

DOI:10.11918/202406018

融合时间约束游走的记忆增强时序图神经网络

金志刚¹, 张浩¹, 苏仁鋆², 赵晓芳³

(1. 天津大学 电气自动化与信息工程学院, 天津 300072; 2. 天津大学 党委宣传部, 天津 300072;
3. 天津城建大学 计算机与信息工程学院, 天津 300072)

摘要:为解决现有时序图表示学习方法难以充分挖掘网络结构特征和依赖关系的问题,提出了一种融合时间约束游走的记忆增强时序图神经网络(time-constrained walk fused memory enhanced temporal graph neural network, TWMTGN)。首先,从交互节点出发构造特定类型的时间约束游走序列,采用记忆模块捕获网络的长期依赖关系,并将游走序列特征融合至节点记忆状态,在事件发生时进行动态更新;其次,根据节点类型和时间间隔设计特征衰减层,建模短期依赖关系,提高模型对关键历史交互节点的识别能力;最后,将目标节点聚合后的历史交互特征输入因果卷积网络,进一步挖掘特征之间的潜在关联。在真实数据集上的实验结果表明:所提网络能够提升时间链路预测任务的效果,且复杂度较小;时间约束游走长度和次数等参数会影响模型的性能。研究提出的时间约束游走序列能够捕捉网络结构特征,节点记忆和特征衰减层有助于捕获网络依赖关系。

关键词: 时序网络; 游走; 记忆; 特征衰减; 因果卷积

中图分类号: TP183

文献标志码: A

文章编号: 0367-6234(2025)09-0029-10

Time-constrained walk fused memory enhanced temporal graph neural network

JIN Zhigang¹, ZHANG Hao¹, SU Renjun², ZHAO Xiaofang³

(1. School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China;

2. Publicity Department, Tianjin University, Tianjin 300072, China;

3. School of Computer and Information Engineering, Tianjin Chengjian University, Tianjin 300072, China)

Abstract: To address the challenges faced by existing temporal graph representation learning methods in fully exploiting structural features and dependencies of the network, a time-constrained walk fused memory enhanced temporal graph neural network (TWMTGN) is proposed. Firstly, a specific type of temporal constraint walk sequence is constructed for the interaction node, and the memory module is used to capture the long-term dependence of the network. The walk sequence features are fused into the memory of interaction node, which is dynamically updated upon the occurrence of events. Secondly, a feature attenuation layer is designed according to the node type and time interval to model short-term dependencies, which improves the model's ability to identify key historical interaction nodes. Finally, the aggregated historical interaction features of the target node are fed into a causal convolutional network to further uncover the potential association between them. Experimental results on real datasets show that the proposed network can improve the performance of temporal link prediction tasks while maintaining relatively low complexity. Parameters such as the length and frequency of time-constrained walks affect the model's performance. The proposed time-constrained walk sequences can effectively capture the structural features of the network, while the node memory and feature attenuation layers help to capture the network dependencies.

Keywords: temporal network; walk; memory; feature attenuation; causal convolution

随着图数据应用的日益广泛,如何对图数据进行建模,并将节点表示为下游任务的低维嵌入向量已成为研究人员关注的重要问题。图神经网络是一种图数据的深度学习方法^[1],是解决图表示学习问

题的有效方法。现实世界的图网络一般都是动态的,拓扑结构随着时间不断变化。对动态网络进行表示学习具有挑战性,但意义重大。动态网络描述了网络如何交互和演化,有助于理解和预测系统的

收稿日期: 2024-06-11; 录用日期: 2024-07-04; 网络首发日期: 2025-03-07

网络首发地址: <https://link.cnki.net/urlid/23.1235.T.20250307.1338.002>

基金项目: 国家自然科学基金(52171337)

作者简介: 金志刚(1972—),男,教授,博士生导师

通信作者: 金志刚, zgin@tju.edu.cn

行为^[2],相关研究成果已应用于欺诈检测^[3]、网络安全^[4]和推荐系统^[5]等领域。

动态图表示学习算法分为离散网络嵌入和连续网络嵌入两大类。离散网络嵌入一般将图分解为几个静态快照,利用静态方法对每个快照进行编码,但这种方法往往忽略了同一快照内的时间信息,无法捕捉复杂的时间动态。连续网络也被称为时序网络,可以表示为交互事件的有序序列,每个事件均与一个连续的时间相关联。时序网络通过在不同交互作用之间传递信息或利用连续时间函数生成嵌入表示来捕捉时间动态^[6],其动态性往往表现在微观和宏观两个层面^[7]。网络的微观变化是一种短期的变化规律,通常体现在节点的交互事件直接对其他节点产生的影响,其会随时间不断衰减;网络的宏观变化是一种长期、稳定的演变规律,表现在网络的拓扑结构遵循一定的变化模式。因此,时序图表示学习模型需要建模短期和长期依赖关系,并进一步挖掘网络演变规律。

在时序网络的相关研究中,Kumar 等^[8]利用循环神经网络(recurrent neural network,RNN)和依赖于时间的投影嵌入模块学习用户-项目交互网络的节点表示;Xu 等^[9]将自注意力应用于时序图的归纳表示学习,设计了一种时间编码器,将连续时间投影为时间向量;Rossi 等^[10]提出一种编码时序图的通用框架,将节点记忆引入时序图注意力网络的时间特征聚合阶段,允许模型学习数据的顺序性。然而,这些方法未明确利用图的拓扑结构信息。Luo 等^[11]设计了一种新的邻域结构特征表示方法,该方法从节点的联合领域中提取关键结构信息;Wang 等^[12]提出了一种因果关系匿名游走策略,该策略通过隐藏节点身份捕捉网络变化规律,采用 RNN 对游走序列进行编码;Zhou 等^[13]将元路径实例抽象为语义单元,考虑其相互作用来挖掘更深层次的语义信息,并利用霍克斯过程(Hawkes process)建模时序信息。但这些方法缺乏对长期依赖关系的捕获。Trivedi 等^[14]考虑网络拓扑结构和节点交互的变化进行网络的表示学习,在事件发生后更新节点表示,对于长期没有交互的节点存在嵌入过时问题。

鉴于此,本文提出了一种融合时间约束游走的记忆增强时序图神经网络(time-constrained walk fused memory enhanced temporal graph neural network,TWMTGN),以挖掘网络的结构特征和依赖关系,联合建模网络异质性和动态性,提高节点表示

