+1-\frac{p}{2}:n} & \boldsymbol{A}_{1:\frac{p}{2},:} & \boldsymbol{A}_{1:\frac{p}{2},1:\frac{p}{2}}\end{bmatrix} \tag{3}$$

$$p=k-1 \tag{4}$$

式中:$n\times n$ 为特征数量,$\boldsymbol{A}^{n\times n}$为输入卷积层的原始特征图矩阵,$k$ 为卷积核尺寸,p 为边界填充尺寸。

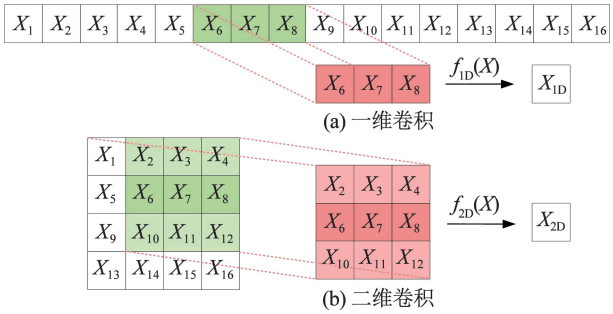


图 1 卷积核感受野对比

Fig. 1 Comparison of receptive fields of convolutional kernels

对于卷积核尺寸为 k 的卷积层,其循环填充后的特征图尺寸见式(5),循环填充示意如图 2 所示。

$$O=(n-k)+p-1 \tag{5}$$

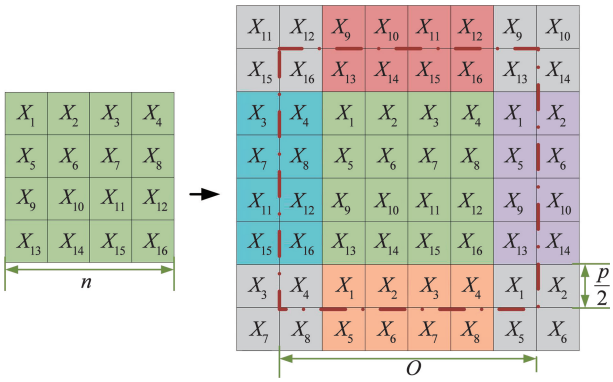


图 2 循环填充示意

Fig. 2 Schematic diagram of circular padding

基于二维 CCAE 的空域特征提取模块的完整结构如图 3 所示。

基于二维 CCAE 的空域特征提取模块包含编码器和解码器两个模块,通过将输入和输出之间的重构误差最小化来学习网络数据的紧凑表示,利用均方误差(mean square error,MSE)衡量输入和输出之间的重构误差。编码器的低维度输出保留了原始数据的关键信息,实现对网络数据空域特征的提取。

2 时域特征融合

注意力机制是一种模拟人类视觉和认知过程的方法,最初由 Vaswani 等[18]提出,被广泛应用于自然语言处理、计算机视觉等领域。注意力机制通过直接建模时域信息捕捉不同数据间的相关性,避免

了传统的循环神经网络的梯度消失或爆炸问题。同时，该机制具有并行计算能力，使得其在具有海量复杂数据，且实时性要求高的工业互联网入侵检测任务中更具优势。

多头注意力机制是注意力机制的一种变体，通过引入多个注意力头对序列信息进行子空间表示，以不同视角提取序列不同部分的相关性。在入侵检测任务中，多头注意力机制通过捕捉网络流量间的相关性来融合时域特征。不同类型的流量之间的相关性信息可能不同，如 DoS 攻击流量在短时间连接频次、单次连接传输字节量两类特征上呈现强相关性，而其他流量在该角度下相关性可能较弱。原始的多头注意力机制为注意力头赋予相同权重，可能导致提取关键时域特征的注意力头贡献不明显，进而误判网络流量的类型。因此，为提高多头注意力机制的精细建模能力，本文提出一种 HAM。在原始机制的基础上，综合注意力分数矩阵和经过加权和计算后的值矩阵计算生成每个注意力头的贡献权重，以抑制相关性较弱信息的干扰。HAM 的计算过程如图 4 所示。

