

基于时空图注意力网络的物联网异常检测

苏宪时, 杨丽平, 马润泽, 江同棒, 曲天歌, 刘洪波

大连海事大学人工智能学院 大连 中国 116026

摘要 异常检测技术已成为维护物联网安全的重要技术之一。物联网设备的流量与一般的互联网流量不同, 具有不同的流量通信模式。应对物联网基础设施的复杂性, 基于机器学习的方法具有自动化检测、环境适应性强以及对大规模数据处理能力, 成为异常检测的主要方法。然而现有的基于机器学习的异常检测模型普遍存在两个问题: 一是忽略了异常信息的复杂演化特征, 缺乏对信息时空依赖关系的总体建模能力; 二是存在噪声数据和冗余特征导致模型计算量大、训练不稳定、检测准确性差等问题。为此, 本文提出了一种基于动态流量图的时空图注意力网络(Spatio-Temporal Graph Attention Network, STGAT)异常检测方法。引入动态流量图的概念, 通过将网络流量数据构建为包含时序关联的动态图快照, 分析网络节点流量的时空演化模式; 通过STGAT从动态图的视角分析网络节点的流量通信, 实现对异常节点的检测; 此外, 还提出一种协同检测模型训练的自适应特征选择机制, 通过筛减噪声和冗余特征, 提升模型训练的稳定性以及检测的准确性。在 UNSW-NB15、CICIDS-2017、CICIDS-001 数据集上的实验结果表明, 相比传统的异常检测模型, 本文的模型在误报率 FAR 与 F1 分数等性能指标上有较大提升; 而且协同学习模型自适应地筛选出和监督任务最紧密相关的特征子集, 更加准确地检测物联网异常。

关键词 可信人工智能; 网络安全; 图神经网络; 异常检测; 自适应特征选择

中图分类号 TP393 DOI号 10.19363/J.cnki.cn10-1380/tn.2025.09.11

IoT Anomaly Detection based on Spatiotemporal Graph Attention Network

SU Xianshi, YANG Liping, MA Runze, JIANG Tongbang, QU Tiange, LIU Hongbo

College of Artificial Intelligence, Dalian Maritime University, Dalian 116026, China

Abstract Anomaly detection technology has become one of the most essential technologies for maintaining the security of the Internet of Things (IoT). The traffic from IoT devices differs from typical internet traffic and follows distinct communication patterns. To deal with the complexity of IoT infrastructure, methods based on machine learning have become the primary method for anomaly detection, owing to their capabilities in automated detection, strong environmental adaptability, and handling of large-scale data. However, the existing machine learning-based anomaly detection models generally suffer from two issues: Firstly, they ignore the complex evolution characteristics of anomaly information, resulting in a lack of modeling capability of spatiotemporal dependencies within the information. Secondly, the presence of noisy data and redundant features often bring problems such as large calculation, instability and poor accuracy to the model's training. To solve these challenges. This paper proposes a Spatiotemporal Graph Attention Network (STGAT) anomaly detection method based on Dynamic Traffic Graph (DTG). This method uses the concept of DTG to construct dynamic graph snapshots, which contain temporal correlation from network traffic data, and analyzes the spatiotemporal evolution pattern of network node traffic; Then, STGAT is used to analyze the traffic communication of network nodes from a dynamic graph perspective, achieving the detection of abnormal nodes; In addition, we propose an adaptive feature selection mechanism for collaborative detection model training to reduce noise and redundant features, which in turn improves the stability of model training and the accuracy of detection. The experimental results on three network security benchmark datasets, the UNSW-NB15, CICIDS-2017, and CICIDS-001, show that compared to traditional anomaly detection models, our model has significant improvements in performance indicators such as False Alarm Rate (FAR) and F1 score. Moreover, the collaborative learning model adaptively filters out feature subsets that are most closely related to the supervisory task, thereby enabling more accurate detection of IoT anomalies.

Key words trustworthy artificial intelligence; network security; graph neural networks; anomaly detection; adaptive feature selection

通讯作者: 杨丽平, 博士后, Email: liping.yang@dlmu.edu.cn。

本课题得到国家自然科学基金项目(No: 62176036, No. 6177210)资助。

收稿日期: 2023-06-28; 修改日期: 2024-03-19; 定稿日期: 2025-08-01

1 引言

随着物联网与智能互联技术的快速发展,网络节点之间的交互与协同变得更加复杂。同时,威胁网络安全的异常与攻击形式也更加多样化,如僵尸网络攻击、分布式拒绝服务攻击(Distributed Denial of Service Attack, DDoS)等。网络威胁的复杂化与规模的扩大化,增加了复杂、异构网络系统所面临的安全风险,如工业物联网^[1]、车辆网络和无人机自组网^[2]等。如何及时地检测出这些复杂的异常行为,避免其造成更严重的损失,已经成为一个非常重要的问题。

入侵检测系统(Intrusion Detection System, IDS),已普遍应用于网络安全领域来检测复杂的异常行为。由于基于签名的入侵检测系统,只能检测已经在数据库中保存签名的异常,所以相关领域研究人员又进一步提出了基于异常的实施方案,理想情况下,基于异常的IDS能够检测到未知的攻击^[3]。传统的异常检测方案主要是一些浅层网络模型,如支持向量机、K邻近算法、逻辑回归、决策树、聚类方法以及混合方法等。这些方法常依赖于先验知识,对复杂的攻击模式很难有效,难以解决在真实应用环境中出现的大规模异常流量数据的情况。由此催生了更多基于深度学习的模型,如深度神经网络(Deep Neural Network, DNN)^[4]、深度置信网络(Deep Belief Network, DBN)^[5]、卷积神经网络(Convolutional Neural Network, CNN)^[6]和递归神经网络(Recurrent Neural Network, RNN)^[7]等。深度学习模型可以直接从原始数据中学习特征,实现端到端的学习,特别是对于大型数据集来说,深度学习方法优势明显。然而,目前基于深度学习的方法只利用数据包中的信息(如数据包长度)或流量的统计信息(如总数据包大小、流字节率等),却没有考虑建模网络节点之间的通信关系。节点之间的通信关系往往对检测复杂的攻击模式至关重要,例如DDoS攻击、僵尸网络^[8]攻击和Sybil攻击等^[9],都具有明显的时空特征。如何了解网络节点之间的拓扑结构和流量的时序演化规律,揭示其潜在的安全风险以实现异常节点的检测,已经成为网络异常检测任务的一个难点。基于图神经网络的异常检测技术,因其处理非欧几里得数据的普适性和高效性,已经成为一种很有前途的方法并逐渐得到了学术界和工业界的广泛关注^[10]。

目前基于图神经网络的方法仍面临众多挑战。例如,在某些情况下异常节点在单个图快照中可能与正常节点表现相同,在时间维度上的动态变化却能彰显出异常节点的不同。另一个挑战是,虽然网络

流量数据包括众多特征,但并不是所有特征都是同等重要的,特征之间可能还存在冗余问题^[11],数据的复杂程度与输入输出需求都会显著影响模型的检测性能。特征选择技术已经成为入侵检测领域研究的基础性任务^[12],为了应对这些挑战,Yin等人^[12]提出了一种使用去噪自编码器(Denoising Auto Encoder, DAE)特征选择的流量异常检测方法,通过最小化含噪声数据的重构向量与原始输入向量的差异,筛选出有较强鲁棒性的流量特征。Liu等人^[13]提出一种多维优化的入侵检测方法,在横向对多数类进行抽样,在纵向通过计算特征和类别的相关性筛选特征。本文提出一种基于动态流量图的时空图注意力网络(Spatio Temporal Graph Attention Network, STGAT)的异常节点检测方法。STGAT作为一种动态图神经网络,可以同时捕获网络节点之间的通信拓扑特征以及节点流量的时序演化特征^[14-15],并分析节点流量通信的全局时空演化模式,实现对异常节点的检测。此外,受注意力机制的启发,该模型引入一种协同学习模型训练的自适应特征选择机制,通过神经网络为每个特征模拟注意力权重,表示特征的重要性,在训练的过程中实现自适应的特征选择。本文工作主要贡献如下:

(1) 构建包含时序关联的动态流量图。将时间流量数据按照固定的时间间隔划分为一系列带有时间戳的离散图快照,使之同时包含网络节点的通信拓扑结构信息以及节点流量的时序演化信息,从而将网络节点的流量通信信息建模为动态图的形式。

(2) 提出一种结合时序注意力网络的动态图异常检测模型STGAT,同时捕获网络节点之间的拓扑结构特征和节点流量的时序演化特征,并以动态图的视角从网络节点的流量通信演化中分析异常的产生,检测节点异常。

(3) 提出一种自适应的特征选择机制,协同学习模型一起训练特征权重矩阵,通过参数的反向传播更新权重,根据检测任务实现自适应的特征选择,优化模型的性能。

2 相关工作

本节介绍了相关的工作,组织如下:2.1节介绍基于序列数据的异常检测方法在网络安全领域的应用。2.2节介绍了从序列数据中构建图数据的技术,并介绍基于图神经网络的异常检测方法。2.3节分析了特征选择技术在网络安全领域中的应用。

