

时间感知图注意力网络在加密货币交易欺诈检测中的应用

1st Zhi ZHENG

School of Finance and Business

Shanghai Normal University

Shanghai, China

Email:hhmcn@outlook.com

2nd Bochuan ZHOU

School of Finance and Business

Shanghai Normal University

Shanghai, China

Email:1000553566@smail.shnu.edu.cn

3rd Yuping SONG

School of Finance and Business

Shanghai Normal University

Shanghai, China

Email:songyuping@shnu.edu.cn

摘要—数字货币交易欺诈检测面临着日益复杂的交易模式和严重的类别不平衡的双重挑战。传统方法依赖于手工特征工程,并且难以捕捉交易网络中的时间性和结构性依赖关系。本文提出了一种增强的时间感知图注意力网络(ATGAT),通过三个模块提升检测性能:(1)设计了一个先进的时间嵌入模块,将多尺度时间差特征与周期位置编码融合;(2)构建了时间感知的三重注意力机制,联合优化结构、时间和全局上下文注意力;(3)使用加权 BCE 损失来解决类别不平衡问题。在 Elliptic++ 数字货币数据集上的实验表明,ATGAT 实现了 0.9130 的 AUC 值,相较于最佳的传统方法 XGBoost 提升了 9.2%,相较于 GCN 提升了 12.0%,相较于标准 GAT 提升了 10.0%。该方法不仅验证了时间感知和三重注意力机制对图神经网络增强的效果,还为金融机构提供了更可靠的欺诈检测工具,并且其设计原则可以泛化到其他时间图异常检测任务。

Index Terms—金融安全,图注意力网络,欺诈检测,不平衡数据,注意力机制。

I. 介绍

在数字金融时代,交易欺诈已成为对全球金融系统安全构成的重大威胁。根据 Nilson Report [1] 的报告,2023 年全球信用卡诈骗损失达到了 338 亿美元。加密货币领域面临的挑战更为严峻——Chainalysis [2] 报告称,2024 年有 409 亿美元的加密货币流向了非法地址,洗钱、庞氏骗局和勒索软件占到了 70% 以上。这些欺诈行为不仅造成了巨大的经济损失,也严重损害了公众对数字金融系统的信任。

早期的欺诈检测研究主要依赖于基于规则的系统 and 传统的机器学习方法。Leonard 等人提出了一个用于信用卡欺诈预警的基于规则的专家系统 [3]。随后,

研究人员将机器学习引入这一领域,包括支持向量机 (SVM) [4],多层感知器 (MLP) [5] 和生成对抗网络 (GAN) [6]。然而,在处理金融交易数据时,这些方法表现出几个局限性:首先,特征提取严重依赖领域专家的知识,阻碍了自动发现数据中的潜在模式;其次,将交易视为独立样本忽视了跨交易的关系;第三,在类别不平衡条件下性能显著下降,而在欺诈检测任务中严重的正负样本比例差异普遍存在 [7]。

近期,图神经网络 (GNNs) 通过将交易建模为图形结构,提供了用于交易欺诈检测的新技术途径,能够同时利用节点特征和网络拓扑信息 [8], [9]。秦等人 [10] 首次将 GNN 应用于银行交易欺诈检测;段等人 [11] 提出了结合因果不变学习的 CaT-GNN;李等人 [12] 设计了异构图自动编码器以增强少数类别的识别。尽管有这些进展,但仍存在一些研究空白:(1) 现有方法未能充分利用交易的时间信息,而时间信息在欺诈模式识别中起着关键作用 [13], [14];(2) 注意力机制的设计仍然相对简单,难以全面考虑局部和全局信息;(3) 缺乏系统解决极端类别不平衡的方案。

基于此分析,本文提出了用于加密货币交易欺诈检测的增强时间感知图注意力网络 (ATGAT)。主要贡献包括:(1) 设计了一个结合多尺度时间差特征和周期编码的时间嵌入模块;(2) 提出了一种从结构、时间和全局维度学习节点表示的时间感知三重注意力机制;(3) 使用加权交叉熵损失函数来缓解类别不平衡问题。在 Elliptic++ 数据集 [15] 上的实验结果表明,所提出的方法在主要评估指标上优于现有方法。

