面向内部用户行为异常检测的图掩码对比学习方法

袁璐[1]，张斌[1]，姜迎畅[1]，房礼国[1]，孙剑文[2]，李红宇[1]，马骏[1]
（1. 信息工程大学密码工程学院，河南 郑州 450001；2. 32708 部队，北京 102300）

摘 要：针对现有内部用户行为异常检测方法在时空关联表征能力较弱、隐蔽异常行为难以识别以及对正常业务波动鲁棒性较弱的问题，提出了一种面向内部用户行为异常检测的图掩码对比学习方法 TGMC。首先，将多源异构日志按时间窗口构建为异构图序列，刻画用户与业务资源实体间的交互关系。然后，设计由图注意力网络与门控循环单元构成的混合编码器，分别提取空间拓扑特征与时序变化特征，实现时空维度的深度耦合。引入掩码重建机制，通过邻域上下文恢复被掩码节点的特征与边结构，挖掘邻域关联模式以提高隐蔽异常行为的检测能力。设计多视图对比学习机制，通过构造增强视图约束同一用户跨视图表示的语义一致性并拉大不同用户间的表示距离，学习稳定的行为表征以提升识别鲁棒性。在 CERT r4.2 与 r5.2 数据集上的实验结果表明，TGMC 方法与基线方法相比，TPR 分别提升 3.59% 和 1.59%，FPR 分别降低 5.45% 和 2.49%。

关键词：图表示学习；对比学习；内部用户行为；鲁棒性
中图分类号：TP393.08
文献标志码：A
DOI：10.11959/j.issn.1000

A Graph Masked Contrastive Learning Method for Insider User Behavior Anomaly Detection

YUAN Lu[1], ZHANG Bin[1], JIANG Yingchang[1], FANG Liguo[1], SUN Jianwen[2], LI Hongyu[1], MA Jun[1]
1. Cryptography Engineering Institute, Information Engineering University, Zhengzhou 450001, China
2. Unit 32708, Beijing 102300, China

Abstract: To address the limitations of existing insider user behavior anomaly detection methods, including weak spatiotemporal correlation representation, difficulty in identifying stealthy anomalous behaviors, and poor robustness to normal business fluctuations, we propose TGMC, a graph masked contrastive learning method for insider user behavior anomaly detection. First, multi-source heterogeneous logs are organized into a sequence of heterogeneous graphs using time windows to characterize interaction relationships between users and business resource entities. Next, a hybrid encoder composed of a graph attention network and a gated recurrent unit is designed to extract spatial topological features and temporal evolution features, respectively, enabling deep coupling across spatiotemporal dimensions. A masked reconstruction mechanism is introduced to recover the features and edge structures of masked nodes from neighborhood context, mining neighborhood correlation patterns to enhance the detection capability for stealthy anomalous behaviors. In addition, a multi-view contrastive learning mechanism is devised: by constructing augmented views, it enforces semantic consistency of the same user's representations across views while enlarging the representation distance between different users, thereby learning stable behavior representations and improving recognition robustness. Experiments on the CERT r4.2 and r5.2 datasets demonstrate that, compared with baseline methods, TGMC improves TPR by 3.59% and 1.59%,

收稿日期：2019–08–16；修回日期：2019–12–17
通信作者：袁璐，yuanlu_l@163.com
基金项目：国家自然科学基金资助项目（No.62477020）；河南省科技攻关项目（No. 252102211040）
Foundation Items：The National Natural Science Foundation of China (No. 62477020), The Science and Technology Research Project of He'nan Province (No. 252102211040)

and reduces FPR by 5.45% and 2.49%.

Keywords: graph representation learning, contrastive learning, insider user behavior, robustness

0 引言

随着信息技术的飞速发展与广泛应用，内部威胁已成为影响信息系统安全的主要因素之一。内部员工拥有合法身份和业务权限，其不当行为可能直接威胁系统的机密性、完整性与可用性[1]。Ponemon 研究所发布的《2025 年 Ponemon 内部威胁报告》指出，2025 年内部威胁事件数量和损失持续上升，单个企业因内部风险造成的年平均损失约 1740 万美元，且事件平均需 80 天以上才能完全遏制[2]。

内部用户异常行为通常可分为显性异常与隐蔽异常两类。显性异常特征明显，易于通过统计特征进行识别，例如短时间内批量拷贝大量敏感文件。隐蔽异常通常表现为将敏感文件的上传或外发操作分散至多个工作时段，或在工作时段访问求职招聘网站等。此类异常从单日行为量与历史基准无显著差异，往往隐匿于大量正常行为之中，基于静态规则的检测方法难以及时、准确识别[3]。因此，如何应对异常行为的隐蔽性并提升检测模型在复杂场景下的鲁棒性，已成为内部用户行为异常检测亟需解决的问题之一。

目前，内部用户行为异常检测方法主要分为基于特征工程、基于序列与基于图表示学习的方法。基于特征工程的检测方法通过统计分析提取用户行为的离散特征来建立基准模式[4-8]。但在处理海量多源的日志数据时，该类方法高度依赖专家经验进行特征设计，且难以捕捉非线性的复杂交互模式[9-10]。基于序列的检测方法将用户活动视为时间序列数据，通常按会话或天为单位划分数据。利用 LSTM、GRU 等循环神经网络捕捉用户行为在时间维度上的演变规律，预测特定时间段内的行为序列是否偏离历史基准[11-17]。尽管此类方法在时序模式挖掘上表现优异，但往往忽略了用户与网络实体之间复杂的交互拓扑结构，难以全面刻画用户行为模式。为了弥补序列方法的不足，近年来研究人员开始引入图神经网络（GNN，graph neural network）来建模实体间的交互关系[18-23]。此类方法将日志数据构建为图，通过聚合邻域信息来学习节点的高阶语义表示，从而有效捕捉用户与资源之间的结构化依赖[24-25]。但在应对现实环境中复杂的内部威胁场景时仍存在以下不足。

1) 内部用户行为具有复杂的空间交互拓扑与时间演化特征。现有方法多侧重单一维度建模[19]，忽略了时空维度的耦合依赖关系，难以有效融合历史时空上下文信息，导致异常行为检测能力不足。

2) 异常用户行为具有高隐蔽性，恶意用户可通过调控访问频次、登录时段等统计特征以伪装成正常行为。现有方法主要依赖于易受操控的用户节点自身特征，忽略了其与业务资源节点间的交互关系，难以有效识别隐藏的异常行为。

3) 用户行为模式容易因合法业务变化而产生波动。现有方法在训练时通常假设数据完整准确且行为模式稳定，学习的表示容易对特定行为模式产生过拟合，在噪声扰动环境下判别能力有限。

针对上述不足，提出一种面向内部用户行为异常检测的图掩码对比学习方法（TGMC，Temporal Graph-based Masked reconstruction and Contrastive learning）。本文的主要贡献如下：

1) 提出了一种时空掩码重建机制。引入 GAT 与 GRU 分别捕获用户与实体间的交互拓扑及行为演化趋势，实现用户行为时空维度的深度耦合。引入随机掩码策略，基于邻域上下文重建缺失信息，使模型深度挖掘用户行为的邻域关联模式，从而有效识别隐蔽异常行为。

2) 设计了多视图对比增强机制。通过构建增强视图使同一用户在不同视图下的行为语义保持一致，同时拉开不同用户行为表示的差异，学习更稳定的用户行为表征，有效缓解了对特定行为模式的过拟合现象，提升了对恶意行为的识别鲁棒性。

3) 对 TGMC 的鲁棒性进行了理论分析，该方法通过时序衰减机制抑制扰动累积，增强对结构缺失的鲁棒性。在 CERT r4.2 和 r5.2 数据集上的实验结果表明，TGMC 在 AUC、TPR 等指标上优于基线方法。