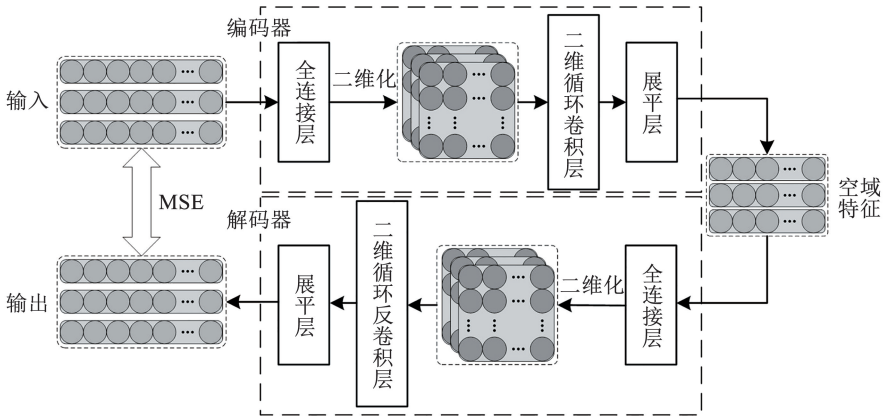


图 3　空域特征提取模块

Fig. 3　Spatial feature extraction module

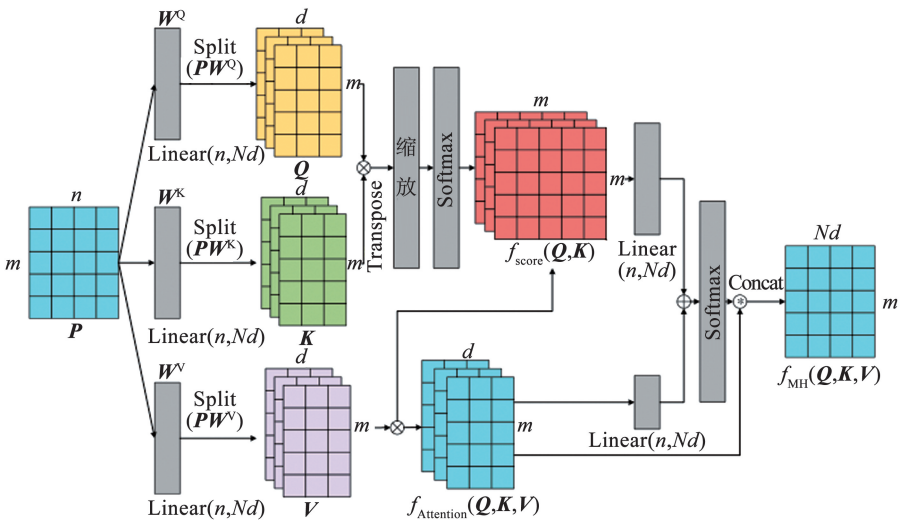


图 4　头部注意力层计算过程

Fig. 4　Calculation process of head attention layer

设输入 HAM 的数据为 $\boldsymbol{P}\in\mathbf{R}^{m\times n}$，将 $\boldsymbol{P}$ 乘以3 个不同的权重矩阵（$\boldsymbol{W}^{\mathrm{Q}}\in\mathbf{R}^{n\times d}$、$\boldsymbol{W}^{\mathrm{K}}\in\mathbf{R}^{n\times d}$、$\boldsymbol{W}^{\mathrm{V}}\in\mathbf{R}^{n\times d}$，其中 d 为后续得到查询矩阵、键矩阵、值矩阵的矩阵维度），得到查询矩阵 $\boldsymbol{Q}\in\mathbf{R}^{m\times d}$、键矩阵 $\boldsymbol{K}\in\mathbf{R}^{m\times d}$、值矩阵 $\boldsymbol{V}\in\mathbf{R}^{m\times d}$，然后将 $\boldsymbol{Q}$ 与 $\boldsymbol{K}$ 进行缩放点积与 Softmax 归一化，得到注意力分数矩阵 $f_{\mathrm{score}}(\boldsymbol{Q},\boldsymbol{K})$，最后将 $f_{\mathrm{score}}(\boldsymbol{Q},\boldsymbol{K})$ 与 $\boldsymbol{V}$ 相乘计算加权和，获得该注意力头下融合时域特征的输出特征矩阵。上述过程计算式为：

$$\begin{cases}\boldsymbol{Q}=\boldsymbol{P}\boldsymbol{W}^{\mathrm{Q}}\\ \boldsymbol{K}=\boldsymbol{P}\boldsymbol{W}^{\mathrm{K}}\\ \boldsymbol{V}=\boldsymbol{P}\boldsymbol{W}^{\mathrm{V}}\end{cases}\tag{6}$$

$$\begin{cases}f_{\mathrm{score}}(\boldsymbol{Q},\boldsymbol{K})=f_{\mathrm{Softmax}}\left(\dfrac{\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d}}\right)\\ f_{\mathrm{Attention}}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V})=f_{\mathrm{score}}(\boldsymbol{Q},\boldsymbol{K})\boldsymbol{V}\end{cases}\tag{7}$$

并行进行多次上述计算，获得多个注意力头下

的 $f_{\text{score}}(\boldsymbol{Q},\boldsymbol{K})$ 和对应的 $f_{\text{Attention}}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V})$。如图 4 所示,将多头注意力产生的编码输出经过两个线性变换层($\boldsymbol{W}^0,\boldsymbol{W}^1\in\mathbf{R}^{d\times N}$,其中 N 为注意力头数),再进行相加和 Softmax 归一化,得到多个注意力头的贡献权重 $\boldsymbol{W}_{\text{Head}}$。计算 $\boldsymbol{W}_{\text{Head}}$ 和 $f_{\text{Attention}}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V})$ 的哈达玛积,输出带注意力头权重的编码输出 $f_{\text{HA}}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V})$。计算式为:

$$\boldsymbol{W}_{\text{Head}}=f_{\text{Softmax}}(\boldsymbol{W}^0 f_{\text{score}}(\boldsymbol{Q},\boldsymbol{K})+\boldsymbol{W}^1 f_{\text{Attention}}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V})) \tag{8}$$

$$f_{\text{HA}}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V})=f_{\text{Attention}}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V})*\boldsymbol{W}_{\text{Head}} \tag{9}$$

式中 * 为哈达玛积。

最后,利用式(10)将多个注意力头进行拼接和线性变换,输出 HAM 的编码输出。

$$\begin{cases} f_{\text{MH}}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V})=\text{Concat}(f_{\text{head}_1},\cdots,f_{\text{head}_n})\boldsymbol{W}^{\text{L}} \\ f_{\text{head}_i}=f_{\text{HA}}(\boldsymbol{P}\boldsymbol{W}_i^{\text{Q}},\boldsymbol{P}\boldsymbol{W}_i^{\text{K}},\boldsymbol{P}\boldsymbol{W}_i^{\text{V}}) \end{cases} \tag{10}$$

式中:$\boldsymbol{W}^{\text{L}}$ 为线性变换的权重矩阵;Concat(·)为拼接函数,用于拼接多个注意力头;f_{head_n} 为不同注意力头的带注意力权重的编码输出;$\boldsymbol{W}_i^{\text{Q}}$、$\boldsymbol{W}_i^{\text{K}}$、$\boldsymbol{W}_i^{\text{V}}$ 为第 i 头注意力头的权重矩阵,在具体实现中,通常由 3 个线性变换层的权重矩阵切分获得,以降低模型复杂度。