2.1 基于序列数据的异常检测

基于序列数据的异常检测通常将网络流量视为

具有时序关联的序列数据, 通过分析异常数据的时序差异模式来实现异常检测。如 Muder 等人^[7]提出了一个由增强版的反向传播算法训练的多层递归神经网络(RNN), 每层网络自适应地调整到不同的参数/超参数, 以增强对特定入侵类型的检测。Shu 等人^[16]提出了一个协作入侵检测系统(Collaborative Intrusion Detection System, CIDS)。利用深度学习与生成对抗网络, 为 VANET 设计协同入侵检测系统(CIDS)。Nguyen 等人^[6]通过遗传算法(GA)和模糊 C 均值聚类(Fuzzy C-means Clustering, FCM)处理以获得高质量特征集, 通过 CNN 和装袋分类器(Bagging Classifier)的混合学习方法显著提高了检测性能。Kanna 等人^[33]提出了一种高准确性的 IDS 模型, 使用经过优化的 CNN(OCNN)和分层多尺度 LSTM(HMLSTM)的统一模型, 实现了空间-时间特征的提取与学习。这些传统的异常检测方法都缺乏对图数据的建模能力。

2.2 基于图数据的异常检测

基于图的异常检测模型, 本质上说, 是将异常检测的焦点从对网络节点的链接, 转移到对网络节点的行为上^[17]。通常将网络流量转化为图结构来表示网络节点的行为。近年来, 相关领域的研究人员提出了多种提取图快照的方法, 如流量分散图(Traffic Dispersion Graph, TDG)^[17]、流量交互图(Traffic Interaction Graph, TIG)^[18]、端点流量图(Endpoint Traffic Graph, ETG)^[19]等。同时, 在过去的几年中, 已经有一些工作对基于图节点行为分析的异常检测任务进行了深入的研究。

如 Zhao 等人^[20]提出了一种协同方法, 通过模式挖掘的技术告知 GNN, 如何通过网络链接聚合局部信息以学习到全局的模式, 实现基于图的异常检测。Wang 等人^[21]提出了一种基于对比的 GNN 的流量异常分析方法, 适用于基于 IoT 智能交通系统的不平衡数据集。Zola 等人^[17]提出了一种基于 GNN 的多步异常节点检测方法。使用两种数据预处理技术 R-hybrid 和 SM-hybrid 平衡图快照数据集, 然后通过多层的图神经网络模型实现对异常节点的分类。虽然这些基于图神经网络的方法考虑了利用图快照的空间特征进行异常检测, 但没有考虑建模图快照之间的时序关联。

2.3 基于特征选择的异常检测

基于特征选择的异常检测方法, 本质是通过减少网络流量特征的数量, 生成最优特征子集。不仅减少了处理、存储、传输数据和快速处理的资源消耗, 还降低了数据特征提取的难度^[24]。网络流量数据往

往由多维特征组成的。如 CICFlowMeter 工具提取的 CICIDS-2017 数据集^[32]有 80 维特征, 而只有少部分特征与学习任务有相关性。特征数较大时, 特征子集组合难以处理^[22]。Ren 等人^[23]提出一种高效的数据优化方法, 采用遗传算法(GA)作为搜索策略来指定候选特征子集, 并将随机森林算法(RF)的分类作为候选特征子集的评价度量。Rashid 等人^[11]介绍了一种用于多元相关的网络异常检测的特征选择方法, 通过主成分分析(PCA)和 Pearson 类标签相关性来选择特征, 提高了没有足够样本生成特征权值的数据集的选择精度。Binbusayyis 等人^[25]评估了四种基于过滤的特征选择方法对网络异常检测任务的优劣。目前异常检测领域缺少自适应的特征选择组件, 由此本文提出了一种混合上下文信息, 协同学习模型一起训练的特征选择机制。

3 基于时空图注意力的异常检测模型

本节将详细介绍基于时空图注意力的异常检测模型, 从动态图的角度分析网络节点的流量时空演化特征, 捕捉复杂攻击模式。模型整体框架如图 1 所示, 主要包含两个核心组件: 一是协同学习模型训练的自适应特征选择模块, 二是基于时空图注意力网络的异常节点检测模块。特征选择模块以图快照的特征矩阵作为输入, 通过基于注意力机制的神经网络产生每个特征的选择权重, 并依据检测任务自适应地选择特征; 异常节点检测模块通过时空图注意力网络 STGAT 抽取流量图快照的时空演化特征, 并根据特征向量实现节点分类。

3.1 动态流量图构建

动态流量图(Dynamic traffic graph, DTG), 是从网络流量数据中提取的包含时空演化信息的图快照序列。网络流量数据的部分示例如图 2 所示。每一条流量数据由{源节点 IP, 目标节点 IP, 流量特征}3 元组表示, 其中每个 IP 地址表示为一个节点, 每条流量数据可看作一组节点之间的通信交互, 有通信的节点之间生成一条边, 某时间间隔内包含流量时空演化信息的图快照可定义为 $G_t = (V_t, A_t)$, 其中 $V_t = \{v_1^t, v_2^t, \dots, v_N^t\} \in R^N$, 表示由 N 个节点组成的顶点集; $A_t = \{a_{ij}^t \mid i, j = 1, 2, \dots, N\} \in R^{N \times F}$, 表示节点链接关系的邻接矩阵, a_{ij}^t 为节点 v_i^t 和 v_j^t 之间的出入度的和。动态流量图由一系列带有时间戳 t 的离散图快照序列组成, 图快照序列如图 3 所示, 黑色为异常边, 灰色为正常边。

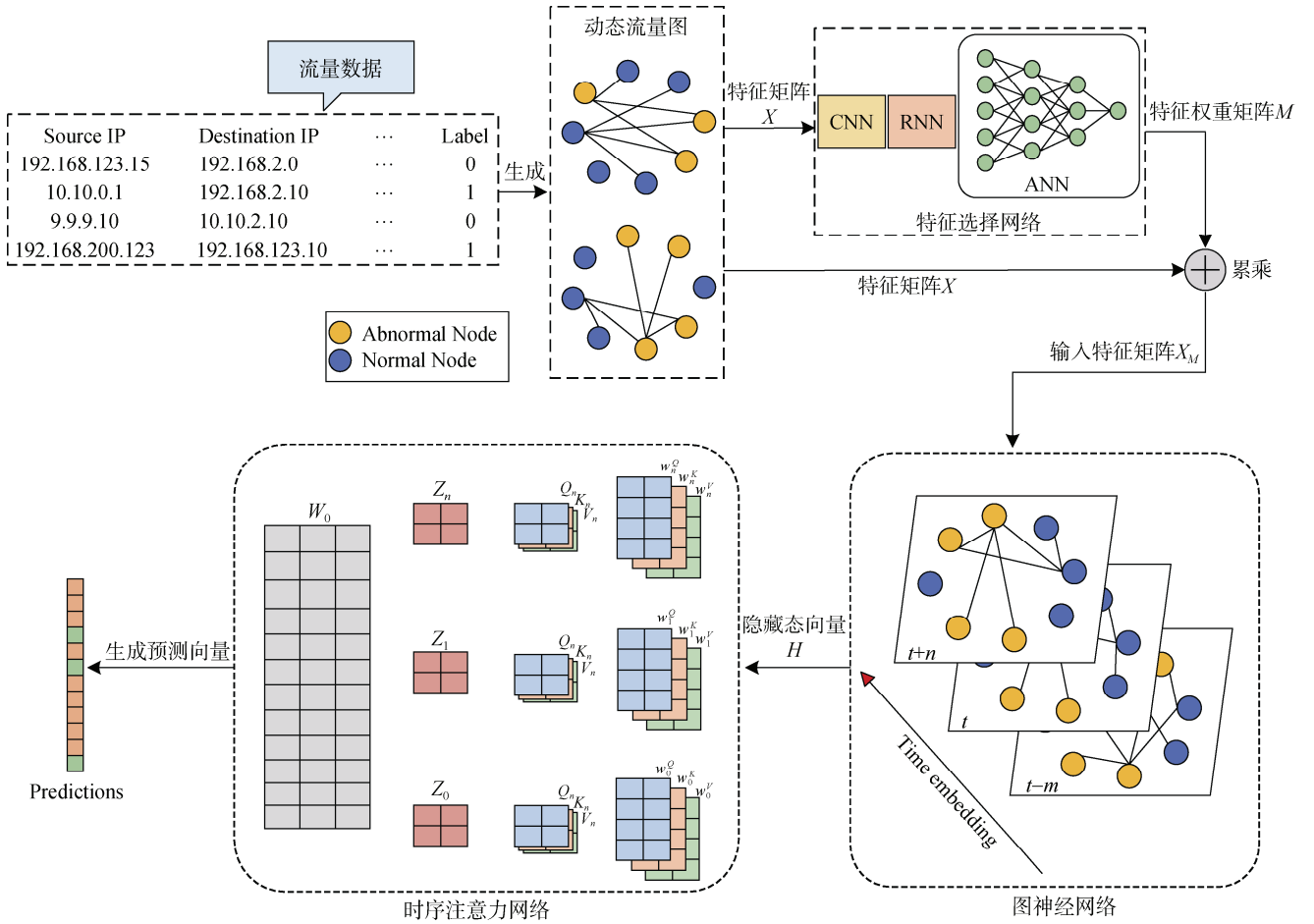


图 1 异常检测模型整体框架图

Figure 1 Overall framework diagram of the anomaly detection model

Source IP	Destination IP	Protocol	Stime	...	During	Label
175.45.176.0	149.171.126.16	tcp	14261926	...	0.121475	0
175.45.176.3	149.171.126.18	udp	14241919	...	1.681642	1

