罗燕燕,刘子傲, 谢媛媛, 等. 基于 CSAtt-GraphSAGE 双通道特征加权融合的企业内部网络威胁检测方法[J]. 智能计算机与应用,2026,16(8):10-17. DOI:10.20169/j.issn.2095-2163.25011501

基于 CSAtt-GraphSAGE 双通道特征加权融合的企业内部网络威胁检测方法

罗燕燕, 刘子傲, 谢媛媛, 纪祥敏
(福建农林大学 计算机与信息学院, 福州 350002)

摘 要: 针对现在研究侧重于用户个体行为分析,忽略了用户之间潜在关联关系的问题,提出一种基于双通道特征加权融合的企业内部网络威胁检测方法。首先,在个体行为通道(Intra)通道采用 CSAtt 分析用户个体行为特征的内在联系和增强关键特征的表达能力;其次,在关联交互通道(Inter)通道将用户行为特征与用户关联关系抽象成异构图,并借助 GraphSAGE 捕捉图中关联用户节点间的群体行为模式;然后,引入可学习权重因子将不同通道获取的特征信息进行加权融合,使后续分类任务不仅能关注到与恶意操作相关的关键特征,还补充了对群体行为模式的分析,以提高整体模型检测性能。在 CMU-CERTr4.2 数据集上的实验表明,提出的双通道特征加权融合方法在解决企业内部网络威胁检测问题上,具备有效性和泛化性用性,最终检测效果的准确率、精确率和 *F*1 值分别达到 95.67%、96.32% 和 95.37%。

关键词: 企业内部网络威胁检测; 自注意力机制; GraphSAGE; 卷积神经网络; 特征融合
中图分类号: TP309.2 **文献标志码**: A **文章编号**: 2095-2163(2026)08-0010-08

Enterprise internal network threat detection method based on CSAtt-GraphSAGE dual-channel feature weighted fusion

LUO Yanyan, LIU Zi'ao, XIE Yuanyuan, JI Xiangmin
(College of Computer and Information Sciences, Fujian Agriculture and Forestry University, Fuzhou 350002, China)

Abstract: Aiming at the problem that many studies focus on user's individual behavior analysis and ignore the potential correlations relationship between users, an enterprise internal network threat detection method based on dual-channel feature weighted fusion is proposed. First, CSAtt is used in the Intra-Channel to analyze the inner connection of individual user behavioral features and enhance the expressive ability of key features. Secondly, in the Inter-Channel, user behavior features and user relationships are abstracted into heterogeneous graphs, and GraphSAGE is used to capture group behavior patterns among correlated user nodes in the graph. Then, we introduce learnable factors to weight and fuse features obtained from different channels, so that the subsequent classification tasks can not only weight focus on the key features related to malicious operations, but also supplement the analysis of group behavior patterns, so as to improve the model detection performance. Experiments on the CMU-CERTr4.2 dataset show that the proposed dual-channel feature weighted fusion method is effective and generalizable in solving the problem of enterprise internal network threat detection, and the final detection effect achieves accuracy, precision, and F1 scores of 95.67%, 96.32% and 95.37%, respectively.

Key words: enterprise internal network threat detection; self-attention mechanism; GraphSAGE; convolutional neural network; feature fusion

0 引 言

随着信息化应用场景的日益复杂,近年来网络威胁逐渐从外部向内部转移。内部威胁指企业内部人员对信息系统和数据进行恶意操作的攻击行为[1]。由于攻击者熟悉内部环境和信息,并持有一

基金项目: 国家自然科学基金重点项目(61332019);福建省自然科学基金(2021J01129)。
作者简介: 罗燕燕(2000—),女,硕士研究生。主要研究方向:网络安全,异常检测; 刘子傲(1997—),男,硕士研究生。主要研究方向:网络安全,加密流量分类; 谢媛媛(2000—),女,硕士研究生。主要研究方向:网络安全,异常检测。
通信作者: 纪祥敏(1971—),男,硕士,副教授。主要研究方向:网络空间安全。Email:jixm168@126.com。
收稿日期: 2025-01-15

定的合法权限访问组织的关键资产，故而能够利用自身身份优势进行隐蔽攻击，且往往难以被及时发现和阻止[2]。一旦内部人员发起攻击，可能会给企业带来重大损失，诸如数据泄露、业务中断等。因此，内部威胁检测成为信息安全领域的一个重要研究方向。

目前，大多数研究基于用户行为数据，借助机器学习或深度学习技术构建内部威胁检测模型[3-4]。然而，这些模型常常忽视了用户间复杂的关联关系，导致检测准确性受限。因此，本文提出了一种双通道特征加权融合检测方法，增强用户关键行为特征的表达能力，同时深入挖掘用户间关联信息，进而提高内部威胁检测模型的性能。