性能。模型采用采样—更新—聚合机制,考虑图结构特征更新节点记忆状态,聚合历史交互节点特征,获得目标节点嵌入表示。具体来说,对于每一条新边,从两个交互节点出发,考虑时间约束对邻居节点进行采样,构造特定类型的游走序列,并对其进行编码,在全局视角下捕捉网络拓扑结构的演变模式,学习网络的长期变化规律;采用记忆模块保存节点历史状态,当事件发生时,更新节点记忆,捕捉网络长期依赖关系;依据节点类型和时间间隔设计特征衰减层,降低时间间隔较远节点的影响,建模网络短期依赖关系;利用因果卷积网络捕获历史交互节点时序特征的潜在关联,最终获得交互节点嵌入表示,应用于时间链路预测任务。

1 融合时间约束游走的记忆增强时序图神经网络

1.1 相关定义

1) 时序图。时序图可以表示为带有时间的边集合,记作 $G = \{(u, v, t) \mid u, v \in X, t \in T\}$,其中, u 与 v 表示交互节点, t 表示对应边的时间, X 为节点集合, T 为时间集合。每条边表示一个事件,具有对应的特征向量。

2) 元路径。时序图的丰富语义可以由元路径获取。元路径是通过节点类型和边类型定义的序列,可以表示一组连接多种类型节点的复合关系^[15]。元路径实例是按照元路径定义,从图中抽取的连续节点序列。

3) 时间链路预测。给定一个时序图 G ,利用 t 之前的历史信息 $\{(u', v', t')\} \in G$ 学习一种揭示网络随时间变化关系和模式的映射函数 $X \rightarrow \mathbf{R}^d, d \ll |X|$,将网络中的节点映射到低维向量空间,利用低维稠密向量预测两个节点在时间 t 是否会有新的边建立。

1.2 TWMTGN

TWMTGN 采用记忆模块存储节点状态,并将图的拓扑结构信息融合至节点记忆的更新过程,框架如图 1 所示。模型利用时间约束游走序列捕获网络的拓扑结构特征,将节点的记忆状态作为节点过去交互的压缩表征。为了更有效地聚合历史交互节点特征,基于节点类型和交互时间间隔设计特征衰减层,提取目标节点的关键历史交互特征,并利用因果卷积网络挖掘其潜在关联;最终形成节点的嵌入向量表示,应用于时间链路预测任务。

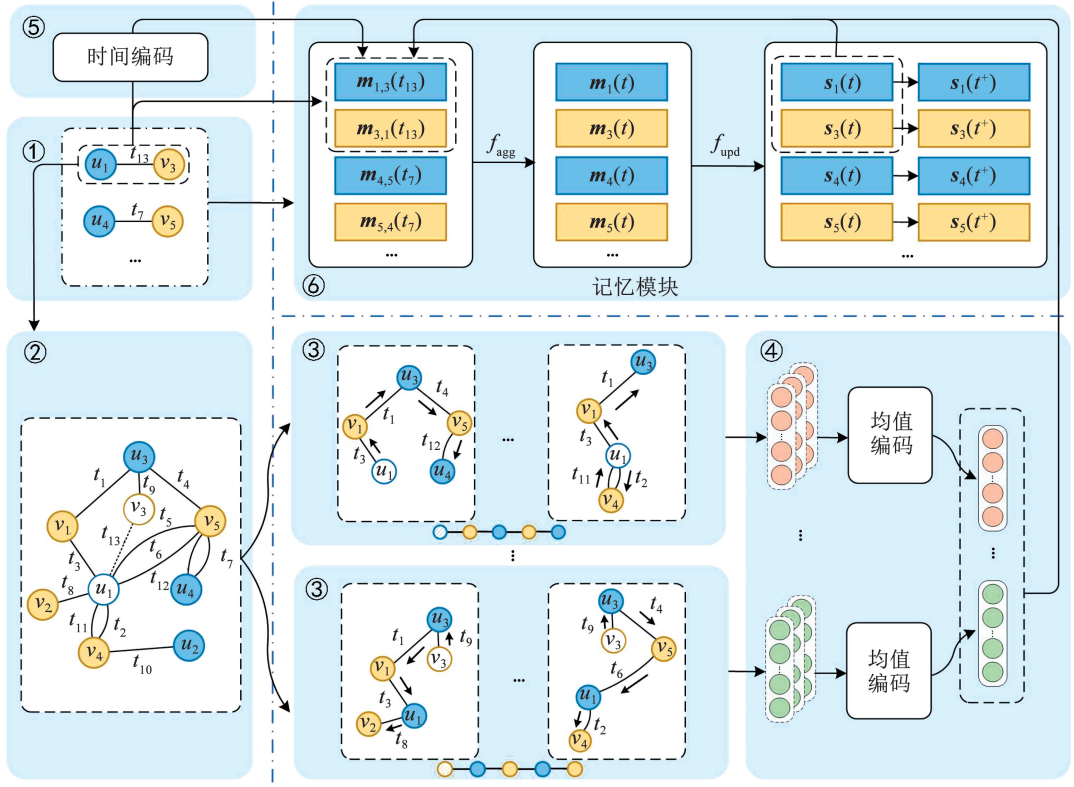


图 1 TWMTGN 框架

Fig. 1 Framework of TWMTGN

1.2.1 构造时间约束游走序列

时间约束游走序列可以确定每个交互节点的影响范围。图 1 中②、③分别给出了在边 (u_1, v_3, t_{13}) 出现时, 以不同类型节点为起点, 构造时间约束游走序列的过程。在时间 t_{13} 分别从交互节点 u_1 和 v_3 出发构造两个路径子集:

$$p_u^t = \{p_u^{t,1}, p_u^{t,2}, \dots, p_u^{t,k}\} \quad (1)$$

$$p_v^t = \{p_v^{t,1}, p_v^{t,2}, \dots, p_v^{t,k}\} \quad (2)$$

式中 k 为游走序列数量。子集中的游走序列长度均相同, 满足相应的元路径定义关系。

游走序列的本质是抽取网络子结构, 更全面地利用网络的拓扑结构信息。表 1 为在 LastFM 数据集上, 以用户 (user, U) 和物品 (item, I) 两种类型节点为起点, 构造不同长度游走序列的结果。

表 1 LastFM 不同游走长度结果

Tab. 1 Results of different walk lengths for LastFM

游走长度	起点为 U 的游走序列	起点为 I 的游走序列
2	U-I	I-U
3	U-I-U	I-U-I
4	U-I-U-I	I-U-I-U
5	U-I-U-I-U	I-U-I-U-I

1.2.2 记忆模块

节点记忆由一组向量组成, 是节点历史交互特

征的压缩表示。对于时间 t 的节点 i , 其记忆向量 $s_i(t)$ 在事件发生时利用 RNN 进行更新, 反映这一事件的影响, 使模型持续学习新出现的节点和边特征。该机制使得模型能够捕获网络长期依赖关系, 对于理解和预测时序图中的复杂行为至关重要^[10]。