Source Entity

Destination Entity

Feature

图 2 网络流量数据示例

Figure 2 Example of network traffic data

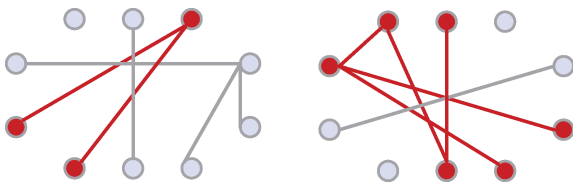


图 3 网络图快照

Figure 3 Network graph snapshot

之后, 输入网络流量数据, 统计 IP 获得图快照的邻接矩阵 A_t , 聚合流量特征获得图快照的特征矩阵 X_t , 统计流量的 Label 获得图快照的标签矩阵 Y_t 。

具体地说, 遍历所有 IP 生成对应的节点编号 n ,

将流量数据按照流量特征中的时间 Stime 排序, 利用源 IP 与目标 IP 查询对应的节点编号 i, j 将对应的 a_{ij}^t 加 1, 循环遍历每条流量数据获得图快照的邻接矩阵 A_t 。然后, 由 F 维的流量特征的加权平均生成每个节点 v_i^t 的特征 $x_i^t \in \mathbb{R}^F$, 则 t 时刻图快照 G_t 的特征矩阵 X_t 为 $X_t = \{x_1^t, x_2^t, \dots, x_N^t\} \in \mathbb{R}^{N \times F}$ 。接着, 通过统计流量特征中的 Label 来获得节点的标签 y_i^t , Label=0 表示正常节点, Label=1 表示异常节点。则图快照的标签矩阵为 $Y_t = \{y_1^t, y_2^t, \dots, y_N^t\} \in \mathbb{R}^N$, 其中 $y_i^t \in \{0, 1\}$ 。最后, 时间间隔为 K 的流量数据转化为一个图快照数据 G_t , 将流量数据抽取为图快照序列 $G_1, G_2, \dots, G_T$,

其中 T 表示时间步长, 每个图快照 G_t 都分别包含节点邻接关系的邻接矩阵 A_t , 节点行为特征的特征矩阵 X_t , 以及节点分类的标签矩阵 Y_t 。

3.2 自适应特征选择

自适应特征选择机制通过特征选择网络求解特征相对于分类任务的相关性, 作为特征选择的评价指标。生成表示每条特征权重的特征向量, 合为特征的权重矩阵, 在训练过程中通过参数反向传播不断修正该权重, 最后通过权重筛选出最优特征子集。自适应特征选择模块如图 4 所示。

具体地, 将图快照序列的特征矩阵 $\{X_1, X_2, \dots, X_T\} \in \mathbb{R}^{N \times F \times T}$ 作为输入, 其中 N 为节点数, F 为特征维度, T 为时间步长, 通过 CNN 网络提取特征之间的抽象关系, 生成特征向量 E , 计算方式如式(1)所示。

$$E = \text{CNN}(X) \quad (1)$$

其中, X 表示每一个时刻的图快照的特征矩阵, $E \in \mathbb{R}^{N \times T \times 32}$ 上述过程利用卷积将低维的特征向量映射到一个更高维的特征空间中, 并且通过压缩特征数量, 剔除噪声和一些冗余特征对权重生成的影响, 本文共使用了 32 个一维卷积以实现特征的映射。流量数据包含时序关系, 只抽取特征间的局部

依赖关系并不能为权重生成网络提供充足的判断依据, 因此, 通过 RNN 网络抽取特征的时序依赖。为每个特征生成一个 r_c 向量, 得到包含时间依赖的特征向量 $R_c = \{r_c^1, r_c^2, \dots, r_c^F\}$, 详细计算过程如式(2)所示:

$$r_c = \text{RNN}([r_{c-1}, E]) \quad (2)$$

其中, c 是 RNN 隐藏层的个数, r_{c-1} 是上一层的隐藏态向量, 本文采用 5 层 RNN, 初始 r_0 为与输入 E 同纬度的全 0 矩阵。为了解决特征权重随着训练过程差值无限制增大的情况^[26], 将每个特征的 r_c 向量的输出值建模两个概率值 p 和 q , 每次对这两个概率权重作 softmax 归一化, 由此生成的特征选择权重可以维持在(0,1)。概率值 p, q 的计算过程为

$$\begin{aligned} p^f &= w_p^f r_c^f + b_p^f \\ q^f &= w_q^f r_c^f + b_q^f \end{aligned} \quad (3)$$

其中, r_c^f 是上一步提取的隐藏层向量。 w^f, b^f 代表概率 p 和 q 的待学习参数。用 m^f 表示特征的选择权重, 那么每个特征 f 的权重为

$$m^f = \frac{\exp(p^f)}{\exp(p^f) + \exp(q^f)} \quad (4)$$

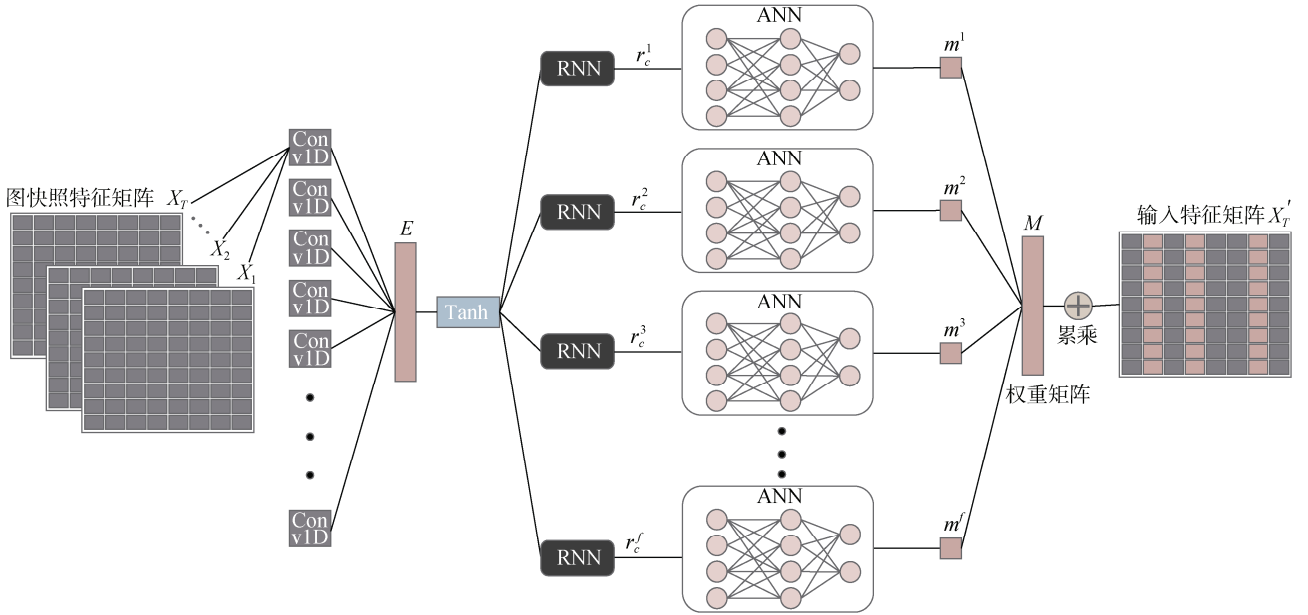


图 4 特征选择网络

Figure 4 Feature selection network

最终, 所有特征的权重共同组成权重矩阵 $M = \{m_n^f | n=1, 2, \dots, N; f=1, 2, \dots, F\} \in \mathbb{R}^{N \times F}$ 。将 M 与特征矩阵 X 累乘, 得到输入特征矩阵 X_M 。通过算法 2 得到最优特征子集。详细步骤如算法 1 所示。

算法 1 自适应特征选择

输入: 图快照序列的特征矩阵 $\{X_1, X_2, \dots, X_T\}$, 时间步长 T , 特征维度 F
输出: 特征选择的最优特征子集 L

```

1: FUNCTION FS( $\{X_1, X_2, \dots, X_T\}$ )
2:   FOR  $f$  in  $F$  do
3:      $E \leftarrow \text{CNN}(X_f)$ ;
4:      $r_c^f \leftarrow \text{RNN}(E)$ ;
5:      $m^f \leftarrow \text{softmax}(\text{ANN}(r_c^f))$ ;
6:   END FOR
7:   RETURN  $M$ 
8: END FUNCTION
9:  $X_M \leftarrow X \cdot M$ ;
10:  $W \leftarrow \text{STGAT}(X_M)$ (算法 2);
11:  $L \leftarrow \max(W)$ ;
12: 输出特征选择的最优特征子集  $L$ 

```

3.3 基于时空图的异常检测

异常检测模型由两个神经网络构成, 分别是注意力网络(GATv2)^[27]和时序注意力网络(AT)。模型以图特征矩阵 $\{X_1, X_2, \dots, X_T\} \in \mathbb{R}^{N \times F \times T}$ 和图邻接矩阵 $\{A_1, A_2, \dots, A_T\} \in \mathbb{R}^{N \times F \times T}$ 作为输入, 通过图注意力网络提取每个图快照中节点的拓扑关系特征, 嵌入到特定的特征向量 g_i 中, 计算过程如式(5)所示:

$$g_i \leftarrow \text{GATv2}(X_i, A_i) \quad (5)$$

详细计算过程如式(6)~(8)所示:

$$e(h_i, h_j) = a^T \text{LeakyReLU}(W \cdot [h_i \| h_j]) \quad (6)$$

$$\alpha_{ij} = \text{softmax}_j(e(h_i, h_j)) \quad (7)$$

$$g_i \leftarrow \sigma \left(\sum_{j \in N_i} \alpha_{ij} \cdot W h_j \right) \quad (8)$$