II. 方法

A. 问题形式化

令 $G = (V, E, X, Y, T)$ 表示一个有向加密货币交易图，其中：

- $V = \{v_i\}_{i=1}^N$ 表示交易节点集，每个节点对应一个比特币交易；
- $E \subseteq V \times V$ 表示有向边集，其中若交易 v_i 向 v_j 发送比特币，则存在边 $(i \rightarrow j)$ 。
- $X \in \mathbb{R}^{N \times d}$ 是节点特征矩阵，其中每个节点具有 d 维原始特征（包括交易金额、时间步信息、统计特征等）。
- $Y \in \{0, 1\}^N$ 表示节点标签，其中 0 表示合法交易，1 表示欺诈交易。训练集仅包含部分节点的标签，其余节点需要进行预测；
- $T = \{t_i\}_{i=1}^N$ 是节点时间戳序列，其中 t_i 表示第 i 笔交易广播到网络的时间（按离散时间步骤索引，例如从 1 到 49）。

我们定义监督节点分类任务：给定一个带有部分已知标签的交易图 (V, E, X, T) ，包括标记节点集 $V_{\text{labeled}} \subset V$ 分为训练集 V_{train} 和验证集 V_{val} ，学习一个映射函数

$$f_\theta : (V, E, X, T) \rightarrow \hat{Y}, \quad (1)$$

以便可以为验证集节点生成准确的预测 $\hat{y}_i \in \{0, 1\}$ 。

B. 模型架构

针对加密货币交易图的时间特性和类别不平衡问题，本文提出了 ATGAT 架构（图 1）。该架构通过三个核心创新模块实现了协同作用：高级时间嵌入捕捉多尺度时间模式，时间感知三重注意力机制融合来自不同维度的信息，加权损失函数处理数据不平衡。

C. 核心模块

1) 高级时间嵌入：为了全面捕捉加密货币交易中的多层次时间关系，本文设计了一个包含三个组件的时间嵌入模块：

a) 基本时间差投影：对于具有时间差 $\Delta t_{ij} = |t_i - t_j|$ 的节点对 (i, j) ，我们通过线性层映射到 d_1 维：

$$\mathbf{f}_\Delta^{(ij)} = \mathbf{W}_\Delta [\Delta t_{ij}] + \mathbf{b}_\Delta \in \mathbb{R}^{d_1}, \quad (2)$$

其中 $\mathbf{W}_\Delta \in \mathbb{R}^{d_1 \times 1}$ 、 $d_1 = \lfloor d_t/4 \rfloor$ 和 d_t 是最终的时间嵌入维度。