1　研究现状

内部威胁检测问题通常被视为一类异常检测任务，需从海量的用户行为审计日志中提取特征，并利用正常和恶意行为特征训练精准的分类模型，以识别数据中不符合预期的异常行为[5]。在早期研究中，机器学习方法在内部威胁检测领域中被广泛应用。Le 等[6-7]提出了基于隐马尔可夫模型（Hidden Markov Model，HMM）的检测方法。孙丹等[8]结合混合 N-Gram 模型和 XGBoost 算法进行特征提取和异常检测。李俊强等[9]结合滑动窗口机制动态调整用户画像模型，并设置异常行为规则来检测异常操作。Bertrand 等[10]利用贝叶斯-高斯混合模型提出了一种基于审计数据的无监督内部威胁检测系统，并采取基于 Word2Vec 的自动特征提取系统来简化实际场景中的使用。Al-Shehari 等[11]提出基于孤立森林算法的异常值识别机制，缓解了 CERT 数据集类别分布过度倾斜的问题。尽管这些方法对用户行为特征表现出良好的学习和分类效果，但机器学习对人工构建的特征有较大依赖性，研究者必须花费大量的时间精力分析日志数据、系统事件和用户行为模式，当选取的特征无法准确反映用户行为属性时，会显著影响算法的检测性能[12]。

深度学习作为一种表示学习方法，其具备学习特征中深层次隐藏信息的优势，缓解了传统机器学习在特征工程方面的局限性。Saaudi 等[13]通过卷积神经网络（Convolutional Neural Network，CNN）从数据中捕获局部特征，然后使用长短期记忆模型（Long Short Term Memory，LSTM）学习这些特征的顺序依赖。黄娜等[14]提出了一种 LSTM 回归模型，通过对时间序列的回归分析，对用户行为序列进行检测。孙小双等[15]采用自动编码器学习正常数据中的用户行为模式，通过计算解码数据与输入数据的重构误差获取异常分数，再结合 Z-score 方法对异常分数进行构架分级。Anju 等[16]提出了堆叠注意力卷积神经网络和双向门控循环单元的 M-EOS 模型，利用用户活动日志和用户信息进行时间序列分类。张光华等[17]提出基于 LSTM-Attention 用户行为的分析模型，结合行为、心理和角色特征对用户的行为习惯进行多角度建模，再通过多层感知机（Multilayer Perceptron，MLP）实施异常判断。

一些研究人员还尝试了基于图的方法。Liu 等[18]提出了一种异构图嵌入模型 Log2vec，以活动为节点构建异构图，以活动之间的关系为边，将正常活动和异常活动区分为不同的组，再应用聚类方法识别内部威胁行为。Jiang 等[19]设计了一个基于图卷积神经网络的异常检测模型，利用员工间的电子邮件通信来构建用户之间的节点连接关系。Li 等[20]使用生成对抗图注意网络模型（Generative Adversarial Graph Attention Networks，GAGAN）解决样本不平衡问题。

尽管深度学习技术在内部威胁检测领域内取得了一定进展，但大多数研究把重点放在对用户行为的分析上，忽略了用户之间的潜在关联关系。然而在企业内部组织架构中，员工之间直接或间接的关系会构成一个复杂的关系网，这也是检测内部威胁行为的关键线索，仅对单一信息源进行分析，会导致检测误报率较高[21]。

为解决上述问题，本文提出了一种基于双通道特征加权融合的检测方法。一方面，在 Intra 通道引入基于自注意力机制的 CSAtt 模型，用于分析用户行为特征的内在关联和增强关键特征的表达能力；另一方面，在 Inter 通道中，将用户行为特征和用户关联关系抽象成异构图，借助图采样与聚合模型（Graph Sample and Aggregate，GraphSAGE）重点挖掘用户之间相互关联所产生的特征信息。然后，采用加权融合策略充分利用双通道输出的特征，使后续分类任务能更全面理解用户行为的复杂性。该方法从多元视角挖掘多维度特征并开展综合评判，提升恶意行为检测的准确率与可靠性。

2　双通道特征加权融合检测模型

本文首先对海量异构日志数据进行系统处理与整合，生成符合模型输入标准的用户行为特征向量。在特征分析阶段采用双通道机制。其中，Intra 通道

引入结合卷积神经网络与自注意力机制的 CSAtt 模型,对用户行为特征的内在关联性进行深入分析,并进一步突出关键特征;Inter 通道结合用户行为特征与用户社交关联关系构建异构图,依托 GraphSAGE 模型迭代聚合相邻用户节点特征,挖掘群体关联关系中潜藏的异常行为模式,从群体行为维度补充个体行为检测信息。模型通过加权求和策略融合双通道输出特征,作为下游分类模块的输入。分类模型选用卷积神经网络学习特征与标签之间的映射关系,实现对恶意用户行为的准确判断。双通道特征加权融合检测模型架构如图 1 所示。