通过特定的特征向量 g_i 获得包含 T 个时间步长的图嵌入向量 $\{g_1, g_2, \dots, g_T\}$ 。接着, 为每个向量嵌入位置编码信息, 得到含有时序信息的隐藏 $S = \{s_i; i \in \mathbb{R}^T\}$, 定义图快照之间的时序关系。然后, 通过时序注意力网络提取出每个图快照中节点的流量时序演化特征 z_i 。

时序注意力网络由一个 Encoder 组成, Encoder 对完整的序列进行编码, 产生每个时间戳的注意力分数, 通过对每个图快照数据 s_i 构建三个矩阵 $Q(\text{query})$ 、 $K(\text{key})$ 、 $V(\text{value})$, 注意力计算过程如式(9)所示:

$$z_i \leftarrow \text{AT}(s_i) \quad (9)$$

详细计算过程如式(10)~式(12)所示, 这里 softmax 激活函数用来归一化注意力权值, 允许我们将矩阵 V 压缩成一个更小的代表性嵌入, 以简化下游神经网络操作中的推理。缩放点积注意通过 $\sqrt{m}$ 来缩放权重, 以减少在训练过程中权重的方差, 促进

稳定的训练。接着我们应用多头自注意^[28]将原始的 Q 、 K 、 V 矩阵映射到不同的子空间, 分别学习不同的注意力关系力特征, 进而得到 Q_i 、 K_i 、 $V_i, i \in \{1, 2, \dots, h\}$, h 是多头注意力的头数, 注意力表示如下公式:

$$\text{Attention}(Q, K, V) = \text{softmax} \left(\frac{QK}{\sqrt{m}} \right) V \quad (10)$$

$$z_i = \text{Attention}(Q_i, K_i, V_i) \quad (11)$$

$$Z = \text{Concat}\{z_1, z_2, \dots, z_h\} \quad (12)$$

则图快照被转化为同时包含节点间拓扑结构信息和节点流量时序变化信息的向量 $Z = \{z_i | i \in \mathbb{R}^T\}$, 然后以此向量为依据, 计算与 $Y = \{y_i | i \in \mathbb{R}^T\}$ 标签的损失训练模型, 实现对异常节点的检测。此外, 针对类别不平衡问题, 采用 Focalloss^[29]损失函数进行优化, 以减少对分类结果的影响。损失计算如式(7)所示:

$$\text{FL}(p_t) = -\alpha_t (1 - p_t)^\gamma \log(p_t) \quad (13)$$

其中

$$\alpha_t = \begin{cases} \alpha, & \text{if } y = 1 \\ 1 - \alpha, & \text{if } y = 0 \end{cases} \quad (14)$$

$$p_t = \begin{cases} p, & \text{if } y = 1 \\ 1 - p, & \text{if } y = 0 \end{cases}$$

$(1 - p_t)^\gamma$ 为调节因子, p_t 是模型预测的结果的类别概率值, γ 是超参数, 平滑地调整了简单样本的降权比例。当 $\gamma=0$ 时, Focalloss 与交叉熵一致, 随着 γ 增加, 调节因子的影响也相应增加。异常节点检测模型的详细训练过程如算法 2 所示。

算法 2 异常节点检测算法(STGAT)

输入: 图快照序列的邻接矩阵 $\{A_1, A_2, \dots, A_T\}$, 图快照序列的特征矩阵 $\{X_1, X_2, \dots, X_T\}$, 图快照序列的标签矩阵 $\{Y_1, Y_2, \dots, Y_T\}$, 时间序列编码 $\{s_1, s_2, \dots, s_T\}$, 时间步长 T

输出: 模型训练结果 θ

```

1: 初始化模型的参数  $\theta$ ;
2: FOR  $i$  in  $T$  do
3:    $g_i \leftarrow \text{GATv2}(X_i, A_i)$ ;
4:    $s_i = \text{Positional\_encoding}(g_i)$ ;
5:    $z_i \leftarrow \text{AT}(s_i)$ ;
6:    $L = \text{Focalloss}(z_i, Y_i)$ ;
7:   根据梯度反向传播更新模型参数  $\theta$ ;
8: END FOR
9: RETURN  $\theta$ 

```

4 实验与结果分析

4.1 实验配置

本节详细描述了数据集的来源、数据预处理的过程以及用于实验评估的评价指标。

4.1.1 数据集

(1) UNSW-NB15 数据集^[30]弥补了网络流量数据集不真实以及过时的缺点, 现在已经成为网络安全领域最常用的基准数据集之一。数据集大致可以分为未处理的原始数据和处理过的训练数据。为了更加真实地模拟应用环境, 本文选择反映网络流量真实规律的原始数据集。

(2) CIDDS-001 数据集^[31]通过 Open-Stack 软件生成的物联网模拟数据集, 模拟了小型工业应用环境, 包括多台服务器、主机和一些硬件设备。此外, 为了保证数据集的真实性, 在其网卡上记录了外部服务器的网络流量, 并将其合并到其他流量中, 其中正常和异常的用户行为都是通过模拟生成, 数据集共包含四周的流量数据。

(3) CICIDS-2017 数据集^[32]是基于 Netflow 的仿真数据集, 数据集共有 80 个特征。模拟的网络中包括一个用户分析代理和 12 台不同的受害者主机以及一个攻击者, 数据集中涵盖了包括 Web Attack, Brute force, DoS, DDoS, Infiltration, Heart-bleed, Bot 和 Scan 等多种攻击类型。

4.1.2 数据预处理

数据预处理主要包括特征转换、特征平衡化和数据归一化三个步骤。特征转换将字符数据转换为数值类型; 特征平衡化通过筛选攻击节点与相关节点的方式, 在保持流量数据时间序列关系的同时平衡数据; 在训练前, 为加速模型训练的速度, 本文采用 min-max 数据归一化方法对数据集进行了归一化。具体如式(8)所示:

$$X_{\text{normalization}} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (15)$$

4.1.3 图快照抽取

如前所述, 在实验之前, 需要从网络流量数据集中提取动态流量图。第一个操作是将数据集以特定时间间隔切片。根据三个数据集背景采样时间的不同, 分布设置了不同的时间间隔, 在 CIDDS-001 数据集中是 30 秒、50 秒、100 秒; 在 UNSW-NB15 数据集中是 80 秒、300 秒、700 秒和 1000 秒; 在 CICIDS-2017 数据集中是 500 秒、1000 秒、1500 秒。如表 1 所示, 在不同的数据集中以不同的时间间隔切片, 获得的图快照的快照数量也不同。如 UNSW-NB15 数据集以 300 秒、700 秒和 1000 秒切片分别产

生 141、62 和 44 个图快照数据。为了避免生成大量仅包含正常节点的图快照, 以及完全空节点的图快照(没有任何节点), 时间间隔不宜设置过大或过小。且因为原始数据集没有作平衡处理, 攻击的发生往往集中在某一时刻或某一段时间。为了数据类别平衡, 仅对发生攻击行为的节点, 及被攻击节点编号, 通过这种方法, 我们可以删掉大量与异常信息无关的节点。初步平衡数据集。如表 1 所示。

4.1.4 评价指标

异常检测是一个分类问题, 准确率(Accuracy)、召回率(Recall)、误报率(False Alarm Rate, FAR)、精度(Precision)和 F1-Score 等 5 个指标已被广泛应用于评价分类模型的性能。此外, 为了对比特征选择的结果, 引入了 Kappa 系数以求更精确地评价模型分类精度。如表 2 所示, 其中 p_o 是每一类正确分类的样本数量之和除以总样本数, 也就是总体分类精度; p_e 是各类真实样本个数与预测出来的样本个数的乘积之和, 再除以总样本个数的平方。

表 1 不同时间间隔产生不同图快照数据

Table 1 Generated different graph snapshot dataset at different time intervals

数据集	时间间隔	数量	正常	异常	比例
CIDDS-001	30	700	1417	1411	1:1
	50	424	851	845	1:1
	100	212	395	453	0.87:1
UNSW-NB15	80	524	1724	23952	0.07:1
	300	142	469	6440	0.07:1
	700	62	200	2838	0.07:1
	1000	44	142	2014	0.07:1
	500	360	616	1544	0.40:1
CICIDS-2017	1000	180	326	754	0.43:1
	1500	120	224	496	0.45:1

表 2 评价指标

Table 2 Evaluation metrics

评价指标	定义
Accuracy	$\frac{TP + TN}{TP + FP + TN + FN}$
Precision	$\frac{TP}{TP + FP}$
Recall	$\frac{TP}{TP + FN}$
False Alarm Rate(FAR)	$\frac{FP}{FP + TN}$
F1-Score	$2 \times \frac{\text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}}$
Kappa	$\frac{p_o - p_e}{1 - p_e}$

4.2 实验结果

本节设计实验分别对模型的各个模块进行评估。首先, 在数据处理过程, 给出流量数据集不同时间间隔切片的对比实验, 评估图快照的时间间隔对模型性能的影响。然后, 比较本文的方法与目前机器学习模型在性能上的差异, 评估本文的方法区分正常和异常网络流的能力; 并通过消融实验, 验证时间注意力模块的效果。其次, 分析不同特征子集对模型效果的影响, 通过比较原始特征全集和特征选择的最优子集的性能差异, 验证本文提出的自适应特征选择机制的有效性。

4.2.1 实验一: 比较不同间隔的图快照

由于流量数据的局部性, 在某时间间隔内攻击可能不发生或密集发生。不合理的切片间隔, 有可能产生仅有正例节点或空节点的图快照^[17], 这种数据会严重降低检测的性能。由此, 本实验以时间间隔对流量数据切片以评估其对模型性能的影响。分别为三个数据集设置不同的时间间隔, CIDDs-001 数据集是 30 秒、50 秒、100 秒, 在 UNSW-NB15 数据集是 80 秒、300 秒、1000 秒, 在 CICIDS-2017 数据集是 500 秒、1000 秒、1500 秒。选择每个数据集的 70% 的流量样本进行训练, 保留 30% 用于测试。为了实验