1 相关工作

1.1 基于特征工程的用户行为异常检测

早期的研究主要采用基于特征工程的内部用户行为异常检测方法。该类方法通常从日志中提取离

散的行为特征，如用户登录频率、文件访问次数、数据拷贝行为等等，随后利用支持向量机（SVM，support vector machine）、随机森林（RF，random forest）或孤立森林（iForest，isolation forest）等传统机器学习算法进行异常分类[30]。

Le 等人[4]基于来自多源日志的用户行为数据，在不同时间粒度（会话级、日级和用户级）下进行特征工程，并采用随机森林、支持向量机等传统机器学习模型对恶意内部人员进行识别。Randive 等人[5]提出一种基于图像的特征表示方法，将用户行为特征转换为图像表示，并利用小波卷积神经网络从频谱和空间维度提取特征以识别恶意内部人员。Duan 等人[6]在此基础上引入上下文感知特征表示机制，通过融合用户行为的上下文信息来增强图像特征表达能力，使用预训练的 ResNet 模型来检测异常用户。Yuan 等人[7]提出了一种基于时间窗口的群体行为异常检测方法，通过在固定时间窗口内聚合个体与群体特征，并比较用户行为与群体基线之间的偏离程度来识别异常用户。Kan 等人[8]提出了一种结合数据调整策略与优化 XGBoost 算法，通过欠采样与样本重加权等策略缓解数据不平衡问题，然后利用改进的 XGBoost 分类器对内部威胁进行识别。该类方法通过特征工程从多源日志中提取并聚合离散的统计特征，以建立检测用户行为的基准。然而，这些方法多依赖人工设计的统计特征或静态特征聚合，难以捕捉用户行为的时序演化和复杂关联关系，在面对隐蔽性较强的内部威胁时鲁棒性和泛化能力仍有限。

1.2　基于序列的用户行为异常检测方法

鉴于用户行为具有显著的时间依赖性，许多研究将用户行为日志建模为时间序列，利用深度学习模型捕捉时序模式[31-32]。此类方法通常按会话或天为单位对用户活动进行切片，利用循环神经网络（RNN）、长短期记忆网络（LSTM）或 Transformer[33]架构来学习正常行为的历史演变规律，并预测当前时刻的行为是否偏离基准。

早期研究多基于 LSTM 的序列预测框架建立正常行为基准，以预测偏差作为异常判据[11-13]。在此基础上，部分研究进一步扩展了序列建模的表达能力。Paul 和 Mishra 提出了 LAC 模型，进一步将 LSTM 自编码器与群体行为信息相结合，以增强对异常行为的判别能力[14]。Manoharan 等人[15]利用 BiLSTM 聚合序列的双向时序依赖以获得用户行为表示，将序列嵌入与人工统计特征拼接，并输入分类器进行内部威胁识别，验证了双向时序特征对复杂行为模式建模的有效性。Chen 等人[16]提出 BST，将 Transformer 的自注意力机制引入用户行为序列建模，通过捕捉行为序列中的长程依赖关系来学习用户的行为模式。Zhou 等人[38]提出 FMLP，以可学习滤波器在频域对行为序列中的噪声交互信号进行过滤，在保持较低模型复杂度的同时有效抑制了噪声干扰。虽然序列模型在捕捉时间依赖方面表现优异，但它们通常将用户行为视为相互独立的单一序列，忽略了不同用户、主机或资源之间的交互关系。

1.3　基于图表示学习的用户行为异常检测方法

为了弥补序列模型在结构信息挖掘上的不足，研究者通过图来建模实体间的交互关系，将用户、主机、文件、进程及其交互行为建模为图结构，通过分析节点属性、边关系及整体拓扑特征来识别异常行为模式[34]。

基于图表示学习的内部用户行为异常检测方法通过建模用户间交互关系，为内部威胁检测提供了有效的结构信息支撑。Jiang 等人[18]和 Fei 等人[19]都采用 GCN 对用户行为图进行建模以识别异常节点与异常行为。尽管在捕捉用户之间的关系依赖方面有效，但在很大程度上忽略了涉及主机、文件和系统事件的更丰富的上下文关系。除 GCN 等图神经网络方法外，采用异构图与图嵌入学习进行威胁检测。Liu 等人[20]提出了 Log2Vec，该方法将企业系统日志建模为一个由用户、主机和事件组成的异构图，并通过随机游走和图嵌入技术学习低维嵌入。进一步地，Li 等人[21]提出了一种双域 GCN 模型，分别捕捉结构关系和基于属性的行为模式，从而实现更准确和自适应的异常检测。为提高检测方法在样本稀缺场景下的适应性与泛化能力，Li 等人[22]提出图元学习框架 GMFITD，使模型能够快速适应新或罕见攻击模式。Kong 等人[23]提出 DPI-ITD 双视角信息驱动框架，将基于图的内部威胁检测扩展至异构且资源受限的物联网系统。Cai 等人[35]提出 LAN，通过动态构建用户关联图并自适应选取邻居节点进行信息聚合，有效捕捉了用户行为的实时演化特征。尽管基于图表示学习的方法在结构关联挖掘上取得了进展，但现有的图方法多基于静态

图快照，难以同时兼顾细粒度的局部拓扑与长期的时序演化。此外，图神经网络往往抗噪能力不足，在面对数据稀疏或存在噪声干扰时，难以学习到稳定的节点表示。

综上所述，现有方法虽在内部用户行为异常检测任务中取得了一定成效，但也存在不同层面的局限性，综合对比结果如表 1 所示。

2　TGMC 方法

2.1　TGMC 框架

本文提出了 TGMC，其总体架构如图 1 所示，主要由 3 个模块组成：动态图构建、时空掩码重建和多视图对比增强。首先，动态图构建模块通过将原始日志划分为时间窗口并提取交互特征，构建有向异构图序列，实现了异构数据的统一结构化表

表 1　　现有方法对比

方法类型	文献	核心技术	优点	局限性
基于特征工程	文献[4]	多粒度分析，机器学习分类	能够从不同粒度挖掘特征，适应性强	依赖人工提取特征，难以捕捉复杂的长序列依赖
	文献[5]	单日行为特征图像化，CNN	将非结构化日志转化为图像，利用 CNN 提取深层模式	仍需手工构造单日特征，图像处理计算开销较大
	文献[6]	上下文感知图像表示，深度残差网络（ResNet）分类	引入上下文信息，增强了特征的抗噪性，特征融合灵活	依赖外部信息，流程更复杂、计算更重
	文献[7]	个体+群体行为建模，自编码器集成	有效识别相对于群体行为的异常突变	群体划分困难，组织结构变更适配成本高
	文献[8]	数据调整策略+优化 XGBoost	兼顾数据清洗与强分类器，适合不平衡场景	依赖特征与规则设计，复杂序关系表达能力有限
基于序列建模	文献[11]	深度神经网络，在线无监督检测	无需标签、可实时	对于超长序列的记忆能力有限，抗噪能力弱
	文献[12]	日志序列建模，LSTM	自动学习日志序列模式，适用于在线检测	对日志模板依赖较强，长序列建模能力有限
	文献[13]	LSTM，序列重构误差计算	可捕捉多日/长时依赖，适用于高维序列	训练推理开销较高，对异常阈值/场景迁移敏感
	文献[14]	LSTM 自编码器，社区发现算法	同时考虑时间异常和群体异常	社区划分依赖强；组织变动敏感
	文献[15]	BiLSTM + 序列上下文理解	利用过去和未来的上下文信息，提高了预测的准确性	窗口与标签质量敏感
	文献[16]	Transformer 建模用户行为序列	能捕捉长程依赖关系，适合复杂行为序列	计算复杂度高，对数据规模和算力要求较高
	文献[38]	基于 MLP 的序列建模，引入频域滤波增强	结构简单高效，训练速度快	对复杂时序依赖建模能力有限
基于图表示学习	文献[18]	同构图构建，GCN 聚合	利用结构信息，减少仅看个体特征导致的信息损失	动态图更新成本高，训练与存储开销较大
	文献[19]	图卷积神经网络（GCNN），邻域特征加权聚合，节点分类优化	用轻量关系特征替代复杂建图，提升效率	关系刻画设计依赖经验，对复杂多跳攻击表达可能不足
	文献[20]	异构图构建，图嵌入算法，基于随机游走的序列采样	异构关系建模更贴近企业环境；无需大量标签	构图规则复杂，维护与更新成本高
	文献[21]	双域图卷积网络（DD-GCN），多通道特征融合	从时空双域挖掘异常，解决了单一视角信息不足的问题	架构复杂，推理速度慢，对硬件资源要求极高
	文献[22]	GNN 特征提取，元学习， MAML 式训练	解决小样本问题，兼顾对抗鲁棒性	元学习训练不稳定，需要设计合理的元任务
	文献[23]	双视角信息驱动，符号化标记与冗余过滤	适应 IoT 异构数据，减少冗余片段干扰	依赖评分机制设计，针对 IoT 设计，通用性受限
	文献[35]	自适应邻居学习，动态图建模用户关系	动态建模用户间关系，增强结构表达能力	图构建与更新复杂，对计算资源要求较高