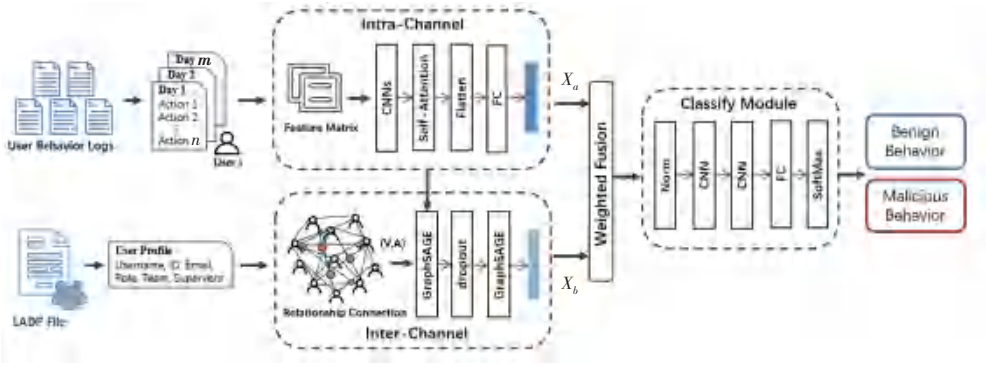


图 1 双通道加权融合检测模型

Fig. 1 Dual-channel weighted fusion detection model

2.1 数据预处理与特征选取

内部威胁数据集由多个子表组成,记录员工各种活动的具体操作以及对应时间戳,或者员工个人信息。通过整合这些多源日志数据,为每个用户构建一个独立的行为序列集合 $U_i(i=1,2,\cdots,m)$,其中 m 表示用户总数,集合 U_i 中包含某个用户的所有操作记录。由于用户的日常行为习惯在一天内的表现更为显著,在完成数据格式转换以及辅助信息补充后,以每日为时间粒度,设计一个全面覆盖用户行为属性的特征选取方法。

本文基于用户行为序列集合 U_i,从数据记录和事件频率两个维度进行特征选取。在数据记录方面,梳理或统计一天内各类活动的操作数据,如首次登录时间、最后登出时间、操作文件类型数量、发送邮件平均大小等。在事件频率方面,统计一段时间内各类活动发生的频次,其中时间段包括一天内、工作时间内和非工作时间内,如一天的登录总数、工作时间内登录次数、非工作时间内外接设备连接次数、非工作时间内邮件发送数量等。若用户在某一天中存在恶意操作,则该天的行为特征被标记成恶意。其中 08:00 至 18:00 为正常工作时间,不在这一时间范围内的活动则被视为非工作时间事件。

基于定义的时间粒度和特征选取方法,生成每位员工各自的行为特征集合 X_{u_i},则某位用户 u_i 的第 t 天的行为特征向量表示为 $x_{u_i}^t$,行为特征向量实例如图 2 所示。

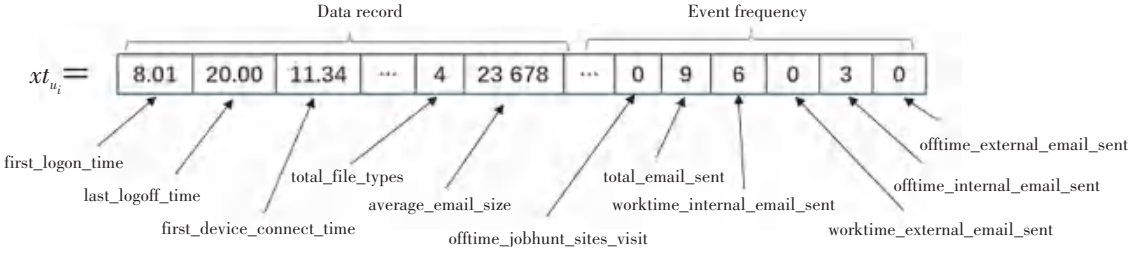


图 2 行为特征向量实例

Fig. 2 Example of behavior feature vector

2.2 基于 CSAtt 的 Intra 通道

在用户行为分析中,登录时长、外接设备使用频率、不同类型网址浏览频率、邮件往来情况等多种因素,均会对用户行为的判断产生影响,并且这些因素之间可能会存在复杂的非线性关系。因此,本文在 Intra 通道中引入一个结合卷积神经网络和自注意力机制的 CSAtt 特征分析模型。卷积神经网络凭借其出色的局部特征捕捉能力,可有效提取用户行为

特征在空间维度上的局部关联模式。例如，当特征向量中同时出现“外接设备使用频率较高”“邮件往来频繁”“浏览了求职网站”等特征时，卷积核可通过学习合适的权重来提取这种局部的高风险行为模式。自注意力机制能够跨越特征中的不同位置，在分析较长跨度的特征元素之间的关系时，克服了卷积操作在长距离依赖处理时的局限性，并且能动态地为各部分特征分配不同的权重，抑制一些对用户行为判断影响较小的特征元素，从而聚焦于更关键的特征部分。两者结合能够捕捉用户行为特征的内在联系，并增强关键特征的表达能力。CSAtt 模型结构如图 3 所示。