公平性, 统一使用图注意力网络(GATv2)在不同数据集上进行实验。实验结果如表 3 所示。

实验结果显示, 使用不同切片间隔产生的图快照数据会对模型效果产生不同的影响。在 CIDDs-001 数据集中, 模型性能在 30 秒和 100 秒间隔时取得了相似的表现, 而以 50 秒为间隔时, 模型的性能却产生了大幅度地下降。在 UNSW-NB15 数据集中, 评估模型的各项指标在 300 秒间隔时达到最佳, 随着时间间隔的扩大, 训练效果逐渐下降; 而在 CICIDS-2017 数据集中, 训练效果随着时间间隔的扩大而提升, 并在 1000~1500 秒间隔保持稳定的结果。这表明提取图快照的时间间隔并非在最大或最小处取得最优, 需根据攻击数据分布差异合理设置。这是因为不同的时间间隔产生的正负样本的比例不同。流量数据具有一定的局部规律, 过长的间隔切片, 可能导致攻击流量特征和正常流量的特征在多次平均之后过于平滑, 过短的时间间隔切片则不能表达节点的统计规律, 影响分类效果。

4.2.2 实验二: 传统的 IDS 的网络安全方法

为了验证本文方法对流量分析任务的有效性, 我们借鉴 Zola 等人^[17]的方式, 比较展示了目前基于机器学习模型的入侵检测方法。在这些研究中, 普遍

表 3 不同时间间隔分类效果
Table 3 Classification effects performance on different intervals

数据集	时间间隔(s)	Accuracy	Recall	Precision	AUC	F1-Score	FAR
CIDDs-001	30	0.86±0.02	0.86±0.02	0.86±0.01	0.90	0.86±0.02	0.13±0.02
	50	0.67±0.05	0.71±0.02	0.55±0.06	0.72	0.64±0.02	0.45±0.03
	100	0.82±0.02	0.80±0.03	0.81±0.03	0.86	0.81±0.02	0.17±0.03
UNSW-NB15	80	0.96±0.01	0.80±0.05	0.74±0.2	0.96	0.76±0.03	0.25±0.03
	300	0.97±0.01	0.86±0.03	0.83±0.01	0.97	0.82±0.01	0.17±0.05
	700	0.95±0.01	0.75±0.01	0.32±0.1	0.83	0.45±0.20	0.58±0.3
CICIDS-2017	1000	0.95±0.01	0.73±0.03	0.43±0.06	0.86	0.54±0.23	0.50±0.15
	500	0.82±0.02	0.65±0.05	0.55±0.03	0.75	0.60±0.05	0.44±0.03
	1000	0.88±0.01	0.81±0.03	0.88±0.04	0.96	0.84±0.01	0.11±0.05
	1500	0.89±0.01	0.85	0.75±0.01	0.88	0.80±0.02	0.24±0.01

表 4 IDS 异常流量分析模型对比
Table 4 Comparison of IDS abnormal traffic analysis models

数据集方法	UNSW-NB15			CICIDS-2017			CIDDs-001		
	Recall	Precision	F1-Score	Recall	Precision	F1-Score	Recall	Precision	F1-Score
DNN ^[4]	0.66	0.86	0.75	0.99	0.97	0.98	0.96	0.93	0.94
RBM ^[5]	0.73	0.70	0.69	0.95	0.93	0.94	0.70	0.83	0.75
CNN ^[6]	0.75	0.85	0.65	0.97	0.96	0.99	0.95	0.87	0.86
RNN ^[7]	0.82	0.81	0.73	0.92	0.99	0.87	0.82	0.81	0.73
CNN+LSTM ^[33]	0.95	0.80	0.77	0.89	0.98	0.90	0.97	0.96	0.97
STGAT(ours)	0.88	0.98	0.92	0.80	0.89	0.85	0.89	0.85	0.87

将网络流量数据作为具有时间序列的数据进行分析, 通过提取流量的统计规律判断模式差异检测异常流量, 维护网络安全。本文的方法与 IDS 的方法类似, 都考虑以时序的方式分析网络流量, 解析异常模式, 且与仅基于时序的流量分析模型相比, STGAT 还具有图结构特征的分析能力。

在网络安全领域, 使用的数据集通常是不平衡的, 所以实验以 Recall、Precision、F1-Score 作为评价模型性能的指标, 区分模型的性能。所有的方法都使用原始的网络安全数据集进行训练与测试, 实验结果如表 4 所示。

结果显示, 有监督的深度学习模型方法能表现出更优秀的性能。相比单一的 CNN 和 RNN, 结合空间模态特征模型和时序模态特征模型特点的 CNN+LSTM^[32]方法具有更优秀的分类性能, 以 UNSW-NB15 数据集的实验结果为例, 召回率从 0.80 左右提升至 0.95, F1-Score 增长至 0.77。此外, STGAT 所获得的分类精度与网络流量进行直接分类所获得的精度非常相似。显然节点分类相比网络流量分类更困难、更复杂^[21], 然而本文的模型在精

度、召回率和 F1-Score 等多数评价指标上仍然优于多数的 IDS 模型。实验结果表明了本文模型对异常流量分析的有效性。

4.2.3 实验三: 异常节点检测的网络安全方法

异常节点检测实验对比了本文模型与其他异常检测模型的优劣。首先, 将图注意力网络和时间注意力网络解耦, 训练了两个模型 GATv2^[27]与 STGAT, 以消融实验验证使时间信息对分类任务的收益。接着与其他两种最先进的异常检测方法 3-GCN^[17]和 GANomaly^[34]对比, 验证本文模型对异常检测任务的有效性。

3-GCN 是 Zola 等人在 2022 年提出的基于双层图卷积神经网络的异常节点检测模型; GANomaly 是基于 GAN 的异常检测模型, 通过对比原始样本与重构样本之间的编码器误差判断异常值。为了保证实验的公平性, 所有的模型都使用统一时间间隔生成的动态流量图快照进行训练。由于 GCN 的计算依赖拉普拉斯矩阵, 统一邻接矩阵为对角矩阵的形式 $A = \{a_{ij} = a_{ji} \mid i, j \in \mathbb{R}^N\}$ 。表 5 展示了各个模型在三个数据集上的性能表现, 实验结果取最优。

表 5 不同异常检测模型的性能对比
Table 5 Performance comparison of different anomaly detection models

方法	评价指标	CICIDS-2017	UNSW-NB15	CIDDS-001
GATv2 ^[27]	Recall	0.81±0.02	0.86±0.05	0.84±0.02
3-GCN ^[17]		0.80±0.01	0.85±0.04	0.83
GANomaly ^[34]		0.80±0.02	0.70±0.03	0.45±0.01
STGAT(ours)		0.87±0.05	0.88±0.01	0.88±0.01
GATv2 ^[27]	Precision	0.88±0.04	0.83±0.04	0.84±0.01
3-GCN ^[17]		0.80±0.02	0.80±0.01	0.82±0.02
GANomaly ^[34]		0.97	0.98	0.88±0.01
STGAT(ours)		0.89±0.04	1±0.03	0.95±0.06
GATv2 ^[27]	F1-Score	0.84±0.05	0.85±0.05	0.84±0.01
3-GCN ^[17]		0.80±0.02	0.82±0.02	0.83±0.02
GANomaly ^[34]		0.83±0.02	0.81±0.01	0.60±0.02
STGAT(ours)		0.89±0.02	0.92±0.01	0.88±0.02
GATv2 ^[27]	FAR	0.11±0.1	0.16±0.03	0.15±0.03
3-GCN ^[17]		0.12±0.3	0.19±0.03	0.16±0.02
GANomaly ^[34]		0.5	0.5	0.7
STGAT(ours)		0.05±0.1	0.01±0.2	0.04±0.03

实验结果表明, 基于图神经网络的模型都可以有效地分类正常与异常行为。三个数据集上的 Precision、Recall 都显著优于其他方法, F1-Score 均高于 0.80, FAR 低于基于 GAN 的异常检测模型。若仅比较基于图神经网络的方法, 在多个数据集上的实验表明 GATv2 模型的性能普遍优于 3-GCN 模型。

在此基础上, 相比单一的 GATv2 模型, 加入时间注意力网络(AT)之后的动态图模型 STGAT 在 Precision、Recall 等指标上都获得了大幅提升, 特别是 F1-Score 和 FAR 有明显改善, F1-Score 在三个数据集上分别提升了 0.08、0.17、0.06, FAR 分别降低了 0.06、0.27、0.11。以上结果证明, 利用节点流量

的时序演化特征有助于提升模型对分类的敏感性,降低模型的 FAR。

4.2.4 实验四: 攻击节点检测

本文的另一个任务目标是检测攻击节点,流量的发送方和接收方都应被明确定义。使用无向边表示节点通信,不能表示发送方和接收方之间的收发关系,需改用单向边定义邻接矩阵 $A = \{a_{ij} \neq a_{ji} \mid i, j \in \mathbf{R}^N\}$ 。规定流量特征只向发送节点聚合,建模流量发送的收发关系。实验结果如表 6 所示。结果表明,面对攻击节点检测任务 STGAT 仍然能取得理想的效果,证