基于 HAM 构建的时域特征融合模块如图 5 所示。

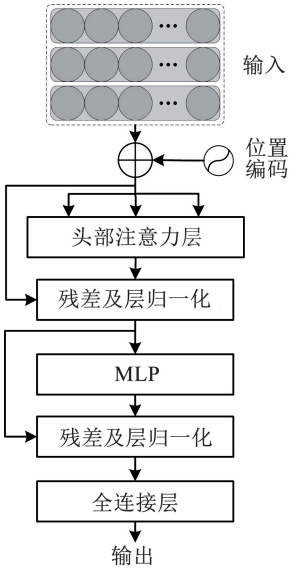


图 5 时域特征融合模块

Fig. 5 Temporal feature fusion module

尽管 HAM 可以融合数据的时域特征,但机制本身无法学习输入数据的位置信息,因此在网络流量进入头部注意力层前,先对其进行位置编码以嵌入位置信息。位置编码嵌入公式为:

$$\boldsymbol{P}=\boldsymbol{Y}+f_{\text{PE}}(r,i) \tag{11}$$

$$\begin{cases} f_{\text{PE}(r,2j)}=\sin(r/10\ 000^{2j/n}) \\ f_{\text{PE}(r,2j+1)}=\cos(r/10\ 000^{2j/n}) \end{cases} \tag{12}$$

式中:$\boldsymbol{Y}\in\mathbf{R}^{m\times n}$ 为输入空域特征样本,r 为当前空域特征样本在输入特征样本 $\boldsymbol{Y}$ 中的位置,j 为当前样本的特征位置,$f_{\text{PE}}(r,i)$ 为位置编码结果,$\boldsymbol{P}\in\mathbf{R}^{m\times n}$ 为嵌入位置信息后的特征样本。

将位置编码后的流量数据传入头部注意力层以融合时域特征,然后应用多层感知机(multilayer perceptron,MLP)巩固模型的建模能力,提高模型的分类性能。由于堆叠多个网络层的深度网络在拟合过程中同样可能出现梯度消失或梯度爆炸的问题,因此在模型中间穿插残差连接和层归一化层。

3 CCAE-HAM

CCAE-HAM 的完整结构如图 6 所示。

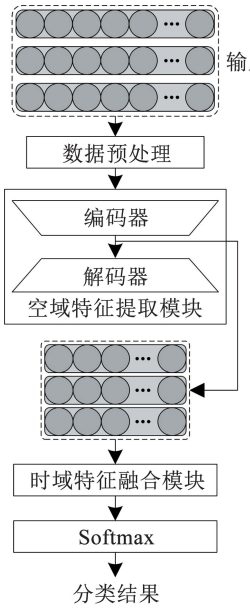


图 6 CCAE-HAM 完整结构

Fig. 6 Complete structure of CCAE-HAM

对原始的入侵检测数据集进行数据预处理操作,使用独热编码方式将数据集中类似传输层协议(如 TCP 和 UDP)等字符型数据特征转换为数值属性。在处理完所有字符型数据特征后,对全局数据按数据特征列进行最大最小值归一化,将数据映射到统一的[0,1],加快模型的收敛速度。归一化公式为

$$x'=\frac{x-x_{\min}}{x_{\max}-x_{\min}} \tag{13}$$

式中:x' 为归一化之后的结果,$x_{\max}$ 代表该特征列中的最大数值,$x_{\min}$ 代表该特征列中的最小数值。

将预处理后的数据传入空域特征提取模块进行训练,待模型训练结束后保存模块参数,以供后续时域特征融合模块调用。

调用训练好的空域特征提取模块，将编码器输出作为空域特征传入时域特征融合模块做进一步处理。最后，将模块输出传入 Softmax 层，对网络流量进行类型判定。

4　实　验

4.1　实验环境与评价指标

实验硬件环境与软件环境见表 1。

表 1　实验环境

Tab. 1　Experimental environment

实验环境	设备/软件	设备型号/软件版本
硬件环境	CPU	12th Gen Intel(R) Core(TM) i5-12400 2.50 GHz
	GPU	NVIDIA GeForce RTX 3070
	机带 RAM	32 GB(3 200 MHz)
软件环境	操作系统	Windows 11
	语言	Python 3.9
	框架	PyTorch 1.13.1

为评估所提方法的性能，采用分类任务常用的评价指标作为参考，主要有准确率 A、精确率 P、召回率 R 和 F_1 分数，其计算式分别为：

$$A=\frac{T_{\mathrm{P}}+T_{\mathrm{N}}}{T_{\mathrm{P}}+T_{\mathrm{N}}+F_{\mathrm{P}}+F_{\mathrm{N}}} \tag{14}$$

$$P=\frac{T_{\mathrm{P}}}{T_{\mathrm{P}}+F_{\mathrm{P}}} \tag{15}$$

$$R=\frac{T_{\mathrm{P}}}{T_{\mathrm{P}}+F_{\mathrm{N}}} \tag{16}$$

$$F_1=\frac{2T_{\mathrm{P}}}{2T_{\mathrm{P}}+F_{\mathrm{P}}+F_{\mathrm{N}}} \tag{17}$$

式中：T_{P}为模型正确识别为攻击样本的样本数量，F_{P}为模型错误将正常样本识别为攻击样本的样本数量，T_{N}为模型正确识别为正常样本的样本数量，F_{N}为模型错误将攻击样本识别为正常样本的样本数量。

4.2　实验数据集

UNSW-NB15 数据集是一个广泛应用于网络入侵检测领域的公开数据集，包含了来自不同攻击类型和正常活动的网络流量数据，每个样本均包含了实际网络通信中的各种特征，如目标端口、协议类型、数据包大小、连接次数、连接建立时间等。该数据集包含多种类型的网络入侵行为，涵盖了工业互联网中常见的攻击类型如 DoS(拒绝服务攻击)、Fuzzers(模糊攻击)等，适配本文研究场景[19]。UNSW-NB15 数据集中各类型样本分布见表 2。

表 2　UNSW-NB15 数据集样本分布

Tab. 2　Sample distribution of UNSW-NB15 dataset

类型	训练集样本数	测试集样本数
Normal	56 000	36 988
Analysis	2 000	677
Backdoor	1 746	583
DoS	12 264	4 089
Exploits	33 393	11 132
Fuzzers	18 184	6 062
Generic	40 000	18 871
Reconnaissance	10 491	3 496
Shellcode	1 133	378
Worms	130	44