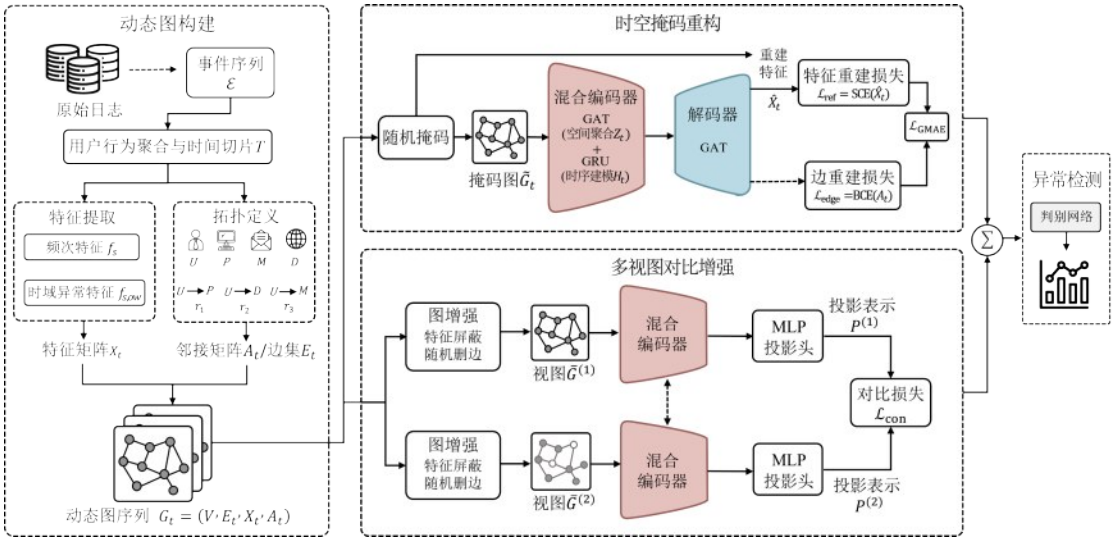


图1　TGMC 方法框架

示。其次，时空掩码重建模块利用 GAT 与 GRU 捕捉时空依赖，并通过掩码恢复任务挖掘节点间的内在语义关联，增强对隐蔽异常的识别能力。最后，多视图对比增强模块通过构造扰动视图并约束同一用户跨视图表示的语义一致性，使模型聚焦于行为本质语义特征，增强特征区分度与鲁棒性。通过这 3 个模块的协同工作，本文方法能够显著提升异常行为识别的准确性与鲁棒性。

2.2　动态图构建

设内部网络环境中的数据源集合为 $\mathcal{S}$，数据流被抽象成事件序列 $\mathcal{E}=\{e_1,e_2,\cdots,e_N\}$，其中，每个事件 e_i 表示为一个四元组 (u,d,t,type)，表明用户 u 在时刻 t 以操作类型 $\text{type}\in\mathcal{S}$ 与实体 d 进行的交互。以用户为中心的行为聚合，将不同的离散事件依据用户 u 进行关联与融合。

将连续时间离散化为固定长度的时间窗口 $T=\{t_1,t_2,\cdots,t_k\}$，由于用户行为具有周期规律性，所以将数据按照天为单位进行切片，落入窗口 t_k 内的事件子集记为 $\mathcal{E}^{(t_k)}$。

为了量化用户的行为模式并捕捉潜在的异常偏离，对于时间窗口 t_k 内的每个用户 u 提取特征向量。首先定义工作时段指示函数：

$$\mathbb{I}_{wh}(t)=\begin{cases}1, & t\in[T_{start},T_{end}]\\ 0, & otherwise\end{cases}\tag{1}$$

主要包含两类关键统计特征：

频次统计特征：针对不同的用户行为，分别计算其日累计发生次数，高频的操作峰值往往暗示着数据窃取或蓄意破坏的风险。对于任意操作类型 $s\in\mathcal{S}$：

$$f_s(u)=\sum_{e\in\mathcal{E}^{(t_k)}}\mathbb{I}(u_e=u\wedge s_e=s)\tag{2}$$

时域异常特征：引入工作时间作为基准，区分标准工作时间与非工作时间操作，分别统计了各类行为在非工作时间段的发生频次。

$$f_{s,ow}(u)=\sum_{e\in\mathcal{E}^{(t_k)}}\mathbb{I}(u_e=u\wedge s_e=s)\cdot\left(1-\mathbb{I}_{wh}(t_e)\right)\tag{3}$$

为了捕捉用户与网络实体之间复杂的交互拓扑结构，将结构化的日志数据转化为一系列按天生成的有向异构图。给定 $G_t=(V,E_t,X_t,A_t)$，其中，$V=U\cup P\cup D\cup M$，表示由用户 U、终端 P、网络域名 D 以及邮件域 M 构成的异构节点集合；E_t 为当日交互边集合，包含 $\mathcal{R}=\{r_1,r_2,r_3\}$ 三种关系类型，分别对应用户-终端、用户-域名及用户-邮件域的交互；$X_t\in\mathbb{R}^{|V|\times d}$ 为节点特征矩阵；$A_t\in\mathbb{R}^{|V|\times|V|}$ 表示静态邻接矩阵，表示网络中任意 2 个节点之间的邻接关系。

对于任意用户节点 u_i 与资源节点 v_j，其在关系 r 下的边权重 w_{ij} 由交互频次聚合得到，能够直观反映用户对特定资源的依赖程度或访问强度，即

$$w_{ij}=\sum_{e\in\mathcal{E}^{(t_k)}}\mathbb{I}(u_e=u_i\wedge v_e=v_j\wedge r_e=r)\tag{4}$$

通过上述构建，图 G_t 同时封装了用户的统计属性与其活动的拓扑结构，为下游的异常检测提供

了完备的数据表征。

2.3　时空掩码重建

基于图表示学习，提出了时空掩码重建机制，其原理流程如图 2 所示。该机制采用“掩码-重建”的自监督任务，学习节点的潜在表示。

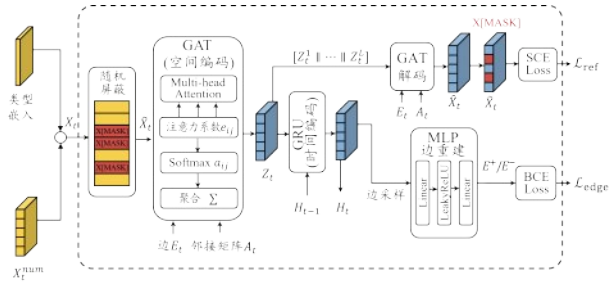


图 2　时空掩码重建

采用随机掩码策略，对于输入特征矩阵 X，第 i 个节点的特征向量 x_i，按比例 r 随机采样一部分良性节点子集 $\tilde{V} \subset V$。采样过程对每个节点独立进行，即各节点以概率 r 被选入 $\tilde{V}$，节点间的采样决策相互独立，以保证掩码分布的随机性与无偏性。将 $\tilde{V}$ 中节点的数值特征替换为可学习的特殊向量 X[MASK]，同时保留其节点类型嵌入不变，掩码后的特征记为 $\tilde{X}$。对每个节点 $v_i \in V$，其掩码后的特征定义为：

$$\tilde{x}_i = \begin{cases} x_{i[M]}, & v_i \in \tilde{V} \\ x_i, & v_i \notin \tilde{V} \end{cases} \tag{5}$$