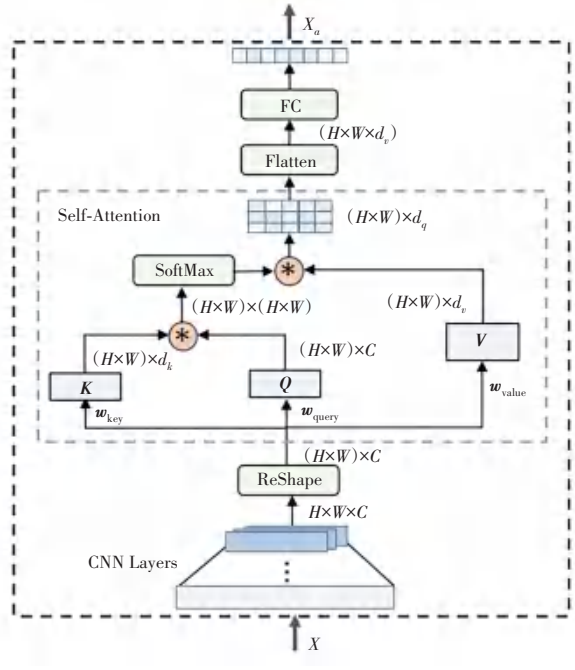


图 3　CSAtt 模型结构

Fig. 3　CSAtt model structure

首先，将用户原始特征向量集合 X 输入卷积层，利用卷积核权重学习特征中的内在结构和模式，每次卷积操作后引入激活函数处理，增强其对复杂特征的表达能力，公式如下：

$$x_{u_i}^{t\,(l)} = \sigma\left(w^{(l)} * x_{u_i}^{t\,(l-1)} + \boldsymbol{b}^{(l)}\right) \tag{1}$$

其中，$x_{u_i}^{t\,(l)}$ 表示 $x_{u_i}^{t}$ 在第 l 层卷积的输出特征，$x_{u_i}^{t\,(0)} = x_{u_i}^{t}$，$x_{u_i}^{t} \in X$；$\sigma(\cdot)$ 表示 ReLU 激活函数；$w^{(l)}$ 和 $\boldsymbol{b}^{(l)}$ 表示第 l 层的卷积核权重和偏置向量。

将卷积层的最终输出重塑为二维特征，以满足与自注意力机制中的 3 个初始权重矩阵进行线性交换的条件，公式如下：

$$\widetilde{x_{u_i}^{t}} = \text{ReShape}\left(x_{u_i}^{t\,(L_c)}, (H \times W) \times C\right) \tag{2}$$

其中，L_c 表示 Intra 通道实际使用的卷积层数；$x_{u_i}^{t\,(L_c)}$ 表示 $x_{u_i}^{t}$ 经过卷积层的最终输出；H、W、C 分别表示 $x_{u_i}^{t\,(L_c)}$ 的高度、宽度和通道数。

自注意力层首先初始化 3 个权重矩阵：$\boldsymbol{w}_{\text{key}}$、$\boldsymbol{w}_{\text{query}}$ 和 $\boldsymbol{w}_{\text{value}}$，随后将这 3 个权重矩阵分别与 $\widetilde{x_{u_i}^{t}}$ 进行矩阵相乘，得到相应的特征权重矩阵，公式如下：

$$\boldsymbol{K} = \widetilde{x_{u_i}^{t}} \cdot \boldsymbol{w}_{\text{key}} \tag{3}$$

$$\boldsymbol{Q} = \widetilde{x_{u_i}^{t}} \cdot \boldsymbol{w}_{\text{query}} \tag{4}$$

$$\boldsymbol{V} = \widetilde{x_{u_i}^{t}} \cdot \boldsymbol{w}_{\text{value}} \tag{5}$$

下一步通过 $\boldsymbol{K}$ 和 $\boldsymbol{Q}$ 的点积衡量特征之间的相关度来计算注意力分数，若两个特征点之间的关联性较高，在计算时会给予较高分数。此时引入一个缩放因子 $\sqrt{d_k}$（d_k 为 K 的维度值）控制点积结果的大小，避免产生梯度问题。然后使用 Softmax 函数将注意力分数进行归一化得到注意力权重矩阵，再将其应用到 $\boldsymbol{V}$ 向量进行加权求和，得到包含注意力权重的特征矩阵，公式如下：

$$\boldsymbol{x}_{\text{SAtt}\,u_i}^{\ t} = \text{Softmax}\left(\frac{\boldsymbol{Q}\cdot\boldsymbol{K}^{\text{T}}}{\sqrt{d_k}}\right)\cdot\boldsymbol{V} \tag{6}$$

最后，将自注意力层输出，引入展平层进行扁平化处理，再使用全连接层将扁平化的特征图重塑为固定大小的一维特征向量，公式如下：

$$x_{a\,u_i}^{\ t} = \boldsymbol{w}_{fc} \cdot \left(\text{flatten}\left(x_{\text{SAtt}\,u_i}^{\ t}\right)\right) + \boldsymbol{b}_{fc},\ \forall x_{a\,u_i}^{\ t} \in X_a \tag{7}$$

其中，$\boldsymbol{w}_{fc}$、$\boldsymbol{b}_{fc}$ 表示全连接层的权重矩阵和偏置向量，X_a 表示 Intra 通道最终输出特征的集合。

2.3　基于 GraphSAGE 的 Inter 通道

在 Inter 通道中，本文采用图神经网络处理用户行为特征和用户关系网络中的结构化信息，核心思想是通过聚合邻居节点的特征信息来更新自身节点的特征表示，旨在分析关联节点之间的特征，并挖掘不同用户间的潜在关联模式与群体行为倾向。