4.3　调参实验

通过调参实验调整模型超参数，以达到模型的最佳性能。实验选取注意力头数、空域特征维度数等关键超参数进行，结果见表 3。

表 3　调参实验结果

Tab. 3　Results of parameter tuning experiment

注意力头数	空域特征维度数	训练集准确率/%
1	16	96.74
1	32	96.06
1	64	95.01
2	16	96.86
2	32	96.96
2	64	96.88
4	16	97.72
4	32	97.04
4	64	97.44
8	16	95.77
8	32	96.04
8	64	96.53

通过实验观察，固定注意力头数为 4 和空域特征维度为 16，在训练集上得到了最佳模型性能。注意力头数表示模型对数据特征关注的角度数量，增加注意力头数通常认为可以提高模型对数据集的拟合能力。但实验结果表明，注意力头数增至 8 时，模型性能反而下降。这是因为在多头注意力机制的具体实现中，为了降低模型复杂度，不同的注意力头共享一个线性变换层，并根据注意力头数对权重矩阵进行切分，这一操作使得注意力头数的增加会导致注意力头维度的减少，尽管模型对数据的关注角度增多，但注意力头提取到的相关性信息却随维度的减少而减少。同时，实验结果表明，空域特征维度为

16 时已充分提取网络流量的关键特征。

在固定上述超参数的基础上,模型其他超参数根据实验以及经验确定,具体设置见表 4。

表 4 模型超参数设置

Tab. 4 Model hyperparameter settings

模型超参数	数值
Batch_size	2 048
学习率	0.01
训练轮数	50
随机丢弃率	0.30
卷积核	3
卷积步长	1

4.4 消融实验

CCAE-HAM 主要由两个模块构成,因此进行两组消融实验。消融实验 1 是固定时域特征融合模块,通过调整空域特征提取模块,观察该模块对整体模型性能的影响;消融实验 2 是固定空域特征提取模块,通过调整时域特征融合模块,观察该模块对整体模型性能的影响。

4.4.1 消融实验 1

保持时域特征融合模块不变,研究无空域特征提取模块与分别基于自编码器(AE)、基于一维卷积自编码器(1DCAE)、基于二维卷积自编码器(2DCAE)、基于 CCAE 的空域特征提取模块对整体模型性能的影响,实验结果见表 5 和图 7。

表 5 消融实验 1

Tab. 5 Ablation experiment 1

模型	*A*/%	*P*/%	*R*/%	MSE
HAM	97.26	97.13	97.93	
AE-HAM	98.21	97.97	98.87	0.005 3
1DCAE-HAM	98.34	98.10	98.89	0.004 7
2DCAE-HAM	98.67	98.24	99.37	0.002 9
CCAE-HAM	99.18	98.59	99.91	0.001 6

根据表 5 可知,与未使用空域特征提取模块的 HAM 相比,CCAE-HAM 的各项指标均有提升。这是因为网络流量中混杂了冗余信息,而空域特征提取模块有效地提炼出了网络流量中的有用信息,一方面提高了数据特征对网络流量类型的表征能力,另一方面使得后续时域特征融合模块能够不受冗余信息干扰,充分学习到网络流量间的相关性,进而对网络流量类型做出正确分类。

由图 7 可以看出,在训练过程中,CCAE 的损失值(Loss)下降速度更快,并且在模型收敛后,loss 值也更低,说明 CCAE 对空域特征的提取能力更强。为验证 CCAE 的训练结果,实验中采用 MSE 作为评估指标。MSE 通过计算原始数据与自编码器解码输出之间的误差来衡量自编码器的特征提取能力,原始数据与解码输出之间的误差越小,说明自编码器提取到的网络流量中的空域特征越有效,因为解码器能将该空域特征更有效地解码成原始数据。本实验分别使用已训练收敛的 AE、1DCAE、2DCAE、CCAE 提取测试集的空域特征,实验结果见表 5。由表 5 可以看出,CCAE 的 MSE 值最低,说明 CCAE 对流量数据空域特征的提取能力最强。CCAE-HAM 的各项指标均优于对比模型也证明了 CCAE 的有效性。