1.2.3 消息函数

消息函数是记忆更新的主要机制。给定时间 t 节点 u 和 v 之间的交互 (u, v, t) , 消息函数计算出两条消息用于更新 u 和 v 的节点记忆。当事件发生时, 将事先构造的、分别以节点 u 和 v 为起点的时间约束游走序列输入一个可训练的嵌入层进行编码, 然后将获得的游走序列特征应用至消息函数中。

$$m_{u,v}(t) = f_{\text{msg}}(w_u(t), s_u(t), s_v(t), \phi(t-t_u), e_v^t) \quad (3)$$

$$m_{v,u}(t) = f_{\text{msg}}(w_v(t), s_v(t), s_u(t), \phi(t-t_v), e_u^t) \quad (4)$$

式中: $m_{u,v}(t)$ 、 $m_{v,u}(t)$ 分别为节点 u 、 v 的消息向量; $w_u(t)$ 、 $w_v(t)$ 为编码后的游走序列特征向量; $s_u(t)$ 、 $s_v(t)$ 分别为节点 u 、 v 的记忆向量; e_u^t 、 e_v^t 分别为节点 u 、 v 用于更新消息的边特征向量; t_u 、 t_v 分别为节点 u 、 v 上次更新记忆的时间; $\phi(\cdot)$ 为一种通用时间编码^[9], 用于捕捉时间依赖关系; $f_{\text{msg}}(\cdot)$ 为可学习的消息函数。为减少计算量, 仅将输入的多特征向量进行拼接。

1.2.4 消息聚合和记忆更新

为提高训练效率,采用批处理的方式训练模型。同一批次中涉及节点 i 的事件可能出现多次,每个涉及节点 i 的事件均会生成一条消息。因此,需要对消息进行聚合并更新节点记忆。

$$\mathbf{m}_i(t) = f_{\text{agg}}(\{\mathbf{m}_{i,p}(t_n) \mid (i,p,t_n) \in J(i,t), t_n \leq t\}) \quad (5)$$

$$\mathbf{s}_i(t^+) = f_{\text{upd}}(\mathbf{s}_i(t), \mathbf{m}_i(t)) \quad (6)$$

式中: $\mathbf{m}_i(t)$ 为聚合后的消息向量; $J(i,t)$ 为在时间 t 涉及节点 i 的事件集; $\mathbf{s}_i(t)$ 、 $\mathbf{s}_i(t^+)$ 分别为节点 i 更新前后的状态向量; $f_{\text{agg}}(\cdot)$ 为消息聚合函数,用于聚合时间 t 同一批次内涉及节点 i 的最新事件消息,生成节点新的记忆状态; $f_{\text{upd}}(\cdot)$ 为记忆更新函数,采用门控循环网络实现。

1.2.5 节点嵌入表示

由于节点记忆仅在涉及该节点的事件发生时才被更新,当长时间没有事件发生时(例如,社交平台用户在某段时间突然停止使用该平台),节点记忆存在过时问题^[10]。为解决该问题,在与目标节点时空邻近的节点上执行图聚合操作,计算任意时间目标节点的嵌入表示。即使一个节点在一段时间内处于非活动状态,但与其邻近的一些节点可能处于活动状态。因此,通过聚合时空邻近节点的特征,并结合时间衰减特性和因果卷积网络,计算目标节点的最新嵌入表示。

与节点 u 在时间 t 及之前产生交互的最近 N 个历史交互节点集可以表示为

$$N(u,t) = \{(v_j, \mathbf{e}_{v_j}^j, t_j) \mid j \in [1, N], t_j \leq t, (u, v_j, t_j) \in G\} \quad (7)$$

式中 $(v_j, \mathbf{e}_{v_j}^j, t_j) \in N(u,t)$ 。下文将 $(v_j, \mathbf{e}_{v_j}^j, t_j)$ 简写为 (v, \mathbf{e}, t') 。

TWMTGN 通过递归计算节点 u 在第 l 层的嵌入过程表示为:

$$\tilde{\mathbf{h}}_u^l(t) = f_{\text{AGG}}(\tilde{\mathbf{h}}_v^{(l-1)}(t), \phi(t-t'), \mathbf{e}) \quad (8)$$

$$\mathbf{h}_u^l(t) = f_{\text{UPD}}(\mathbf{h}_u^{(l-1)}(t), \tilde{\mathbf{h}}_u^l(t)) \quad (9)$$

式中: $\tilde{\mathbf{h}}_u^l(t)$ 、 $\tilde{\mathbf{h}}_v^{(l-1)}(t)$ 分别为第 l 层和第 $l-1$ 层历史交互节点聚合后的特征表示; $\mathbf{h}_u^l(t)$ 、 $\mathbf{h}_u^{(l-1)}(t)$ 分别为第 l 层和第 $l-1$ 层目标节点的嵌入表示; t' 为上次交互的时间; $\phi(t-t')$ 为编码后的时间; $f_{\text{UPD}}(\cdot)$ 为更新函数,采用多层感知机实现; $f_{\text{AGG}}(\cdot)$ 为聚合函数,聚合历史交互节点特征。目标节点的初始表示采用其记忆状态,即 $\mathbf{h}_u^0(t) = \mathbf{s}_u(t^+)$ 。

聚合函数 $f_{\text{AGG}}(\cdot)$ 采用多头注意力机制实现,表示为:

$$\tilde{\mathbf{h}}_u^l(t) = f_{\text{ATT}}(\mathbf{Q}^l(t), \mathbf{K}^l(t), \mathbf{V}^l(t)) \quad (10)$$

$$\mathbf{Q}^l(t) = \mathbf{h}_u^{(l-1)}(t) \parallel \phi(0) \quad (11)$$

$$\begin{aligned} \mathbf{K}^l(t) = \mathbf{V}^l(t) = & \\ & [\mathbf{h}_{v_1}^{(l-1)}(t) \parallel \mathbf{e}_{v_1}^{t_1} \parallel \phi(t-t_1), \dots, \\ & \mathbf{h}_{v_N}^{(l-1)}(t) \parallel \mathbf{e}_{v_N}^{t_N} \parallel \phi(t-t_N)] \end{aligned} \quad (12)$$

式中: $f_{\text{ATT}}(\cdot)$ 为多头注意力网络, $\{t_1, \dots, t_N\}$ 为每个历史交互节点对应的时间, $\{\mathbf{e}_{v_1}^{t_1}, \dots, \mathbf{e}_{v_N}^{t_N}\}$ 为对应的边特征, $\{\mathbf{h}_{v_1}^{(l-1)}(t), \dots, \mathbf{h}_{v_N}^{(l-1)}(t)\}$ 为截至时间 t 与节点 u 产生交互的最近 N 个历史交互节点的表示, $\mathbf{Q}^l(t)$ 为聚合了目标节点自身特征的向量, $\mathbf{K}^l(t)$ 和 $\mathbf{V}^l(t)$ 为聚合了目标节点历史交互特征的向量。