为同时捕捉空间拓扑和时间依赖，设计了由图注意力网络（GAT）和门控循环单元（GRU）组成的混合编码器。利用 GAT 对多跳邻域节点的信息进行加权聚合，从而学习具有结构感知能力的节点表示。对于任意节点 v_i 及其邻居节点 v_j，注意力系数为 e_{ij}，通过 Softmax 函数对注意力系数进行归一化，得到最终的归一化权重 a_{ij}：

$$\alpha_{ij} = softmax\left(e_{ij}\right) = \frac{exp\left(e_{ij}\right)}{\sum_{k \in \mathcal{N}\left(v_i\right)} exp\left(e_{ik}\right)} \tag{6}$$

然后，利用多头注意力机制（Multi-head Attention）捕捉拓扑结构。设共有 K 个注意力头，则节点 v_i 的空间聚合表示为

$$z_i = \mathbin{/\!/}_{k=1}^{K} \sigma\left(\sum_{j \in \mathcal{N}\left(v_i\right)} \alpha_{ij}^{k} \boldsymbol{W}^{k} h_j \right) \tag{7}$$

其中，$\boldsymbol{W}^k$ 为第 k 个头的线性变换矩阵，$\sigma(\cdot)$ 为非线性激活函数。基于该机制，即使在部分节点特征被掩码时，GAT 仍可利用邻域上下文推断节点语义，从而捕捉图快照内的拓扑结构和邻居信息。GAT 编码为：

$$Z_t = \mathrm{GAT}\left(\tilde{X}_t, E_t, A_t\right) \tag{8}$$

在此基础上，引入 GRU 对节点表征进行时序建模，通过更新门与重置门动态调控历史信息的保留与当前输入的融合。以当前时刻 GAT 的输出 Z_t 和前一时刻的隐藏状态 H_{t-1} 作为当前输入，生成更新后的隐藏状态 H_t：

$$H_t = \mathrm{GRU}\left(Z_t, H_{t-1}\right) \tag{9}$$

解码器将编码后的输出映射回原始特征空间，以重建被掩码的节点特征及图结构。基于图神经网络架构，聚合邻域上下文信息为 R_t：

$$R_t = g\left(\left[Z_t^1 \mathbin{/\!/} \cdots \mathbin{/\!/} Z_t^L\right]\right) \tag{10}$$

基于聚合后的表征 R_t 及当前图结构，重建后的特征矩阵 $\hat{X}_t$：

$$\hat{X}_t = \mathrm{GAT}_{dec}\left(R_t, E_t, A_t\right) \tag{11}$$

为衡量原始节点特征和重建特征之间的差异，定义了特征重建损失 $\mathcal{L}_{\mathrm{ref}}$ 为

$$\mathcal{L}_{\mathrm{ref}} = \frac{1}{\left|\tilde{V}\right|} \sum_{i \in \tilde{V}} \mathrm{SCE}\left(\hat{x}_i, x_i\right) = \frac{1}{\left|\tilde{V}\right|} \sum_{i \in \tilde{V}} \left(1 - \frac{\hat{x}_i \cdot x_i}{\mathbin{/\!/} \hat{x}_i \mathbin{/\!/} \mathbin{/\!/} x_i \mathbin{/\!/}}\right)^{\gamma} \tag{12}$$

其中，x_i 与 $\hat{x}_i$ 分别为节点 i 的原始特征与重建特征，$\gamma > 0$ 为缩放因子。该损失考虑到高维特征空间中方向一致性的重要性，采用缩放余弦误差（SCE）[36] 解决数据不平衡问题，更好地捕捉特征向量的语义方向。

此外，为了显式地保留图的拓扑结构，引入了边重构损失 $\mathcal{L}_{\mathrm{edge}}$，通过二元交叉熵（BCE）约束模型对图中存在的边进行预测：

$$\mathcal{L}_{\mathrm{edge}} = -\frac{1}{\left|E^{+} \cup E^{-}\right|} \sum_{(i,j) \in E^{+} \cup E^{-}} \begin{pmatrix} \left(1 - y_{ij}\right) log\left(1 - prob_{ij}\right) \\ + y_{ij} log\left(prob_{ij}\right) \end{pmatrix} \tag{13}$$

其中，y_{ij} 表示边的存在性，真实边记为 E^{+}，通过负采样得到的伪负边为 E^{-}；当且仅当 $(i,j) \in E^{+}$ 时

$y_{ij}=1$，否则$y_{ij}=0$；$prob_{ij}$表示预测节点i和j之间的连接概率。

基于上述定义，该机制的总损失函数由特征重建损失与边重构损失加权构成：

$$\mathcal{L}_{\mathrm{GMAE}}=\lambda_1\mathcal{L}_{\mathrm{ref}}+\lambda_2\mathcal{L}_{\mathrm{edge}} \tag{14}$$

其中，λ_1和λ_2为平衡系数，用于调节不同损失项的相对贡献。在节点特征噪声较大时，可适当增大λ_1以提升语义表示的鲁棒性；在用户交互行为稀疏的场景下，可适当增大λ_2以强化对拓扑结构的学习。综上，通过自监督的时空掩码重建任务，模型获得了对节点语义与图拓扑的整体理解，为后续的异常判别奠定了基础。

2.4　多视图对比增强

为增强模型对用户行为正常波动的适应，缓解对统计特征的过拟合问题，设计了多视图对比增强机制。该机制要求同一节点在随机扰动生成的不同视图下保持语义一致的表示，而不同节点的表示则应在潜在空间中具有足够的可区分性。

具体而言，对于同一张图G，构造两个随机增强视图$\bar{G}^{(1)}$和$\bar{G}^{(2)}$。如图3所示，每个视图均采用了两种随机扰动策略：

1) 特征遮蔽：随机将输入特征矩阵中的部分元素置零，模拟特征缺失或噪声。

2) 随机删边：随机移除图中一定比例的边，以概率q从边集合E随机丢弃部分边，得到两份边索引与边属性$(\bar{E}^{(1)},\bar{A}^{(1)})$、$(\bar{E}^{(2)},\bar{A}^{(2)})$，增强模型对结构不完整的鲁棒性。

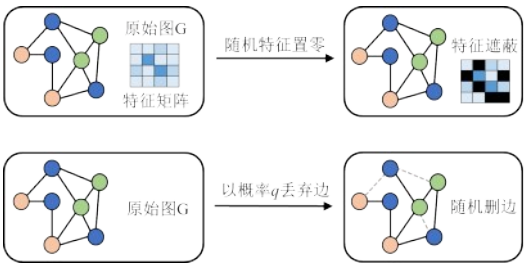


图3　对比视图的扰动策略

将两个增强视图分别输入共享参数的混合编码器，得到对应的节点表示，并通过多层感知机（MLP）构成的投影头映射至对比学习空间。记第i个节点在视图$\bar{G}^{(1)}$与视图$\bar{G}^{(2)}$中的投影表示分别为$p_i^{(1)}$和$p_i^{(2)}$。对于每个节点而言，$(p_i^{(1)},p_i^{(2)})$构成正样本对，而与其他节点$(j\neq i)$的表示构成负样本对。