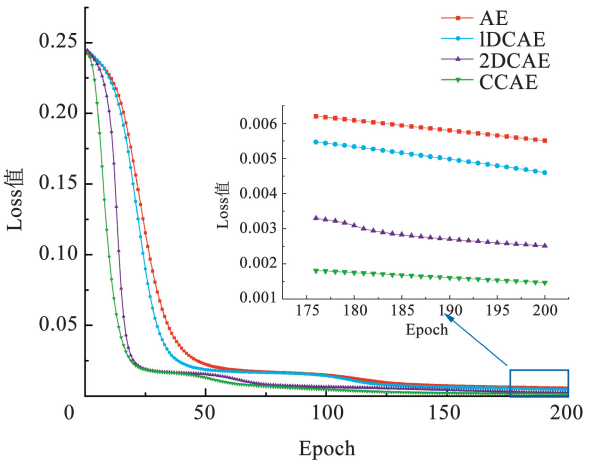
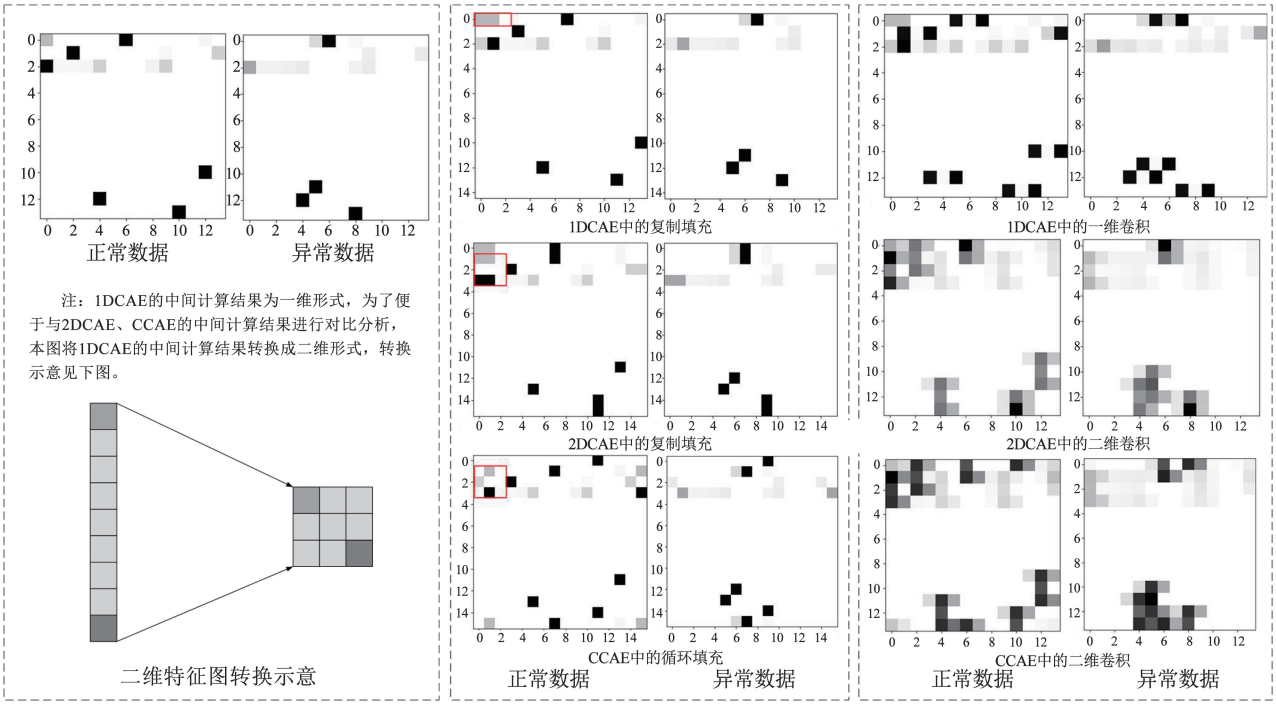


图 7 空域特征提取模块损失曲线对比

Fig. 7 Comparison of loss curves of spatial feature extraction module

为进一步分析 CCAE 中对原始数据的升维处理和循环填充操作对流量数据空域特征提取的影响,选取正常数据和异常数据在 CCAE、1DCAE、2DCAE 计算下的中间计算结果,并生成可视化特征灰度图,如图 8 所示。

从图 8 可以看出,网络流量通过 CCAE 计算得到的中间计算结果的特征图信息相比于 1DCAE、2DCAE 计算所得到的特征图信息更加丰富。在 1DCAE 中,网络流量为一维数据形式,其中的每个特征仅与其他两个特征相邻,而在 CCAE 中,网络流量为二维数据形式,其中的每个特征与其他 8 个特征相邻。CCAE 对原始数据的升维处理扩大了卷积核的感受野,提高了卷积操作对数据特征的提取能力。2DCAE 对特征图使用传统的复制填充操作,当卷积核对特征图边缘特征进行卷积时,会作用于相同数据,导致计算冗余。而 CCAE 引入了循环填充方式,通过对特征图进行拼接延展,使得卷积核在对特征图边缘特征进行卷积时能够跨越原始特征图的边界,提取到原始特征图两端特征之间的相关性信息。综上所述,CCAE 中对原始数据的升维处理和循环填充操作提高了模型对网络流量特征提取的有效性,提升了整体模型对网络流量的分类性能。



(a) 原始数据特征　　(b) 填充后的数据特征图　　(c) 卷积后的数据特征图

图 8　可视化特征灰度图

Fig. 8　Visualized feature grayscale image

4.4.2　消融实验 2

保持空域特征提取模块不变，研究时域特征融合模块对整体模型性能的影响，实验结果见表 6。

表 6　消融实验 2

Tab. 6　Ablation experiment 2　　%

模型	A	P	R	F_1
CCAE	92.63	93.22	93.41	93.32
CCAE-MAM	97.85	95.61	99.24	97.39
CCAE-HAM	99.18	98.59	99.91	99.25

根据表 6 可知，与未使用时域特征融合模块的 CCAE 相比，CCAE-HAM 的各项指标均有提升。这是因为 CCAE-HAM 在对网络流量类型进行判定时并非单独考虑当前网络流量，还综合考虑了不同网络流量之间的相关性信息，使得 CCAE-HAM 在对网络流量进行分析时具备时空多角度的分析能力，对于某些异常特征不明显的网络流量，CCAE-HAM 能够综合网络流量的时空信息对其类型做出正确判定，提高了模型的召回率，减少了漏报的情况。在时域特征融合模块中，使用 HAM 的模型性能优于使用多头注意力机制的模型。因为 HAM 为贡献度更高的注意力头赋予了更大的权重，从而提升了模型对网络流量间相关性的提取能力。

为进一步分析网络流量相关性以及 HAM 对注意力头贡献度的学习能力，选取测试集某段连续时间段的流量数据，使用训练好的 HAM 权重计算并输出该时间段中最后一条流量对其他流量的注意力分数、该条流量经过加权和计算后的值矩阵分量以及 HAM 综合注意力分数和值矩阵分量所计算生成的注意力头贡献权重，结果见表 7、表 8 和图 9。

表 7　某时间段最后一条流量的值矩阵分量

Tab. 7　Value matrix components of the last traffic in a certain period

注意力头	值矩阵分量			
	矩阵分量 1	矩阵分量 2	矩阵分量 3	矩阵分量 4
注意力头 1	−0.015 7	−0.487 5	−0.292 4	0.270 4
注意力头 2	0.908 1	−1.687 6	−0.618 4	−1.744 9
注意力头 3	0.454 5	0.277 3	0.306 3	−0.766 7
注意力头 4	1.125 1	−3.754 4	3.314 8	1.017 4