1.2.6 特征衰减和因果卷积

从时序图学习的网络嵌入不仅应该捕捉网络结构特征,而且要反映网络演化过程。为了模拟过去事件对现在产生的影响,更准确地建模网络演化过程,本文考虑节点类型和时间衰减特性设计了特征衰减层,以捕捉目标节点关键历史交互节点的特征。

$$\mathbf{h}_{v_n}^*(t) = \mathbf{h}_{v_n}(t) \cdot f_{\text{de}}(\sigma(\alpha_{\psi(n)}) \cdot (t-t_n)) \quad (13)$$

式中: $n \in \{1, \dots, N\}$; $f_{\text{de}}(\cdot)$ 为衰减函数,定义为 $f_{\text{de}}(x) = 1/\log(e+x)$; $\alpha_{\psi(n)}$ 为特定于节点类型的可训练参数; $\sigma(\cdot)$ 为激活函数, $\sigma = 1/(1+e^{-x})$ 是为了约束 $(t-t_n)$ 的系数在 $(0, 1)$ 内。更新后的 $\mathbf{K}^l(t)$ 和 $\mathbf{V}^l(t)$ 计算式为

$$\begin{aligned} \mathbf{K}^l(t) = \mathbf{V}^l(t) = & \\ & [\mathbf{h}_{v_1}^{*(l-1)}(t) \parallel \mathbf{e}_{v_1}^{t_1} \parallel \phi(t-t_1), \dots, \\ & \mathbf{h}_{v_N}^{*(l-1)}(t) \parallel \mathbf{e}_{v_N}^{t_N} \parallel \phi(t-t_N)] \end{aligned} \quad (14)$$

式中 $\{\mathbf{h}_{v_1}^{*(l-1)}(t), \dots, \mathbf{h}_{v_N}^{*(l-1)}(t)\}$ 为节点 u 最近 N 个历史交互节点经过特征衰减层后的表示。

因果卷积网络结合了因果卷积和膨胀卷积技术,能够有效捕捉时间序列数据中的因果关系和时间序列中的潜在依赖关系^[16]。在时序图中,节点之间的交互也可以被看作是一种时间序列数据,节点的特征可能随着时间的推移而变化。在进行多头注意力计算时,将因果卷积网络应用至 $\mathbf{K}^l(t)$ 和 $\mathbf{V}^l(t)$ 的计算中,得到因果卷积网络输出后的值:

$$\mathbf{K}^{*l}(t) = \mathbf{V}^{*l}(t) = f_{\text{causal}}(\mathbf{K}^l(t)) \quad (15)$$

式中 $f_{\text{causal}}(\cdot)$ 为因果卷积网络。

聚合历史交互节点特征的过程如图 2 所示。将历史交互节点经过特征衰减层和时间编码层后获得的特征表示与边特征进行拼接,输入因果卷积网络,获得聚合后的历史交互节点特征。

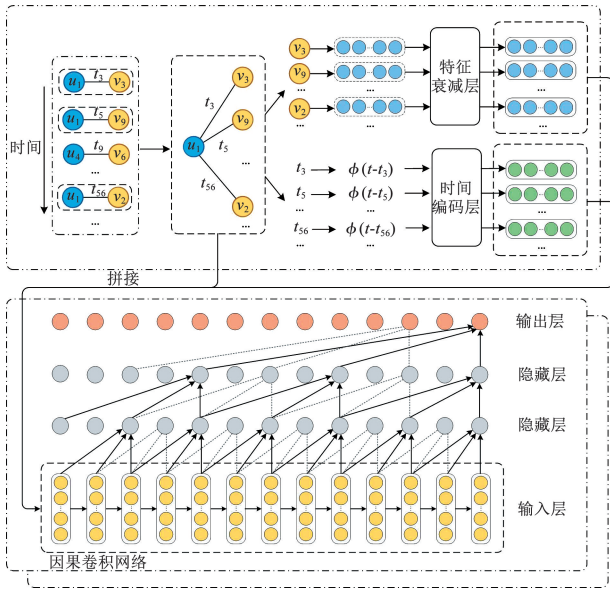


图2 聚合目标节点历史交互特征

Fig. 2 Aggregate the historical interaction features for target node

1.2.7 模型训练

对于时间链路预测任务,通过构造正负样本将其转化为二分类任务。将给定时序图中存在的边均视为正样本(即正边),不存在的边均视为负样本(即负边),利用负采样策略^[17]解决数据不平衡问题,并利用二元交叉熵损失(BCE-loss, L_{BCE})函数训练。 L_{BCE} 表达式为

$$L_{BCE} = \sum_{(u,v,t) \in G} -\log[\sigma(-\mathbf{h}_u^l(t) \mathbf{h}_v^l(t))] - FE_{v_q \sim P_n(v)}[\log \sigma(\mathbf{h}_u^l(t) \mathbf{h}_q^l(t))] \quad (16)$$

式中: F 为负样本数量, $P_n(v)$ 为节点空间中的负采样分布, $E_{v_q \sim P_n(v)}$ 为期望值。

TWMTGN 的训练过程见算法1。

算法1 TWMTGN 的训练过程:

input: 时序图 $G = \{(u, v, t)\}$; 边特征 $\{e\}$; 历史交互节点数量 N ; 网络层数 L

output: 正负样本得分 S_{pos}, S_{neg}

1. **for** (u, v, t) in G **do**
2. Generate p_u^t and p_v^t 为 u 和 v 构造时间约束游走序列
3. **end for**
4. **for** each batch (u, v, t) in G **do**
5. $\mathbf{m}_{u,v}(t) \leftarrow f_{msg}(p_u^t, s_u(t), t, t_u, e)$ 计算消息
6. $\mathbf{m}_u(t) \leftarrow f_{agg}(\mathbf{m}_{u,v}(t))$ 聚合同一节点消息
7. $s_u(t^+) \leftarrow f_{upd}(s_u(t), \mathbf{m}_u(t))$ 更新节点记忆
8. **for** $l = 1, \dots, L$ **do**
9. $\mathbf{h}_{v_n}^{*(l-1)}(t) \leftarrow \mathbf{h}_{v_n}^{(l-1)}(t), n \in \{1, \dots, N\}$ 特征衰减层
10. $\mathbf{K}^l(t), \mathbf{V}^l(t) \leftarrow f_{causal}(\mathbf{h}_{v_n}^{*(l-1)}(t), \mathbf{e}_{v_n}^l, t - t_n), n \in \{1, \dots, N\}$ 因果卷积层
11. $\bar{\mathbf{h}}_u^l(t) \leftarrow f_{ATT}(\mathbf{Q}^l(t), \mathbf{K}^l(t), \mathbf{V}^l(t))$ 聚合目标节点和历

