

融合图信息瓶颈与Transformer的时序知识图谱推理方法

辛奕辰, 李时冲, 陈斌, 程章桃, 李耶, 周帆

引用本文

辛奕辰, 李时冲, 陈斌, 程章桃, 李耶, 周帆. 融合图信息瓶颈与Transformer的时序知识图谱推理方法[J]. 计算机科学, 2026, 53(4): 393-405.

XIN Yichen, LI Shichong, CHEN Bin, CHENG Zhangtao, LI Ye, ZHOU Fan. Enhancing Temporal Knowledge Graph Reasoning Method with Graph Information Bottleneck and Transformer [J]. Computer Science, 2026, 53(4): 393-405.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[基于Transformer架构的RNA二级结构预测方法](#)

Prediction Method of RNA Secondary Structure Based on Transformer Architecture
计算机科学, 2026, 53(3): 375-382. <https://doi.org/10.11896/jsjcx.250100005>

[基于双分支融合与分段域适应迁移学习的疲劳驾驶检测](#)

Fatigue Driving Detection Based on Dual-branch Fusion and Segmented Domain Adaptation Transfer Learning
计算机科学, 2026, 53(3): 78-87. <https://doi.org/10.11896/jsjcx.250500025>

[基于大语言模型和深度网络的认知评估量表自动诊断](#)

Large Language Model and Deep Network Based Cognitive Assessment Automatic Diagnosis
计算机科学, 2026, 53(3): 41-51. <https://doi.org/10.11896/jsjcx.250600034>

[聚焦边界和多尺度特征融合的脑卒中病灶分割](#)

Boundary-focused Multi-scale Feature Fusion Network for Stroke Lesion Segmentation
计算机科学, 2026, 53(2): 264-272. <https://doi.org/10.11896/jsjcx.250300137>

[融合多尺度特征和注意力机制的时间序列预测模型](#)

Time Series Forecasting Model Integrating Multi-scale Features and Attention Mechanism
计算机科学, 2026, 53(2): 180-186. <https://doi.org/10.11896/jsjcx.250100113>

融合图信息瓶颈与 Transformer 的时序知识图谱推理方法

辛奕辰¹ 李时冲¹ 陈斌¹ 程章桃¹ 李耶^{1,2} 周帆¹

¹ 电子科技大学信息与软件工程学院 成都 610054

² 喀什地区电子信息产业技术研究院 新疆 喀什 844099

(ycx@std.uestc.edu.cn)

摘要 时序知识图谱以四元组(主体,关系,客体,时间)的形式动态记录事件知识,能够有效刻画现实世界中知识的动态演化特性,因此在推荐系统、大语言模型和知识问答等领域得到了广泛应用。然而,其固有的知识不完备性限制了进一步的拓展和应用。时序知识图谱推理任务旨在预测(补充)图谱中缺失的事件知识,因而受到学术界和工业界的高度关注。现有的时序知识图谱推理方法主要聚焦于挖掘图谱快照内部的结构信息以及快照之间的时序依赖,仍存在以下两个主要问题:1)在建模过程中未充分考虑图谱快照中潜在的噪声和冗余信息;2)过度依赖局部时间窗口内的序列模式,忽视了图谱全局时序依赖的建模。为解决上述问题,提出了一种融合图信息瓶颈理论与 Transformer 的时序知识图谱推理框架 GIBformer。该框架首先引入图信息瓶颈理论,对时序知识图谱中的结构信息进行压缩,保留与下游预测任务强相关的关键信息,同时有效抑制噪声和冗余信息的干扰;其次,利用 Transformer 的多头注意力机制,捕捉跨图谱快照的全局时序依赖模式,并融合局部的时序演化信息,实现对缺失事件知识的精准预测。在 4 个主流基准数据集上进行的大量实验证明了该模型的有效性。

关键词: 时序知识图谱;图信息瓶颈;Transformer;时序知识图谱推理

中图分类号 TP391

Enhancing Temporal Knowledge Graph Reasoning Method with Graph Information Bottleneck and Transformer