表 8　某时间段最后一条流量对其他流量的注意力分数

Tab. 8　Attention scores of the last traffic to other traffic in a certain period

流量类型	注意力分数			
	注意力头 1	注意力头 2	注意力头 3	注意力头 4
正常	0.039 9	0.024 4	0.016 9	0.016 2
正常	0.040 9	0.012 1	0.023 5	0.018 3
正常	0.048 2	0.013 6	0.027 8	0.020 6
正常	0.099 2	0.015 4	0.003 8	0.068 6
正常	0.117 7	0.243 9	0.003 6	0.074 3

表 8(续)

流量类型	注意力分数			
	注意力头 1	注意力头 2	注意力头 3	注意力头 4
攻击	0.018 0	0.252 4	0.505 4	0.201 6
正常	0.123 6	0.143 4	0.005 3	0.105 7
正常	0.113 8	0.129 5	0.006 6	0.109 6
正常	0.062 5	0.013 3	0.031 7	0.024 9
正常	0.058 8	0.014 0	0.027 7	0.021 4
正常	0.064 4	0.011 4	0.031 0	0.017 1
正常	0.076 3	0.010 1	0.038 9	0.022 8
正常	0.065 5	0.008 3	0.035 4	0.020 8
正常	0.048 0	0.006 9	0.037 9	0.020 3
攻击	0.014 2	0.054 9	0.107 9	0.162 6
攻击	0.009 1	0.046 4	0.096 7	0.095 2

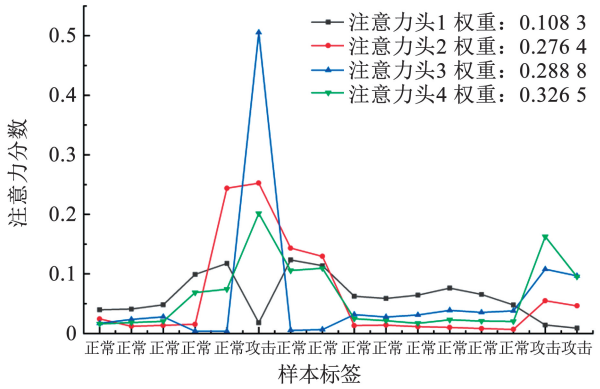


图 9　注意力分数及注意力头贡献权重分析

Fig. 9　Analysis of attention scores and contribution weights of attention heads

HAM 通过学习获得每条流量数据在不同的注意力头下与其他流量数据的相关性信息,即每条流量数据对其他流量数据的注意力分数,然后根据注意力分数对值矩阵进行加权求和,得到融合时域信息的流量数据,赋予了模型在网络流量分类时考虑时域信息的能力。结合表 7、表 8 和图 9 可以看出,最后一条攻击数据在注意力头 1 的关注角度下,对其他流量数据的注意力分数分布最为均衡,与其他数据的关联不明显,甚至对同为攻击类型的流量数据也没有明显的关注,说明注意力头 1 并没有有效提取到流量之间的相关性。因此,HAM 通过学习该攻击数据在注意力头 1 下的值矩阵分量和注意力分数分布为注意力头 1 赋予了较低的权重。在其他注意头的关注角度下,注意力分数表现出明显的差异,特别是注意力头 3 和注意力头 4。注意力头 3 在注意力分数上分布差异最明显,但 HAM 学习到注意力头 3 下该攻击数据的值矩阵分量对于判定网络流量类型贡献较弱,因此综合考虑两方面因素,为注意力头 3 赋予了相应的权重。而注意力头 4 在关注其他攻击数据的同时,还考虑了其他攻击数据相邻位置的流量信息,说明注意力头提取到了流量之间显著的相关性信息,因此 HAM 为其赋予了较高的权重。综合分析表 7、表 8 和图 9 数据可知,不同的注意力头在提取流量间的相关性时表现不同。HAM 能够通过学习注意力头的贡献度,为其分配不同的权重,提高注意力机制对时序相关性的提取能力,从而实现对当前流量数据的精准分类。

4.5　二分类实验

将本文所提方法与随机森林(random forest, RF)、CNN[20]、RNN[21]、LSTM[22]、HA[23](hierarchical attention)进行对比,其中 RF、CNN 为传统入侵检测方法,RNN、LSTM 为基于传统时序神经网络的入侵检测方法,HA 为基于注意力机制的入侵检测方法,对比结果见表 9。其中,本文搭建了 RF 方法,并收集了实验数据,对于采用与本文不同数据集的方法[20-22],本文对其进行了复现实验,以便于与本文方法进行对比,对于采用与本文相同数据集的方法[23],实验结果引用自原论文。

表 9　性能对比结果

Tab. 9　Comparison results of performance　%

检测方法	A	P	R	F_1
RF	82.17	75.84	99.22	85.97
CNN	86.96	81.48	98.76	89.29
RNN	89.59	87.56	94.53	90.90
LSTM	92.68	89.71	97.93	93.64
HA[23]	98.76	99.35	98.94	99.14
CCAE-HAM	**99.18**	**98.59**	**99.91**	**99.25**

由表 9 可以看出,CCAE-HAM 和传统入侵检测方法、基于传统时序神经网络的入侵检测方法在召回率上表现良好,其中 CCAE-HAM 的表现最佳。在保持高召回率的同时,CCAE-HAM 的精确率、准确率及 F_1 分数远优于传统方法,进一步分析发现,RF、CNN 等方法主要根据流量数据自身的特征对该流量做出分类,而流量数据中存在的冗余信息可能会干扰方法的判断,特别是 RF 会在建立决策树时随机选取数据特征,导致无法有效利用流量数据中的有用信息,方法表现较差。同样是提取网络流量时序性,CCAE-HAM 相较于 RNN、LSTM 具备更强的提取能力,所提取到的数据相关性能够更好地提升方法的分类性能。与基于注意力机制的入侵检测方法相比,CCAE-HAM 在准确率和 F_1 分数上优于 HA,说明了方法在整体性能上优于 HA。因为准确率考虑了方法对流量正确分类和错误分类的总数,而 F_1 分数综合考虑了方法的精确率和召回率表现,准确率

