基于增强关系图卷积网络的数据违规转售检测方法

王宇翔[1,2,3]，张玲翠[1,3]，侯雨桥[1,3]，杨倩[1,3]，牛犇[1,3]
（1.中国科学院信息工程研究所，北京 100085；2.中国科学院大学网络空间安全学院，北京 100049；
3.网络空间安全防御全国重点实验室，北京 100085）

摘　要：为解决数据流通交易场景下的数据违规转售问题，基于交易上下文信息相似度和交易因果顺序约束，对关系图卷积网络的消息传递和特征聚合过程进行改进，提出一种增强关系图卷积网络模型，可有效学习复杂交易关系下的违规转售行为特征。基于该模型设计一种数据违规转售检测方法，预测数据交易拓扑图中节点是否为违规转售交易节点。构造带有违规转售样本的模拟数据交易数据集并展开对比实验，结果证明了所提方法的有效性。

关键词：数据流通交易；数据违规转售；关系图卷积网络；注意力机制
中图分类号：TN92
文献标志码：A
DOI：10.11959/j.issn.1000–436x.2026083

Illicit data resale detection via an enhanced relational graph convolutional network

Wang Yuxiang[1,2,3], Zhang Lingcui[1,3], Hou Yuqiao[1,3], Yang Qian[1,3], Niu Ben[1,3]
1. Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100085, China
2. School of Cyber Security, University of Chinese Academy of Sciences, Beijing 100049, China
3. State Key Laboratory of Cyberspace Security Defense, Beijing 100085, China

Abstract: Illicit data resale in data trading scenarios exhibited strong concealment and was difficult to detect. An enhanced relational graph convolutional network was proposed by optimizing message passing and feature aggregation with transaction contextual similarity and causal temporal order constraints, enabling effective representation of illicit resale behaviors under complex transaction relations. Based on this model, a detection method was developed to predict the existence of illicit resale behaviors in transaction topology graphs. A simulated data trading dataset containing anomalous resale samples was constructed, and comparative experiments were performed. The results indicate that the proposed method provides an effective solution for illicit data resale detection in data trading scenarios.

Keywords: data trading, illicit data resale, relational graph convolutional network, attention mechanism

0　引言

数据交易是指数据供方和需方之间进行的，以特定形态数据为标的，以货币或者其他等价物作为对价的交易行为[1]。作为数据要素流通中的关键一环[2]，数据交易是充分释放数据要素潜在价值、促进数字经济高质量发展的重要引擎。目前，数据交易产业呈高速发展态势。仅2024年，我国数据市场交易规模就已超过1 600亿元[3]。到2035年，全球数据交易市场规模预估将达到8 120亿美元[4]。

收稿日期：2026–02–02；修回日期：2026–03–22
通信作者：杨倩，yangqian@iie.ac.cn
基金项目：国家重点研发计划基金资助项目（No.2023YFB3106505）；国家自然科学基金资助项目（No.U24A20240, No.62441226）
Foundation Items：The National Key Research and Development Program of China (No.2023YFB3106505), The National Natural Science Foundation of China (No.U24A20240, No.62441226)

由于数据具有复制无成本、使用无损耗等固有属性，以数据产品为交易标的的交易行为面临区别于其他交易场景的安全威胁。恶意的数据买家受利益驱动，可能会不按照交易合同协议使用数据，而是将数据复制后转卖给第三方，这就是数据违规转售交易行为[5]。该行为直接损害相关数据卖家的合法权益与经济利益，破坏数据交易市场的公平竞争秩序，给数据交易生态带来严重危害。

数据确权被认为是解决数据违规转售问题的有效手段。从法律角度来看，数据确权是通过对数据主体赋权，使其对数据享有相应的法律保障，从而在一定程度或范围内针对数据具有排除他人侵害的效力[6]。从技术角度来讲，相应的数据确权方案一般通过数据所有权确认和交易溯源性保障来实现。数据所有权确认最直接的方法是对所有权主张方和原始所有者的数据相似性进行度量[7]。区块链技术因具有不可篡改等特性，也常用于数据所有权确认[8-9]。智能合约技术因其可对交易行为进行程序化约束，是保障数据交易可溯源的有效工具[10-11]。

上述数据确权方案大多依赖于提取数据的权属属性并将其进行对比或上链，但是这些权属属性只在单一数据交易平台内部有效。如果恶意转售者将买到的数据产品转移到另一个数据交易平台进行出售，由于新平台无法验证原平台中的确权信息，恶意转售者很可能被视为数据原始所有者而令其违规转售行为得逞。另外，如果数据交易平台未部署数据确权方案，那么数据违规转售行为更容易实施。因此，数据确权方案在应对数据违规转售问题上存在局限性。

从数据交易行为出发对数据违规转售行为进行检测是一种更可行的技术方案。首先，这种方案不依赖于数据交易平台相关的数据权属属性，即便恶意转售者跨平台实施数据转售，其交易行为所呈现出的特征仍可被识别。其次，即便数据交易平台未部署违规转售检测系统，恶意转售者的相关交易行为仍会留存完整记录，为事后追责提供有效依据。对于任一笔数据交易，交易平台都有义务记录交易信息，相关交易记录完整记录了交易参与方的交互过程。因此，可以通过数据交易记录刻画数据交易行为，展开相应的数据违规转售检测。

基于神经网络的检测模型能通过数据交易记录学习到复杂的交易行为模式，与基于规则、统计方法或特征工程的传统方法相比，能更精准地捕捉到违规转售行为特征。对于交易记录这种结构化数据来说，通过图建模方式，将原本相互独立的交易条目构建为交易拓扑结构，可以呈现交易间的内在关联，挖掘深层次的交易行为规律。因此，针对数据交易记录，设计一种基于图神经网络（graph neural network, GNN）的异常交易行为检测方法，可以为解决数据违规转售问题提供新的思路。

1　相关工作

金融交易具有与数据交易相似的业务环境与交易行为模式，该场景下的异常检测研究可以为本文提供良好借鉴。

1.1　基于GNN的金融交易异常检测

金融交易场景下基于GNN的异常检测一般将交易主体作为节点，将主体之间的交易作为边，检测是否存在异常行为[12]。该场景面临的最主要异常行为是金融欺诈。Dou等[13]提出CARE-GNN（camouflage-resistant GNN）来检测经过伪装的欺诈者。Liu等[14]提出PC-GNN（pick and choose GNN）来解决欺诈检测场景下交易数据集标签不平衡问题（即欺诈样本数量远少于正常样本）。Xiang等[15]则将时序特征融入交易拓扑，提出GTAN（gated temporal attention network）来检测信用卡欺诈。

在各种GNN模型中，图卷积网络（graph convolutional network, GCN）[16]是较早出现且使用较为广泛的一种，其核心是消息传递机制，即通过边来传递经过加权变换的邻居特征，在中心节点处聚合邻居节点与自身特征来得到新特征，是一种直推式学习。现有基于GCN的金融交易异常检测可归纳为节点分类[17]、边分类[18]和图分类[19]3种范式，基于节点分类的研究工作占主流。针对GCN的改进模型主要是拓展到归纳式学习的GraphSAGE (sample and aggregate)和增加了注意力机制的GAT（graph attention network），如Duan等[20]和Hu等[21]分别利用GraphSAGE和GAT来检测以太坊钓鱼欺诈和电信诈骗。

但是经典GCN和相应改进都假设图中的节点只通过同一种关系的边相连，而金融交易拓扑更适合通过含有多种关系边的图来描述，比如两个交易主体间的买入或卖出就是两种关系。因此，相较于

GCN，关系图卷积网络（relational graph convolutional network, RGCN）[22]更适用于金融交易场景下的行为建模。

1.2 金融交易场景下的RGCN研究

RGCN在消息传递过程中为不同关系的边分配独立权重，能够有效描述多关系图结构[19]，因而被越来越多的研究者应用于关联关系多样化的金融交易场景。比如，RGCN可以建模信贷场景下各个企业之间的多类型关系，提取企业信用嵌入。Mitra等[23]基于该信用嵌入，结合随机森林对印度中小微企业信用进行分类；Jiang等[24]则利用该信用嵌入对我国新三板中小微企业信用进行评估。除此之外，RGCN还可以建模各只股票之间的多关系拓扑，Li等[25]将其用于预测股票开盘价相较于前一日收盘价的涨跌。

