基于自监督图神经网络和混合神经网络的入侵检测

王　明
（河北科技师范学院　网络技术中心，河北　秦皇岛　066000）

摘　要：为了解决现有网络入侵检测方法在特征提取单一、数据依赖强以及模型泛化能力差等方面的问题，提出了一种基于自监督图神经网络和混合神经网络的入侵检测方法。首先，通过自监督学习策略，利用图卷积网络提取网络流量数据中的结构特征，增强模型在无标签数据上的特征学习能力，从而降低对标注数据的依赖并提升泛化能力。其次，使用卷积神经网络提取网络流量中时间序列的空间特征，并通过长短时记忆网络建模时间依赖性，进行多视角特征提取，提高检测的全面性。最后，设计了一种特征融合策略，丰富模型特征表示，提升模型鲁棒性。在公开数据集上的实验结果表明，所提方法具有更高的准确率和 $F1$ 值。

关键词：自监督学习；图神经网络；混合神经网络；入侵检测
中图分类号：TP393.08；TP18　　**文献标识码：**A　　**DOI：**10.19358/j.issn.2097-1788.2024.09.004
引用格式：王明. 基于自监督图神经网络和混合神经网络的入侵检测［J］. 网络安全与数据治理，2024，43(9)：21-25.

Intrusion detection based on self-supervised graph neural networks and hybrid neural networks

Wang Ming
(Network Technology Center, Hebei Normal University Of Science & Technology, Qinhuangdao 066000, China)

Abstract: To address the issues of limited feature extraction, strong data dependency, and poor generalization ability in existing network intrusion detection methods, this paper proposes an intrusion detection method based on self-supervised graph neural networks and hybrid neural networks. Firstly, through a self-supervised learning strategy, a graph convolutional network is employed to extract structural features from network traffic data, enhancing the model's ability to learn features from unlabeled data. This reduces dependence on labeled data and improves generalization ability. Secondly, a convolutional neural network is used to extract spatial features from the time series of network traffic, and a long short-term memory network is employed to model temporal dependencies, enabling multi-view feature extraction and improving detection comprehensiveness. Finally, a feature fusion strategy is designed to enrich the model's feature representation and enhance its robustness. Experimental results on public datasets demonstrate that the proposed method achieves higher accuracy and $F1$ score.

Key words: self-supervised learning; graph neural network; hybrid neural network; intrusion detection

0　引言

网络入侵检测系统（Intrusion Detection System，IDS）[1]主要用于监控网络活动，以便迅速识别潜在的恶意行为、攻击事件以及违反系统安全策略的行为。网络入侵检测在现代信息安全体系中具有举足轻重的地位。随着网络技术的飞速发展和信息化程度的不断提高，网络攻击手段变得日益复杂和多样化，给企业和个人的信息安全带来了严峻的挑战。网络入侵检测技术通过对网络流量和系统日志的实时监控和分析，能够及时发现和阻止潜在的安全威胁，有效防范数据泄露、服务中断和资源滥用等安全事件的发生。随着人工智能技术的进步，近年来研究人员已经不断应用深度学习技术来解决网络入侵检测中的若干复杂问题。这些研究不仅着眼于提高检测准确率，还致力于降低误报率，以提升整体系统的效率。特别是卷积神经网络（CNN）[2]和递归神经网络（RNN）[3]在处理大量复杂数据时表现出色，被广泛用于

特征提取和异常行为识别。此外，自监督学习[4]、时间卷积网络（TCN）[5]等新兴方法的引入，也为网络入侵检测领域带来了新的希望和发展方向。

Li[6]等人将 GRU－RNN 网络模型引入入侵检测任务中，提升了模型数据时序特征的检测能力。Imrana[7]等人提出了基于双向长短时记忆网络（LSTM）的入侵检测方法，利用正反两个方向的 LSTM 网络捕捉正反时序特征，并对提取的双向特征进行融合，显著提升了检测性能，但对于训练数据标注具有较高的要求，模型泛化能力不足。张安琳[8]等人将卷积神经和双向门控循环结合，对融合后的特征进行时序特征的提取，提升了模型的检测能力。Halbouni[9]等人提出了一种混合神经网络，利用 CNN 来提取网络流量数据的空间特征，并结合 LSTM 来捕捉时间特征。这种混合模型在处理复杂网络流量数据方面具有高效性和准确性。但此种方法计算复杂性高，尤其是处理大规模数据集时，需要大量的计算资源和时间。Wang[10]提出了一种自监督学习的入侵检测方法，该方法无需标签数据？大大减少了数据标注的成本。通过数据增强、特征表示、特征投影和对比学习等步骤，提高模型的检测能力，但该方法在复杂的网络攻击场景下，模型可能会产生较高的误报率。