和 F_1 分数能够从整体上反映方法性能。进一步地，虽然 CCAE-HAM 和 HA 同样使用注意力机制提取流量间相关性，但 CCAE-HAM 能对流量间相关性进行特异性提取，根据数据相关性的重要程度分配注意力头的权重，对流量间相关性有更强的提取能力。因此，CCAE-HAM 在召回率上优于 HA。实际上，相较于精确率，工业互联网中的入侵检测任务对召回率有着更高的敏感性。在入侵检测任务中，召回率高意味着方法能够识别出更多的攻击流量，漏报入侵事件的情况更少。在实际应用中，方法漏报入侵事件（低召回率）往往比误报入侵事件（低精确率）更加危险。一旦方法漏报了某些入侵事件，这些入侵行为便可能对工业互联网造成重大危害，而误报入侵事件则可以通过后续验证和调查来排除。因此，相比于 HA，CCAE-HAM 在工业互联网入侵检测任务中更具有优势。

4.6　多分类实验

为检验 CCAE-HAM 对具体攻击类别的检测能力，进行了多分类实验。CCAE-HAM 在各类别流量数据上的性能表现见表 10。

表 10　CCAE-HAM 多分类性能表现

Tab. 10　Multi-classification performance of CCAE-HAM　%

类别	P	R	F_1
Normal	98.38	98.84	98.61
Fuzzers	75.11	63.28	68.69
Analysis	0	0	0
Backdoor	0	0	0
DoS	100.00	15.51	26.85
Expliots	57.96	93.31	71.50
Generic	99.82	96.24	97.99
Reconnaissance	70.54	65.68	68.02
Shellcode	100.00	11.38	20.43
Worms	0	0	0

从表 10 可以看出，CCAE-HAM 对于 Normal、Generic 类数据具有最好的检测性能，对 Fuzzers、Reconnaissance 等类表现欠佳，对 Analysis、Backdoor、Worms 类的检测效果最差。其原因在于 UNSW-NB15 数据集数据不平衡，各类别数据的数量差别过大，Normal、Generic 类数据在数据集中有大量实例且通常连续出现，流量之间的时序规律较为固定，方法能够充分学习该类流量自身特征以及与其他流量之间的相关性。Backdoor、Worms 等类在数据集中数量较少且穿插在各类数据中，导致方法无法很好地学习 Backdoor、Worms 等类的时空域特征，因此无法准确判断出其类型。

对多分类实验结果综合考虑，CCAE-HAM 整体准确率达到 87.36%，将 CCAE-HAM 与其他入侵检测方法进行对比，结果如图 10 所示。对比方法实验数据来源于复现[20-22]以及对原论文结果的直接引用[24-25]。

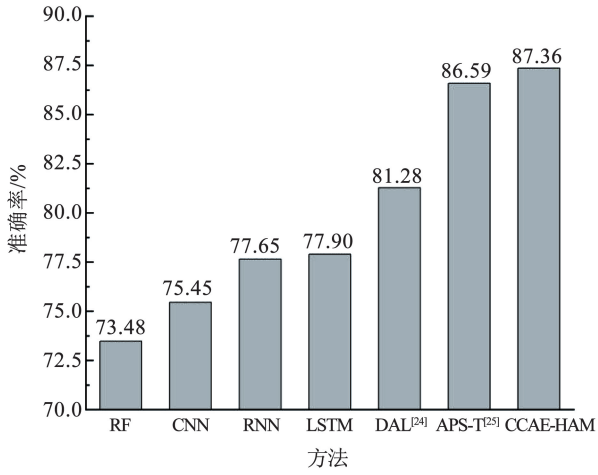


图 10　不同方法多分类性能对比

Fig. 10　Comparison of multi-classification performance of different methods

多分类任务较二分类任务更具有挑战性，因为流量数据之间的关联性更加复杂和多样。因此，在多分类任务下，方法性能相对于二分类任务可能会有所下降。然而，与传统入侵检测方法（RF、CNN[20]）、基于时序神经网络的入侵检测方法（RNN[21]、LSTM[22]）以及其他基于注意力机制的入侵检测方法（DAL[24]、APS-T[25]）相比，本文提出的方法仍然具有优势。原因是本文方法对于流量特征的提取能力更强，且能够从时空域角度更加全面地分析网络流量数据。

5　结　论

1）本文提出了一种堆叠二维循环卷积自编码器和头部注意力机制的入侵检测方法，其由二维循环卷积自编码器和头部注意力机制组成。

2）针对工业互联网中数据海量、信息冗余的挑战，引入二维卷积层和循环填充方式，有效提取了网络数据空域特征，避免了冗余信息的干扰。

3）考虑到网络流量具有复杂的相关性，提出一种头部注意力机制，为每个注意力头赋予不同的权重，提高了多头注意力机制的时序特征提取能力。

4）实验结果显示，本文提出的工业互联网入侵检测方法在二分类任务中准确率达 99.18%，在多分类任务中准确率达 87.36%，优于传统入侵检测方法。表明该方法能够充分提取网络流量的空域特征，减少冗余信息，并差异性提取流量之间的相关性

信息,进而对网络流量类型进行判断,可以有效解决工业互联网领域入侵检测应用面临的问题。

参考文献

[1]智晓欢, 刘伟, 张立彤. 工业互联网及其面临的安全问题[J]. 智能制造, 2023(4): 20
Zhi Xiaohuan, Liu Wei, Zhang Litong. Industrial internet and its security issues[J]. Intelligent Manufacturing, 2023(4): 20

[2]王军, 王华琳, 黄博文, 等. 基于联邦学习和自注意力的工业物联网入侵检测[J]. 吉林大学学报(工学版), 2023, 53(11): 3229
Wang Jun, Wang Hualin, Huang Bowen, et al. Intrusion detection for industrial internet of things based on federated learning and self-attention [J]. Journal of Jilin University (Engineering and Technology Edition), 2023, 53(11): 3229. DOI: 10.13229/j.cnki.jdxbgxb.20221027