尽管RGCN可以有效建模交易拓扑中的不同类型关系，但仍存在两个问题：1) 在消息传递过程中无法对同一关系类型下邻居节点特征差异性或相似性进行建模，对不同邻居节点都采用同质化的消息传递策略；2) 无法对交易行为的时间顺序进行建模。有些研究工作已经注意到这些问题，例如，针对问题1)，Jiang等[24]利用约束注意力机制挖掘RGCN节点的上下文信息；针对问题2)，Li等[25]利用长短期记忆网络（long-short term memory, LSTM）来对时间维度特征进行建模。但是，现有工作都只关注一方面问题，缺乏一种从RGCN消息传递与特征聚合过程出发，对同类型邻居节点特征差异与交易时序关系进行统一建模的研究方法。

1.3 本文工作

针对上述问题，本文提出一种基于增强RGCN模型的数据违规转售检测方法，主要工作如下。

1) 设计了一种结合了注意力的交易上下文信息挖掘机制，根据相邻交易节点间的上下文信息相似程度，利用注意力机制对消息传递过程中同类型关系下的邻居特征进行差异化处理，实现对交易上下文异常关联信息的捕捉，增强对可疑违规转售交易行为的发现能力。

2) 设计了一种结合了衰减时序门的交易因果顺序约束机制，根据邻居节点间交易时序发现交易间因果顺序，结合交易时间差衰减，利用门控机制对消息传递过程中的邻居特征进行方向性约束，实现对转售交易因果顺序的发现与利用，提升对正常交易与违规转售交易的区分能力。

3) 基于真实的在线交易数据集，生成可建模数据交易行为的数据交易记录，并构造和注入违规转售交易样本，构建面向数据违规转售检测的数据集，在该数据集及其他通用数据集上对数据违规转售检测方法的有效性进行验证与分析。

2 问题的提出

2.1 基于数据交易拓扑图的数据违规转售检测

数据交易拓扑图根据数据交易记录构建，将拓扑图中的节点定义为交易记录中的一笔交易，这样可以直接通过图神经网络根据交易特征定位违规转售交易。根据两笔交易的买卖双方之间有无交互关系，确定相应两节点之间是否有边相连，并将图中的边按照关系分为两种类型：1)角色共享型，两笔交易具有相同的买方或卖方；2)角色交换型，一笔交易的买方（或卖方）是另一笔交易的卖方（或买方）。显然，违规转售交易节点必定有一条角色交换型边连接。这种将交易拓扑图中的边按关系划分为两类的思路尚未见诸同类金融交易异常检测研究，其优势是可以强化同类交易的内在关联性，提升图神经网络对违规转售交易的检测效能。基于上述规则生成的交易拓扑图示例如图1所示。

序号	卖方	买方
1	参与者1	参与者4
2	参与者1	参与者5
3	参与者2	参与者5
4	参与者2	参与者6
5	参与者1	参与者7
6	参与者7	参与者8
7	参与者6	参与者9
8	参与者3	参与者5
9	参与者4	参与者10
10	参与者3	参与者11
11	参与者11	参与者10
12	参与者2	参与者9

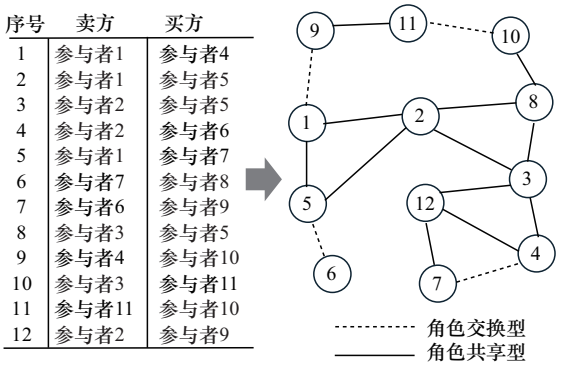


图1 交易拓扑图示例

定义1 数据交易拓扑图。令$G=\{V,R,E,\boldsymbol{X},\boldsymbol{Y}\}$为一张双关系无向图，其中，$V=\{v_1,v_2,\cdots,v_N\}$为节点集合，节点数量为$N$，每个节点$v_i$对应一笔数据交易记录；$R=\{r_1,r_2\}$为关系类型集合，其中$r_1$为角色共享型关系，即两笔交易在同一交易角色（如相同卖方或相同买方）下发生，r_2为角色交换型关系，即两笔交易在买卖角色上发生交换（如前一交易的买方在后一交易中作为卖方出现）；$E=$

$\{\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_N\}$为边集合，$E\subseteq V\times R\times V$，每条边$\varepsilon_i=(u,v,r)\in E$表示节点$v\in V$和节点$u\in V$之间存在一种类型为$r\in R$的无向关系；$\boldsymbol{X}\in\mathbb{R}^{N\times D}$是所有节点的特征矩阵，其中$\boldsymbol{x}_i$为节点$v_i$的特征向量，特征数量为$D$；$\boldsymbol{Y}\in\{0,1\}^N$是所有节点的标签向量，$y_i=1$表示节点$v_i$是违规转售交易，$y_i=0$则是正常交易。

问题描述：在给定数据交易拓扑图$G=\{V,R,E,\boldsymbol{X},\boldsymbol{Y}\}$的条件下，学习一个有效的图表示模型，描述不同关系类型下交易节点之间的结构与语义关联，从而预测交易节点是否为违规转售交易。

2.2　数据违规转售检测任务中RGCN的局限性分析

处理数据交易拓扑图这种多关系异构图需要用到关系图卷积网络RGCN。不论何种图神经网络，其核心都是聚合邻居节点特征的消息传递机制。作为图卷积网络GCN的扩展模型，RGCN的核心是在消息传递过程中为每种关系分配独立权重矩阵，表示为

$$\boldsymbol{h}_i^{(l+1)}=\sigma\left(\sum_{r\in R}\sum_{j\in\mathcal{N}_r(i)}\frac{1}{c_{i,r}}\boldsymbol{W}_r^{(l)}\boldsymbol{h}_j^{(l)}+\boldsymbol{W}_0^{(l)}\boldsymbol{h}_i^{(l)}\right)\tag{1}$$

其中，$\boldsymbol{h}_i^{(l)}$为节点i在第l层的特征向量，$\mathcal{N}_r(i)$是节点i在关系r下的邻居集合，$c_{i,r}$是缓解邻居数量影响的归一化常数（通常为$|\mathcal{N}_r(i)|$），$\boldsymbol{W}_r^{(l)}$是关系r对应的第l层权重矩阵，$\boldsymbol{W}_0^{(l)}$是节点i自连接的权重矩阵，$\sigma(\cdot)$为激活函数（如ReLU）。

基于RGCN的数据违规转售检测任务本质上是利用L层堆叠的RGCN对数据交易拓扑图中的节点进行二分类预测，对应标签概率计算式为

$$\hat{y}_i=\sigma_{\text{sigmoid}}\left(\boldsymbol{W}_{\text{cls}}\boldsymbol{h}_i^{(L)}\right)\tag{2}$$

其中，$\hat{y}_i\in\{0,1\}$是预测节点i属于违规转售交易的概率，$\boldsymbol{W}_{\text{cls}}$是分类层的权重矩阵，$\boldsymbol{h}_i^{(L)}$是经过$L$层堆叠的RGCN处理后的节点$i$的最终特征。

然而，经典RGCN无法适配数据交易场景的两大固有特点，在数据违规转售检测任务中存在先天性局限。

1) RGCN无法对邻居节点间的时序关系进行有效表征。对于相邻交易节点来说，交易时间先后对于判断两笔交易之间的关系、区分正常与违规转售交易具有重要作用。假设交易j的时间晚于i，如果二者通过角色交换型边相连，那么i与j之间就是同一主体完成买方、卖方转换后所形成的“先买后卖”或“先卖后买”触发关系；如果二者通过角色共享型边相连，那么i与j之间就是同一主体先后买入或卖出形成的关联关系。在数据违规转售检测任务中，需要额外关注“先买后卖”这类触发关系，后发生的交易有可能就是违规转售行为。