史交互节点特征

12. $\mathbf{h}_u^l(t) \leftarrow f_{UPD}(\mathbf{h}_u^{(l-1)}(t), \bar{\mathbf{h}}_u^l(t))$ % 更新目标节点特征
13. **end for**
14. **end for**
15. Calculate L_{BCE} , **do** back-propagation and update parameters
16. **return** S_{pos}, S_{neg}

2 实验结果与分析

采用时间链路预测任务评估模型的有效性,具体采用直推式和归纳式两种实验设置全面评估模型在预测未知链路方面的能力。直推式设置即利用训练集中的数据预测那些已知的节点在未来的链接情况;归纳式设置即基于模型所学习到的网络特性,预测那些在训练集中未曾出现过的节点的未来链接。实验在 Intel (R) Core (TM) i5-12600KF CPU 和 NVIDIA GeForce RTX 3060 GPU 上运行。

2.1 评价指标和数据集

参考文献[6,18],实验采用受试者工作特征曲线下面积 A 和平均准确率 P 作为评价指标。 A 衡量模型在不同阈值下对正负样本的区分能力, P 衡量模型在不同召回率下的平均准确率。 A, P 值越大代表性能越好。

选取 LastFM、MOOC、UCI 和 Enron 4 个数据集进行实验,详细信息见表2。

1) LastFM。该数据集记录了用户听歌的历史数据信息,包含 1 000 名用户和 1 000 首最常听的歌曲,产生了 1 293 103 次互动。

2) MOOC。该数据集记录了用户点击在线课程或回答的数据,包含 7 047 名用户(user)和 97 个项目(item,包括 video, answer 等),共有 411 749 次互动。

3) UCI。该数据集记录了大学生在论坛上发表的帖子,包含 1 899 名用户,产生了 59 834 次互动。

4) Enron。该数据集记录了公司员工之间发送的电子邮件,包含 184 名用户,产生了 125 235 次互动。

表2 数据集统计

Tab. 2 Statistics of datasets

数据集	节点数量	时间边数量	节点特征维度	边特征维度	训练集-验证集-测试集/%
LastFM	1 980	1 293 103	172	172	70-15-15
MOOC	7 144	411 749	172	172	70-15-15
UCI	1 899	59 834	100	100	70-15-15
Enron	184	125 235	32	32	70-15-15

2.2 基线模型

为评估 TWMTGN 的有效性,采用时序图注意力网络(temporal graph attention network, TGAT)^[9]、时序图网络(temporal graph network, TGN)^[10]、邻居感知时序网络(neighborhood-aware temporal network, NAT)^[11]、基于因果匿名游走的网络(causal anonymous walk, CAW)^[12]、位置编码映射时序图网络(position-encoding injective temporal graph net, PINT)^[19] 5 种先进的时序图网络模型作为基线比较模型。其中, TGAT 和 TGN 通过聚合事件触发的时间子图建模节点动态; CAW 和 NAT 是构造邻域结构特征的模型; PINT 在 TGN 基础上进行优化,探索了时序图的表达能力。

1) TGAT 设计一种时间感知图注意力网络,用于时序图上的归纳表示学习,采用新的时间编码方法,结合自注意力机制处理连续时间,捕捉节点和拓扑结构的时序特征。

2) TGN 提出一种对以时间事件序列表示的动态图进行深度学习的通用框架,利用一种新的训练策略允许模型从数据的顺序性中学习,并保持高效的并行处理能力。

3) CAW 提出一种因果关系匿名游走策略,通过时间随机游走自动提取并表示时序网络中的动态模式,采用新颖的匿名化策略保持方法的归纳性,建立网络模式之间的相关性。

4) NAT 提出一种邻域感知的时序网络模型,采用基于字典类型的邻域表示,允许快速构建多个节点联合邻域的结构特征,并设计一种 N-缓存(N-cache)数据结构支持并行访问和更新邻域表示。

5) PINT 提出一种新的时序图网络架构,采用注入式时序信息聚合和更新函数捕捉时间优先级,引入相对位置特征增强模型对图结构的理解,在图的表达能力方面得到强化。

2.3 对比实验

TWMTGN 利用 PyTorch 框架实现,模型参数设置参考文献[10]。采用早停法以避免过拟合,并设置耐心值为 5,训练轮次为 50,批大小为 200,注意力头数为 2,学习率为 0.000 1,随机失活率为 0.1。为公平地进行比较,所有数据预处理后的输入格式一致,多次运行取平均值得到对比实验结果,见表 3。由表 3 可以看出, TWMTGN 在 4 个数据集上的指标优于绝大多数基线模型,说明 TWMTGN 在不同数据集上均有很好的泛化能力。

在捕获邻域特征方面, TWMTGN 在直推式设置下, A 和 P 在 4 个数据集上比 CAW 分别平均提高了

11.55% 和 9.39%, 比 NAT 分别平均提高了 4.38% 和 2.40%; 在归纳式设置下, A 和 P 在 4 个数据集上比 CAW 分别平均提高了 11.35% 和 9.48%, 比 NAT 分别平均提高了 4.84% 和 2.68%。虽然 CAW 和 NAT 分别通过因果匿名游走和获得一定跳数内的邻域表示捕获网络拓扑结构特征,但二者没有明确地区分挖掘出的语义类型,主要关注网络的局部结构,并且忽略了节点之间的长期依赖关系。因此,与 TWMTGN 相比, CAW 和 NAT 在时间链路预测任务上显现出不足。

在建模节点动态方面, TWMTGN 在直推式设置下, A 和 P 在 4 个数据集上比 TGN 分别平均提高了 6.64% 和 7.27%; 在归纳式设置下, A 和 P 在 4 个数据集上比 TGN 分别平均提高了 8.43% 和 8.44%。虽然 TGN 和 TWMTGN 均通过记录节点记忆建模节点动态,但 TGN 没有明确利用网络的拓扑结构特征,并且忽略了事件发生时间间隔的影响,也未进一步挖掘历史交互节点不同特征的内部关联。因此, TGN 与 TWMTGN 相比表现出劣势。

在 LastFM、Enron 等交互密集的时序图上,聚合事件触发的时间子图建模节点动态的方法不如构建邻域结构特征的方法,这是由于后者为每个节点的邻居保存一种单独的表示,旧的交互不会被新的交互压缩成一个单一的表示。TWMTGN 结合二者优势,进行时间约束游走构建邻域结构特征,结合记忆模块建模长期依赖关系,考虑事件影响随时间衰减的特性,挖掘历史交互特征的潜在关联,更好地捕捉时间动态。