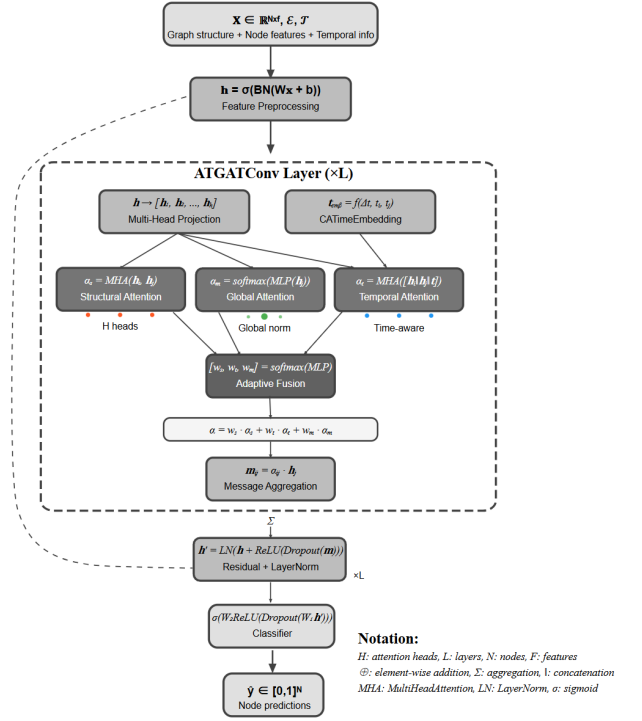


图 1. ATGAT 的整体架构。

b) 多尺度时间特征提取：我们提取了三个非线性变换特征的尺度： $f_{\text{short}}(\Delta t_{ij}) = \Delta t_{ij}$ 、 $f_{\text{medium}}(\Delta t_{ij}) = \log(\Delta t_{ij} + 1)$ 、 $f_{\text{long}}(\Delta t_{ij}) = \sqrt{\Delta t_{ij} + 1}$ 。拼接后，这些特征通过线性层映射到 $d_3 = \lfloor d_t/4 \rfloor$ 维：

$$\mathbf{f}_{\text{ms}}^{(ij)} = \mathbf{W}_{\text{ms}} [f_{\text{short}}, f_{\text{medium}}, f_{\text{long}}]^T + \mathbf{b}_{\text{ms}} \in \mathbb{R}^{d_3}. \quad (3)$$

c) 周期位置编码：受 Transformer [16] 的启发，我们为每个节点的时间戳 t_i 构建了 d_{pos} 维的正弦位置编码：

$$\begin{aligned} \text{PE}_{(t_i, 2k)} &= \sin\left(\frac{t_i}{10000^{2k/d_{\text{pos}}}}\right), \\ \text{PE}_{(t_i, 2k+1)} &= \cos\left(\frac{t_i}{10000^{2k/d_{\text{pos}}}}\right). \end{aligned} \quad (4)$$

对于节点对 (i, j) ，我们将两种编码连接起来，并将维度减少到 $d_2 = \lfloor d_t/2 \rfloor$ 维。

在连接上述特征后，我们应用线性变换、LayerNorm、ReLU 和 Dropout 进行融合，得到最终的时间嵌入 $\mathbf{e}_{\text{time}}^{(ij)} \in \mathbb{R}^{d_t}$ 。

2) 时间感知三重注意力：为了全面建模拓扑结构、时间依赖性和全局上下文，本文在每个 GAT 层中并行计算三种类型的注意力：

- **结构注意力** $\alpha_{ij}^{\text{struct}}$ ：基于节点特征的多头点积注意力

- **时间注意力** $\alpha_{ij}^{\text{temp}}$: 将时间嵌入 $e_{\text{time}}^{(ij)}$ 融入键/值以捕捉时间相关性
- **全局上下文注意力** $\alpha_{ij}^{\text{global}}$: 所有边特征的全局标准化

一个自适应融合网络动态计算权重 $[w_s, w_t, w_g]$ 以获得最终的注意力:

$$\alpha_{ij} = w_s \alpha_{ij}^{\text{struct}} + w_t \alpha_{ij}^{\text{temp}} + w_g \alpha_{ij}^{\text{global}}. \quad (5)$$

最后, 执行多头聚合和消息传递以更新节点表示。

3) 成本敏感损失优化: 在交易图中, 合法交易约占 98%, 而欺诈交易仅占约 2%。直接应用标准二元交叉熵 (BCE) 会使模型偏向预测合法类别。因此, 本文采用加权 BCE 损失函数:

具有正样本权重 $w_{\text{pos}} = N_{\text{neg}}/N_{\text{pos}}$ 和负样本权重 $w_{\text{neg}} = 1$, 我们有:

$$\mathcal{L}_{\text{WBCE}} = -[w_{\text{pos}} y \log(\hat{y}) + w_{\text{neg}} (1-y) \log(1-\hat{y})]. \quad (6)$$

III. 实验

A. 数据集和实验设置

本研究利用了来自 Elliptic++ 的交易子图 [15], 将比特币交易映射到合法类别 (交易所、钱包提供商、合法服务) 中的真实实体与非法类别 (诈骗、恶意软件、恐怖组织、勒索软件) 中进行对比。数据集包含 203,769 个节点和 234,355 条边, 其中 2% (4,545 个) 节点被标记为非法, 21% (42,019 个) 被标记为合法, 剩下 77% 没有标签。