在训练过程中，通过最大化正样本对的一致性、最小化与负样本对的相似性，使同一节点在不同扰动视图下的表示相互靠近，同时将其与其余节点的表示拉开距离，从而获得更具判别性的表示空间。在异常样本占比较低的场景下，该一致性约束促使模型充分利用大量良性节点的互补视图信息学习稳定表示，从而降低类别不平衡对表示学习的影响。对比学习与模型参数同步更新，以保证节点表示能够在不同视图之间保持表征稳定，而对不同节点保持必要的区分性。两个视图的投影结果记为$P^{(1)}=\{p_1^{(1)},\cdots,p_N^{(1)}\}$与$P^{(2)}=\{p_1^{(2)},\cdots,p_N^{(2)}\}$，温度系数为$\tau$，定义相似度评分：

$$\sin(a,b)=\exp(\cos(a,b)/\tau) \tag{15}$$

损失函数定义如下：

$$\mathcal{L}_i^{(1\to 2)}=-\log\frac{\sin(p_i^{(1)},p_i^{(2)})}{\sum_{k\neq i}\sin(p_i^{(1)},p_k^{(1)})+\sum_{j=1}^{N}\sin(p_i^{(1)},p_j^{(2)})} \tag{16}$$

同理，定义视图$2\to 1$的损失为$\mathcal{L}_i^{(2\to 1)}$。最终计算对比损失$\mathcal{L}_{\mathrm{con}}$公式为：

$$\mathcal{L}_{\mathrm{con}}=\frac{1}{2N}\sum_{i=1}^{N}\left(\mathcal{L}_i^{(1\to 2)}+\mathcal{L}_i^{(2\to 1)}\right) \tag{17}$$

通过最小化该损失，模型能够有效过滤合法行为波动引发的噪声，使正常行为模式在潜在空间聚集，将真实异常凸显为离群点，从而显著增强样本可分性，降低误报率并提升了异常检测的准确性。

2.5　鲁棒性分析

为反映模型鲁棒性，下面将从理论角度分析TGMC的鲁棒性。

定义1　（Lipschitz连续性）函数$f:\mathcal{X}\to\mathcal{Y}$称为$L$-Lipschitz连续的，若对任意$x_1,x_2\in\mathcal{X}$，满足

$$/\!/\, f(x_1)-f(x_2)\,/\!/\,_{\mathcal{Y}}\leqslant L\,/\!/\,x_1-x_2\,/\!/\,_{\mathcal{X}} \tag{18}$$

引理1　设GAT的Lipschitz常数为L_{GAT}，GRU隐藏状态满足收缩稳定性，即存在遗忘因子$\rho\in(0,1)$与常数L_{GRU}。若输入图序列的特征扰动和结构扰动分别满足$/\!/\,\Delta X_t\,/\!/\,_F\leqslant\epsilon_X$与$/\!/\,\Delta A_t\,/\!/\,_F\leqslant\epsilon_A$，则时刻$T$的隐藏状态扰动应满足：

$$/\!/\,\tilde{H}_T-H_T\,/\!/\,_F\leqslant\frac{L_{\mathrm{GAT}}\cdot L_{\mathrm{GRU}}}{1-\rho}\cdot\left(\epsilon_X+C_A\epsilon_A\right) \tag{19}$$

其中，C_A为结构传播系数。

证明　由GAT的Lipschitz连续性，对于任意

时刻t的空间编码输出扰动有

$$\begin{aligned}&/\!/\ \Delta Z\ /\!/\ _F \leqslant L_{\mathrm{GAT}} \cdot (\ )\\&\leqslant L_{\mathrm{GAT}}\left(\epsilon_X + C_A\epsilon_A\right)\end{aligned} \tag{20}$$

记 $\delta = L_{\mathrm{GAT}}\left(\epsilon_X + C_A\epsilon_A\right)$。由GRU的收缩稳定性，对任意扰动$\Delta H_{t-1}$和$\Delta Z_t$有

$$\begin{aligned}/\!/\ \Delta H_t\ /\!/\ _F &\leqslant \rho\ /\!/\ \Delta H_{t-1}\ /\!/\ _F + L_{GRU}\ /\!/\ \Delta Z_t\ /\!/\ _F\\&\leqslant \rho\ /\!/\ \Delta H_{t-1}\ /\!/\ _F + L_{GRU}\delta\end{aligned} \tag{21}$$

时刻t的输入扰动对时刻T的影响衰减为ρ^{T-t}，展开递推式并求和可得：

$$\begin{aligned}/\!/\ \Delta H_T\ /\!/\ _F &\leqslant L_{GRU}\sum_{t=1}^{T}\ \rho^{T-t}\delta\\&= \frac{L_{GRU}\left(1-\rho^T\right)}{1-\rho}\delta\\&\leqslant \frac{L_{GRU}}{1-\rho}\delta\end{aligned} \tag{22}$$

代入$\delta = L_{\mathrm{GAT}}\left(\epsilon_X + C_A\epsilon_A\right)$，可得

$$/\!/\ \tilde{H}_T - H_T\ /\!/\ _F \leqslant \frac{L_{\mathrm{GAT}} \cdot L_{\mathrm{GRU}}}{1-\rho}\left(\epsilon_X + C_A\epsilon_A\right) \tag{23}$$

式中左边为时序编码器的输出扰动，右边为输入扰动经Lipschitz传播和时序衰减后的累积界，确保了长时序依赖下的稳定性。证毕。

引理 2 设多视图对比增强的特征遮蔽率为p_m，丢边率为q，编码器Lipschitz常数为L_f。若遮蔽矩阵$M \in \{0,1\}^{N\times d}$中每个元素以$p_m$置零，删边后邻接矩阵为$\tilde{A}$，则节点表示对增强扰动的期望偏差满足：

$$\mathbb{E}[\ ] \leqslant L_f\left(\sqrt{p_m}\ /\!/\ X\ /\!/\ _F + C_A\sqrt{q}\ /\!/\ A\ /\!/\ _F\right) \tag{24}$$

其中，$P^{(0)}$为原始表示，$P^{(1)}$为增强视图表示。

证明 特征遮蔽后有$\tilde{X} = X \odot M$，其中$\odot$表示逐元素乘积。由于M的每个元素独立同分布，且$\mathbb{E}\left[M_{ij}\right] = 1 - p_m$，由Jensen不等式可得

$$\begin{aligned}\mathbb{E}[\ ] &\leqslant \sqrt{\mathbb{E}[\ ]}\\&= \sqrt{p_m}\ /\!/\ X\ /\!/\ _F\end{aligned} \tag{25}$$

类似地，随机删边导致的结构扰动满足：

$$\mathbb{E}[\ ] \leqslant \sqrt{q}\ /\!/\ A\ /\!/\ _F \tag{26}$$

由编码器Lipschitz连续性可得：

$$\begin{aligned}\mathbb{E}[\ ] &\leqslant L_f(\ )\\&= L_f\left(\mathbb{E}[\ ] + C_A\mathbb{E}[\ ]\right)\\&\leqslant L_f\left(\sqrt{p_m}\ /\!/\ X\ /\!/\ _F + C_A\sqrt{q}\ /\!/\ A\ /\!/\ _F\right)\end{aligned} \tag{27}$$

式中左边为增强视图与原始表示的期望偏差，右边为遮蔽率和删边率控制的扰动上界。当对比损失收敛后，模型在$\sqrt{p_m}$和$\sqrt{q}$量级扰动下仍能保持表示稳定。证毕。

由引理1和2可知，TGMC的表示扰动随输入特征与结构扰动呈线性可控，并在时序递推中呈指数衰减。同时，对比增强进一步提升了在缺失与结构变化情形下的稳定性，为TGMC的鲁棒性提供理论支撑。