首先，基于用户行为特征与企业员工组织关系网络构建异构图，该过程主要分 3 步：(1)将用户 u_i 第 t 天的行为视为一个独立的节点 $v_{u_i}^{t}$；(2)节点 $v_{u_i}^{t}$ 对应属性特征为 $x_{a\,u_i}^{\ t}$，即用户 u_i 第 t 天的行为特征 $x_{u_i}^{t}$ 经过 Intra 通道处理后得到的输出特征；(3)结合员工信息与表 1 中的规则为节点之间构建无向边，得到邻接矩阵 $\boldsymbol{A}$。经过以上过程，完成了图的构建以及节点特征的嵌入，得到 GraphSAGE 模型的输入 $\boldsymbol{G} = (V, \boldsymbol{A}, X_a)$。

表 1 边构建规则

Table 1 Edge construction rules

编号	规则描述
规则 1	同一用户 u_a,对于不同日期的第 m 天和第 n 天,在 $V_{u_a}^m$ 和 $V_{u_a}^n$ 之间添加无向边
规则 2	若 u_a、u_b 为相同的岗位角色,对于任意第 m 天和第 n 天,在 $V_{u_a}^m$ 和 $v_{u_b}^n$ 之间添加无向边
规则 3	若 u_a、u_b 属于同一团队,对于任意第 m 天和第 n 天,在 $V_{u_a}^m$ 和 $v_{u_b}^n$ 之间添加无向边
规则 4	若 u_a、u_b 有相同的直属主管,对于任意第 m 天和第 n 天,在 $v_{u_a}^m$ 和 $v_{u_b}^n$ 之间添加无向边
规则 5	若 u_a 是 u_b 的直属主管,对于任意第 m 天和第 n 天,在 $v_{u_a}^m$ 和 $v_{u_b}^n$ 之间添加无向边

考虑实际场景中海量的用户数据和复杂的用户关系,本文使用基于归纳式学习法的 GraphSAGE 模型[22],该模型主要包含采样和聚合两大模块。在 GraphSAGE 的每一层迭代中,针对图 $G(V,\boldsymbol{A},X_a)$ 中任意目标节点 $v_{u_i}^t$ 执行如下操作:(1)对其邻居节点进行采样,得到采样后的邻居集合;(2)采用均值聚合方法对采样集合内的所有节点特征做平均运算,得到邻居聚合特征;(3)将邻居聚合特征与目标节点特征进行拼接融合,再经过线性变换与激活函数处理,得到目标节点在该层新表征。当层数进一步增加时,目标节点会重复上述步骤持续更新其在图中的特征表示。经过 l 层 GraphSAGE 迭代后,节点 $v_{u_i}^t$ 的特征更新过程公式如下:

$$x_{N_{u_i}}^{t(l)} = \frac{1}{|N(v_{u_i}^t)^{(l)}|} \cdot \sum_{v_{u_j}^k \in N(v_{u_i}^t)^{(l)}} x_{gu_j}^{k\ (l-1)} \tag{8}$$

$$x_{g_{u_i}}^{t(l)} = \sigma(\mathrm{concat}(W_1^{(l)} \cdot x_{g_{u_i}}^{t(l-1)}, W_2^{(l)} \cdot x_{N_{u_i}}^{t(l)})) \tag{9}$$

$$x_{b_{u_i}}^t = x_{g_{u_i}}^{t(\mathrm{Lg})}, \forall x_{b_{u_i}}^t \in X_b \tag{10}$$

其中,$N(v_{u_i}^t)^{(l)}$ 表示 $v_{u_i}^t$ 在第 l 层时的采样邻居节点集合;特征初始化满足 $x_{g_{u_i}}^{t(0)} = x_{a_{u_i}}^t, x_{a_{u_i}}^t \in X_a$;$W_1^{(l)}$、$W_2^{(l)}$ 分别表示第 l 层聚合器的融合权重;$\sigma(\cdot)$ 表示 ReLU 激活函数;L_g 表示该通道中 GraphSAGE 的实际使用层数;X_b 表示 Inter 通道所有输出特征的集合。

两层 GraphSAGE 模型的特征提取过程如图 4 所示。在第一层中,目标节点通过聚合其采样邻居的原始特征来更新自身的特征表示,同时,目标节点的邻居也同步执行相同的采样与聚合操作。在第二层中,由于目标节点的邻居已经聚合来自其邻居的信息,当目标节点再次实行采样和聚合操作时,便能够实现将远处间接关联的节点信息融合至目标节点上。图中所有节点在每层均会同步进行上述的操作,最终生成新的图嵌入表示 $\widetilde{G} = (V,\boldsymbol{A},X_b)$。