尽管这些研究在网络入侵检测方面取得了一定的成果，但仍然存在一些问题。首先，现有方法在对网络数据特征提取单一，学习数据的本质特征不全面。同时在不同任务上表现差异较大，模型泛化能力差，缺少各种场景下的适应能力。为了解决上述问题，本文提出了一种基于自监督图神经网络和混合神经网络的入侵检测。该方法准确地对网络数据进行特征提取，并对时序特征进行了高效的上下文处理，提高网络入侵检测的性能。

本文的主要贡献如下：

（1）提出了一种基于自监督学习的图神经网络（GNN），利用图卷积网络（GCN）有效地捕捉节点之间的复杂关系，提高入侵检测的性能。同时自监督学习可以在没有标注数据的情况下，通过设计预训练任务来学习数据的潜在特征，增强了模型的泛化能力。

（2）开发了一种混合神经网络模型，结合 CNN 和 LSTM，能够有效提取时间序列的空间特征和时间依赖性，提升对时序特征的处理能力。

（3）设计了一种特征融合策略，将 GNN 和混合卷积神经网络－循环神经网络（CNN-LSTM）模型的输出进行融合，可以具有丰富特征表示，通过融合不同模型的特征，可以降低模型对特定特征模式的过拟合，提高了网络入侵检测系统的准确性和鲁棒性。

1　本文算法概述

本文结合自监督 GNN 和 CNN-LSTM 模型，此算法旨在利用图结构数据的特征和时间序列数据的依赖性，提升入侵检测系统的性能。本文算法整体框图如图 1 所示。

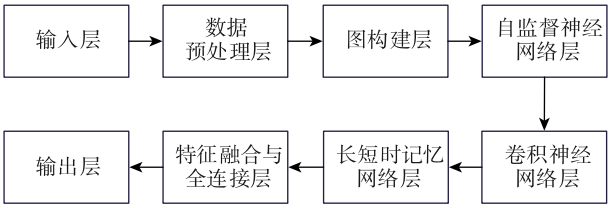


图 1　本文算法的整体框图

1.1　图构建

首先，收集网络流量数据并进行预处理，将其转换为图格式 $G=(V, E)$，具体表示为：

$$G=(V, E) \tag{1}$$

$$V=\{v_1, v_2, \cdots, v_n\} \tag{2}$$

$$E=\{(v_i, v_j) \mid 流_i 和流_j 共享相同的源\ IP\ 或目的\ IP\} \tag{3}$$

其中，V 是节点集合，每一个节点代表一个网络流；E 是边集合，表示流之间的关系，例如共享源 IP 或目的 IP。

1.2　自监督表征学习

1.2.1　图卷积网络（GCN）

本文算法通过 GCN[11]提取网络流量数据中的结构特征。GCN 通过邻接节点的信息聚合来更新每个节点的表示。第 l 层的节点表示更新公式为：

$$\boldsymbol{h}_v^{l+1}=\sigma\left(\sum_{u\in N(v)}\frac{1}{c_{vu}}\mathrm{W}^{(l)}h_u^{(l)}+\boldsymbol{b}^{(l)}\right) \tag{4}$$

其中，$\boldsymbol{h}_v^{l+1}$ 表示第 $l+1$ 层节点 v 的表示向量；$N(v)$ 表示节点 v 的邻居节点集合；$\mathrm{W}^{(l)}$ 是第 l 层的权重矩阵；c_{vu} 是归一化常数，通常取节点度的平方根的倒数；$\boldsymbol{b}^{(l)}$ 是第 l 层的偏置向量；σ 是非线性激活函数。