明了本文的模型在时空演化复杂环境异常检测领域中的有效性。

4.2.5 实验五: 比较不同的特征相关性排序

为验证特征选择机制的有效性,在两个数据集上进行了对比实验,其中 UNSW-NB15 数据集包含 45 个特征, CICIDS-2017 数据集包含 75 个数据特征。首先,为不同的数据集生成特征的权重矩阵,将权重矩阵与训练样本聚合,在训练的过程中通过参数反向传播来获得每个特征的权值,然后根据权重形成降序的特征集合。特征权重结果如图 5 所示。

表 6 攻击节点检测基于 STGAT

Table 6 Attack node detection based on STGAT

数据集	Accuracy	Recall	Precision	AUC	F1-Score	FAR
CIDDS-001	0.83±0.02	0.84±0.03	0.79±0.04	0.87	0.81±0.05	0.20±0.10
UNSW-NB15	0.96±0.01	0.85±0.02	0.91±0.01	0.98	0.89±0.01	0.01±0.01

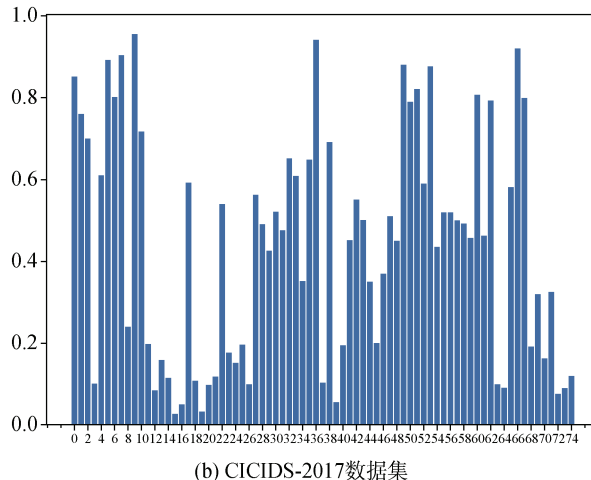
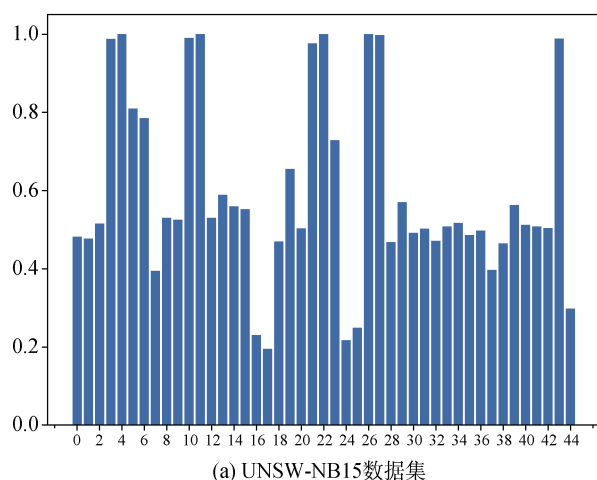


图 5 特征相关性权重

Figure 5 Feature correlation weights

将特征按照权重的数值排序后,就获取了按数据特征与监督目标相关性排序的特征集合。集合中排名靠前的特征,可被认为是对监督任务贡献较大的特征。本实验比较了两种通过计算特征相关性进行特征选择的算法,分别是 Kendall's tau 相关性算法^[11]和基于随机森林分类的算法^[25]。首先,所有算法都对特征进行相关性排序,生成相应的排序集合。之后每次选择一个排序靠前的特征构成新的特征子集,输入异常检测模型训练,验证该特征对模型性能的影响。图 6 和图 7 展示了不同相关性排序算法生成的特征排序集合的质量。可以直观地观察到,所有计算特征相关性构建特征子集的方法,都对训练模型有促进效果。均能以更少的特征数量取得更高的训练指标。如 UNSW-NB15 数据集上的实验显示,仅使用相关

性排序筛选出的前 4 个特征构成特征子集进行训练,异常检测模型的 Accuracy、Precision、Recall、Kappa 系数等指标就已经提升至 0.80 以上。同样的条件下,使用未经过相关性排序的原始特征集合训练的异常检测模型,需要更多特征才能达到相似效果。并且使用排序后的特征集合作为训练数据,模型往往能取得更高的性能峰值,在 UNSW-NB15 数据集上,使用原始与随机森林方法计算相关性排序的质量较差,使用此排序构建的特征集合,模型训练效果都不佳,各个训练指标的峰值较低,且训练的结果更不稳定。

Kendall's tau 相关性算法的训练结果和本文的方法有相似的性能表现。在 CICIDS-2017 数据集上的训练结果显示,所有相关性方法排序后的特征集合的训练模型性能,普遍优于使用未排序的原始特

征集合训练的模型性能。这说明特征选择机制能够筛选出与监督分类任务更具有相关性的特征。此外, 在多个特征排序集合上的训练结果都表明训练性能的递减性, 这说明并不是所有的特征都对模型的训

练效果有正面的促进作用。冗余或与分类任务弱相关的特征可能导致模型分类性能的下降。最终, 以较少的特征数量达到较高分类精度的特征子集就是最优特征子集。

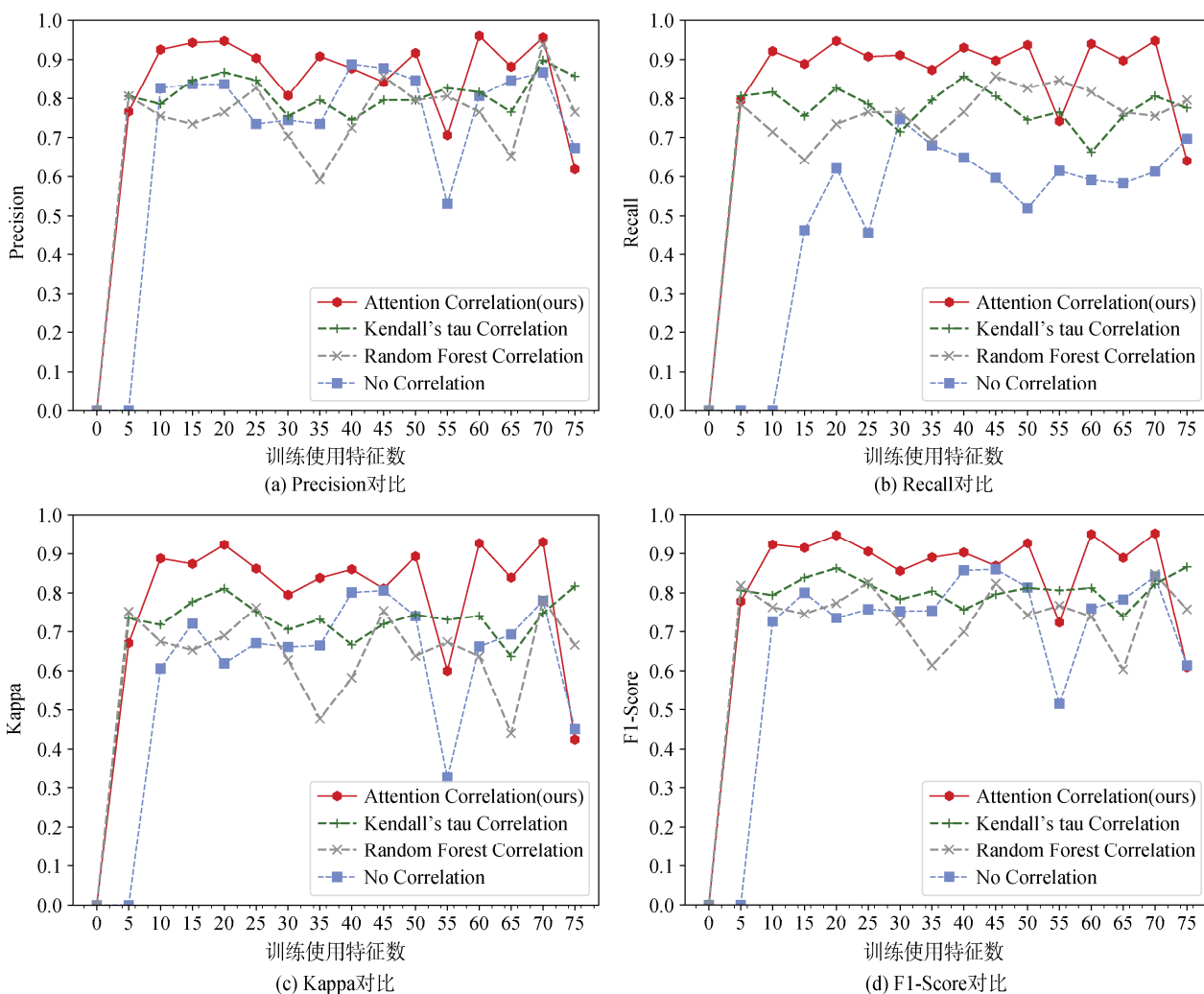
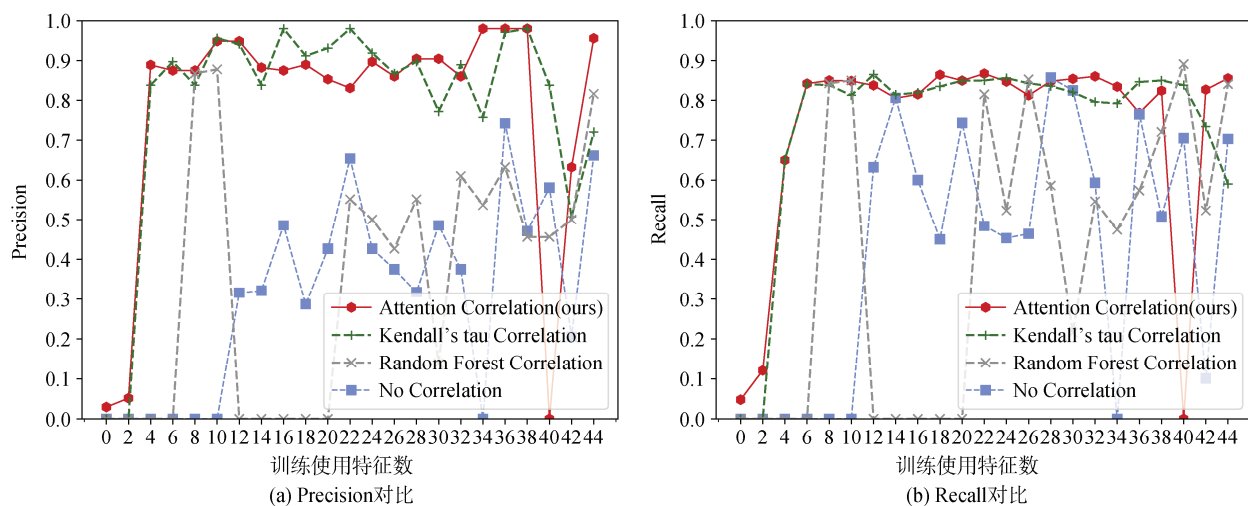