TWMTGN 在不同类型数据集上的表现有所差异。不同数据集的实体和关系类型有所差异,其通常会以不同的模式影响当前的各种交互^[20]。因此,语义特征对不同数据集上时间链路预测任务的重要性有所差异。通常,异质图包含的语义信息更丰富, TWMTGN 能够有效应对网络的异质性,在同质图上也能取得良好的效果。

2.4 复杂度分析

为比较模型的复杂度,统计模型参数量的大小,在批大小为 200 的情况下,对 UCI 数据集上的训练、推理阶段的速度进行比较,结果见表 4。由表 4 可知,虽然 NAT 在训练和推理速度上具有优势,但其以牺牲精度为代价,且 NAT 参数量为 TWMTGN 的 8 倍左右,所需存储空间更大。TWMTGN 与 CAW、PINT 等模型相比具有显著优势,因为构造邻域结构特征或计算相对位置特征的开销较高。总体来看, TWMTGN 具有良好的预测性能和可接受的开销。

表 3 时间链路预测任务对比实验结果

Tab. 3 Comparative experimental results for temporal link prediction tasks

%

任务	模型	LastFM		MOOC		Enron		UCI	
		A	P	A	P	A	P	A	P
直推式	CAW	71.65	75.39	72.26	72.05	81.59	83.11	87.80	89.64
	NAT	81.98	85.01	87.77	86.82	81.38	83.56	90.85	92.73
	TGAT	51.08	53.61	68.40	67.35	54.51	60.45	68.46	65.03
	TGN	74.62	73.49	88.74	86.46	84.02	82.41	85.57	86.29
	PINT	67.87	70.32	88.96	87.08	55.58	57.88	73.68	77.87
	TWMTGN	84.54	85.07	91.71	90.25	89.52	87.74	93.72	94.67
归纳式	CAW	64.60	68.88	73.65	73.77	83.67	84.37	86.57	88.58
	NAT	83.95	86.61	84.18	83.43	79.09	82.49	87.30	90.28
	TGAT	51.90	53.95	66.80	65.99	53.85	59.07	61.43	60.12
	TGN	82.36	81.82	88.73	86.34	72.34	72.13	76.73	79.47
	PINT	31.65	41.40	90.15	88.58	30.35	41.73	63.86	69.94
	TWMTGN	88.64	89.27	91.35	89.91	83.96	82.98	89.92	91.37

表 4 复杂度对比

Tab. 4 Complexity comparison

模型	参数量/ 10^6	每轮训练时间/s	每批推理时间/ms
CAW	17.84	64.46	81.42
NAT	8.13	2.94	6.94
TGAT	12.64	8.14	14.95
TGN	0.60	3.09	13.76
PINT	0.49	13.25	98.50
TWMTGN	1.02	5.45	14.54

即去除时间约束的随机游走过程; 2) TWMTGN w/o decay, 即去除特征衰减层; 3) TWMTGN w/o causal, 即去除因果卷积网络。

图 3 为消融实验结果。与原始模型相比, 3 种变体的性能指标均有不同程度的下降, 在不同数据集上的表现也有所差异。在 LastFM 数据集上, 去除时间约束随机游走策略后模型效果下降最多, 原因是 LastFM 是异质数据集, 时间约束随机游走策略能有效提取不同类型的异质信息, 更好地利用网络的结构特征。在 Enron 数据集上, 去除特征衰减层后模型效果下降幅度较大, 这说明时间衰减特性对聚合历史交互节点特征具有重要影响。去除因果卷积层后, 实验指标在 4 个数据集上均有所下降, 证明了其对挖掘历史交互特征潜在时序关联的有效性。

2.5 消融实验

为进一步验证 TWMTGN 引入网络拓扑结构特征和设计历史交互节点特征聚合方法的有效性, 设置 3 种变体进行消融实验: 1) TWMTGN w/o walk,

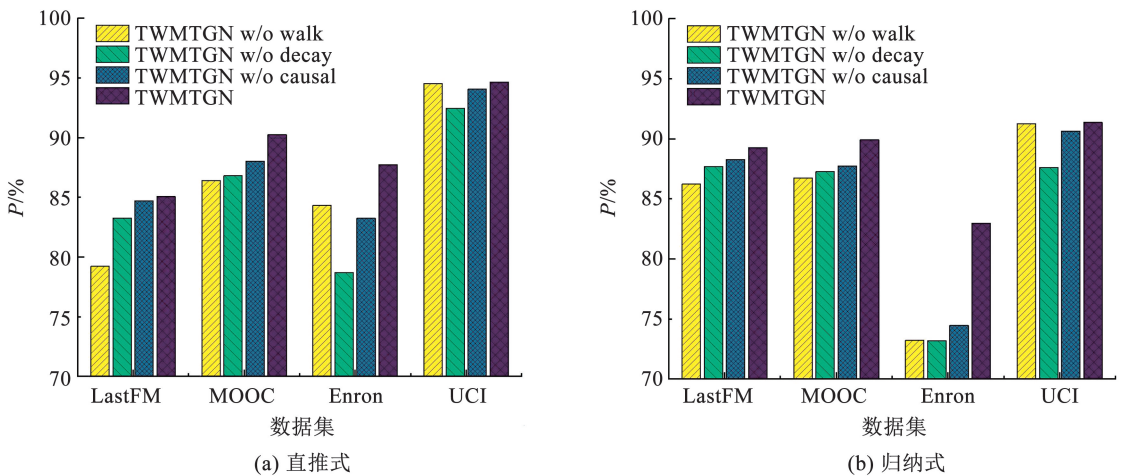


图 3 消融实验结果

Fig. 3 Ablation experiment results

2.6 参数灵敏度分析

为了对不同时间约束随机游走序列进行聚合,在 UCI 数据集上选取平均值、标准差、最小值、最大值 4 种编码方式进行测试,结果如图 4 所示。由图 4 可知,均值编码方式效果最好。在聚合历史交互节点特征获得更新后的目标节点嵌入表示时,历史交互节点的数量是重要的影响因素。因此,在 UCI 数据集上测试了历史交互节点数量对模型效果的影响,结果如图 5 所示。由图 5 可知,随着节点数量的增加, P 先提升后下降。原因是随着目标节点的信息感受野扩大,模型获取的有效信息增加,但当历史交互节点数量进一步扩大时,出现了更多噪声和冗余信息,干扰模型对关键特征的学习,导致性能下降。