每个节点包含 183 维特征: (1) 94 个局部特征描述交易属性 (输入/输出数量、费用); (2) 72 个从一跳邻居统计聚合的特征 (最大值、最小值、标准差、相关性); (3) 17 个由 Elliptic++ 添加的增强特征。

交易跨越了 49 个离散时间步骤, 间隔为 $\approx 2\text{-week}$, 涵盖了将近两年的时间。

考虑实验可控性和计算约束, 我们采用仅使用 46,564 个标记节点的监督学习框架。这些数据被划分为 8:1:1 的比例用于训练、验证和保留集, 其中保留部分在我们的实验中预留但未使用。模型使用 AdamW 优化器 ($\text{lr}=0.005$) 并采用余弦退火训练了 170 个周期。

B. 基线方法

为了全面评估所提出方法的性能, 我们选择了以下两类基准方法进行比较:

传统机器学习方法: (1) 逻辑回归 (LR): 线性分类器的代表; (2) 决策树 (DT): 能够捕捉非线性决策边界; (3) 随机森林 (RF): 集成学习方法的典型代表; (4) XGBoost (XGB): 在众多竞赛中表现出色的梯度提升方法。这些方法直接使用节点特征进行分类, 而不考虑图结构信息。

GNN 方法: (1) 使用二元交叉熵损失的标准图卷积网络 (GCN); (2) 带有加权损失函数以处理类别不平衡的 GCN (GCN-W)。这些方法可以利用图形结构信息, 但缺乏时间建模能力。

C. 评估指标

考虑欺诈检测任务的特殊性和数据集不平衡的严重性, 本文采用以下评估指标: AUC (ROC 曲线下的面积) 作为主要指标, 因为它不受类别分布的影响, 并且在不同的阈值下全面评估模型性能; Accuracy、Precision、Recall 和 F1-Macro 作为辅助指标, 提供多维度性能评估。

D. 基准实验

表 I 展示了各种方法在 Elliptic++ 数据集上的综合性能比较。实验结果表明, 所提出的 ATGAT 方法在多个指标上实现了最优或接近最优的性能。

表 I
椭圆++ 数据集上的性能比较

Model	Accuracy	Precision	Recall	F1-Macro	AUC
LR	0.9195	0.3586	0.6556	0.7101	0.7949
DT	0.8872	0.2875	0.7612	0.6774	0.8279
RF	0.9810	0.9708	0.6622	0.8887	0.8305
XGB	0.9803	0.9358	0.6755	0.8872	0.8364
GCN	0.9153	0.3421	0.6456	0.7007	0.8144
GCN-W	0.8997	0.3036	0.6887	0.6832	0.8683
ATGAT-W	0.9715	0.7924	0.6278	0.8427	0.9130

从实验结果中, 我们观察到以下现象: (1) 在传统机器学习方法中, XGBoost 的表现最佳, AUC 值为 0.8364, 但仍显著低于图神经网络方法; (2) 标准 GCN 达到的 AUC 值为 0.8144, 在使用加权损失后提升至 0.8683, 验证了处理类别不平衡的必要性; (3) ATGAT 达到了 0.9130 的 AUC 值, 比最好的传统方法 XGBoost 提高了 9.2%, 比标准 GCN 提高了 12.1%, 比加权 GCN 提高了 5.1%。

值得注意的是，尽管 RandomForest 在精度 (0.9708) 和 F1-Macro (0.8887) 方面表现出色，但其 AUC 仅为 0.8305，这表明它在固定阈值下表现良好，但在不同阈值下的综合性能不如 ATGAT。这一现象进一步验证了将 AUC 作为主要评估指标的合理性。

E. 消融研究

为了全面了解每个模型组件的贡献，我们设计了系统的消融实验。通过逐步移除或替换关键模块，我们定量评估每个组件的重要性。所有消融实验都在相同的设置下使用不同的随机种子重复 10 次以确保结果稳定性。