1.2.2　自监督任务

自监督学习的目标是通过设计自监督任务，利用无标签数据进行特征学习，提升模型的泛化能力。通过数据本身的内在结构生成伪标签，使得模型在无需人工标注的情况下也能学习到有价值的表示。通过利用大量可用的无标签数据，自监督学习可以揭示复杂的模式和底层结构，大大丰富了特征空间。不仅减轻了对标注数据的依赖，还为模型在不同数据场景中的有效泛化提供了坚实基础。

本文采用基于节点相似性的自监督任务，其损失函数定义为：

$$L_{\text{self}} = -\sum_{v \in V}\sum_{u \in N(v)} \log P(u \mid v) \tag{5}$$

其中，L_{self}是自监督任务的损失函数，$P(u \mid v)$ 表示给定节点 v 的情况下预测节点 u 的概率。

概率 $P(u \mid v)$ 采用 softmax 函数计算。

$$P(u \mid v) = \frac{\exp(\boldsymbol{h}_u \cdot \boldsymbol{h}_v)}{\sum_{k \in N(v)} \exp(\boldsymbol{h}_k \cdot \boldsymbol{h}_v)} \tag{6}$$

其中，h_v 是节点 v 的表示向量，$\boldsymbol{h}_u$ 是节点 u 的表示向量，k 是节点 v 的邻居节点。

1.3 空间特征提取

在时间序列数据的处理过程中，使用 CNN 提取空间特征。卷积层的目的是捕捉局部的时空模式，提高模型对短期变化的响应能力。

对于时间步 t 的输入特征向量 $\boldsymbol{x}_t$，卷积操作公式为：

$$\boldsymbol{z}_t = \text{ReLU}(\boldsymbol{W}_c * \boldsymbol{x}_t + \boldsymbol{b}_c) \tag{7}$$

其中，$\boldsymbol{z}_t$ 表示时间步 t 的输出特征向量，$\boldsymbol{W}_c$ 是卷积核，$\boldsymbol{b}_c$ 是偏置向量，$*$ 表示卷积操作；$\boldsymbol{x}_t$ 是时间步 t 的输入特征向量，ReLU 是激活函数。通过卷积操作，可以提取局部时间序列特征，使模型能够识别短期的模式和变化。

1.4 时间特征学习

为了捕捉时间序列数据中的长时依赖性，将卷积层提取的空间特征输入到长 LSTM 网络中进行处理。

LSTM 网络能够有效地捕捉时间序列中的长期依赖性，其核心由三个门控机制组成：输入门 i_t、遗忘门 f_t 和输出门 o_t。每个门的计算公式如下：

$$i_t = \sigma(\boldsymbol{W}_i \cdot [h_{t-1}, \boldsymbol{z}_t] + \boldsymbol{b}_i) \tag{8}$$

其中，i_t 表示输入门，σ 表示 sigmoid 激活函数，$\boldsymbol{W}_i$ 是输入门的权重矩阵，$\boldsymbol{h}_{t-1}$ 表示上一个时间步的隐藏状态，$\boldsymbol{z}_t$ 表示当前时间步的输入特征，$\boldsymbol{b}_i$ 是输入门的偏置向量。

$$f_t = \sigma(\boldsymbol{W}_f \cdot [h_{t-1}, \boldsymbol{z}_t] + \boldsymbol{b}_f) \tag{9}$$

其中，f_t 表示遗忘门，σ 表示 sigmoid 激活函数，$\boldsymbol{W}_f$ 是遗忘门的权重矩阵，$\boldsymbol{b}_f$ 是遗忘门的偏置向量。

$$o_t = \sigma(\boldsymbol{W}_o \cdot [h_{t-1}, \boldsymbol{z}_t] + \boldsymbol{b}_o) \tag{10}$$

其中，o_t 表示输出门，σ 表示 sigmoid 激活函数，$\boldsymbol{W}_o$ 是输出门的权重矩阵，$\boldsymbol{b}_o$ 是输出门的偏置向量。

细胞状态更新公式为：

$$\tilde{c}_t = \tanh(\boldsymbol{W}_C \cdot [h_{-1}, \boldsymbol{z}_t] + \boldsymbol{b}_c) \tag{11}$$