定义2　交易因果顺序。相邻交易节点间以时序先后为前提所表现出的前序交易行为对后序交易行为的触发或关联关系。交易因果顺序对于发现违规转售交易行为非常重要，然而RGCN不具备表征节点间时序关系的能力，因此无法利用交易因果顺序来检测违规转售交易。

2) RGCN无法对同一关系类型下邻居节点间各部分特征的相似性进行建模。交易节点特征的核心部分是产品类别、行业、领域等交易上下文信息。在数据违规转售检测任务中，检测效果高度依赖对交易上下文信息的有效挖掘，邻居节点间交易上下文信息相似程度对于发现违规转售交易节点具有重要作用。例如，i与j由角色交换型边连接，如果i与j的交易上下文信息相似程度很高，那么两笔交易之间有一笔是违规转售交易的可能性就很高。但是，RGCN对同一关系类型下邻居节点间的各部分特征采取均等化处理方式，无法有效提取节点特征中的交易上下文信息。

因此，为提升数据交易场景中违规转售交易的检测效果，必须构建一种可融合建模交易因果顺序与交易上下文信息的增强RGCN模型。

3　方案设计

本文设计了一种融合交易上下文信息相似度与交易因果顺序约束的增强关系图卷积网络模型——CCA-RGCN（context-chrono aware RGCN），并基于该模型提出了一种数据违规转售检测方法。本方法由4个模块组成：1)数据交易拓扑图构建模块；2)基于交易上下文信息相似度的注意力加权模块；3)基于交易因果顺序约束的衰减时序门控模块；4)违规转售交易节点判别模块。其中，模块2)和模块3)是CCA-RGCN的核心组成部分。

基于CCA-RGCN的数据违规转售检测方法架构如图2所示。首先，基于多个跨平台数据交易记录构建数据交易拓扑图，将原始历史时序交易记录转化为包含两种交易关系的图结构表示。随后，计

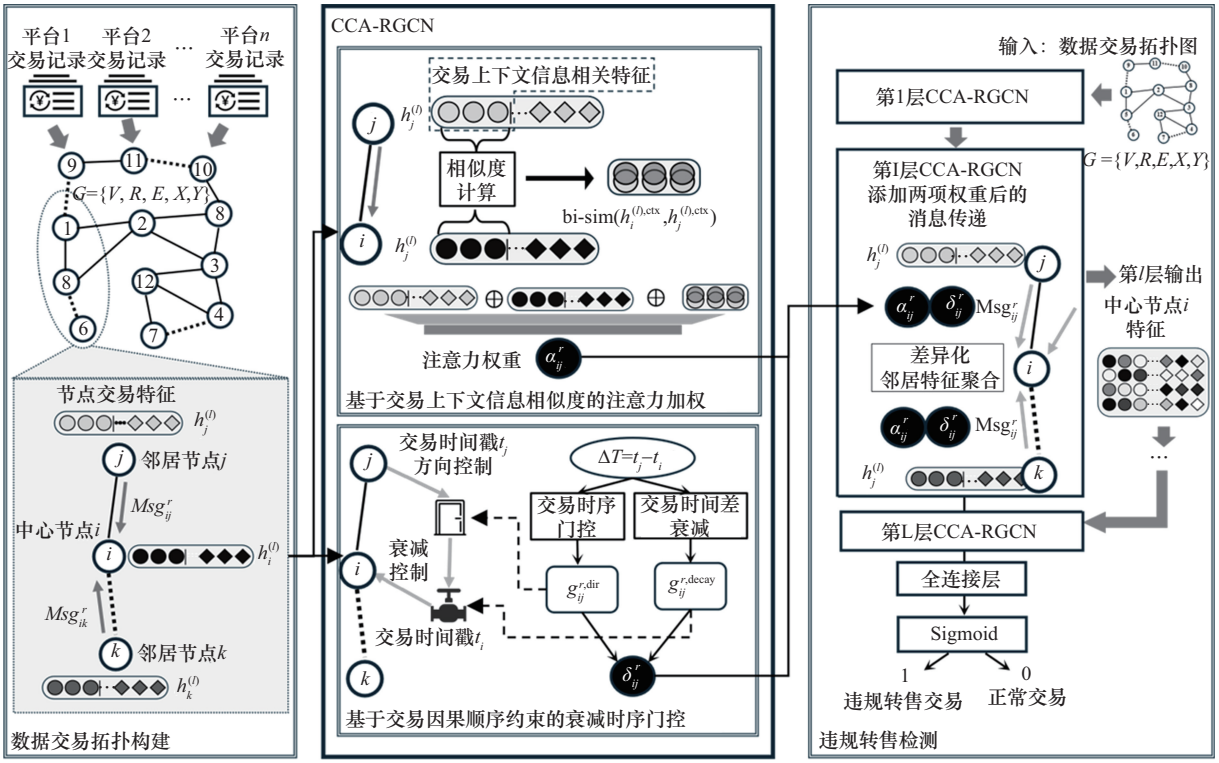


图2　基于CCA-RGCN的数据违规转售检测方法架构图

算相邻交易节点间的上下文信息相似度，基于该相似度，通过注意力机制为消息传递过程中的邻居特征分配动态权重；同时，根据邻居节点之间的交易时间差，通过时序门控发现交易之间的因果顺序，继而结合时间差衰减，对消息传递过程中邻居特征的方向性和权重进行控制。最后，通过堆叠多层CCA-RGCN学习交易节点的行为特征高阶表示，并结合全连接层与判别函数，对交易节点进行分类，实现对违规转售交易的检测。

3.1　数据交易拓扑图构建

如算法1所示，本文提出一种基于买卖双方身份标识的数据交易拓扑图构建算法，根据多个数据交易平台的带标记历史交易记录，创建双关系无向图 $G=\{V,R,E,\boldsymbol{X},\boldsymbol{Y}\}$。首先，生成图的节点、节点特征和标记（1)~3)行），分别根据买家标识（6)~11)行）和卖家标识（12)~15)行），找到其参与的所有交易，并在两两交易之间连接边，得到角色共享型边集合；然后，以此类推根据所有身份标识得到混合边集合（16)~19)行），后者与前者的差集就是角色交换型边集合（20)行）；

最后，根据所属集合确定每条边的关系（21)行）。

算法1　基于身份的数据交易拓扑图构建

输入　数据交易记录df_trade

输出　双关系无向图 $G=\{V,R,E,\boldsymbol{X},\boldsymbol{Y}\}$

1) V=df_trade中所有交易的索引
2) $\boldsymbol{X}$=df_trade中所有交易的特征
3) $\boldsymbol{Y}$=df_trade中所有交易的标记
4) 初始化 角色共享型边集合E_RS, 角色交换型边集合E_RX, 混合边集合E_ALL, 边对应关系集合Rel
5) 从df_trade提取买家身份标识集合ID_BUY和卖家身份标识集合ID_SELL
6) for id in ID_BUY do:
7) 初始化身份标识到交易索引的映射ID_TXN, 边集合edge
8) 遍历df_trade, 找到所有与id相关的交易索引, 添加到ID_TXN
9) 遍历ID_TXN中的所有交易索引, 两两之间生成一条无向边, 添加到edge
10) end for
11) 将edge添加到E_RS
12) for id in ID_SELL do:
13) 重复7) ~9)
14) end for

15) 将edge添加到E_RS

16) for id in (ID_BUY+ID_SELL) do:

17)　重复7) ~9)

18) end for

19) 将edge添加到E_ALL

20) E_RX=E_ALL-E_RS

21) 根据E_RS和E_RX，确定每条边的关系，更新Rel

22) R=Rel

23) V=E_RS + E_RX

3.2　基于交易上下文信息相似度的注意力加权

节点的特征可分为交易基础特征（交易时间、买卖双方、交易金额等）和交易上下文信息特征（产品类别、行业、更新频率、交易类型等）。据此，将式(1)中通过关系r类型边相连接的邻居节点特征$\boldsymbol{h}_j^{(l)}$拆分为交易基础特征$\boldsymbol{h}_j^{(l),\text{core}} \in \mathbb{R}^{H-H'}$和交易上下文信息特征$\boldsymbol{h}_j^{(l),\text{ctx}} \in \mathbb{R}^{H'}$，表示为

$$\boldsymbol{h}_j^{(l)} = \boldsymbol{h}_j^{(l),\text{core}} \oplus \boldsymbol{h}_j^{(l),\text{ctx}} \tag{3}$$