XIN Yichen¹, LI Shichong¹, CHEN Bin¹, CHENG Zhangtao¹, LI Ye^{1,2} and ZHOU Fan¹

¹ School of Information and Software Engineering, University of Electronic Science and Technology of China, Chengdu 610054, China

² Kash Institute of Electronics and Information Industry, Kash, Xinjiang 844099, China

Abstract Temporal knowledge graphs (TKGs) dynamically record event knowledge in the form of quadruples (subject, relation, object, timestamp), effectively capturing the dynamic evolution of knowledge in the real world. As a result, they have been widely applied in various domains such as recommender systems, large language models, and knowledge-based question answering. However, their inherent incompleteness poses significant challenges for further development and application. Temporal knowledge graph reasoning aims to predict missing event knowledge in TKGs, and has thus attracted considerable attention from both academia and industry. Existing methods for TKG reasoning mainly focus on extracting structural information within graph snapshots and modeling temporal dependencies between them. Nonetheless, they still suffer from two major limitations: 1) insufficient handling of noise and redundancy present in the snapshots during the modeling process; 2) an overreliance on local temporal patterns within short time windows, while ignoring global temporal dependencies across the entire TKG. To address these issues, this paper proposes GIBformer, a novel temporal knowledge graph reasoning framework that integrates the graph information bottleneck principle with a Transformer architecture. Specifically, it first introduces the graph information bottleneck to compress structural information in TKGs, preserving key information that is highly relevant to downstream prediction tasks while effectively filtering out noise and redundancy. Then, a Transformer with multi-head attention is employed to capture global temporal dependencies across snapshots, while also incorporating local temporal dynamics to enhance the prediction of missing event knowledge. Extensive experiments conducted on four widely-used benchmark datasets demonstrate the effectiveness of the proposed model.

Keywords Temporal knowledge graph, Graph information bottleneck, Transformer, Temporal knowledge graph reasoning

到稿日期:2025-04-11 返修日期:2025-07-04

基金项目:国家自然科学基金(62176043, 62072077, U22A2097);新疆维吾尔自治区重点研发计划(2024B03041)

This work was supported by the National Natural Science Foundation of China(62176043, 62072077, U22A2097) and Xinjiang Uygur Autonomous Region Key Research and Development Program(2024B03041).

通信作者:李耶(liyeuestc@uestc.edu.cn)

1 引言

知识图谱(Knowledge Graph, KG)通过图的形式组织存储事实,在各个领域得到广泛应用,如自然语言理解^[1]和推荐系统^[2]。然而,传统的静态知识图谱在表达现实世界中事实的动态演化方面存在显著局限^[3]。例如,面对事实三元组(巴西,成为,金砖国家主席国),无法判断该信息是否仍然有效,抑或是巴西已经不再担任主席国。这种缺乏时间维度的表示难以反映事实随时间推移所发生的变化。为克服上述问题,时序知识图谱(Temporal Knowledge Graph, TKG)引入时间维度,将事实表示为四元组(主体,关系,客体,时间),从而构建能够动态反映知识演化过程的图谱。例如,上述事实按时序知识图谱中可表示为(巴西,成为,金砖国家主席国,2025/01/01),明确表明该事件发生于2025年,有效捕捉了事实的时效性。这一动态建模能力使得时序知识图谱在事件预测、动态推荐、时序问答等领域展现出更强的表达力与应用潜力。

鉴于现实世界事件的复杂性,时序知识图谱往往是不完备的,因此需要借助推理来推断特定时间点缺失的事件。时序知识图谱推理^[4-5]旨在基于历史事实推断缺失的历史事件或未来事件,主要可以分为插值(补全历史缺失事实)和外推(预测未来事件)两种任务设置^[6]。其中,外推任务由于需要对未来事件进行推理,挑战性更大。本文主要对外推任务进行研究。现有方法主要采用图神经网络对时序知识图谱快照内的结构信息进行建模,并结合循环神经网络以捕捉快照之间的时序依赖关系^[7-8]。例如,RE-NET^[9]通过局部邻域聚合建模相邻时间事实,但难以捕捉跨快照的全局周期性规律;TLogic^[10]虽引入时间逻辑规则,却无法动态权衡不同历史模式的重要性;文献^[7]的方法通过图神经网络建模时序知识图谱快照内结构信息,并结合循环神经网络捕捉时序知识图谱快照间实体演化的时序依赖信息。