其中，$\tilde{c}_t$表示候选细胞状态，tanh 是双曲正切激活函数。$\boldsymbol{W}_C$ 是候选细胞状态的权重矩阵，$\boldsymbol{b}_c$ 是候选细胞状态的偏置向量。

$$c_t = f_t \cdot c_{-1} + i_t \cdot \tilde{c}_t \tag{12}$$

其中，c_t 表示当前时间步的细胞状态，c_{-1} 表示上一个时间步的细胞状态。

隐藏状态更新公式为：

$$h_t = o_t \cdot \tanh(c_t) \tag{13}$$

其中，h_t 表示当前时间步的隐藏状态。

1.5 特征融合与分类

在特征提取和时间特征学习之后，将 GNN 和 CNN-LSTM 模型的输出进行融合，通过全连接层进行分类。将 GNN 提取的结构特征 $\boldsymbol{h}_v$ 和 CNN-LSTM 提取的时序特征 h_t 进行融合，融合后的输出表示为：

$$\boldsymbol{h}_{\text{fusion}} = \boldsymbol{h}_v \oplus \boldsymbol{h}_t \tag{14}$$

其中：$\boldsymbol{h}_v$ 表示 GNN 提取的结构特征向量，$\boldsymbol{h}_t$ 表示 CNN-RNN 提取的时序特征向量，⊕表示向量连接操作。

融合后的特征向量 $\boldsymbol{h}_{\text{fusion}}$通过全连接层进行分类：

$$y = \varphi(\boldsymbol{W}_f \boldsymbol{h}_{\text{fusion}} + \boldsymbol{b}_f) \tag{15}$$

其中，y 表示输出的分类结果，$\boldsymbol{W}_f$ 表示全连接层的权重矩阵，$\boldsymbol{b}_f$ 表示全连接层的偏置向量，φ 表示激活函数。

1.6 损失计算与模型训练

采用联合损失函数来优化模型，包括自监督任务的损失 $\mathscr{L}_{\text{self}}$和入侵检测任务的损失 $\mathscr{L}_{\text{cls}}$。

联合损失函数定义为：

$$\mathscr{L}_{\text{total}} = \alpha \mathscr{L}_{\text{self}} + \beta \mathscr{L}_{\text{cls}} \tag{16}$$

其中，$\mathscr{L}_{\text{total}}$表示总损失函数；$\mathscr{L}_{\text{self}}$表示自监督任务的损失函数；$\mathscr{L}_{\text{cls}}$表示入侵检测任务的损失函数；$\alpha$ 和 β 是超参数，控制两个任务的损失权重。

入侵检测任务的损失通常采用交叉熵损失，定义为：

$$\mathscr{L}_{\text{cls}} = \text{Ny}_i \cdot log(\hat{\angle}\text{y}_i) \tag{17}$$

其中，N 表示样本数量。y_i 表示第 i 个样本的真实标签，$\hat{\angle} y_i$ 表示模型的第 i 类的预测概率。

$\hat{\angle} y_i$ 的计算公式如下：

$$\hat{\angle} y_i = \frac{\exp(z_i)}{\sum_j \exp(z_j)} \tag{18}$$

其中，z_i 表示第 i 类的未归一化预测值，j 表示类的数量。

通过优化联合损失函数训练模型，提高模型的准确率和鲁棒性。训练过程包括前向传播计算损失、反向传播更新权重，直至损失函数收敛。

2 实验结果分析

2.1 实验设置介绍

2.1.1 数据集介绍

（1）NSL-KDD 数据集[12]是对 KDD99 数据集的改进版，包含四个子集：KDDTest + 、KDDTest-21、KDDTrain + 和 KDDTrain + _ 20Percent。相较于 KDD99 数据集，NSL-KDD 在数据分布和类别平衡方面进行了优化。

（2）UNSW-NB15 数据集[13]是网络入侵检测研究中

常用的数据集，包含约 250 万条数据记录，涵盖正常流量和 9 种攻击类型。该数据集由澳大利亚新南威尔士大学创建，旨在提供多样且现实的网络流量，以评估和开发入侵检测系统的性能。

2.1.2 评价指标介绍

本文使用准确率（Accuracy，Ac）、查准率（Precision，Pr）、查全率（Recall，Re）和 $F1$ 值（FScore，$F1$）进行评价。评价指标定义如下：