3 实验与结果分析

3.1 实验设置

本文在CERT r4.2和CERT r5.2[37]公开数据集上进行实验。CERT数据集由卡内基梅隆大学软件工程研究所（CMU SEI）构建，模拟了大型企业环境下的用户行为日志，在该领域被广泛应用。如表1所示，CERT r4.2和CERT 5.2的规模不同，分别模拟了拥有1000名和2000名员工的公司，都包含五种活动的日志文件：登录/注销、电子邮件通信、设备连接、文件处理和HTTP浏览，包含恶意用户越权访问外发、云存储上传等隐蔽威胁场景。CERT r4.2包括三种威胁场景中的70个恶意内部人员，有32770227个活动和7323个恶意活动。CERT r5.2包含79856664项活动，其中10306个恶意活动。数据集详细信息如表1：

表1 数据集统计信息

数据集	正常用户	异常用户	正常活动	异常活动
CERT r4.2	930	70	32,762,904	7,323
CERT r5.2	1901	99	79,846,358	10,306

实验硬件环境：NVIDIA RTX 4090 GPU，运行于Ubuntu 22.04系统。软件环境： Python 3.12，PyTorch框架，CUDA 12.1。TGMC方法参数设置，相关信息如表2所示。

3.2 基线方法

为了评估TGMC的检测性能，将TGMC与6种基线方法进行了比较。RNN通过循环结构捕获用户行为序列的时序依赖关系；Transformer基于自注意力机制建模行为模式；DeepLog[12]利用LSTM

表2　实验参数

参数	参数说明	取值
optimizer	优化器	Adam
learning rate	学习率	0.005
epochs	最大轮次	100
dropout	丢弃层	0.3
hidden_dim	隐藏层维度	64
dp	丢边率	0.3
γ	掩码率	0.2

从系统日志中学习正常行为模式并检测异常；BST[16]融合用户行为序列与上下文特征进行行为预测；FMLP[38]采用滤波增强的MLP架构对行为序列建模；LAN[35]通过自适应邻居学习实现实时用户异常行为检测。

3.3　评估指标

为评估TGMC的性能，分别使用AUC（Area Under Curve）、真阳性率（TPR/Recall）、假阳性率（FPR）和准确率（Accuracy）作为内部威胁检测任务的评估指标。AUC作为ROC曲线的量化指标，对类别分布的不平衡具有相对稳健性，能够评估模型在极端不平衡数据中的综合性能。TPR直接反映模型对内部威胁的检出能力，是威胁检测场景中的核心指标之一。FPR可以很好地评估模型将正常数据归类为威胁的程度，量化模型对良性行为的误报程度。上述评估指标的计算式分别为：

$$\mathrm{AUC}=\int_0^1 \mathrm{TPR}(\mathrm{FPR})\,\mathrm{d}(\mathrm{FPR}) \tag{28}$$

$$\mathrm{TPR}=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} \tag{29}$$

$$\mathrm{FPR}=\frac{\mathrm{FP}}{\mathrm{FP}+\mathrm{TN}} \tag{30}$$

$$\mathrm{Accuracy}=\frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TP}+\mathrm{TN}+\mathrm{FN}+\mathrm{FP}} \tag{31}$$

3.4　性能比较

TGMC及其基线的检测有效性如表3所示。从结果可知，TGMC在CERT r4.2数据集上实现了0.9428的AUC和0.0866的FPR，在CERT r5.2数据集上实现了0.9465的AUC和0.0618的FPR。与其他基线方法相比，TGMC实现了最高的召回率，表明其对用户异常行为的有效识别能力。

表3　TGMC和基线方法的性能比较

数据集	方法	AUC	TPR	FPR
CERT r4.2	RNN	0.7521	0.6934	0.3622
	Transformer	0.7981	0.7195	0.2799
	DeepLog[12]	0.7469	0.7152	0.3767
	BST[16]	0.6777	0.6554	0.3451
	FMLP[38]	0.8526	0.7983	0.2027
	LAN[35]	0.9369	0.8875	0.1411
	TGMC	0.9428	0.9254	0.0866
CERT r5.2	RNN	0.8641	0.8286	0.2361
	Transformer	0.8628	0.7621	0.1985
	DeepLog[12]	0.8549	0.7767	0.2336
	BST[16]	0.8162	0.7417	0.2301
	FMLP[38]	0.8435	0.8757	0.2889
	LAN[35]	0.9439	0.8814	0.0867
	TGMC	0.9465	0.8973	0.0618

RNN、Transformer、DeepLog及BST等基于序列的方法主要侧重于对日志的时间依赖性进行建模，TGMC相比之下，AUC提升了约8%至26%，FPR显著降低。FMLP通过频域滤波增强了MLP的序列建模能力，在CERT r5.2上取得了0.8757的TPR，但其FPR高达0.2889。这些方法忽略了实体间的交互结构，导致模型在面对复杂的用户行为模式时难以区分良性变化与恶意行为，对噪声较为敏感，容易产生误报。相比之下，LAN通过学习自适应邻居引入了图结构信息，弥补了序列模型的不足，其在两个数据集上的AUC均超过0.93，但在处理细粒度的语义特征时仍显不足。与LAN相比，TGMC在两个数据集上的TPR分别提升了约3.8%和1.6%，FPR分别降低了约5.4%和2.5%。这表明TGMC通过引入图掩码重建机制与对比学习，不仅能够有效捕获长时序依赖，还能从扰动的结构中重构深层的结构语义，从而增强了节点表示的判别性。

综上所述，TGMC能够更有效地融合时序动态与结构拓扑信息，在保证高检出率（TPR）的同时显著降低了误报率（FPR），在内部用户行为异常检测任务中表现出较好的鲁棒性与有效性。

3.5　超参数实验

为分析模型对超参数的敏感性及其对异常检测性能的影响，围绕丢边率（dp）、掩码率（γ）和学习率（lr）进行了网格搜索实验。图4展示了在

CERT r4.2 数据集上不同参数设置下的 AUC、Recall 和 Accuracy。

如图 4(a)所示，随着对比增强机制中丢边率的增长，丢边率为 0.3 时性能最好，AUC 和 TPR 到达峰值。这表明过小的边扰动提供的对比信号不足，提高丢边率后适度的结构扰动能够有效地防止对局部图拓扑的过度拟合。

如图 4(b)所示，当 γ 从 0.1 增加到 0.2 时，AUC 分数逐渐上升。这表明较低的掩码率会降低自监督重建的难度，导致编码器过度依赖原始特征。当 γ 超过 0.2 时，整体性能下降。这是因为较高的掩码率会将更多噪声引入图中，信息损失过多，从而降低性能。

如图 4(c)所示，极小的学习率值会导致模型欠拟合，随着学习率的增大整体性能上升。当学习率 lr 为 0.005 时，综合效果最佳。lr 继续增大时，AUC 和 Recall 下降，说明过大的 lr 会破坏重建和对比目标联合优化的协同性。

实验结果表明，所提出的模型在学习率 lr = 0.005、掩码率 γ = 0.2 以及丢边率 dp = 0.3 的设置下达到了最佳性能。

图 5 展示了不同参数设置下的 FPR 分数。固定其中一个参数的值，观察其他参数配置对模型性能的影响。由图可以看出，根据图 4 分析所得参数配置所对应的 FPR 值为 0.0886，相比其他参数配置处于较低范围。

3.6　消融实验

为了进一步验证各个模块的作用效果，在完整 TGMC 方法基础上进行消融实验。本实验主要验证三个核心组件对模型性能的贡献，即时序机制、图掩码重建以及多视图对比增强。因此，在消融实验中，将完整的 TGMC 方法与移除时序依赖（w/o Temporal）、移除掩码重建（w/o Masking）和移除对比增强（w/o CL）这 3 种情况进行异常检测效果对比，消融实验结果如表 4 所示。