尽管时序知识图谱推理领域已取得诸多研究进展,但仍面临以下两项关键挑战:其一,时序知识图谱中普遍存在大量噪声和冗余信息,例如无关实体或低频关系可能干扰结构信息的有效传播,降低推理精度;其二,现有方法多侧重于局部时间窗口内的时序模式建模,忽略了跨快照的全局时序依赖关系,如周期性事件或长期演化规律,限制了模型对复杂时序动态的建模能力。

针对上述挑战,有效压缩时序知识图谱快照中的结构信息并建模快照间的全局时序依赖,成为提升推理性能的关键。信息瓶颈理论通过最大化地保留任务相关信息并最小化冗余信息,为信息压缩提供了坚实的理论基础^[11]。同时,Transformer具备强大的全局注意力机制,能够高效建模长距离依赖关系与整体时序特征^[12]。此外,历史交互的重复性建模可以通过检索全局历史事实,来辅助当前预测,避免关键线索的遗漏。基于上述思路,本文提出了一种融合图信息瓶颈(Graph Information Bottleneck, GIB)与Transformer的时序知识图谱推理框架GIBformer。具体而言,首先利用图信息瓶颈对每个快照内的结构信息进行压缩,去除冗余与噪声;其次,引入Transformer对快照序列进行建模,从而捕捉长期依赖与周期性演化模式;然后,进一步设计重复历史交互模块,

检索全局历史事实以约束预测空间;最后,通过时间感知的解码器融合多源信息,提升对未来事件的预测准确性。

本文的主要贡献总结如下:

1)首次将信息瓶颈理论引入时序知识图谱推理任务中,提出了一种图信息瓶颈增强的推理框架,有效压缩了历史知识图谱中的冗余交互信息与噪声信息,同时保留了与推理任务高度相关的子图结构信息;

2)设计了一种全局-局部联合建模的时序依赖建模方法,该方法利用Transformer捕获图谱快照间的全局时序演化依赖,并通过门控循环单元建模局部时序序列依赖,提升模型对复杂事件演化规律的建模能力;

3)在多个基准数据集上进行了大量实验,结果表明,所提模型显著优于现有方法,进一步验证了信息压缩与全局时序依赖建模的有效性。

2 相关工作

2.1 时序知识图谱推理

时序知识图谱推理任务近年来受到广泛关注,其主要可以分为两类任务设置:插值设置和外推设置^[6,13]。

插值设置下的时序知识图谱推理方法旨在补全缺失的历史事实。早期方法多为对静态知识图谱模型的直接扩展,通常将时间戳视为独立的附加属性,与实体和关系进行关联。例如,文献^[14]将经典的TransE模型^[15]扩展至时序知识图谱,通过替换其评分函数以适应时间信息,从而实现按时序三元组的建模。然而,单纯地将时间戳作为额外属性并未充分挖掘其潜在的动态结构性与语义信息,因而在刻画事件随时间演化的动态特性方面存在一定局限。为此,后续研究进一步将时间戳嵌入潜在空间,结合实体的结构语义信息进行建模,通常通过设计专门的时间嵌入函数来实现。其中较为典型的嵌入方式包括时间旋转函数^[16]、时间超平面函数^[17]以及非线性嵌入函数^[18]等。具体而言,文献^[16]将时间信息表示为四元数向量空间中的旋转操作,通过旋转机制将时间嵌入融入到实体和关系的四元数表示中;文献^[17]则设计了时间超平面机制,通过构造潜在空间中的时间超平面来划分不同时间段的知识图谱子图,并在超平面上对实体和关系进行投影和平移;文献^[18]提出了四元组分配网络,分别对实体、关系和时间戳进行语义建模,从而促进跨空间信息的交互融合。