（1）准确率：正确分类的样本占总样本的比例。

$$\mathrm{Ac}=\frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TP}+\mathrm{TN}+\mathrm{FP}+\mathrm{FN}} \tag{19}$$

（2）查准率：实际为正例的样本被正确分类的比例。

$$\mathrm{Pr}=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}} \tag{20}$$

（3）查全率：被正确分类的正例占所有正例的比例。

$$\mathrm{Re}=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} \tag{21}$$

（4）$F1$ 值：表示查准率和查全率的调和平均值，用于综合评价模型的性能。

$$\mathrm{F1}=\frac{2\cdot \mathrm{Pr}\cdot \mathrm{Re}}{(\mathrm{Pr}+\mathrm{Re})} \tag{22}$$

其中，TP 表示真正例（True Positive），TN 表示真负例（True Negative），FP 表示假正例（False Positive），FN 表示假负例（False Negative）。

2.2 实验结果分析

为了验证本文所提算法的有效性，本文将其与其他几种主流入侵检测方法进行了对比，结果如表 1、表 2 所示。本节将对实验结果进行详细分析。

表 1 KDDTest + 数据集中 5 种算法多分类对比（%）

方法	准确率	查准率	查全率	$F1$ 值
AlertNet [14]	78.50	81.00	78.50	76.50
CNN[15]	81.75	82.43	82.71	82.57
MDNN[16]	77.55	81.23	77.55	75.43
CNN-BiLSTM[17]	83.58	85.82	84.49	85.14
本文	86.42	84.00	88.84	86.35

表 1 展示了在 KDDTest + 数据集上 5 种算法的对比结果。可以看出，本文算法在准确率、查全率均优于其他方法。本文算法的查准率低于文献［17］的方法，但高于其他三种算法，且具有更高的 $F1$ 值，说明本文算法的综合检测性能更好。

表 2 UNSW-NB15-test 数据集中 5 种算法多分类对比（%）

方法	准确率	查准率	查全率	$F1$ 值
AlertNet [14]	66.00	62.30	66.00	59.60
CNN[15]	74.61	81.01	75.65	78.24
MDNN[16]	62.87	76.01	63.10	64.15
CNN-BiLSTM[17]	77.16	82.63	79.91	81.25
本文	86.49	83.00	81.30	82.15

表 2 展示了在 UNSW-NB15-test 数据集上 5 种算法的对比结果。本文算法的各项指标均高于其他算法。相较于 NSL-KDD 数据集，UNSW-NB15 数据集检测难度更大，检测结果证明了本文算法的有效性。

3 结论

本文提出了一种基于自监督图神经网络和混合神经网络的入侵检测方法。本文算法引入自监督学习机制和图结构网络，降低了模型对数据的依赖并提升了模型的泛化能力。GNN 擅长处理图结构数据，能够有效地捕捉节点之间的复杂关系，提高模型的检测效果。同时，本文算法提出了一种混合网络模型，结合 CNN 和 LSTM 的优势，提取不同类型的特征，从多个视角分析数据，提升了模型检测的全面性。最后，本文算法设计了一种特征融合策略，将 GNN 和 CNN-LSTM 模型的输出进行融合，将来自不同模型的特征组合在一起，形成更加全面和丰富的特征表示，使模型在不同环境下表现更稳定。实验结果证明了本文算法的优越性。

参考文献

[1] 蹇诗婕，卢志刚，杜丹，等. 网络入侵检测技术综述［J］. 信息安全学报，2020，5（4）：96－122.

[2] WANG H，CAO Z，HONG B. A network intrusion detection system based on convolutional neural network［J］. Journal of Intelligent & Fuzzy Systems，2020，38（6）：7623－7637.

[3] LE T T H，KIM Y，KIM H. Network intrusion detection based on novel feature selection model and various recurrent neural networks［J］. Applied Sciences，2019，9（7）：1－29.

[4] 梁欣怡，行鸿彦，侯天浩. 基于自监督特征增强的 CNN-BiLSTM 网络入侵检测方法［J］. 电子测量与仪器学报，2022，36（10）：65－73.