经典RGCN无法利用节点的交易上下文信息来捕捉违规转售交易行为特征。对于式(1)中的经典RGCN消息传递机制来说，$\boldsymbol{h}_j^{(l),\text{core}}$与$\boldsymbol{h}_j^{(l),\text{ctx}}$并无本质上的区别，均通过统一的权重$\boldsymbol{W}_r^{(l)}$完成无差异的特征变换。然而$\boldsymbol{h}_j^{(l),\text{ctx}}$相较于$\boldsymbol{h}_j^{(l),\text{core}}$而言包含更多的违规转售判断信息。举例来说，如果中心节点i和邻居节点j之间由角色交换型边相连，且二者的$\boldsymbol{h}_i^{(l),\text{ctx}}$和$\boldsymbol{h}_j^{(l),\text{ctx}}$相似性很高，那么$i$和$j$之间更有可能构成潜在的转售关联；反之，$i$和$j$之间是正常交易的可能性更大。

由此可见，邻居节点间的交易上下文信息相似程度对于发现违规转售行为具有重要意义。交易上下文信息越相似，这一对交易存在违规转售关联的可能性就越大。因此需要为RGCN增加一种通过交易上下文信息相似性捕捉违规转售关联信息的机制。

本文为CCA-RGCN设计了一种结合了注意力的交易上下文信息挖掘机制，对邻居节点相关特征的相似度进行量化和表征，并将其作为影响因子融入邻居消息传递过程，从而实现根据邻居节点交易上下文信息相似度对邻居特征进行差异化加权聚合，提升对违规转售交易节点的检测能力。

首先，将节点的交易上下文信息特征$\boldsymbol{h}_j^{(l),\text{ctx}}$映射到$Z$维嵌入空间

$$\hat{\boldsymbol{h}}_i^{(l),\text{ctx}} = \boldsymbol{W}^{\text{ctx}} \boldsymbol{h}_i^{(l),\text{ctx}} \tag{4}$$

其中，$\boldsymbol{W}^{\text{ctx}} \in \mathbb{R}^{Z \times H'}$是可训练的映射权重矩阵。

然后，对邻居节点i和j之间的交易上下文信息相似度进行测量和表征。本文采用双线性相似度来表示$\hat{\boldsymbol{h}}_i^{(l),\text{ctx}}$和$\hat{\boldsymbol{h}}_j^{(l),\text{ctx}}$之间的相似度，计算式为

$$\text{bi-sim}\left(\hat{\boldsymbol{h}}_i^{(l),\text{ctx}}, \hat{\boldsymbol{h}}_j^{(l),\text{ctx}}\right) = \frac{\hat{\boldsymbol{h}}_i^{(l),\text{ctx}\top} \boldsymbol{M} \hat{\boldsymbol{h}}_j^{(l),\text{ctx}}}{\left\| \hat{\boldsymbol{h}}_i^{(l),\text{ctx}} \right\|_2 \cdot \left\| \hat{\boldsymbol{h}}_j^{(l),\text{ctx}} \right\|_2} \tag{5}$$

其中，$\left\| \hat{\boldsymbol{h}}_i^{(l),\text{ctx}} \right\|_2$和$\left\| \hat{\boldsymbol{h}}_j^{(l),\text{ctx}} \right\|_2$表示对$\hat{\boldsymbol{h}}_i^{(l),\text{ctx}}$和$\hat{\boldsymbol{h}}_j^{(l),\text{ctx}}$进行$L2$归一化。$\boldsymbol{M} \in \mathbb{R}^{Z \times Z}$是可训练的双线性权重矩阵，可以学习$\hat{\boldsymbol{h}}_i^{(l),\text{ctx}}$和$\hat{\boldsymbol{h}}_j^{(l),\text{ctx}}$两个向量的特征之间应如何交叉、组合和加权。因此，采用双线性相似度能够更精准地捕捉邻居节点之间交易上下文信息特征的组合关联模式，从而生成更有意义的相似度分数。

接着，针对得到的邻居节点交易上下文信息相似度，利用注意力机制来量化其对邻居消息的影响。将中心节点特征$\boldsymbol{h}_i^{(l)}$、邻居节点特征$\boldsymbol{h}_j^{(l)}$和二者之间的交易上下文信息相似度$\text{bi-sim}\left(\hat{\boldsymbol{h}}_i^{(l),\text{ctx}}, \hat{\boldsymbol{h}}_j^{(l),\text{ctx}}\right)$拼接，然后通过一个可训练的注意力向量$\boldsymbol{a}^r \in \mathbb{R}^{3D}$映射为标量空间中的注意力得分$e_{ij}^r$，表示为

$$e_{ij}^r = \text{LeakyReLU}\left(\boldsymbol{a}^r \left[\boldsymbol{W}_{\text{att}} \boldsymbol{h}_i^{(l)} \| \boldsymbol{W}_{\text{att}} \boldsymbol{h}_j^{(l)} \| \text{bi-sim}\left(\hat{\boldsymbol{h}}_i^{(l),\text{ctx}}, \hat{\boldsymbol{h}}_j^{(l),\text{ctx}}\right)\right]\right) \tag{6}$$

其中，$\boldsymbol{W}_{\text{att}} \in \mathbb{R}^{H \times H}$是可训练的权重矩阵；$\text{LeakyReLU}(\cdot)$是引入非线性的激活函数；$\boldsymbol{a}^r$和$e_{ij}^r$上标中的$r \in R$代表边的类型，即针对每种边计算相应的注意力得分。

之所以要在交易上下文信息相似度之外还要将邻居节点特征和与中心节点特征一同拼接，是因为这3个部分对注意力得分具有不同的意义。拼接$\text{bi-sim}\left(\hat{\boldsymbol{h}}_i^{(l),\text{ctx}}, \hat{\boldsymbol{h}}_j^{(l),\text{ctx}}\right)$的目的是增加邻居节点之间的交易上下文信息特征相似性对注意力得分的影响。而将$\boldsymbol{h}_i^{(l)}$与$\boldsymbol{h}_j^{(l)}$拼接，可以使注意力得分能够额外捕捉中心节点与邻居节点之间所有交易特征的交互模式，发现正常与异常交易之间的行为模式偏离，提升对违规转售交易的检测精度。

求得注意力得分e_{ij}^r后，针对每一种类型边，分别利用softmax函数对注意力得分进行归一化，可得到相应注意力权重，将其定义为邻居节点交易

上下文信息相似度权重α_{ij}^r，表示为

$$\alpha_{ij}^r = \mathrm{softmax}\left(\exp\left(e_{ik}^r\right)\right) = \frac{\exp\left(e_{ik}^r\right)}{\sum\limits_{k \in \mathcal{N}_r(i)} \exp\left(e_{ik}^r\right)} \tag{7}$$

其中，i和r满足

$$\forall i,r: \sum_{j \in \mathcal{N}_r(i)} \alpha_{ij}^r = 1 \tag{8}$$

其中，α_{ij}^r是邻居节点交易上下文信息相似度在邻居特征聚合过程中对所传递邻居消息产生的影响因子，可以实现对邻居特征的差异化加权聚合，让CCA-RGCN更关注与中心节点交易上下文信息高度相似的邻居特征，从而更有效地识别可疑的数据转售行为。

3.3 基于交易因果顺序约束的衰减时序门控

交易因果顺序对于检测违规转售行为具有重要作用，“先买后卖”交易因果顺序可作为识别违规转售交易节点的重要依据。除交易因果顺序外，邻居节点之间的交易时间差也能为违规转售检测任务提供额外判别信息。其原因在于大部分数据产品会周期性更新，违规转售者因此会选择尽快卖出所购入的数据产品；若转售周期过长，数据产品的真正所有者将完成更新，潜在买家不会购买违规转售者手中的过时数据产品，转售行为也将失去价值。