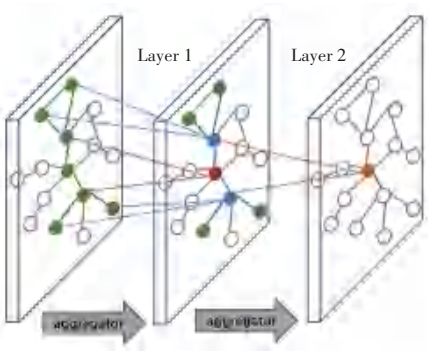


图 4 两层 GraphSAGE 特征提取过程

Fig. 4 Two-layer GraphSAGE feature extraction process

2.4 特征融合与内部威胁检测

在获取双通道各自提取的特征后,本文采用加权特征融合策略,相较于传统的简单相加融合,引入了 γ_1 和 γ_2 两个可学习权重参数,公式如下:

$$\widetilde{\boldsymbol{X}} = \gamma_1 X_a + \gamma_2 X_b \tag{11}$$

其中,$\widetilde{\boldsymbol{X}}$ 为双通道特征加权融合后得到的综合特征向量;X_a 为单通道网络提取的用户个体行为特征;X_b 为 GraphSAGE 挖掘得到的用户群体关联图特征。

该自适应加权融合策略不仅有效融合两个通道的关键信息,还能在模型训练的过程中根据特征对模型分类的影响程度动态调整 X_a 和 X_b 的权重占比。为确保特征数据的稳定性与一致性,进一步执行特征归一化处理。随后采用 CNN 分类模型挖掘融合特征的深层次信息,学习融合特征与标签之间的映射关系,完成最终的分类任务。

3 实验与结果分析

3.1 数据集

卡耐基梅隆大学发布的 CERT 内部威胁数据集[23]是内部威胁检测领域使用最广泛的公开数据集。该数据集包含企业内部员工的网络操作行为日志和个人信息,具体内容见表 2。

CERT r4.2 包含更为丰富的威胁场景以及较高的恶意用户比例,该数据集记录了企业中 1 000 名员工在 18 个月内的网络操作行为,共 322 万条日志记录,其中仅包括 70 名恶意内部人员,7 423 条恶意

操作日志。经过预处理得到正常标记数据 355 046 条，异常标记数据仅 956 条。鉴于整体数据高度不平衡，本文使用结合随机过采样和降采样技术的混合采样策略，构建一个均衡的样本数据集，减少类别不平衡对模型性能的影响。该样本包含 3 826 条用户行为特征数据，按照 7∶3 划分训练集和测试集，基于此样本构建了包含 785 624 条边的无向图，其中图节点平均度为 205。

表 2　CMU-CERTr4.2 数据集描述

Table 2　CMU-CERTr4.2 data set description

文件名称	内容描述
logon.csv	用户登录/注销信息
device.csv	用户连接/断开外接驱动器信息
file.csv	用户复制/修改文件操作信息，包括文件类型、大小、内容等
email.csv	用户发送内部/外部邮件的信息，包括收件人、邮件大小、内容等
http.csv	用户浏览网站的 URL、网页内容等
LDAP	企业所有员工信息，包括岗位角色、部门、工作团队、主管等

3.2　实验环境与参数设置

实验使用 Windows11 操作系统、GeForce RTX 4060(16 GB)显卡、Python 语言和 Pytorch2.1 框架，CUDA 版本为 12.1。本文 CSAtt 中有 4 个参数相同的卷积层，其中 filter 个数均为 16，卷积核大小为 3，步长为 1，全连接层输出维度为 20；GraphSAGE 层数为 2，输出维度分别为 64 和 20，邻居采样数量为 100，dropout 比率为 0.2；CNN 分类模型有 2 个卷积层，其中 filter 个数分别为 8 和 16，卷积核大小均为 3，步长为 1，采用最大池化。在训练过程中，采取全批次训练，学习率设置为 0.01，迭代次数为 500 次，使用 Adam 优化器和交叉熵损失函数。

3.3　对比实验

本文将所提出的检测方法与传统机器学习和深度学习方法进行对比分析，实验结果见表 3。

通过对比发现，决策树(Decision Tree，DT)、随机森林(Random Forest，RF)和高斯朴素贝叶斯(Gaussian Naive Bayes，GNB)三种经典机器学习方法均呈现良好的分类性能，但与表现较好的 RF 分类器相比，本文模型在准确率、精确率和 $F1$ 值上分别提升了 3.66%、7.49%和 3.56%，说明本文模型检测效果更精准可靠。

在深度学习技术方面，CNN 方法的学习分类效果优于机器学习方法，但缺乏对群体行为的分析，导致其检测效果受限。GCN 和 GraphSAGE 均使用图神经网络考虑节点之间关系并分析节点特征。在较大规模的用户关联图数据中，GCN 需要考虑所有邻居节点进行卷积操作，计算开销大且容易出现过拟合现象，导致分类性能不佳。GraphSAGE 能够通过合理的采样，关注有代表性的邻居节点，在保证性能的同时进行有效学习，但是在图传播更新的过程中，会削弱原始特征中与恶意行为高相关特征的表达能力。CNN-GCN 方法分别提取用户个人特征和关联特征信息并融合，后续完成恶意行为分类。实验结果表明，本文模型检测效果在准确率、精确率和 $F1$ 值上分别达到 95.67%、96.32% 和 95.37%，优于单一深度学习方法以及基于 CNN-GCN 特征融合的检测方法，说明本文提出的基于 CSAtt-GraphSAGE 双通道特征加权融合的检测方法在特征分析和恶意行为检测方面更具优势。