在外推设置下,时序知识图谱推理旨在根据已知的历史事实推断未来事件的发生,因而具有更高的实用价值与挑战性。这类方法通常将具有相同时间戳的历史事实构建为历史快照子图,以挖掘其中蕴含的结构与语义信息。在建模策略上,外推方法多采用深度学习框架,尤其是基于图神经网络(GNN)与循环神经网络(RNN)的方法。具体而言,文献^[19]将关系信息引入图卷积网络(GCN),用于捕捉实体间关系随时间演化的动态变化,并据此推断未来事实。文献^[20]将图卷积网络与门控循环单元(GRU)结合,通过融合关系上下文与词语上下文,实现对多类型并发事件及其候选参与者的联合预测。文献^[16]构建了自回归的循环事件网络,联合局部图结构聚合与全局事件编码,实现未来事件的多步推理。此

外,文献[6]融合实体的静态属性以增强其演化表示,采用递归机制建模图谱序列的结构依赖与相邻事件间的时间序列模式。文献[21]综合局部时序递归编码与全局重复历史建模,通过周期性时间解码器实现多模态历史特征的联合推理。文献[22]引入逻辑规则与时间随机游走机制,在保证时间一致性的同时增强模型可解释性。文献[23]基于对比学习框架,捕捉历史与非历史事件之间的依赖关系,从而提升模型的判别能力。文献[24]通过子图采样构建推理子图,并引入时间关系注意机制,引导模型提取更具预测性的封闭子图。

然而,现有的大多数推理模型主要聚焦于图谱快照内的结构建模与快照间的时序模式挖掘,通过联合使用图神经网络与循环神经网络来捕捉事件演化过程中的局部依赖。然而,这类方法普遍忽略了两个关键问题:一是图谱快照内部存在大量与预测无关的噪声和冗余信息,这些噪声和冗余信息可能会干扰图谱内结构信息的有效传播;二是缺乏对时序知识图谱整体演化过程的建模能力,难以捕捉全局时序模式。为此,本文提出了一种融合信息瓶颈和 Transformer 架构的时序知识图谱推理框架 GIBformer。与上述方法不同,GIBformer 模型在核心设计上融合了图信息瓶颈理论与 Transformer 架构,一方面通过信息瓶颈机制压缩快照内部冗余噪声,强化关键信息提取,另一方面,利用 Transformer 的全局建模能力,系统性建模图谱快照间的长期依赖,显著提升了推理的鲁棒性与时序建模能力。

2.2 信息瓶颈

近年来,信息瓶颈理论(Information Bottleneck, IB)在机器学习与深度学习领域引发了广泛关注。其核心思想是在保留与目标任务相关的关键信息的前提下,压缩输入信息,获得简洁且高效的表示^[11]。文献[25]首次将变分信息瓶颈方法引入深度学习框架,推动了该理论在实际模型中的应用。现有研究主要从两个方向将信息瓶颈机制引入图神经网络。一类方法将信息瓶颈用于图神经网络提取后的特征表示,通过选择输入特征的子集,估计和优化信息瓶颈目标。例如,文献[26]将信息瓶颈机制引入图学习中,以筛选冗余节点和边;文献[22]采用无监督方式建模异构信息网络中的一致性假设;文献[27]则将信息瓶颈应用于关键子图识别任务,挖掘任务相关的局部结构。另一类方法则将图结构信息直接纳入信息压缩过程,使结构信息参与特征选择。例如,文献[28]在信息瓶颈查询空间中引入结构与特征正则项,从而增强表示的判别能力;文献[29]则将信息瓶颈原则扩展至动态图表示学习任务,通过构建最小充分性与共识性通道,实现对时空结构中冗余信息的压缩与关键信息的提取。在此基础上,本文在模型设计中引入图信息瓶颈理论,对时序知识图谱中每一个图谱快照进行结构信息压缩,旨在有效保留与推理任务相关的重要子图结构和邻居语义信息,同时显著抑制快照内部的噪声干扰。与现有将信息瓶颈应用于静态图表征学习的工作不同,GIBformer 首次将信息瓶颈机制引入时序知识图谱推理框架中,针对动态演化的图快照,设计了适配时序特性的结构压缩模块,从而在时序推理过程中兼顾局部子图筛选与关键演化模式捕捉。