然而，经典RGCN无法对交易行为的时间顺序进行建模。因此，本文为CCA-RGCN设计了一种衰减时序门机制，用于发现中心节点和邻居节点之间的交易因果顺序，并将其作为影响因子融入邻居消息传递过程，对邻居特征进行方向性约束。该方法可分为交易时序门控部分和交易时间差衰减部分：交易时序门控部分在邻居消息传递过程中主要起到方向滤波器的作用，通过判断邻居节点交易是否发生于中心节点之前，限制可能发生的从未来交易向过去交易的消息传递；交易时间差衰减部分起到了对时序门控部分的调控作用，可以给短期内发生的转售交易对应的邻居消息赋予更高的权重。由于需要建模的时序关系仅是邻居节点间的交易顺序，而非长时间序列，因此不必采用类似于LSTM这样解决长时序依赖的网络，避免在稀疏时间样本上引发过拟合，同时避免计算开销的增加。

针对交易时序门控部分，本文提出了一种与边类型相关的时序方向门，其核心是边类型相关因果评分函数。该评分函数的设计原则是根据交易时间差限制邻居特征消息的传递方向，对沿时间正向传递的消息予以奖励，反之则予以惩罚，可以显著增强“先买后卖”这种疑似违规转售的交易因果顺序所对应的邻居特征。其具体定义式为

$$f_{ij}^r(\Delta T) = d_{ij}^r \cdot \Delta T - c_{ij}^r \cdot |\Delta T| - b_{ij}^r \tag{9}$$

其中，$\Delta T = T_j - T_i$，T_i和T_j分别是邻居节点i和j的交易时间戳，$\Delta T < 0$说明邻居节点j向中心节点i传递的消息是从未来交易传向过去交易，反之亦然；d_{ij}^r是可训练的时序方向惩罚参数，$d_{ij}^r \cdot \Delta T$奖励从过去交易向未来交易的消息传递，惩罚从未来交易向过去交易的消息传递，其数值越大则对消息传递方向的约束性越强；c_{ij}^r是可训练的时间差惩罚参数，$-c_{ij}^r \cdot |\Delta T|$对“过去交易→未来交易”以及“未来交易→过去交易”两个方向上较大的交易时间差都进行惩罚；b_{ij}^r是可训练的交易间隔参数，其作用是学习到交易之间的最小时间间隔，可以惩罚过快的交易行为；所有参数上标中的$r \in R$代表边的类型，即针对每种边可训练不同的时序方向惩罚参数、时间差惩罚参数和交易间隔参数。然后，利用激活函数将$f_{ij}^r(\Delta T)$的函数值限制在(0,1)内，即得到

$$g_{ij}^{r,\mathrm{dir}} = \sigma\left(f_{ij}^r(\Delta T)\right) = \mathrm{sigmoid}\left(d_{ij}^r \cdot \Delta T - c_{ij}^r \cdot |\Delta T| - b_{ij}^r\right) \tag{10}$$

其中，$g_{ij}^{r,\mathrm{dir}} \in (0,1)$就是最终的边类型相关时序方向门，当$g_{ij}^{r,\mathrm{dir}}$趋近于0时，对邻居节点$j$向中心节点$i$传递的消息起限制作用。

交易时间差衰减部分的设计原则是根据交易时间间隔长短对消息传递过程中的邻居特征施加权重。时间间隔越长则违规转售的可能性越小，权重也就越小；反之则权重增大。因此，采用一阶指数衰减的形式来计算交易时间差的权重，计算式为

$$g_{ij}^{r,\mathrm{decay}} = \exp\left(-\frac{|\Delta T|}{\tau_r}\right) \tag{11}$$

其中，τ_r是与边类型相关的可训练参数，可针对不同边类型学习到不同的衰减速率，影响交易时间差权重。

最终形式的衰减时序门融合了$g_{ij}^{r,\mathrm{dir}}$与$g_{ij}^{r,\mathrm{decay}}$，将其输出结果定义为邻居节点交易因果顺序约束权重$\delta_{ij}^r$，表示为

$$\delta_{ij}^r = g_{ij}^{r,\mathrm{dir}} \cdot g_{ij}^{r,\mathrm{decay}} \tag{12}$$

其中，δ_{ij}^r就是邻居节点交易因果顺序在邻居特征聚

合过程中对所传递消息施加的方向性约束影响因子，使CCA-RGCN优先聚合与中心节点存在“先买后卖”交易因果顺序触发关系的邻居信息，可以强化对违规转售交易节点的识别能力。

3.4　CCA-RGCN模型

基于结合了注意力的交易上下文信息挖掘机制和结合了衰减时序门的交易因果顺序约束机制，对式(1)中的经典RGCN消息传递机制添加相应的邻居节点交易上下文信息相似度权重和交易因果顺序约束权重，这就构成了本文CCA-RGCN模型的核心部分。在经典RGCN中，邻居节点j经过r类型边向中心节点i传递的消息，就是经过关系r对应权重矩阵变换后的自身特征，即式(1)中的$\boldsymbol{W}_r^{(l)}\boldsymbol{h}_j^{(l)}$；而在CCA-RGCN中，$j$向$i$传递的消息则添加了上述两项权重，表示为

$$\mathrm{Msg}_{ij}^{r}=\alpha_{ij}^{r}\cdot\delta_{ij}^{r}\cdot\boldsymbol{W}_r^{(l)}\boldsymbol{h}_j^{(l)} \tag{13}$$

其中，α_{ij}^{r}和δ_{ij}^{r}分别是式(7)和式(12)中定义的两项权重。α_{ij}^{r}与δ_{ij}^{r}相结合，通过乘积耦合的方式共同为传递的邻居消息分配权重，在特征聚合过程中对违规转售节点特征进行联合强化，有效整合了衰减时序门控模块和注意力加权模块的输出，使CCA-RGCN能够更高效地捕捉数据违规转售的行为模式。将这两个模块分开设计，还可以根据两项权重数值，为模型的检测结果提供一定的解释性。

除了添加两项权重因子外，CCA-RGCN针对RGCN消息传递公式中的自连接项$\boldsymbol{W}_0^{(l)}\boldsymbol{h}_i^{(l)}$进行了改进。$\boldsymbol{h}_i^{(l)}$是中心节点$i$自身的特征向量，将其变换后予以保留的初衷是避免中心节点被异质邻居的特征“稀释”，其前提是$\boldsymbol{h}_i^{(l)}$与聚合后的邻居消息具有一致的信息量。然而在数据交易拓扑图中，与边类型相关的邻居节点消息Msg_{ij}^{r}与$\boldsymbol{h}_i^{(l)}$异质性程度较高，如果还使用经典RGCN中的自连接项，中心节点保有的有价值信息将被稀释。因此，借鉴残差网络思想，CCA-RGCN利用门控残差连接来替换原有的自连接项，即

$$\boldsymbol{W}_0^{(l)}\boldsymbol{h}_i^{(l)}\Rightarrow(1-\gamma^{(l)})\boldsymbol{W}_0^{(l)}\boldsymbol{h}_i^{(l)} \tag{14}$$

其中，$\gamma^{(l)}\in(0,1)$是可训练的标量门，可以有效调节中心节点特征和聚合后邻居节点特征之间的信息占比。

基于上述改进，最终形式的CCA-RGCN消息传递计算式由式(1)变为

$$\boldsymbol{h}_i^{(l+1)}=\sigma\left(\sum_{r\in R}\sum_{j\in\mathcal{N}_r(i)}\alpha_{ij}^{r}\cdot\delta_{ij}^{r}\cdot\boldsymbol{W}_r^{(l)}\boldsymbol{h}_j^{(l)}+(1-\gamma^{(l)})\boldsymbol{W}_0^{(l)}\boldsymbol{h}_i^{(l)}\right) \tag{15}$$

CCA-RGCN从交易上下文信息和交易因果顺序约束两个层面分别对违规转售行为进行特征建模，并在消息传递过程中添加乘积耦合权重，实现对违规转售特征的联合强化表征，弥补了经典RGCN无法对同一关系类型下邻居节点特征差异性以及时间顺序进行建模的双重缺陷。与T-GCN、EvolveGCN这种解决节点特征或图结构参数随时间演化问题的神经网络不同，CCA-RGCN更关注的是数据交易拓扑中节点间短期内的交易时间先后顺序和交易上下文信息相似度。