表 3　不同方法对比实验结果

Table 3　Comparative experimental results of different methods

Method	Accuracy/%	Precision/%	$F1$/%
DT	91.48	88.91	91.20
RF	92.01	88.83	91.81
GNB	89.25	86.39	90.98
CNN	92.26	90.34	91.95
GCN	73.91	66.32	76.68
GraphSAGE	93.32	91.64	93.02
CNN-GCN	94.17	93.61	93.61
Proposed	**95.67**	**96.32**	**95.37**

3.4　消融实验

为验证各模块对整体检测性能的影响，进行了一系列消融实验，实验结果见表 4。首先，移除 Intra 通道，将原始特征输入 Inter 通道进行分析提取，再将通道输出特征与原始特征进行加权融合。其次，移除 Inter 通道，将经过 Intra 通道处理的特征直接输入检测模块。最后，将 Intra 通道和 Inter 通道全部移除，原始特征直接输入卷积神经网络中进行分类检测。实验结果表明，两个特征提取通道对于模型的性能均有提升作用，其中 Inter 通道的添加对模型精确率的提升最为显著，说明对群体行为模式的补充分析很有必要，在提升模型整体检测性能的同时，减少了对正常样本的误判。

根据计算得到的 FPR 和 TPR 绘制了消融实验中不同模型结构下的 ROC 曲线(见图 5)，其中，采用了双通道特征加权融合方法的模型对应的曲线

AUC = 0. 976 1,表现最优。移除双通道中任意模块,AUC 值均会产生明显下降,这与表 4 的数据表现一致。

表 4 消融实验结果

Table 4 Ablation experiment results

Model structure	Accuracy/%	Precision/%	*F*1/ %
Proposed	**95. 67**	**96. 32**	**95. 37**
Without Intra-Channel	94. 62	94. 18	94. 31
Without Inter-Channel	93. 05	92. 29	92. 67
Without Dual-Channel	92. 26	90. 34	91. 95

常见的特征融合策略有拼接和相加。为深入探究不同融合策略在本文所提出的检测方法中的表现,本文通过实验对比了简单特征拼接、特征相加以及加权特征拼接、相加这几种融合方法。结果见表 5 所示,采用加权相加策略时模型的检测效果最优。

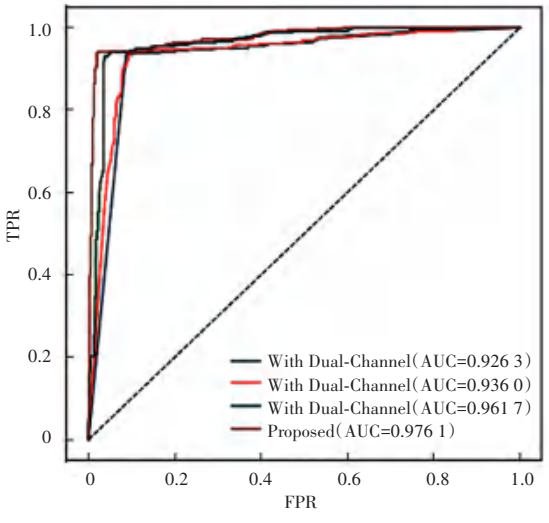


图 5 不同方法下的 ROC 曲线

Fig. 5 ROC curves under different methods

表 5 不同融合策略的比较

Table 5 Comparison of different fusion methods

Fusion method	Accuracy/%	Precision/%	*F*1/ %
Concatenate	95. 02	95. 48	94. 67
Weight Concatenate	95. 54	95. 78	95. 25
Add	95. 15	96. 27	94. 78
Weight Add	**95. 67**	**96. 32**	**95. 37**

另外,本文采用了在一维特征向量处理领域被广泛认可的深度学习分类模型,其中包括多层感知机(MLP)、循环神经网络(RNN)和长短时记忆网络(LSTM),对比各模型在是否使用双通道特征加权融合方法的情况下的检测效果,结果见表 6。从中可见,在引入双通道进行特征加权融合后,各个模型的检测性能得到了显著提升。从而证明双通道特征加权融合在提升异常检测性能方面具有普遍适用性。

表 6 不同分类模型的效果

Table 6 Effect of different classification models

Method	Dual-Channel status	Accuracy/%	Precision/%	*F*1/ %
MLP	without Dual-Channel	90. 03	85. 67	89. 97
	with Dual-Channel	**93. 97**	**93. 61**	**93. 61**
RNN	without Dual-Channel	89. 77	85. 78	89. 65
	with Dual-Channel	**94. 62**	**95. 71**	**94. 22**
LSTM	without Dual-Channel	90. 17	85. 89	90. 09
	with Dual-Channel	**95. 15**	**94. 98**	**94. 85**

4 结束语

针对现有内部威胁检测方法普遍忽略用户之间潜在关联关系的问题,本文提出了基于双通道特征加权融合的企业内部网络威胁检测方法。其中,Intra 通道采用基于自注意力机制的 CSAtt 模型分析和提取有价值的关键特征,Inter 通道借助 GraphSAGE 模型挖掘用户间关联特征与群体行为倾向,双通道输出特征经过加权融合,使后续分类任务能充分结合个体与群体两个方面,理解用户行为的复杂性。基于 CERTr4. 2 数据集开展对比实验,结果验证了本文 CSAtt-GraphSAGE 双通道特征加权融合检测方法的有效性,从多元视角挖掘特征并综合分析判断,提升了恶意行为检测的准确性,同时降低了误报率,为内部威胁检测领域提供了新的思路和解决方案。

参考文献

[1] Homoliak I, Toffalini F, Guarnizo J, et al. Insight into insiders and IT: A survey of insider threat taxonomies, analysis, modeling, and countermeasures[J]. ACM Computing Surveys, 2020, 52(2): 1-40.