当移除时序机制时，模型在两个数据集上的 AUC 和 TPR 均出现下降，且 FPR 升高。这表明仅依赖静态图表示难以有效区分持续异常与良性行为变化，时序建模对于捕捉长期行为动态及缓解概念漂移具有必要性。去除图掩码机制导致模型性能显著下滑，在 CERT r4.2 上 AUC 由 0.9428 降至 0.6215，且 FPR 大幅增加，在 CERT r5.2 上也出现一致的趋势。这表明掩码重建机制是通过利用结构和时间上下文推断节点语义，降低了模型对原始特

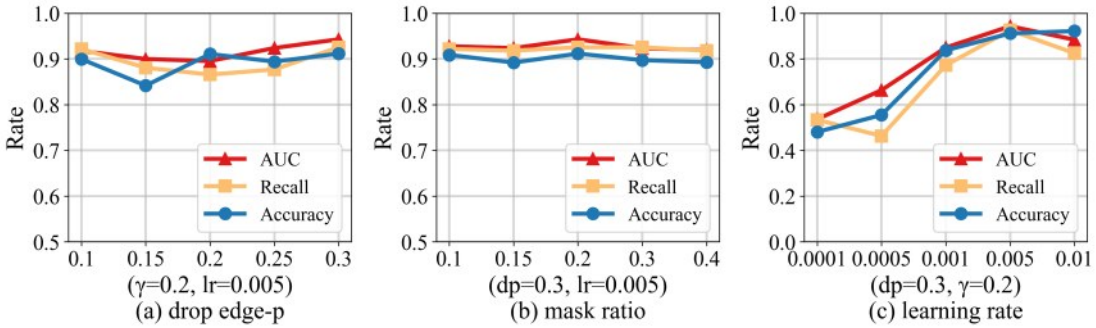


图 4　TGMC 在不同超参数下的性能

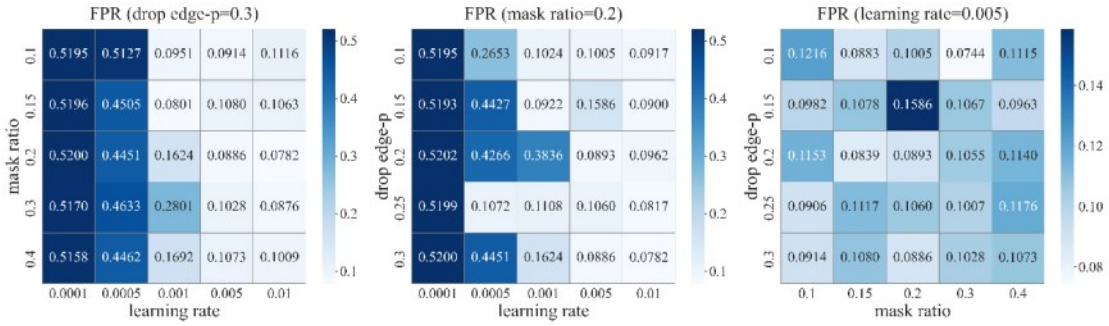


图 5　TGMC 在不同超参数下的 FPR 值

(a) (b) (c)

表 4

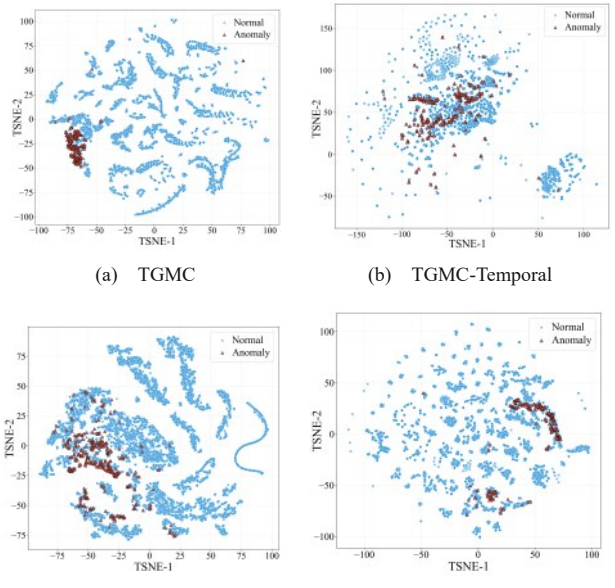
TGMC 方法消融性能对比

数据集	方法	AUC	TPR	FPR	ACC
CERT r4.2	w/o Temporal	0.8104	0.7351	0.1212	0.8785
	w/o Masking	0.6215	0.6816	0.4727	0.5279
	w/o CL	0.9247	0.9067	0.1095	0.8905
	TGMC	0.9428	0.9254	0.0866	0.9115
CERT r5.2	w/o Temporal	0.9233	0.7622	0.0867	0.9130
	w/o Masking	0.7205	0.7733	0.5001	0.5006
	w/o CL	0.7677	0.5027	0.1313	0.8680
	TGMC	0.9465	0.8973	0.0618	0.9382

征的过度依赖，从而提升了鲁棒性。去除对比增强机制后，CERT r5.2 上的 AUC 下降约 0.18，TPR 也大幅降低。这表明对比学习对于确保增强视图之间的表示一致性以及提高节点间可区分性至关重要。综上，TGMC 在两个数据集上的各项指标均优于其他变体。

为直观分析各个机制对模型性能的影响，采用 TSNE 技术分别对 TGMC 模型与去除时序（TGMC- Temporal）、掩码重建（TGMC-Masking）和多视图对比增强机制（TGMC-CL）的节点表示进行二维可视化投影。如图 6 所示，蓝色表示良性样本，而红点表示异常样本。

对于 TGMC，正常行为与异常行为在低维嵌入空间中呈现出更清晰的分离结构，异常样本形成相对紧凑且可区分的簇。相比之下，其他方法的可视化结果中异常与正常样本存在重叠，异常样本分布更加分散且边界模糊，反映出模型对原始特征噪声和局部结构扰动的敏感性增加。这表明时序、掩码重建和多视图对比增强机制有助于模型从结构与上下文中学习判别性语义表示，增强异常行为在潜在空间中的可分性。该可视化结果进一步验证了三个机制在提升异常识别能力与表示鲁棒性方面的有效性。



(a) TGMC　(b) TGMC-Temporal

(c) TGMC-Masking　(d) TGMC-CL

图 6　基于 CERT r4.2 的 TSNE 可视化

4　结束语

本文针对内部用户行为检测中时空耦合建模不足、隐蔽异常行为识别困难以及行为模式波动等挑战，提出了一种面向内部用户行为异常检测的图掩码对比学习方法 TGMC。首先，构建了用户-实体异构交互图，为异常判别提供输入。通过设计 GAT 与 GRU 混合编码器，聚合邻域信息并融合历史时空状态以更好的表征用户行为。其次，引入掩码重建机制，通过随机掩码并重建，挖掘更细粒度的局部结构依赖，有效识别表层特征伪装下的隐蔽异常行为。设计多视图对比增强机制，通过约束用户行为语义表示一致性，并扩大不同用户行为表征差异，增强了特征空间的区分度，增强对行为模式波动与噪声的鲁棒性。在 CERT r4.2 和 r5.2 数据集上的实验结果表明，TGMC 在 AUC 等关键指标上均优于现有基线方法，验证了本文方法的有效性。后续的研究中，可引入大语言模型（LLM）以增强对异常行为底层语义的解析能力，进一步提升模型的泛化能力与适应性。未来可以考虑将 TGMC

部署于企业安全运营中心，对员工的多模态行为日志进行实时建模与异常预警，辅助快速定位潜在内部威胁，降低人工排查成本。

参考文献：

[1] 孙德刚,刘美辰,李梅梅,等. 内部威胁分析与防御综述[J]. 信息安全学报,2025,10(1):176-193.

[2] Ponemon Institute. (2025). 2025 Ponemon Insider Threat Report. Dtex Systems.