2.3 基于 Transformer 的知识图谱推理

现有基于 Transformer 的知识图谱推理方法主要通过 Transformer 模块学习实体的结构信息与上下文语义嵌入。文献[30]利用 Transformer 编码实体邻居信息,提升结构表示能力;文献[31]则通过 Transformer 挖掘实体及其上下文之间的语义联系。然而,这些方法大多未考虑知识图谱中的时间信息,因而难以适用于时序知识建模任务。为弥补这一不足,部分研究开始尝试将时间信息引入 Transformer 框架中。文献[32]将与时间相关的文本信息输入 Transformer,以增强时序知识图谱嵌入的时间表达能力;文献[33]设计了两种 Transformer 变体,分别用于捕捉瞬时结构信息与时间演化模式,并引入关系连续时间编码函数,以削弱与查询无关的事实干扰,提升模型对长期演化过程的建模能力;文献[34]则结合预训练 Transformer,联合建模结构特征与时间上下文,并构建事件演化树,进一步增强了时序推理能力。尽管 Transformer 在传统知识图谱补全任务中得到广泛应用,但在时序知识图谱推理领域的系统性探索仍相对有限。现有研究多聚焦于结构化知识与非结构化文本的融合,而对图谱快照间全局时序依赖的建模关注不足。与此不同,GIBformer 旨在利用 Transformer 强大的全局建模能力,有效捕捉时序知识图谱中快照之间的长期依赖关系与演化规律,从而提升对未来事件的预测能力。GIBformer 充分利用 Transformer 架构的长距离依赖建模优势,不仅对快照内部提取出的关键信息进行统一编码,还显式建模快照序列的长期演化规律,从而实现对未来事件的更准确推理,突破了现有方法局限于局部时序建模的瓶颈。

3 问题定义及预备知识

3.1 时序知识图谱

令 \mathcal{E} , \mathcal{R} 和 T 分别表示实体、关系和时间戳的集合。一个事件知识则表示为四元组 (s, r, o, t) 的形式,其中 $s, o \in \mathcal{E}$ 分别表示主体和客体, $r \in \mathcal{R}$ 表示关系类型, $t \in T$ 表示事件发生的时间戳。所有发生在时间窗口 $(t - \Delta t, t]$ 的事件元组组成了一个静态知识图谱 G_t , 而一个时序知识图谱 $\mathcal{G} = \{G_1, G_2, \dots, G_T\}$ 则表示一系列静态知识图谱快照的集合。

3.2 问题定义

时序知识图谱外推推理的目标是基于一系列历史知识图谱快照来预测未来可能发生的事件知识。进一步地,时序知识图谱外推推理任务依据预测对象的不同,又可以划分为实体预测和关系预测。其中,实体预测旨在预测四元组查询中缺失的实体,比如四元组查询 $(s, r, ?, t)$ 中缺失的客体或四元组查询 $(?, r, o, t)$ 中缺失的主体。类似地,关系预测旨在预测四元组查询 $(s, ?, o, t)$ 中缺失的关系。本研究可同时进行实体预测和关系预测任务,即基于海量的历史事件知识预测出给定查询中缺失的实体和关系。

3.3 信息瓶颈

信息瓶颈理论使用互信息(Mutual Information, MI)作为优化目标,以平衡模型数据拟合能力和模型的鲁棒性。经典信息瓶颈理论^[6]设定数据 \mathbf{X} 、表示 \mathbf{Z} 和目标 \mathbf{Y} 服从马尔可夫链 $(\mathbf{Y} \rightarrow \mathbf{X} \rightarrow \mathbf{Z})$, 即表示 \mathbf{Z} 依赖于数据 \mathbf{X} , 并且在已知数据 \mathbf{X} 的