3.5　违规转售交易节点检测

违规转售交易节点检测网络由L层CCA-RGCN和1层全连接层堆叠构成，上一层的输出作为下一层的输入。检测网络的初始输入除了生成的数据交易拓扑图$G=\{V,R,E,\boldsymbol{X},\boldsymbol{Y}\}$外，还包括两个索引变量：$I_T$指出节点交易特征$\boldsymbol{x}_i$中交易时间戳的索引位置，用于衰减时序门中邻居节点交易时间差的计算；I_C指出节点特征$\boldsymbol{x}_i$中交易上下文信息特征（即$\boldsymbol{x}_i^{\mathrm{ctx}}$）的索引范围。第$L$层CCA-RGCN的输出$\boldsymbol{h}_i^{(L)}$经过全连接层进行线性变换，再通过sigmoid函数将输出映射到[0,1]之间，这就是检测网络的输出，即交易i是违规转售交易的概率p_i。

在检测网络训练过程中使用加权二元交叉熵损失，其损失计算式为

$$\mathcal{L}_{\mathrm{WBCE}}=-\frac{1}{N}\sum_{i=1}^{N}\left[w\cdot y_i\ln(p_i)+(1-y_i)\ln(1-p_i)\right] \tag{16}$$

其中，w是正样本的权重系数，$y_i\in\{0,1\}$是第i个节点的真实标签（1表示违规转售，0表示合规交易）。

在每一层CCA-RGCN计算输出时，沿用了经典RGCN对权重矩阵$\boldsymbol{W}_r^{(l)}$进行的基分解方法，即

$$\boldsymbol{W}_r^{(l)}=\sum_{b=1}^{B}a_{r,b}^{(l)}\cdot\boldsymbol{V}_b^{(l)} \tag{17}$$

其中，$\boldsymbol{V}_b^{(l)}$是基矩阵；B是基矩阵的数量；$a_{r,b}^{(l)}$是第l层中，关系r对应第b个基矩阵的组合系数（标量），每个关系仅学习B个系数。采用基分解的目

的是在不损失关系特征建模能力的前提下，减少多种边类型关系导致的参数膨胀问题。

4 实验与结果分析

4.1 数据集

本文基于公开的在线交易数据集Olist Brazilian E-Commerce[26]，构建了面向数据违规转售检测（data trading resale detection）的数据集DTRD。针对真实交易场景中违规转售行为具有强隐蔽性，且受商业隐私约束导致标注样本极度稀缺的问题，本文通过异常注入的方式生成带标注的违规转售样本。DTRD的字段设计参考了本文研究合作的数据交易所实际业务流程与数据结构，从而使仿真环境更加贴近真实数据交易场景。

首先对原始数据进行预处理，剔除缺失或异常交易记录，并对数值型特征进行归一化处理。将在线交易商品及其交易过程等价构建为数据产品及数据交易行为，保留交易价格、交易时间、买卖双方等基础属性，并将商品类别、商品规格等映射为行业、产品类型等交易上下文信息。

在异常注入阶段，根据预设的异常比例，从构建的正常数据交易中选取样本作为初始交易，然后按以下条件生成转售交易：1) 产品ID、交易ID重新生成；2) 初始交易的买家成为转售交易的卖家；3) 转售交易时间晚于初始交易时间；4) 转售交易买家、转售交易与初始交易之间的时间差通过自助法从正常交易的买家和交易时间差中抽取；5) 转售价格在原始价格的±5%区间内均匀浮动；6) 行业、产品类型等特征保持一致。

最终构建的DTRD数据集包含44 375笔有效交易。为系统评估模型在不同异常比例下的鲁棒性，本文分别构建了异常交易注入比例约为1%、5%和15%的3个数据集，分别记为DTRD-1、DTRD-5和DTRD-15。

为评估CCA-RGCN在除数据违规转售检测外其他通用任务上的泛化性能，本文还在公开数据集YelpChi[27]上进行了对比实验。

4.2 实验设置

本文选取多种具有代表性的方法作为对比模型进行实验验证。对比方法包括仅利用交易节点特征进行判别的非图模型，以及能够建模交易关系的图神经网络模型。

多层感知机（multilayer perceptron, MLP）仅利用交易节点自身特征进行判别，用于评估在不引入交易关系情况下的违规转售检测性能。

GCN[16]通过在图结构上聚合邻居节点特征来更新节点表示，可以建模交易节点之间的交易关系。

RGCN[22]在GCN的基础上引入多类型关系，可以建模交易节点之间不同类型的交易关系。

GAT[21]通过注意力机制对不同邻居特征分配自适应权重，可以聚焦关键交易节点。

GraphSAGE[20]采用基于采样的邻居聚合策略，可以对未知交易节点和图结构进行归纳式推理。

CARE-GNN[13]采用了标签感知相似性度量、邻居选择等机制，可以检测经过伪装的欺诈交易节点。

PC-GNN[14]设计了标签平衡采样策略与邻居采样策略，可以缓解图中交易节点类别不平衡问题。

为保证对比实验的公平性，所有方法均采用相同的交易特征输入及相同的交易关系图结构进行训练与评估。数据集按照4:2:4的比例划分为训练集、验证集和测试集，分别用于模型训练、参数选择和性能评估。模型训练采用小批量子图采样策略，并统一使用加权二元交叉熵损失函数以缓解类别不平衡问题，其中类别权重根据训练集中样本比例设置。优化器选用Adam，学习率为1×10^{-3}，权重衰减为5×10^{-4}，隐藏维度设为64；训练过程中基于验证集性能进行模型参数选择，并采用早停策略防止过拟合。对比模型均按照原论文或官方实现中的推荐配置进行设置。

4.3 评价指标

数据交易违规转售检测任务中异常样本数量远少于正常样本，传统以准确率为代表的评价指标容易受到多数类别样本主导，难以真实反映模型对违规转售行为的检测性能。因此，本文选用ROC-AUC、F1-macro及AP这3个对类别偏差具有较强鲁棒性的评价指标。

ROC-AUC表示受试者操作特征（receiver operating characteristic, ROC）曲线下的面积，其可形式化定义为随机选取一个异常交易样本x^+与一个正常交易样本x^-，模型对异常样本给出更高预测得分的概率，计算式为

$$\text{ROC - AUC} = P\left(s(x^+) > s(x^-)\right) \tag{18}$$

其中，$s(\cdot)$表示模型输出的预测得分。AUC不依赖具体判别阈值，能够反映模型对异常交易样本的整体排序能力。

AP（average precision）为Precision-Recall曲线下的面积，其定义为在不同召回率区间内对精确率进行加权平均，计算式为

$$\mathrm{AP}=\sum_{k=1}^{N}\left(R_k-R_{k-1}\right)P_k \tag{19}$$

其中，P_k与R_k分别表示第k个判别阈值下的精确率与召回率。与ROC-AUC相比，AP更关注模型在高召回区域的性能表现，能够更直接反映模型在少数异常样本识别与排序方面的有效性。

$F1$分数定义为精确率与召回率的调和平均，计算式为

$$F1_c=\frac{2\cdot \mathrm{Precision}_c\cdot \mathrm{Recall}_c}{\mathrm{Precision}_c+\mathrm{Recall}_c} \tag{20}$$

其中，c表示类别标签。F1-macro为各类别$F1$分数的未加权平均，表示为

$$\mathrm{F1\text{ - }macro}=\frac{1}{C}\sum_{c=1}^{C}\mathrm{F1}_c \tag{21}$$

其中，C为类别数，该指标能够避免评价结果被多数类别样本主导，用于衡量模型在各类别上的整体分类平衡性。

4.4　对比实验结果

不同异常比例下各模型在DTRD数据集上的性能对比如表1所示，每组实验均进行3次取平均值以避免偶然性。从表1可以看出，本文CCA-RGCN的性能明显优于所有对比方法，证明了其有效性。同时，可以得出以下几个结论。

1) 在不利用图结构信息的情况下，MLP在各异常比例设置下的检测性能始终处于较低水平，说明仅依赖节点自身特征难以有效识别违规转售行为。当进一步引入图结构后，GCN、GAT与GraphSAGE的性能虽有所改善，但整体提升仍然有限，尤其在DTRD-1这一极端不平衡数据集上，其检测能力下降更为明显。其中，GCN与GAT的F1-macro和AP均低于0.5，表明在交易关系稀疏且异质的条件下，仅依赖同质性假设难以实现对正常交易与违规转售交易的有效区分。