[3]Alazab M, Khurma R A, Awajan A, et al. A new intrusion detection system based on Moth-Flame Optimizer algorithm [J]. Expert Systems with Applications, 2022, 210: 118439. DOI: 10.1016/j.eswa.2022.118439

[4]丁红卫, 万良, 龙廷艳. 深度自编码网络在入侵检测中的应用研究[J]. 哈尔滨工业大学学报, 2019, 51(5): 185
Ding Hongwei, Wan Liang, Long Tingyan. Research on the application of deep auto-encoder network in intrusion detection[J]. Journal of Harbin Institute of Technology, 2019, 51(5): 185. DOI: 10.11918/j.issn.0367-6234.201809115

[5]Imrana Y, Xiang Yanping, Ali L, et al. A bidirectional LSTM deep learning approach for intrusion detection[J]. Expert Systems with Applications, 2021, 185: 115524. DOI: 10.1016/j.eswa.2021.115524

[6]Mushtaq E, Zameer A, Umer M, et al. A two-stage intrusion detection system with auto-encoder and LSTMs[J]. Applied Soft Computing, 2022, 121: 108768. DOI: 10.1016/j.asoc.2022.108768

[7]Zhao Ruijie, Gui Guan, Xue Zhi, et al. A novel intrusion detection method based on lightweight neural network for internet of things[J]. IEEE Internet of Things Journal, 2021, 9(12): 9960. DOI: 10.1109/JIOT.2021.3119055

[8]郭志民, 周劼英, 王丹, 等. 基于 Transformer 神经网络模型的网络入侵检测方法[J]. 重庆大学学报, 2021, 44(11): 81
Guo Zhimin, Zhou Jieying, Wang Dan, et al. Network intrusion detection method based on Transformer neural network model[J]. Journal of Chongqing University, 2021, 44(11): 81. DOI: 10.11835/j.issn.1000-582X.2021.11.010

[9]Basset M A, Hawash H, Chakrabortty R K, et al. Semi-supervised spatiotemporal deep learning for intrusions detection in IoT networks [J]. IEEE Internet of Things Journal, 2021, 8(15): 12251. DOI: 10.1109/JIOT.2021.3060878

[10]Nazir A, Khanr A. A novel combinatorial optimization based feature selection method for network intrusion detection[J]. Computers & Security, 2021, 102: 102164. DOI: 10.1016/j.cose.2020.102164

[11]Zhao Ruijie, Huang Yiteng, Deng Xianwen, et al. A novel traffic classifier with attention mechanism for industrial internet of things [J]. IEEE Transactions on Industrial Informatics, 2023, 19(11): 10799. DOI: 10.1109/TII.2023.3241689

[12]Amutha S, Kavitha R, Srinivasan R, et al. Secure network intrusion detection system using NID-RNN based deep learning [C]//2022 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI). Chennai: IEEE, 2022: 1. DOI: 10.1109/ACCAI53970.2022.9752526

[13]Kaushik S, Bhardwaj A, Alomari A, et al. Efficient, lightweight cyber intrusion detection system for IoT ecosystems using MI2G algorithm[J]. Computers, 2022, 11(10): 142. DOI: 10.3390/computers11100142

[14]Tavara S. Parallel computing of support vector machines: a survey [J]. ACM Computing Surveys (CSUR), 2019, 51(6): 1. DOI: 10.1145/3280989

[15]Basati A, Faghih M M. PDAE: efficient network intrusion detection in IoT using parallel deep auto-encoders[J]. Information Sciences, 2022, 598: 57. DOI: 10.1016/j.ins.2022.03.065

[16]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 (MilCIS). Piscataway: IEEE, 2015: 7348942. DOI: 10.1109/MilCIS.2015.7348942

[17]Cheng Lei, Khalitov R, Tong Yu, et al. Classification of long sequential data using circular dilated convolutional neural networks [J]. Neurocomputing, 2023, 518: 50. DOI: 10.1016/j.neucom.2022.10.054

[18]Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need [C]//Neural Information Processing Systems. La Jolla: MIT Press, 2017: 30

[19]孙海丽, 龙翔, 韩兰胜, 等. 工业物联网异常检测技术综述[J]. 通信学报, 2022, 43(3): 196
Sun Haili, Long Xiang, Han Lansheng, et al. Overview of anomaly detection techniques for industrial internet of things[J]. Journal on Communications, 2022, 43(3): 196. DOI: 10.11959/j.issn.1000-436x.2022032

[20]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 (ICACCI). Udupi: IEEE, 2017: 1222. DOI: 10.1109/ICACCI.2017.8126009

[21]Yin Chuanlong, Zhu Yuefei, Fei Jinlong, et al. A deep learning approach for intrusion detection using recurrent neural networks [J]. IEEE Access, 2017, 5: 21954. DOI: 10.1109/ACCESS.2017.2762418

[22]Althubiti S A, Jones E M, Roy K. LSTM for anomaly-based network intrusion detection [C]//2018 28th International Telecommunication Networks and Applications Conference (ITNAC). Sydney: IEEE, 2018: 8615300. DOI: 10.1109/ATNAC.2018.8615300

[23]Liu Chang, Liu Yang, Yan Yu, et al. An intrusion detection model with hierarchical attention mechanism[J]. IEEE Access, 2020, 8: 67542. DOI: 10.1109/ACCESS.2020.2983568

[24]Cao Ke, Zhu Jinqi, Feng Weijia, et al. Network intrusion detection based on dense dilated convolutions and attention mechanism[C]// 2021 International Wireless Communications and Mobile Computing (IWCMC). Harbin: IEEE, 2021: 463. DOI: 10.1109/IWCMC51323.2021.9498652

[25]Wang Xinlei, Wang Xiaojuan, He Mingshu, et al. Spatial-temporal graph model based on attention mechanism for anomalous IoT intrusion detection [J]. IEEE Transactions on Industrial Informatics, 2023, 20(3): 3497. DOI: 10.1109/TII.2023.3308784