条件下与目标 \mathbf{Y} 无关。其核心目标是在保证 \mathbf{Z} 对 \mathbf{Y} 具有最大预测能力的同时,最小化 \mathbf{Z} 对 \mathbf{X} 的依赖,即学习最小充分表示 \mathbf{Z} 。这一优化目标正式定义为以下互信息优化问题:

$$\arg \min_{\mathbf{Z}} (-I(\mathbf{Z}; \mathbf{Y}) + \beta I(\mathbf{Z}; \mathbf{X})) \quad (1)$$

其中, $I(\mathbf{Z}; \mathbf{X})$ 约束 \mathbf{Z} 仅保留输入 \mathbf{X} 中与目标任务 \mathbf{Y} 相关的信息,从而减少冗余与噪声; $I(\mathbf{Z}; \mathbf{Y})$ 确保 \mathbf{Z} 仍然包含对 \mathbf{Y} 进行预测所需的关键信息; β 作为调节参数,用于平衡信息压缩与模型预测能力。

在图学习领域,信息瓶颈用于在最小化信息损失的前提下对图的属性信息进行压缩。给定图 $G = (\mathbf{X}, \mathbf{A})$ 及其标签 \mathbf{Y} , 其中 \mathbf{X} 表示节点特征, \mathbf{A} 表示邻接矩阵, 信息瓶颈目标是学习一个最优的压缩图表示 $G_{\text{IB}} = (\mathbf{X}_{\text{IB}}, \mathbf{A}_{\text{IB}})$ 。这一优化过程可以形式化表达为:

$$\arg \min_{G_{\text{IB}}} (-I(G_{\text{IB}}; \mathbf{Y}) + \beta I(G_{\text{IB}}; G)) \quad (2)$$

其中, $I(G_{\text{IB}}; G)$ 约束压缩后的子图并去除无关的结构噪声; $I(G_{\text{IB}}; \mathbf{Y})$ 确保压缩后的子图仍包含对 \mathbf{Y} 进行预测所需的关键结构信息; β 是调节因子。

图信息瓶颈理论(GIB)在时序知识图谱推理任务中具有天然的适用性与优势。首先,时序知识图谱中的快照通常包含大量与目标推理查询无关的冗余实体与关系,这些噪声信息不仅增加了推理的难度,还可能引入误导性特征。图信息瓶颈通过最大化地保留与推理目标相关的信息,同时最小化与预测无关的信息,理论上能够提升推理的准确性与鲁棒性。其次,时序知识图谱推理不仅需要捕捉局部结构,还需关注随时间演化的动态模式,图信息瓶颈能够促使模型关注关键演化路径与重要实体交互,从而增强对长期演化规律的建模能力。此外,从理论基础上看,图信息瓶颈理论本质上是

一种最小充分表示(Minimal Sufficient Representation)学习机制,能够在不丢失关键信息的前提下,实现输入数据的最优压缩。这一特性对于时序知识图谱中高噪声、稀疏、高动态性的特征分布具有天然的适配性。基于上述分析,本文将图信息瓶颈理论引入时序知识图谱推理,有望在提高推理精度、强化模型泛化能力与降低推理复杂度等方面发挥重要作用。

本文引入图信息瓶颈机制,对时序知识图谱快照中的结构信息进行压缩,旨在有效过滤冗余结构与噪声干扰,同时保留与推理任务高度相关的关键子图结构,从而学习更加精确的实体关系表示,进一步提升模型的推理性能与鲁棒性。