消融研究包括五个模型变体：(1) 基线 GAT(B-GAT)：仅标准图注意力网络，不包含时间信息或增强模块；(2) 结构 GAT(S-GAT)：仅保留结构注意力机制；(3) 时间 GAT(T-GAT)：仅保留时间注意力机制；(4) 使用 BCELoss 的 ATGAT(ATGAT)：完整的三重注意力网络；(5) 使用加权 BCELoss 的 ATGAT(ATGAT-W)：带有额外类别权重策略的 ATGAT。

表 II
消融研究结果 (均值 $\pm$ 标准差)

Model Variant	Accuracy	Precision	Recall	F1-Macro	AUC
B-GAT	0.9110 $\pm$ 0.0313	0.2984 $\pm$ 0.1177	0.5844 $\pm$ 0.2100	0.6712 $\pm$ 0.0722	0.8295 $\pm$ 0.0260
S-GAT	0.9714 $\pm$ 0.0015	0.7982$\pm$0.0268	0.6166 $\pm$ 0.0044	0.8403 $\pm$ 0.0063	0.8993 $\pm$ 0.0046
T-GAT	0.9713 $\pm$ 0.0011	0.7948 $\pm$ 0.0208	0.6189 $\pm$ 0.0076	0.8403 $\pm$ 0.0048	0.8989 $\pm$ 0.0078
ATGAT	0.9709 $\pm$ 0.0011	0.7806 $\pm$ 0.0201	0.6298$\pm$0.0078	0.8409 $\pm$ 0.0049	0.9111 $\pm$ 0.0043
ATGAT-W	0.9715$\pm$0.0012	0.7924 $\pm$ 0.0235	0.6278 $\pm$ 0.0087	0.8427$\pm$0.0046	0.9130$\pm$0.0055

表 II 展示了关键发现：基线 GAT 表现不佳 (AUC=0.8295 $\pm$ 0.0260)，方差较高，表明标准 GAT 不适合处理不平衡任务。单一注意力机制 (结构/时间 GAT) 将 AUC 提高到约 0.899，验证了它们的有效性。完整的 ATGAT 达到了 AUC=0.9111 $\pm$ 0.0043，并具有最高的召回率 (0.6298 $\pm$ 0.0078)，确认了三重注意力的协同作用。ATGAT-W 通过类别加权进一步提高到 AUC=0.9130 $\pm$ 0.0055，验证了成本敏感策略的有效性。

F. 效应量分析

为了全面评估模型改进的实际意义，我们计算了 Cohen's d 效应量来量化不同模型变体之间的性能差异。根据科恩的标准， $|d| < 0.2$ 表示效果可忽略；0.2-0.5 为小效应；0.5-0.8 为中等效应； > 0.8 为大效应。

为了量化模型改进的实际意义，计算了关键比较的 Cohen's d 效应量。AUC 的结果为 Baseline GAT 与 ATGAT-Weighted 之间的 4.442，Structural GAT 与

表 III
COHEN'S D 效应量用于 AUC 差异

	B-GAT	S-GAT	T-GAT	ATGAT	ATGAT-W
B-GAT	0.000	-3.734	-3.616	-4.378	-4.442
S-GAT	3.734	0.000	0.048	-2.653	-2.704
T-GAT	3.616	-0.048	0.000	-1.929	-2.081
ATGAT	4.378	2.653	1.929	0.000	-0.381
ATGAT-W	4.442	2.704	2.081	0.381	0.000

ATGAT 之间的 2.653，以及 0.381。这些结果表明，从 Baseline GAT 到完整模型的改进，以及从单个到三重注意力机制的改进具有极高的实际意义，而类别加权策略尽管带来了相对较小的提升，但仍具有实际应用价值。

IV. 结论

本文提出了一个增强的时间感知图注意力网络 (ATGAT)，用于加密货币交易欺诈检测。该方法通过三个关键模块解决了现有方法的局限性：首先，设计的高级时间嵌入模块通过融合多尺度时间差特征与周期位置编码，有效地捕捉了交易中的动态时间模式；然后，三重注意力机制从结构、时间和全局维度学习节点表示，显著增强了模型的表现力；最后，加权损失策略系统地解决了欺诈检测中极端类别不平衡的问题。