2) 针对欺诈场景特征进行改进的图模型在检测性能上表现出一定优势。PC-GNN通过引入不平衡感知采样，在低异常比例场景下保持了相对稳定的检测效果，表明针对类别分布进行建模有助于缓解异常样本稀缺带来的影响。然而，该方法主要从样本分布层面进行调整，对交易上下文信息与交易时序利用仍然有限。CARE-GNN通过引入邻居筛选与权重调整机制，在一定程度上提升了针对噪声与伪装行为的鲁棒性，但其性能对邻居质量较为敏感，在不同异常比例设置下波动明显。

3) RGCN通过显式建模多类型交易关系，在不同异常比例设置下均表现出较为稳定的检测性能，验证了关系信息在违规转售检测中的重要作用，在DTRD-15数据集中，其AP值达到0.89。然而RGCN无法对同一关系类型下邻居节点特征差异性以及时间顺序进行建模，限制了违规转售检测效果。

4)在多关系建模基础上，CCA-RGCN进一步改进了邻居信息的传递方式，在各异常比例设置下

表1　　**不同异常比例下各模型在DTRD数据集上的性能对比**

模型	DTRD-1			DTRD-5			DTRD-15		
	F1-macro	ROC-AUC	AP	F1-macro	ROC-AUC	AP	F1-macro	ROC-AUC	AP
MLP	0.616 4	0.666 3	0.312 7	0.709 0	0.826 7	0.515 4	0.638 5	0.847 5	0.633 0
GCN	0.486 5	0.785 4	0.432 0	0.502 0	0.858 3	0.597 6	0.587 5	0.911 4	0.687 6
GAT	0.474 5	0.826 2	0.370 0	0.587 3	0.920 2	0.664 2	0.649 7	0.919 9	0.675 4
GraphSAGE	0.534 1	0.817 7	0.471 6	0.625 4	0.877 2	0.707 7	0.697 0	0.914 1	0.788 2
RGCN	0.666 0	0.885 9	0.638 3	0.727 7	0.931 3	0.828 4	0.788 5	0.958 8	0.889 7
CARE-GNN	0.496 0	0.738 8	0.401 1	0.717 1	0.844 9	0.403 1	0.765 1	0.864 3	0.460 4
PC-GNN	**0.686 7**	0.844 6	0.432 8	0.730 7	0.862 4	0.671 4	0.754 2	0.887 0	0.719 2
CCA-RGCN	0.664 2	**0.935 9**	**0.783 1**	**0.772 1**	**0.978 3**	**0.884 6**	**0.791 0**	**0.984 1**	**0.933 5**

均表现出明显优势。在异常比例最低的DTRD-1数据集中，其AP为0.78，显著高于RGCN的0.64，提升幅度达到14.48%，说明在极端不平衡场景下模型对异常样本的风险排序能力显著增强。CCA-RGCN对交易上下文信息相似度高且存在“先买后卖”交易因果顺序的交易赋予更高权重，从而提升异常风险评分的区分度。在DTRD-5与DTRD-15数据集中，其在AP与ROC-AUC等阈值无关指标上的持续优势进一步说明，改进效果体现在整体排序质量的提升，而非阈值调节带来的局部优化。

为了分析模型在不同判定阈值下的性能变化情况，本文在DTRD-5数据集上对比了各模型在违规转售检测任务上的PR曲线，如图3所示。

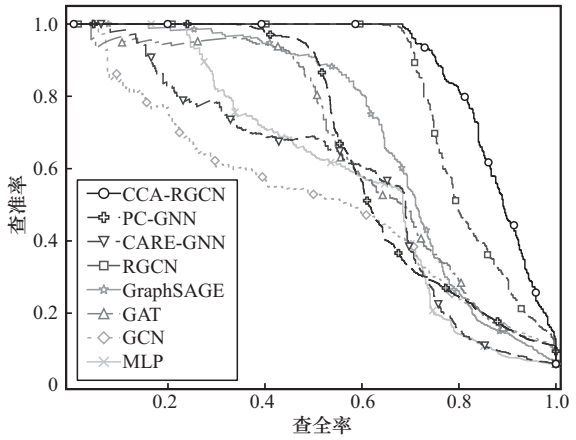


图3 不同模型在违规转售检测任务上的PR曲线对比

由图3可以看出，CCA-RGCN的曲线整体位于其他模型之上。当召回率超过0.6时，其精确率下降趋势相对缓和，在较高召回条件下仍维持较好的误报控制能力。相比之下，其余模型在召回率提升过程中精确率下降更为明显，在召回率超过0.8区间内精确率已降至较低水平。该结果表明，CCA-RGCN在实际违规转售检测场景中更有利于在保证较高异常覆盖率的同时维持稳定的检测性能。

此外，为验证本文方法的泛化能力，本文进一步在公开数据集YelpChi上进行了对比实验，各模型的检测性能如表2所示，每组实验均进行3次取平均值以避免偶然性。尽管在YelpChi数据集上CCA-RGCN并非表现最优，但其整体性能仍处于中上水平，未出现明显性能退化，表明该方法并未依赖特定异常构造假设，具备一定的跨数据集泛化能力。

表2 各模型在YelpChi数据集上的检测性能

模型	F1-macro	ROC-AUC	AP
MLP	0.461 1	0.615 3	0.208 0
GCN	0.420 9	0.631 5	0.202 1
GAT	0.607 0	0.760 4	0.393 3
GraphSAGE	0.450 8	0.536 4	0.431 2
RGCN	0.598 4	0.743 2	0.460 0
CARE-GNN	0.649 3	0.793 4	0.426 8
PC-GNN	**0.679 7**	**0.818 1**	**0.480 3**
CCA-RGCN	0.594 4	0.748 9	0.462 6

4.5 消融实验

本节通过消融实验对CCA-RGCN模型中各模块的有效性进行验证。为分析各模块对整体性能的贡献，基于完整模型构造了3个变体方法，分别去除其中的某一模块，其余结构与参数设置保持一致。

1) CCA-RGCN-w/o Dir：针对CCA-RGCN模型中“基于交易因果顺序约束的衰减时序门控模块”，去除交易时序门控部分，即去除式(12)中的$g_{ij}^{r,\mathrm{dir}}$，用于分析在消息传递过程中对邻居特征所施加的方向性权重对于违规转售行为学习的影响。

2) CCA-RGCN-w/o Decay：针对CCA-RGCN模型中“基于交易因果顺序约束的衰减时序门控模块”，去除交易时间差衰减部分，即去除式(12)中的$g_{ij}^{r,\mathrm{decay}}$，用于分析在消息传递过程中对邻居特征所施加的交易时间间隔权重对于违规转售行为学习的影响。

3) CCA-RGCN-w/o Contextual：整体去除CCA-RGCN模型中的“基于交易上下文信息相似度的注意力加权模块”，用于分析消息传递过程中的交易上下文信息相似度权重对违规转售行为特征建模的影响。

在数据集DTRD-5上的消融实验结果如表3所示。由表3可知，移除任一模块后模型性能均出现不同程度下降，表明各模块在违规转售检测任务中均对模型性能产生了不可替代的影响。

去除“基于交易因果顺序约束的衰减时序门控模块”中的交易时序门控部分后，CCA-RGCN的F1-macro、ROC-AUC和AP分别由0.77、0.97、0.88降为0.74、0.95、0.83；这说明在学习违规转售行为时，交易因果顺序在邻居特征聚合过程中施

加的方向性约束最为关键。这一结果符合直观判断：在数据交易场景下，同一主体“先买后卖”的行为涉嫌违规转售的可能性最大。

表3　　消融实验结果

模型	F1-macro	ROC-AUC	AP
CCA-RGCN	**0.772 1**	**0.978 3**	**0.884 6**
CCA-RGCN-w/o Dir	0.741 5	0.951 0	0.832 0
CCA-RGCN-w/o Decay	0.764 5	0.962 0	0.863 3
CCA-RGCN-w/o Contextual	0.755 1	0.958 9	0.841 5

整体去除“基于交易上下文信息相似度的注意力加权模块”后，CCA-RGCN的F1-macro、ROC-AUC和AP分别降为0.76、0.96、0.84；这说明在建模违规转售行为特征时，交易上下文信息相似度权重至关重要。在数据交易场景下，如果同一主体实施了“先买后卖”的行为，且买卖的数据交易上下文信息相似度很高，则其涉嫌违规转售的可能性进一步上升。