[5] Tao Zhiyong，Yan Minghao，Liu Ying. Channel coding closed set recognition based on temporal convolutional networks［J］. Journal of Huazhong University of Science and Technology（Natural Science Edition），2022，50（3）：12－17.

[6] LI J，XIA S Z，LAN H Y，et al. Network intrusion detection method based on GRU-RNN［J］. Journal of Harbin Engineering

University, 2021, 42 (6): 879 – 884.

[7] IMRANA Y, XIANG Y P, ALI L, et al. A bidirectional LSTM deep learning approach for intrusion detection [EB/OL]. (2021 – 12 – 05) [2023 – 09 – 26]. https://doi.org/10.1016/j.eswa.2021.115524.

[8] 张安琳, 张启坤, 黄道颖, 等. 基于 CNN 与 BiGRU 融合神经网络的入侵检测模型 [J]. 郑州大学学报 (工学版), 2022, 43 (3): 37 – 43.

[9] HALBOUNI A, GUNAWAN T S, HABAEBI M H, et al. CNN-LSTM: hybrid deep neural network for network intrusion detection system [J]. IEEE Access, 2022 (10): 99837 – 99849.

[10] WANG Z, LI Z, WANG J, et al. Network intrusion detection model based on improved BYOL self-supervised learning [J]. Security and Communication Networks, 2021, 2021: 1 – 23.

[11] ZHANG H, LU G, ZHAN M, et al. Semi-supervised classification of graph convolutional networks with Laplacian rank constraints [J]. Neural Processing Letters, 2022: 1 – 12.

[12] DEVARAKONDA A, SHARMA N, SAHA P, et al. Network intrusion detection: a comparative study of four classifiers using the NSL-KDD and KDD' 99 datasets [C]//Journal of Physics: Conference Series. IOP Publishing, 2022, 2161 (1): 012043.

[13] MOUALLA S, KHORZOM K, JAFAR A. Improving the performance of machine learning-based network intrusion detection systems on the UNSW-NB15 dataset [J]. Computational Intelligence and Neuroscience, 2021: 1 – 13.

[14] VINAYAKUMAR R, ALAZAB M, SOMAN K, et al. Deep learning approach for intelligent intrusion detection system [J]. IEEE Access, 2019 (7): 41525 – 41550.

[15] WU K E, CHEN Z G, LI W. A novel intrusion detection model for a massive network using convolutional neural networks [J]. IEEE Access, 2018 (6): 50850 – 50859.

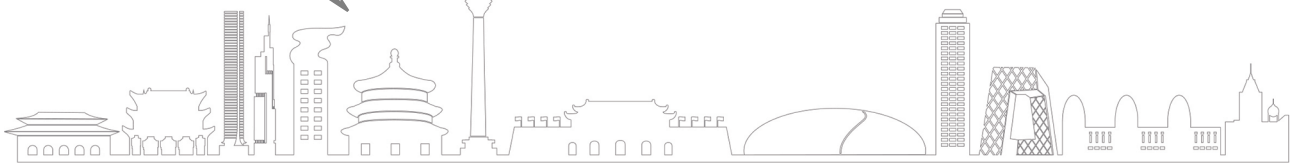
[16] ALTWAIJRY N, ALQAHTANI A, ALTURAIKI I. A deep learning approach for anomaly-based network intrusion detection [C]// Proceedings of the Big Data and Security: First International Conference, 2020: 603 – 615.

[17] JIANG K, WANG W, WANG A, et al. Network intrusion detection combined hybrid sampling with deep hierarchical network [J]. IEEE Access, 2020 (8): 32464 – 32476.

(收稿日期: 2024 – 09 – 14)

作者简介:

王朋 (1994 –), 男, 硕士, 工程师, 主要研究方向: 深度学习、网络安全。



版权声明

凡《网络安全与数据治理》录用的文章，如作者没有关于汇编权、翻译权、印刷权及电子版的复制权、信息网络传播权与发行权等版权的特殊声明，即视作该文章署名作者同意将该文章的汇编权、翻译权、印刷权及电子版的复制权、信息网络传播权与发行权授予本刊，本刊有权授权本刊合作数据库、合作媒体等合作伙伴使用。同时，本刊支付的稿酬已包含上述使用的费用，特此声明。

《网络安全与数据治理》编辑部