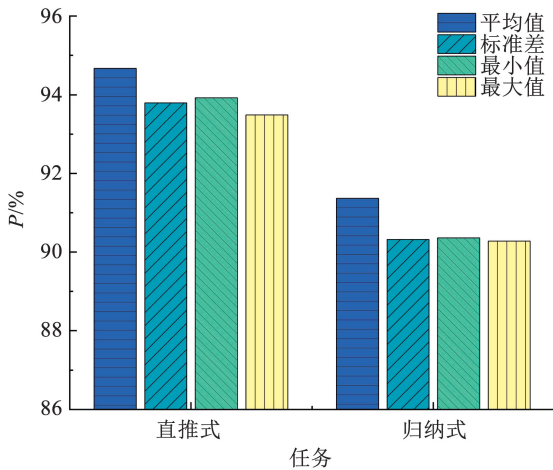


图 4 游走序列编码方式的影响

Fig. 4 Effect of walk sequence encoding methods

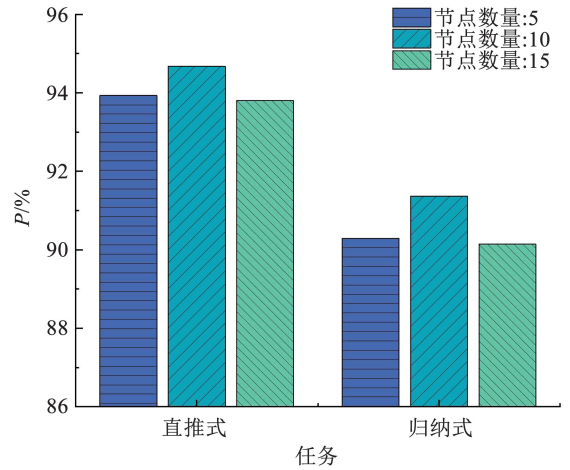


图 5 历史交互节点数量的影响

Fig. 5 Effect of the number of historical interaction nodes

时间约束游走的长度和次数对于捕获网络拓扑结构特征有重要影响。游走长度的增加能够更好地捕捉网络全局结构信息,然而过长的游走长度可能会导致冗余信息的产生。过多的游走次数也会增加模型的计算负担。因此,在 UCI 数据集上测试了不同游走长度和次数对模型效果的影响,结果如图 6 所示。在直推式设置下,当游走长度和次数均为 4 时,结果最优;在归纳式设置下,当游走长度为 2、游走次数为 6 时,结果最优。因此,可以通过平衡实验精度要求和计算复杂度来选取游走长度和次数。

为检验超参数对模型的影响,在 UCI 数据集上选取批大小和注意力头数进行实验,结果如图 7、8 所示。结果显示,最优的批大小为 200,最优的注意力头数为 2。

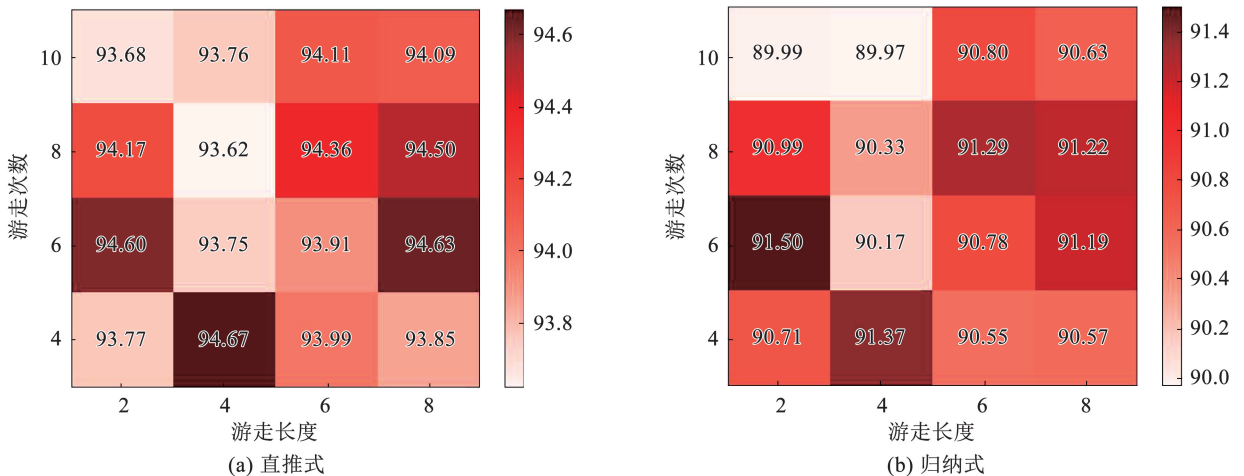


图 6 游走长度和次数对指标 P 的实验结果

Fig. 6 Experimental results of walk length and frequency on P

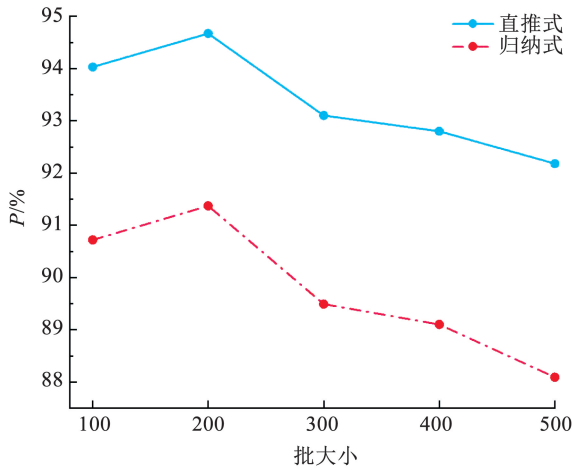


图7 批大小实验结果

Fig. 7 Experimental results for batch size

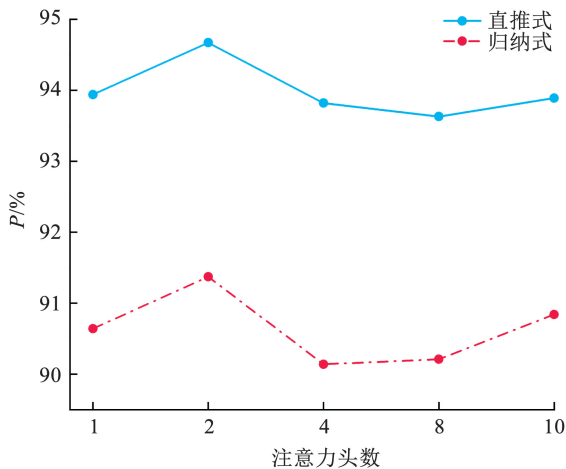


图8 注意力头数实验结果

Fig. 8 Experimental results of attention head count

3 结论

本文针对时序图表示学习提出了一种融合时间约束游走的记忆增强时序图神经网络,联合提取网络的异质性和演化模式,主要结论如下:

1) 构造的时间约束游走序列可以有效捕捉时序网络邻域结构特征。

2) 记录节点记忆状态能够捕获网络长期依赖关系。将编码后的游走序列融合至节点记忆的更新过程,能够利用网络拓扑结构信息学习网络长期演变规律。

3) 利用节点类型和时间衰减特性设计特征衰减层,能够提高模型对关键历史交互节点的识别能力,降低发生时间较长的历史事件对目标节点产生的影响,有效聚合了历史交互节点特征。