[2] Al-Shehari T, Kadire M, Al-Mhiqani M N, et al. Comparative evaluation of data imbalance addressing techniques for CNN-based insider threat detection[J]. Scientific Reports, 2024, 14(1): 24715.

[3] 崔景洋, 陈振国, 田立勤, 等. 基于机器学习的用户与实体行为分析技术综述[J]. 计算机工程, 2022, 48(2):10-24.

[4] Yuan Shuhan, Wu Xintao. Deep learning for insider threat detection: review, challenges and opportunities[J]. Computers & Security, 2021, 104: 102221.

[5] Gong Yiru, Cui Susu, Liu Song, et al. Graph-based insider threat detection: A survey[J]. Computer Networks, 2024, 25: 110757.

[6] Le D C, Zincir-Heywood A N. Evaluating insider threat detection

workflow using supervised and unsupervised learning [C]// Proceedings of 2018 IEEE Security and Privacy Workshops. Piscataway, NJ: IEEE, 2018: 270-275.
[7] Liu Chenchen, Zhong Yi, Wang Yuanli. Improved detection of user malicious behavior through log mining based on IHMM[C]// Proceedings of the 5th International Conference on Systems and Informatics. Piscataway, NJ: IEEE, 2019: 1193-1198.
[8] 孙丹，饶兰香，施炜利，等. 基于混合 N-Gram 模型和 XGBoost 算法的内部威胁检测方法[J]. 计算机与现代化，2022(8):99-105.
[9] 李俊强，黄洪，周子云. 基于用户画像的自适应内部威胁检测模型[J]. 网络安全与数据治理，2022，41(11):43-48.
[10] Bertrand S, Desharnais J, Tawbi N. Unsupervised user-based insider threat detection using Bayesian Gaussian mixture models [C]//Proceedings of the 20th Annual International Conference on Privacy, Security and Trust. Piscataway, NJ: IEEE, 2023: 1-10.
[11] Al-Shehari T, Al-Razgan M, Alfakih T, et al. Insider threat detection model using anomaly-based isolation forest algorithm [J]. IEEE Access, 2023, 11: 118170-118185.
[12] 郭世泽，张磊，潘雨，等. 内部威胁发现检测方法研究综述[J]. 数据采集与处理，2022，37(3):488-501.
[13] Saaudi A, Al-Ibadi Z, Tong Y, et al. Insider threats detection using CNN-LSTM model[C]//Proceedings of 2018 International Conference on Computational Science and Computational Intelligence. Piscataway, NJ: IEEE, 2018: 94-99.
[14] 黄娜，何泾沙，吴亚飚，等. 基于 LSTM 回归模型的内部威胁检测方法[J]. 信息网络安全，2020，20(9):17-21.
[15] 孙小双，王宇. 基于自动编码器的内部威胁检测技术 [J]. 计算机工程与设计，2022，43 (10): 2725-2730.
[16] Anju A, Krishnamurthy M. M-EOS: Modified-equilibrium optimization-based stacked CNN for insider threat detection[J]. Wireless Networks, 2024, 30: 2819-2838.
[17] 张光华，闫风如，张冬雯，等. 基于 LSTM-Attention 的内部威胁检测模型[J]. 信息网络安全，2022，22(2):1-10.
[18] Liu Fucheng, Wen Yu, Zhang Dongxue, et al. Log2vec: A heterogeneous graph embedding based approach for detecting cyber threats within enterprise [C]// Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security. New York: ACM, 2019: 1777-1794.
[19] Jiang Jianguo, Chen Jiuming, Gu Tianbo, et al. Anomaly detection with graph convolutional networks for insider threat and fraud detection [C]// Proceedings of 2019 IEEE Military Communications Conference. Piscataway, NJ: IEEE, 2019: 109-114.
[20] Li Chaoyang, Li Fenghua, Yu Mingjie, et al. Insider threat detection using generative adversarial graph attention networks [C]// Proceedings of 2022 IEEE Global Communications Conference. Piscataway, NJ: IEEE, 2022: 2680-2685.
[21] Tian Tian, Gong Yiru, Jiang Bo, et al. Insider threat detection based on heterogeneous graph neural network[C]//Proceedings of 2023 IEEE 22nd International Conference on Trust, Security and Privacy in Computing and Communications. Piscataway, NJ: IEEE, 2023: 628-635.
[22] Hamilton W L, Ying R, Leskovec J. Inductive representation learning on large graphs[J]. arXiv, 1706.02216, 2017.
[23] Lindauer B. Insider threat test dataset [DB/OL]. (2020-09-30)[2026-06-10]. DOI: 10.1184/R1/12841247.v1.