去除“基于交易因果顺序约束的衰减时序门控模块”中的交易时间差衰减部分后，CCA-RGCN的F1-macro、ROC-AUC和AP分别降为0.76、0.96、0.86；这说明在学习违规转售行为时，交易时间间隔权重产生了一定的作用。在数据交易场景下，“先买后卖”两笔交易之间的时间间隔越短，违规转售的可能性越大。

消融实验结果表明，CCA-RGCN对于数据违规转售的检测性能依赖于多个模块和部分的协同作用，交易因果顺序约束和交易上下文信息相似度对于违规转售行为特征的学习建模缺一不可。

5　结束语

针对数据交易场景中的数据违规转售问题，设计了一种增强关系图卷积网络CCA-RGCN，并提出了基于CCA-RGCN的违规转售检测方法。CCA-RGCN主要由基于交易上下文信息相似度的注意力加权模块和基于交易因果顺序约束的衰减时序门控模块构成。违规转售检测方法以数据交易记录为输入，构建以交易为节点的双关系交易拓扑，经过多层堆叠的CCA-RGCN处理，可以高效识别具有违规转售行为的交易节点。在实验环节，构建了带有违规转售样本的模拟数据交易数据集DTRD。基于DTRD的实验结果表明，本文方法在整体检测性能上均优于现有方法；在公开数据集YelpChi上的实验结果证明，CCA-RGCN具备一定的跨数据集适应能力。

参考文献：

[1] GB/T 37932-2025 数据安全技术 数据交易服务安全要求[S].
GB/T 37932-2025 Data security technology — security requirements for data transaction service[S].

[2] 李凤华, 李晖, 牛犇, 等. 数据要素流通与安全的研究范畴与未来发展趋势[J]. 通信学报, 2024, 45(5): 1-11.
Li F H, Li H, Niu B, et al. Research category and future development trend of data elements circulation and security[J]. Journal on Communications, 2024, 45(5): 1-11.

[3] 国家数据局. 2024年全国数据市场交易规模超1600亿元[EB]. (2025-04-02)[2026-01-29].
National Data Administration. China’s 2024 Data market transaction volume exceeded 160 billion yuan[EB]. (2025-04-02)[2026-01-29].

[4] 华为技术有限公司. 智能世界2035[EB]. (2025-09)[2026-01-29].
Huawei. Intelligent world 2035[EB]. (2025-09)[2026-01-29].

[5] Zhang J Y, Bi Y R, Cheng M Y, et al. A survey on data markets[PP]. V1. (2024-11-09)[2026-02-02]. arXiv: arXiv.2411.07267.

[6] 王利明. 数据何以确权[J]. 法学研究, 2023, 45(4): 56-73.
Wang L M. Confirmation of data rights: why and how[J]. Chinese Journal of Law, 2023, 45(4): 56-73.

[7] Jung T, Li X Y, Huang W C, et al. AccountTrade: accountability against dishonest big data buyers and sellers[J]. IEEE Transactions on Information Forensics and Security, 2019, 14(1): 223-234.

[8] Gupta P, Dedeoglu V, Kanhere S S, et al. TrailChain: Traceability of data ownership across blockchain-enabled multiple marketplaces[J]. Journal of Network and Computer Applications, 2022, 203: 103389.

[9] Sahoo S, Halder R. Traceability and ownership claim of data on big data marketplace using blockchain technology[J]. Journal of Information and Telecommunication, 2021, 5(1): 35-61.

[10] Dai W Q, Dai C K, Choo K R, et al. SDTE: a secure blockchain-based data trading ecosystem[J]. IEEE Transactions on Information Forensics and Security, 2020, 15: 725-737.

[11] Chen F, Wang J H, Jiang C K, et al. Blockchain based non-repudiable IoT data trading: simpler, faster, and cheaper[C]//Proceedings of the IEEE INFOCOM 2022 - IEEE Conference on Computer Communications. Piscataway: IEEE Press, 2022: 1958-1967.

[12] Motie S, Raahemi B. Financial fraud detection using graph neural networks: a systematic review[J]. Expert Systems with Applications, 2024, 240: 122156.

[13] Dou Y T, Liu Z W, Sun L, et al. Enhancing graph neural network-based fraud detectors against camouflaged fraudsters[C]//Proceedings of the 29th ACM International Conference on Information & Knowledge Management. New York: ACM Press, 2020: 315-324.

[14] Liu Y, Ao X, Qin Z D, et al. Pick and choose: a GNN-based imbalanced learning approach for fraud detection[C]//Proceedings of the Web Conference 2021. New York: ACM Press, 2021: 3168-3177.

[15] Xiang S, Zhu M Z, Cheng D W, et al. Semi-supervised credit card fraud detection via attribute-driven graph representation[J]. Proceed-

ings of the AAAI Conference on Artificial Intelligence, 2023, 37(12): 14557-14565.

[16] Kipf T N, Welling M. Semi-supervised classification with graph convolutional networks[PP]. V4. (2017-02-22) [2026 – 02 – 02]. arXiv: arXiv.1609.02907.

[17] Weng Y P, Chen X, Chen L, et al. GAIN: graph attention & interaction network for inductive semi-supervised learning over large-scale graphs[J]. IEEE Transactions on Knowledge and Data Engineering, 2022, 34(9): 4257-4269.

[18] Wang H W, Hooi B, He D, et al. EGNN-ad: an effective graph neural network-based approach forAnomaly detection onEdge-Attributed graphs[C]//Database Systems for Advanced Applications. Berlin: Springer, 2024: 321-331.

[19] Song C C, Lin X X, Shen H Y, et al. UniFORM: towards unified framework for anomaly detection on graphs[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2025, 39(12): 12559-12567.

[20] Duan X, Yan B, Dong A, et al. Phishing frauds detection based on graph neural network on Ethereum[C]//Proceedings of the 17th International Conference on Wireless Algorithms, Systems, and Applications. Berlin: Springer, 2022: 351-363.

[21] Hu X X, Chen H T, Zhang J J, et al. GAT-COBO: cost-sensitive graph neural network for telecom fraud detection[J]. IEEE Transactions on Big Data, 2024, 10(4): 528-542.

[22] Schlichtkrull M, Kipf T N, Bloem P, et al. Modeling relational data with graph convolutional networks[C]//The Semantic Web. Berlin: Springer, 2018: 593-607.

[23] Mitra R, Dongre A, Dangare P, et al. Knowledge graph driven credit risk assessment for micro, small and medium-sized enterprises[J]. International Journal of Production Research, 2024, 62(12): 4273-4289.

[24] Jiang C X, Zheng J Y, Xu Q F. Mapping creditworthiness for Chinese small and medium-sized enterprises: integrating knowledge graphs and graph neural networks[J]. Applied Intelligence, 2026, 56(1): 46.

[25] Li W, Bao R H, Harimoto K, et al. Modeling the stock relation with graph network for overnight stock movement prediction[C]//Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence. San Francisco: Margan Kaufmann, 2020: 4541-4547.

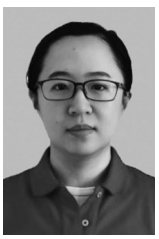
[26] Sionek A. Brazilian e-commerce public dataset by Olist[EB]. (2018) [2026-01-29].

[27] Rayana S, Akoglu L. Collective opinion spam detection: bridging review networks and metadata[C]//Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM Press, 2015: 985-994.

[作者简介]

王宇翔（1991– ），女，山西吕梁人，博士，中国科学院信息工程研究所工程师，主要研究方向为数据安全。

张玲翠（1986– ），女，河北故城人，博士，中国科学院信息工程研究所高级工程师，主要研究方向为网络与系统安全、数据安全。



侯雨桥（1989– ），女，北京人，博士，中国科学院信息工程研究所工程师，主要研究方向为数据安全。

杨倩（1989– ），女，山东聊城人，博士，中国科学院信息工程研究所工程师，主要研究方向为数据安全。

牛犇（1984– ），男，陕西西安人，博士，中国科学院信息工程研究所研究员、博士生导师，主要研究方向为隐私计算、数据安全。