图 6 UNSW-NB15 数据集不同特征子集的训练效果

Figure 6 Training effect performance of different feature subsets based on UNSW-NB15



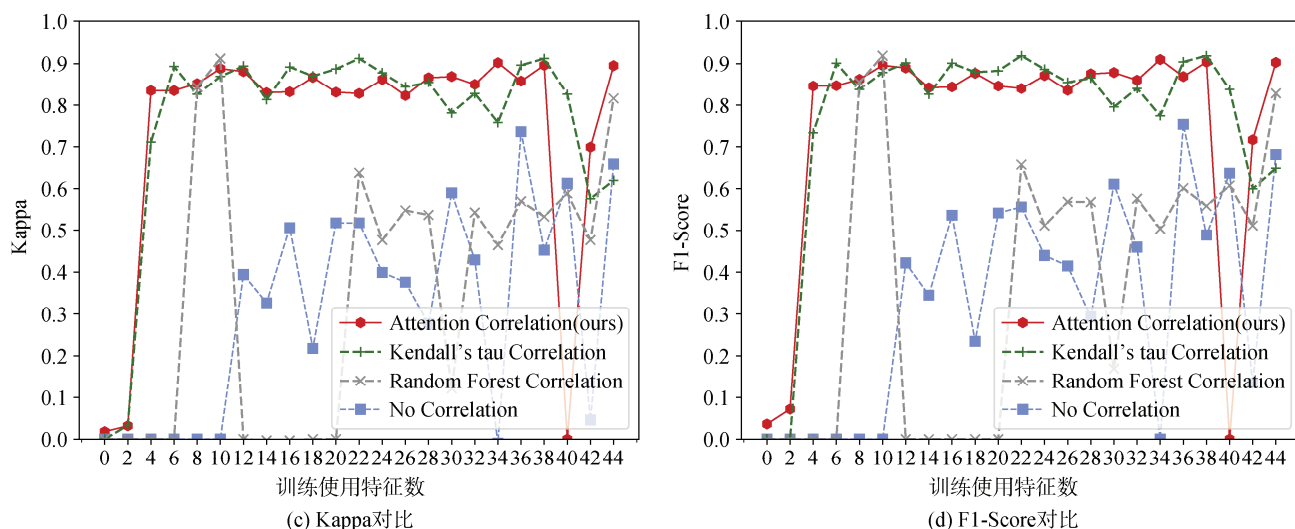


图7 CICIDS-2017数据集不同特征子集的性能

Figure 7 Training effect performance of different feature subsets based on CICIDS-2017

由上述步骤进行特征选择之后, UNSW-NB15 数据集和 CICIDS-2017 数据集中, 我们分别选择出排序前 10 的特征、前 20 的特征作为两个数据集的最优特征子集。最后, 本文还展示了模型分别使用特征选择筛选出的最优特征子集与原始特征全集的训练结果。黑色条柱代表使用最优特征子集的训练结果, 灰色表示使用原始特征全集的训练结果。实验结果

如图 8 所示。在 Precision、FAR、F1-Score、Kappa 系数等所有的评价指标上, 使用最优特征子集的异常检测模型均取得了更理想的训练结果, 且使用的特征数量更少。实验结果表明, 本文的特征选择机制可选择出训练效果最佳的前 K 个特征组成最优特征子集以减少冗余和弱相关性的特征对训练性能的损耗。

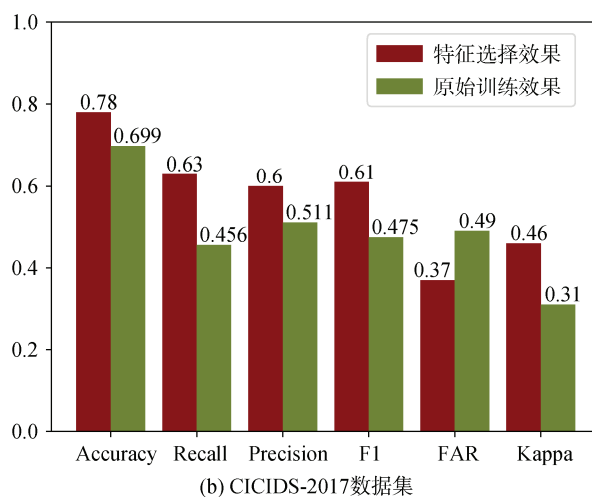
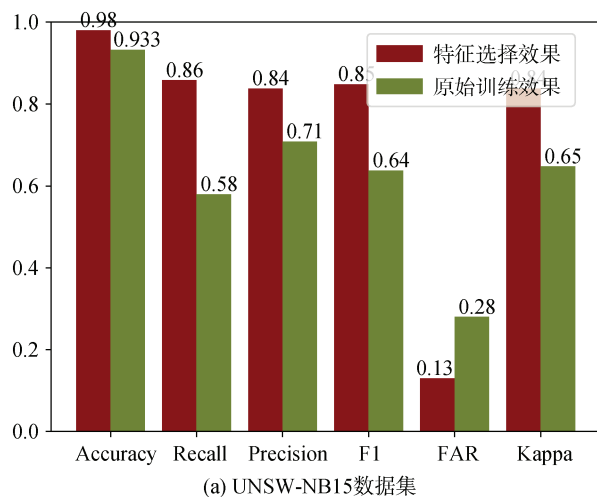


图8 基于最优子集和原始子集的性能对比

Figure 8 Model performance comparison based on optimal subset and original subset

5 结论

本文提出了一种基于动态流量图的时空图注意力网络(STGAT)异常检测方法, 解决了使用深度学习方法进行网络节点异常检测的潜在的局限性, 即没有分析网络节点的流量时序演化规律。该方法首先从网络流量数据中提取动态流量图的信息, 将流

量分类问题转换为网络节点的行为分类问题, 利用时空图注意力网络STGAT同时训练图的结构特征以及统计特征。此外, 通过自适应特征选择对模型进行了优化。在多个物联网的网络安全数据集上的实验结果表明, 本文的方法在所有实验中均取得了优异的性能, 在多个数据集上 F1-Score 均能达到 0.9 左右。特别地, 基于不同的时间间隔和样本密度, 在

CIDDS-001、UNSW-NB15、CICIDS-2017 数据集中, 分别以 30 秒、300 秒和 1000 秒为时间间隔抽取的图快照数据更能发挥模型的优势。此外, 相较于图注意力网络模型 GATv2, 加入时间注意力网络(AT)的动态图模型在各类指标上表现更好, 特别是 Precision 和 F1-Score 两个指标, 普遍提升了 30%~40%; FAR 也有明显下降, 如在 UNSW-NB15 数据集上, FAR 从 0.16 降低至 0.01, 证明了使用时间注意力模块对提升模型对分类的敏感度的收益。最后, 在 UNSW-NB15 和 CICIDS-2017 两个基准数据集上研究了特征选择前后的不同特征子集对异常检测性能的影响。实验结果表明, 使用特征选择的最优子集, 能够提升模型收敛速度, 并能以更少的特征数量达到更好的训练效果。目前待解决的问题主要有: (1) 本文使用求均值的方式生成节点特征, 不可避免地存在特征过平滑的潜在风险, 未来需要考虑更合理的方式生成节点特征; (2) 当前模型使用现有的数据集训练, 如何将模型在工业物联网环境中实际部署并与真实流量数据交互, 仍然是一个开放性的问题。