4) 因果卷积网络能够有效捕获历史交互节点特征的潜在关联。

参考文献

- [1] 王天保, 刘昱, 郭继昌, 等. 图卷积神经网络行人轨迹预测算法 [J]. 哈尔滨工业大学学报, 2021, 53(2): 53
WANG Tianbao, LIU Yu, GUO Jichang, et al. Pedestrian trajectory prediction algorithm based on graph convolutional network [J]. Journal of Harbin Institute of Technology, 2021, 53(2): 53. DOI: 10.11918/202006051
- [2] YANG Menglin, ZHOU Min, KALANDER M, et al. Discrete-time temporal network embedding via implicit hierarchical learning in hyperbolic space [C]//Proceedings of the 27th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Singapore: ACM, 2021: 1975. DOI:10.1145/3447548.3467422
- [3] LI Xiaoxiao, SAUDE J, REDDY P, et al. Classifying and understanding financial data using graph neural network [C]//AAAI Workshop on Knowledge Discovery from Unstructured Data in Financial Services. New York: AAAI, 2020
- [4] KING I J, HUANG H H. Euler: detecting network lateral movement via scalable temporal link prediction [J]. ACM Transactions on Privacy and Security, 2023, 26(3): 35. DOI:10.1145/3588771
- [5] WU Cheng, WANG Chaokun, XU Jingcao, et al. Instant representation learning for recommendation over large dynamic graphs [C]//39th IEEE International Conference on Data Engineering, ICDE 2023. Anaheim: IEEE, 2023: 82. DOI: 10.1109/ICDE55515.2023.00014
- [6] LI Longhai, DUAN Lei, WANG Junchen, et al. Memory-enhanced transformer for representation learning on temporal heterogeneous graphs [J]. Data Science and Engineering, 2023, 8(2): 98. DOI: 10.1007/s41019-023-00207-w
- [7] LU Yuanfu, WANG Xiao, SHI Chuan, et al. Temporal network embedding with micro-and macro-dynamics [C]//28th ACM International Conference on Information and Knowledge Management. Beijing: ACM, 2019: 469. DOI:10.1145/3357384.3357943
- [8] KUMAR S, ZHANG Xikun, LESKOVEC J. Predicting dynamic embedding trajectory in temporal interaction networks [C]//Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Anchorage: ACM, 2019: 1269. DOI:10.1145/3292500.3330895
- [9] XU Da, RUAN Chuanwei, KORPEOGLU E, et al. Inductive representation learning on temporal graphs [C]//8th International Conference on Learning Representations, ICLR 2020. Addis Ababa: ICLR, 2020
- [10] ROSSI E, CHAMBERLAIN B, FRASCA F, et al. Temporal graph networks for deep learning on dynamic graphs [EB/OL]. (2020-06-18) [2024-06-02]. <https://arxiv.org/pdf/2006.10637>
- [11] LUO Yuhong, LI Pan. Neighborhood-aware scalable temporal network representation learning [C]//1st Learning on Graphs Conference. Mountain View: PMLR, 2022: 1
- [12] WANG Yanbang, CHANG Yenyu, LIU Yunyu, et al. Inductive representation learning in temporal networks via causal anonymous walks [C]//9th International Conference on Learning Representations. Vienna: ICLR, 2021
- [13] ZHOU Wei, HUANG Hong, SHI Ruize, et al. Temporal

heterogeneous information network embedding via semantic evolution [J]. IEEE Transactions on Knowledge and Data Engineering, 2023, 35(12): 13031. DOI:10.1109/TKDE.2023.3287260

[14] TRIVEDI R, FARAJTABAR M, BISWAL P, et al. Representation learning over dynamic graphs[EB/OL]. (2018-03-12)[2024-06-02]. <https://arxiv.org/pdf/1803.04051>

[15] 杨应修, 陈红梅, 周丽华, 等. 基于数据增强的异质图注意力网络[J/OL]. 计算机科学. (2024-05-28)[2024-06-02]. <https://link.cnki.net/urlid/50.1075.tp.20240527.1600.015>
YANG Yingxiu, CHEN Hongmei, ZHOU Lihua, et al. Heterogeneous graph attention network based on data augmentation [J/OL]. Computer Science. (2024-05-28)[2024-06-02]. <https://link.cnki.net/urlid/50.1075.tp.20240527.1600.015>

[16] BAI Shaojie, KOLTER J Z, KOLTUN V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling[EB/OL]. (2018-03-04)[2024-06-02]. <https://arxiv.org/pdf/1803.01271>

[17] MIKOLOV T, SUTSKEVER I, CHEN Kai, et al. Distributed representations of words and phrases and their compositionality [C]//27th Conference on Neural Information Processing Systems. Nevada: Morgan Kaufmann Publishers, 2013: 3111

[18] ZHANG Lin, GUO Jiawen, BAI Qijie, et al. Dynamic heterogeneous graph representation learning with neighborhood type modeling[J]. Neurocomputing, 2023, 533: 46. DOI:10.1016/j.neucom.2023.02.060

[19] SOUZA A H, MESQUITA D, KASKI S, et al. Provably expressive temporal graph networks [C]//36th Conference on Neural Information Processing Systems. New Orleans: Morgan Kaufmann Publishers, 2022: 32257

[20] JI Yugang, JIA Tianrui, FANG Yuan, et al. Dynamic heterogeneous graph embedding via heterogeneous hawkes process[C]//European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases. Bilbao: Springer, 2021: 388. DOI:10.1007/978-3-030-86486-6_24

封面图片说明

封面图片来自本期论文“第一类投毒攻击及其特征分析”,是上海交通大学图像处理与模式识别研究所黄晓霖教授团队关于深度神经网络鲁棒性与安全性的研究成果。研究团队提出了第一类投毒攻击的新型范式,揭示其区别于现有“干净标签”与特征碰撞等第二类投毒攻击的本质特征。图片展示了第一类投毒攻击的原理:攻击者基于监督变分自编码器对训练样本进行结构性篡改,尽管在视觉上呈现显著变化,但模型却因未学习到关键判别特征,仍将其误分类为目标类别。该过程在特征空间中造成标签正确但特征扰乱的效果,从而诱导深度模型形成错误决策边界。研究结果表明:该攻击手法具备高度隐蔽性与破坏性,可在多种主流神经网络结构上稳定诱导第一类分类错误,且主流数据增强防御机制在此攻击下普遍失效。研究成果为神经网络安全性系统评估与新型防御策略设计提供了重要理论支撑。

(图文提供:王鹏博,陈思哲,黄晓霖。上海交通大学图像处理与模式识别研究所)