4 融合图信息瓶颈与 Transformer 的时序知识图谱推理方法

GIBformer 的整体架构如图 1 所示。它主要由 4 部分组成,分别是 GIB 卷积编码模块、时间序列演化模块、重复事件历史挖掘模块和时间依赖预测网络模块。其中,GIB 卷积编码模块通过对历史时序知识图谱快照进行信息压缩与卷积处理,捕获图谱快照内与推理任务有关的子图结构信息和实体语义信息,同时尽可能压制图谱快照内无关的结构交互信息和噪声;时间序列演化模块使用门控循环单元与多头注意力机制来建模隐藏在图谱快照间的时序依赖模式;重复事件历史挖掘模块,通过查询与当前查询有关的重复性历史事实,以强调历史重复信息对当前预测的影响;模型通过时间依赖预测网络模块对查询的时间信息进行编码,并寻找周期性、非周期性的时序模式,最后通过解码器完成实体和关系预测任务。

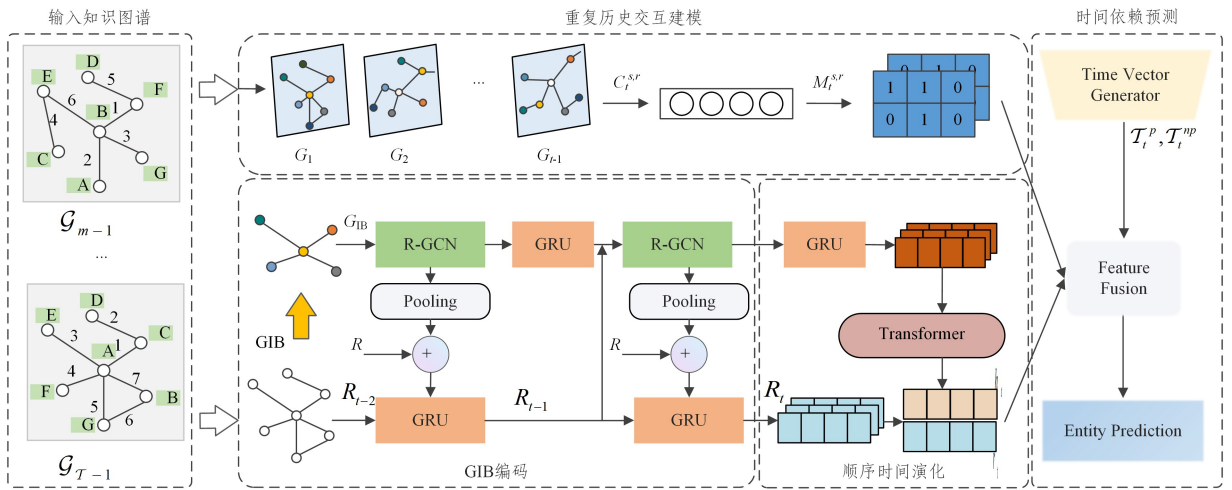


图 1 GIBformer 模型整体框架

Fig. 1 Overall structure of the GIBformer model

4.1 GIB 卷积编码模块

GIB 卷积编码模块旨在对历史知识图谱快照中的冗余事件结构与交互信息进行压缩与编码,挖掘与推理任务高度相关的子图结构特征与实体语义表示。同时,通过抑制图谱快照中无关的结构交互信息与噪声干扰,学习更加紧凑且准确的实体与关系表示。针对给定的查询 $(s, r, ?, t)$, 考虑利用最

近 m 个相邻时间戳的知识图谱快照 $\{G_{t-m}, \dots, G_{t-1}\}$ 来进行结构信息的压缩与编码。具体而言,引入图信息瓶颈机制,对每个图谱快照中的结构信息进行压缩,以去除冗余与噪声,同时保留与目标查询相关的关键子图结构。随后,采用图卷积操作对压缩后的子图结构进行编码,从而提取准确且高质量的实体语义表示与子图结构特征,为后续时序建模提供高质量输入。

受先前研究^[18]的启发,为抑制与推理任务无关的实体语义特征,在卷积编码模块中设计了特征掩码机制。具体地,掩码机制在实体特征表示上应用,通过可学习参数 $\mathbf{M} \in R^d$,控制不同特征维度的重要性。掩码应用方式如下:

$$\mathbf{X}_{\text{IB}} = \mathbf{X}_r + (\mathbf{X} - \mathbf{X}_r) \odot \mathbf{M} \quad (3)$$