在 Elliptic++ 加密货币数据集上的实验评估完全验证了所提出方法的有效性。ATGAT 实现了 0.9130 的 AUC，相比最佳的传统方法 XGBoost 提升了 9.2%，相比标准图神经网络 GCN 提升了 12.1%。系统性的消融实验表明，时间嵌入、三重注意力机制和加权策略都对性能提升做出了显著贡献。值得注意的是，Cohen's d 效应量分析显示模型改进具有极高的统计学意义和实际应用价值。

尽管这项研究取得了成就，但仍存在一些限制和未来需要调查的领域。首先，受计算资源的限制，本研究采用基于经验的固定超参数设置，没有进行系统性的超参数搜索，并使用了训练-验证两个阶段的评估方法，这种方法可能具有进一步优化的潜力。其次，所使用的数据集主要集中在比特币交易领域，数据多样性有限，这可能会对模型在其他金融交易场景中的泛化能力产生影响。此外，虽然 ATGAT 在 AUC 性能上表现出色，但在一些基于阈值的指标上并未达到最先进的结果，可以通过调整阈值来进一步优化，以充分发挥该模型的潜

力。通过这些努力，我们将进一步推进金融欺诈检测技术的发展，并为金融安全和风险控制提供更有效的技术支持。

参考文献

- [1] T. N. Report, “Card fraud losses worldwide in 2023,” Nilson Report, Tech. Rep., January 2024.
- [2] Chainalysis, “The chainalysis 2025 crypto crime report,” Chainalysis Inc., Tech. Rep., 2025.
- [3] K. J. Leonard, “The development of a rule based expert system model for fraud alert in consumer credit,” *European Journal of Operational Research*, vol. 80, no. 2, pp. 350–356, 1995.
- [4] Y. Xu, “Credit card fraud detection based on support vector machines(in chinese),” *Computer Simulation*, vol. 28, no. 08, pp. 376–379+384, 2011.
- [5] A. H. Elmi, S. bt. Ibrahim, and R. Sallehuddin, “Detecting sim box fraud using neural network,” in *International Conference on IT Convergence and Security*, 2013.
- [6] C. Jian, “Credit card fraud detection based on self-encoder and adversarial generative networks(in chinese),” Master’s thesis, Shanghai Jiao Tong University, 2019.
- [7] H. He and E. A. Garcia, “Learning from imbalanced data,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 9, pp. 1263–1284, 2009.
- [8] T. N. Kipf and M. Welling, “Semi-supervised classification with graph convolutional networks,” *CoRR*, vol. abs/1609.02907, 2016.
- [9] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, “Graph attention networks,” 2018.
- [10] Z. Qin, Y. Zhou, and Z. Li, “Graph neural network-based bank transaction fraud detection method(in chinese),” *Computer Science*, vol. 51, no. S2, pp. 921–928, 2024.
- [11] Y. Duan, G. Zhang, S. Wang, X. Peng, W. Ziqi, J. Mao, H. Wu, X. Jiang, and K. Wang, “Cat-gnn: Enhancing credit card fraud detection via causal temporal graph neural networks,” 2024.
- [12] M. T. Singh, R. K. Prasad, G. R. Michael, N. K. Kaphungkui, and N. H. Singh, “Heterogeneous graph auto-encoder for creditcard fraud detection,” 2024.
- [13] E. Rossi, B. Chamberlain, F. Frasca, D. Eynard, F. Monti, and M. Bronstein, “Temporal graph networks for deep learning on dynamic graphs,” 2020.
- [14] D. Xu, C. Ruan, E. Korpeoglu, S. Kumar, and K. Achan, “Inductive representation learning on temporal graphs,” 2020.
- [15] Y. Elmougy and L. Liu, “Demystifying fraudulent transactions and illicit nodes in the bitcoin network for financial forensics,” *arXiv preprint arXiv:2306.06108*, 2023.
- [16] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, “Attention is all you need,” 2023.