参考文献

- [1] Ding X O, Yu S J, Wang M X, et al. Anomaly Detection on Industrial Time Series Based on Correlation Analysis[J]. *Journal of Software*, 2020, 31(3): 726-747.
(丁小欧, 于晟健, 王沐贤, 等. 基于相关性分析的工业时序数据异常检测[J]. *软件学报*, 2020, 31(3): 726-747.)
- [2] Srivastava A, Prakash J. Future FANET with Application and Enabling Techniques: Anatomization and Sustainability Issues[J]. *Computer Science Review*, 2021, 39: 100359.
- [3] Khraisat A, Gondal I, Vamplew P, et al. Survey of Intrusion Detection Systems: Techniques, Datasets and Challenges[J]. *Cybersecurity*, 2019, 2(1): 20.
- [4] Sarhan M, Layeghy S, Portmann M. Evaluating Standard Feature Sets towards Increased Generalisability and Explainability of ML-Based Network Intrusion Detection[J]. *Big Data Research*, 2022, 30: 100359.
- [5] Mayuranathan M, Murugan M, Dhanakoti V. RETRACTED ARTICLE: Best Features Based Intrusion Detection System by RBM Model for Detecting DDoS in Cloud Environment[J]. *Journal of Ambient Intelligence and Humanized Computing*, 2021, 12(3): 3609-3619.
- [6] Nguyen M T, Kim K. Genetic Convolutional Neural Network for Intrusion Detection Systems[J]. *Future Generation Computer Systems*, 2020, 113: 418-427.
- [7] Almiani M, AbuGhazleh A, Al-Rahayfeh A, et al. Deep Recurrent Neural Network for IoT Intrusion Detection System[J]. *Simulation Modelling Practice and Theory*, 2020, 101: 102031.
- [8] Zhao H, Shu H, Liu C G, et al. Research on Dynamic Self-Organizing P2P Botnet[J]. *Journal of Cyber Security*, 2023, 8(2): 28-41.
(赵昊, 舒辉, 刘潮歌, 等. 动态自组织 P2P 僵尸网络的构建及其防御[J]. *信息安全学报*, 2023, 8(2): 28-41.)
- [9] Yin J, Cui X, Fang B X, et al. A Pollution-Resilient Hybrid P2P Botnet[J]. *Journal of Cyber Security*, 2018, 3(1): 68-82.
(尹捷, 崔翔, 方滨兴, 等. 一种抗污染的混合 P2P 僵尸网络[J]. *信息安全学报*, 2018, 3(1): 68-82.)
- [10] Chen B F, Li J D, Lu X J, et al. Survey of Deep Learning Based Graph Anomaly Detection Methods[J]. *Journal of Computer Research and Development*, 2021, 58(7): 1436-1455.
(陈波冯, 李靖东, 卢兴见, 等. 基于深度学习的图异常检测技术综述[J]. *计算机研究与发展*, 2021, 58(7): 1436-1455.)
- [11] Rashid A N M B, Ahmed M, Sikos L F, et al. Anomaly Detection in Cybersecurity Datasets via Cooperative Co-Evolution-Based Feature Selection[J]. *ACM Transactions on Management Information Systems*, 2022, 13(3): 1-39.
- [12] Yin Z N, Ma H L, Hu T. A Traffic Anomaly Detection Method Based on the Combination of DAE and GRU[J]. *Journal of Cyber Security*, 2023, 8(2): 11-27.
(尹梓诺, 马海龙, 胡涛. 基于 DAE 和 GRU 组合的流量异常检测方法[J]. *信息安全学报*, 2023, 8(2): 11-27.)
- [13] Liu X Q, Shan C, Ren J D. An intrusion detection method based on multi-dimensional optimization of traffic anomaly analysis[J]. *Journal of Cyber Security*, 2019, 4(1): 14-26.
(刘新倩, 单纯, 任家东, 等. 基于流量异常分析多维优化的入侵检测方法[J]. *信息安全学报*, 2019, 4(1): 14-26.)
- [14] Yao R, Xia S X, Zhou Y, et al. Spatial-Temporal Video Object Segmentation with Graph Convolutional Network and Attention Mechanism[J]. *Journal of Image and Graphics*, 2021, 26(10): 2376-2387.
(姚睿, 夏士雄, 周勇, 等. 时空图卷积网络与注意机制的视频目标分割[J]. *中国图象图形学报*, 2021, 26(10): 2376-2387.)
- [15] Cheng K Y, Wu J X, Wang W S, et al. Multi-Person Interaction Action Recognition Based on Spatio-Temporal Graph Convolution[J]. *Journal of Image and Graphics*, 2021, 26(7): 1681-1691.
(成科扬, 吴金霞, 王文杉, 等. 融合时空图卷积的多人交互行为识别[J]. *中国图象图形学报*, 2021, 26(7): 1681-1691.)
- [16] Shu J G, Zhou L, Zhang W Z, et al. Collaborative Intrusion Detection for VANETs: A Deep Learning-Based Distributed SDN Approach[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2021, 22(7): 4519-4530.
- [17] Zola F, Seguro-Gil L, Bruse J L, et al. Network Traffic Analysis through Node Behaviour Classification: A Graph-Based Approach with Temporal Dissection and Data-Level Preprocessing[J]. *Computers & Security*, 2022, 115: 102632.
- [18] Shen M, Zhang J P, Zhu L H, et al. Accurate Decentralized Application Identification via Encrypted Traffic Analysis Using Graph Neural Networks[J]. *IEEE Transactions on Information Forensics and Security*, 2021, 16: 2367-2380.
- [19] Li Y Z, Li R J, Zhou Z, et al. GraphDDoS: Effective DDoS Attack Detection Using Graph Neural Networks[C]. *2022 IEEE 25th International Conference on Computer Supported Cooperative Work in Design*, 2022: 1275-1280.
- [20] Zhao T, Jiang T W, Shah N, et al. A Synergistic Approach for Graph Anomaly Detection with Pattern Mining and Feature

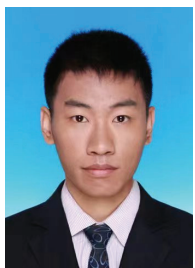
- Learning[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2022, 33(6): 2393-2405.
- [21] Wang Y, Lin X, Wu J, et al. Contrastive GNN-Based Traffic Anomaly Analysis Against Imbalanced Dataset in IoT-Based ITS[C]. *GLOBECOM 2022 - 2022 IEEE Global Communications Conference*, 2022: 3557-3562.
- [22] Li J D, Cheng K W, Wang S H, et al. Feature Selection: A Data Perspective[J]. *ACM Computing Surveys*, 2017, 50(6): 1-45.
- [23] Ren J D, Guo J W, Qian W, et al. Building an Effective Intrusion Detection System by Using Hybrid Data Optimization Based on Machine Learning Algorithms[J]. *Security and Communication Networks*, 2019, 2019(1): 7130868.
- [24] Saber A, Abbas M, Fergani B. Two-Dimensional Intrusion Detection System: A New Feature Selection Technique[C]. *2020 2nd International Workshop on Human-Centric Smart Environments for Health and Well-being*, 2021: 69-74.
- [25] Binbusayyis A, Vaiyapuri T. Comprehensive Analysis and Recommendation of Feature Evaluation Measures for Intrusion Detection[J]. *Heliyon*, 2020, 6(7): e04262.
- [26] Gui N, Ge D N, Hu Z Y. AFS: An Attention-Based Mechanism for Supervised Feature Selection[J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2019, 33(1): 3705-3713.
- [27] Brody S, Alon U, Yahav E. How attentive are graph attention networks [J]. *Proceedings of the International Conference on Learning Representations*, 2022: 1-26.
- [28] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[J]. *Advances In Neural Information Processing Systems*, 2017, 30.
- [29] Lin T Y, Goyal P, Girshick R, et al. Focal Loss for Dense Object Detection[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, 42(2): 318-327.
- [30] Moustafa N, Slay J, Creech G. Novel Geometric Area Analysis Technique for Anomaly Detection Using Trapezoidal Area Estimation on Large-Scale Networks[J]. *IEEE Transactions on Big Data*, 2017, 5(4): 481-494.
- [31] Ring M, Wunderlich S, Grödl D, et al. Flow-based benchmark data sets for intrusion detection[C]. *Proceedings of the 16th European Conference on Cyber Warfare and Security*, 2017: 361-369.
- [32] 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.
- [33] Rajesh Kanna P, Santhi P. Unified Deep Learning Approach for Efficient Intrusion Detection System Using Integrated Spatial-Temporal Features[J]. *Knowledge-Based Systems*, 2021, 226: 107132.
- [34] Yan K. Chiller Fault Detection and Diagnosis with Anomaly Detective Generative Adversarial Network[J]. *Building and Environment*, 2021, 201: 107982.



苏宪时 于 2019 年在山东理工大学软件工程专业获得学士学位。现在大连海事学校计算机技术专业攻读硕士学位。研究领域为网络安全、图神经网络。研究兴趣包括: 强化学习、可信人工智能。Email: sxs@dlnu.edu.cn



杨丽平 于 2022 年在大连海事大学计算机应用技术专业获得博士学位。现为大连海事大学人工智能学院博士后研究员, 研究兴趣包括: 可信人工智能、动态图对抗防御、图神经网络。Email: liping.yang@dlnu.edu.cn



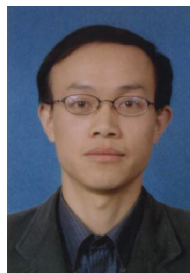
马润泽 于 2015 年在北京理工大学信息与电子学院获得学士学位。现在大连海事大学计算机技术专业攻读硕士学位。研究领域为网络安全、图神经网络。研究兴趣包括: 网络安全、可信人工智能。Email: mrz@dlnu.edu.cn



江同棒 于 2016 年在内蒙古工业大学获学士学位, 现在大连海事大学计算机科学与技术专业攻读博士学位, 研究领域为可信人工智能、机器学习、网络安全。Email: tongbangjiang@foxmail.com



曲天歌 于 2021 年于大连海事大学智能科学技术专业获得学士学位。现在大连海事大学电子信息专业攻读硕士学位。研究领域为智能机器人、机器视觉。研究兴趣包括: 智能机器人、计算机视觉。Email: qtgdlnu@qq.com



刘洪波 于 2005 年在大连理工大学计算机科学与技术专业取得博士学位。现为大连海事大学人工智能学院院长、教授、博士生导师, 研究兴趣包括: 认知计算、可信人工智能、网络安全、大数据。Email: lhb@dlnu.edu.cn