[3] 郭渊博,刘春辉,孔菁,等. 内部威胁检测中用户行为模式画像方法研究 [J]. 通信学报, 2018, 39 (12): 141-150.

[4] Le D. C., Zincir-Heywood N., Heywood M. I., 2020. Analyzing data granularity levels for insider threat detection using machine learning. IEEE Trans. Netw. Serv. Manag. 17 (1), 30–44.

[5] Randive K., Mohan R., Sivakrishna A. M., 2023. An efficient pattern-based approach for insider threat classification using the image-based feature representation. J. Inf. Secur. Appl. 73, 103434.

[6] Duan S.-M., Yuan J.-T., Wang B., 2024. Contextual feature representation for image-based insider threat classification. Comput. Secur. 140, 103779.

[7] Yuan L.-P., Choo E., Yu T., Khalil I., Zhu S., 2021. Time-window based groupbehavior supported method for accurate detection of anomalous users. In: 2021 51st Annual IEEE/IFIP International Conference on Dependable Systems and Networks. DSN, IEEE, pp. 250–262.

[8] Kan X, Fan Y, Zheng J, et al. Data adjusting strategy and optimized XGBoost algorithm for novel insider threat detection model[J]. Journal of the Franklin Institute, 2023, 360(16): 11414-11443.

[9] 李涛,毕悦,胡爱群.面向智能系统的内部威胁多源日志分析与检测方法[J].信息网络安全,2025,25(04):509-523.

[10] 胡向东,张琴.基于特征组合优化的工业互联网恶意行为实时检测方法[J].电子学报,2024,52(09):3075-3085.

[11] Tuor A, Kaplan S, Hutchinson B, et al. Deep learning for unsupervised insider threat detection in structured cybersecurity data streams[C]// AAAI Workshops. 2017: 224-231.

[12] Du M, Li F, Zheng G, et al. Deeplog: Anomaly detection and diagnosis from system logs through deep learning[C]//Proceedings of the 2017 ACM SIGSAC conference on computer and communications security. 2017: 1285-1298.

[13] Villarreal-Vasquez M, Modelo-Howard G, Dube S, et al. Hunting for insider threats using LSTM-based anomaly detection[J]. IEEE Transactions on Dependable and Secure Computing, 2021, 20(1): 451-462.

[14] Paul S, Mishra S. Lac: Lstm autoencoder with community for insider threat detection[C]//proceedings of the 4th International Conference on Big Data Research. 2020: 71-77.

[15] Manoharan P, Hong W, Yin J, et al. Optimising insider threat prediction: exploring BiLSTM networks and sequential features[J]. Data Science and Engineering, 2024, 9(4): 393-408.

[16] Chen Q, Zhao H, Li W, et al. Behavior sequence transformer for e-commerce recommendation in alibaba[C]//Proceedings of the 1st international workshop on deep learning practice for high-dimensional sparse data. 2019: 1-4.

[17] 杨梦华,易军凯,朱贺军.基于CNN-LSTM算法的内部威胁检测方法[J].信息网络安全,2025,25(02):327-336.

[18] Jiang J, Chen J, Gu T, et al. Anomaly detection with graph convolutional networks for insider threat and fraud detection[C]//MILCOM 2019-2019 IEEE military communications conference (MILCOM). IEEE, 2019: 109-114.

[19] Fei K, Zhou J, Su L, et al. A graph convolution neural network based method for insider threat detection[C]//2022 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking (ISPA/BDCloud/SocialCom/SustainCom). IEEE, 2022: 66-73.

[20] Liu F, Wen Y, Zhang D, 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. 2019: 1777-1794.

[21] Li X, Li X, Jia J, et al. A high accuracy and adaptive anomaly detection model with dual-domain graph convolutional network for insider threat detection[J]. IEEE Transactions on Information Forensics and Security, 2023, 18: 1638-1652.

[22] Li X, Li L, Li X, et al. GMFITD: Graph meta-learning for effective few-shot insider threat detection[J]. IEEE Transactions on Information Forensics and Security, 2024.

[23] Kong K, Jin X, Liu D, et al. DPI-ITD: A Dual-Perspective Information-Driven Framework for Insider Threat Detection in IoT Systems[J]. IEEE Internet of Things Journal, 2025.

[24] 严莉,张凯,徐浩,等. 基于图注意力机制和 Transformer 的异常检测[J]. 电子学报, 2022, 50 (04): 900-908.

[25] Qi Y, Yan C, Wang Z, et al. ATHITD: Attention-based temporal heterogeneous graph neural network for insider threat detection[J]. Computers & Security, 2025: 104587.

[26] Wang C, Zhu H. Wrongdoing monitor: A graph-based behavioral anomaly detection in cyber security[J]. IEEE Transactions on Information Forensics and Security, 2022, 17: 2703-2718.

[27] Li Z, Cheng X, Sun L, et al. A hierarchical approach for advanced persistent threat detection with attention-based graph neural networks[J]. Security and Communication Networks, 2021, 2021(1): 9961342.

[28] Kotb H M, Gaber T, AlJanah S, et al. A novel deep synthesis-based insider intrusion detection (DS-IID) model for malicious insiders and AI-generated threats[J]. Scientific Reports, 2025, 15(1): 207.

[29] Zhou S, Wang L, Yang J, et al. Sitd: Insider threat detection using siamese architecture on imbalanced data[C]//2022 IEEE 25th International Conference on Computer Supported Cooperative Work in Design (CSCWD). IEEE, 2022: 245-250.

[30] Lv B, Wang D, Wang Y, et al. A hybrid model based on multi-dimensional features for insider threat detection[C]//International Conference on Wireless Algorithms, Systems, and Applications. Cham: Springer International Publishing, 2018: 333-344.

[31] 冯冠云,付才,吕建强,等.基于操作注意力和数据增强的内部威胁检测[J].网络与信息安全学报,2023,9(03):102-112.

[32] Pal P, Chattopadhyay P, Swarnkar M. Temporal feature aggregation with attention for insider threat detection from activity logs[J]. Expert Systems with Applications, 2023, 224: 119925.

[33] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[J]. Advances in neural information processing systems, 2017, 30.

[34] Gong Y, Cui S, Liu S, et al. Graph-based insider threat detection: A sur-

vey[J]. Computer Networks, 2024, 254: 110757.

[35] Cai X, Wang Y, Xu S, et al. Lan: learning adaptive neighbors for real-time insider threat detection[J]. IEEE Transactions on Information Forensics and Security, 2024.

[36] Hou Z, Liu X, Cen Y, et al. Graphmae: Self-supervised masked graph autoencoders[C]//Proceedings of the 28th ACM SIGKDD conference on knowledge discovery and data mining. 2022: 594-604.

[37] Lindauer B. Insider Threat Test Dataset[DS]. Pittsburgh: Carnegie Mellon University, 2020. DOI:10.1184/R1/12841247.v1.

[38] Zhou K, Yu H, Zhao W X, et al. Filter-enhanced MLP is all you need for sequential recommendation[C]//Proceedings of the ACM web conference 2022. 2022: 2388-2399.

袁璐（2000− ），女，辽宁丹东人，博士，信息工程大学博士生，主要研究方向为网络信息防御、异常行为检测等。

孙剑文（1988− ），女，北京人，博士，32708部队、工程师，主要研究方向为异常检测、深度学习等。

姜迎畅（2000− ），男，河南洛阳人，博士，信息工程大学博士生，主要研究方向为威胁情报分析、大模型应用等。

张斌（1969- ），男，河南南阳人，博士，信息工程大学教授、博士生导师，主要研究方向为信息系统安全等。

马骏（1981− ），男，河北安国人，博士，信息工程大学副教授、硕士生导师，主要研究方向为态势感知、网络攻防。

李红宇（2001- ），男，河南商丘人，信息工程大学博士生，主要研究方向为源代码漏洞检测，机器学习算法等。