其中, \odot 表示 Hadamard 积(逐元素积), \mathbf{X}_r 是从实体输入特征 \mathbf{X} 的先验分布中采样的随机变量。通过与掩码 \mathbf{M} 相乘,模型在训练过程中能够自主调节实体各特征维度的重要性。在训练过程中,掩码 \mathbf{M} 被作为可学习参数,随着后续实体/关系预测任务和消息压缩任务的目标函数同步优化,无需单独监督信号。若某一特征维度与推理任务无关,其在掩码 \mathbf{M} 中对应的权重将在训练过程中逐渐收敛至接近零,从而有效实现噪声特征的过滤与有用特征的保留。经过特征掩码压缩后,实体 u 的表示更新为 $\mathbf{X}_{\text{IB}}(u)$ 。通过上述设计,GIBformer 不仅在特征表示级别上实现了信息压缩,还有效提升了实体表示的任务相关性与推理精度。

同时,为有效筛选与实体/关系预测任务高度相关的快照子图结构信息,将图谱快照中所有关系边建模为一组相互独立的伯努利随机变量,形式如下:

$$\mathbf{A}_{\text{IB}} = \bigcup_{u,v \in \mathcal{V}} \{a_{u,v} \sim \text{Ber}(\pi_{u,v})\} \quad (4)$$

其中, $\pi_{u,v}$ 表示实体节点对 (u, v) 之间关系边的采样概率,作为可学习的注意力权重,用于衡量图谱中关系边与当前推理任务的相关性。采样概率 $\pi_{u,v}$ 越大,表示该关系边越有可能携带与实体/关系预测紧密相关的子图结构信息,因此在子图压缩过程中将被优先保留。具体而言,针对任意一个四元组中的实体节点 (u, v) ,其关系边的采样概率 $\pi_{u,v}$ 由输入节点的特征表征生成,计算方式如下:

$$\mathbf{Z}(u) = \varphi(\mathbf{X}_{\text{IB}}(u)) \quad (5)$$

$$\pi_{u,v} = \text{sigmoid}(\mathbf{Z}(u)\mathbf{Z}(v)^{\text{T}}) \quad (6)$$

其中, $\varphi(\cdot)$ 表示一个两层感知机(MLP),用于进行实体表示的非线性映射。为了在训练过程中实现端到端可微优化,避免因伯努利分布的离散导致梯度中断,采用连续松弛技术^[35],将伯努利分布近似为连续可导形式。

$$\text{Ber}(\pi_{u,v}) \approx \text{sigmoid}\left(\frac{1}{t} \left(\log\left(\frac{\pi_{u,v}}{1-\pi_{u,v}}\right) + \log\left(\frac{\epsilon}{1-\epsilon}\right) \right)\right) \quad (7)$$

其中, $\epsilon \sim \text{Uniform}(0, 1)$ 是从均匀分布采样的噪声变量; $t \in R^+$ 为温度参数,用于调节松弛程度,当 t 较小时,连续松弛更接近原始的伯努利分布,当 t 较大时,分布趋于平滑。

为了进一步保持压缩后子图的稀疏性与计算可行性,在连续松弛后的权重图上施加阈值掩码操作:仅保留松弛权重大于预设非负阈值 a_0 的关系边,其余边被置零。经过掩码处理后,得到一个对称且稀疏的邻接矩阵,作为最终用于子图编码的压缩图结构。通过上述设计,关系边的采样机制不仅实现了与实体/关系预测任务紧密相关的动态子图结构过滤,同时在保证模型可微性的前提下,有效抑制了无关冗余结构噪声,进一步提升了实体的表示学习。

此外,采用变分推断方法^[18]对式(2)进行优化,得到上界表达式:

$$-I(G_{\text{IB}}; \mathbf{Y}) + \beta I(G_{\text{IB}}; G) \leq \frac{1}{N} \sum_{i=1}^N \left\{ -\log q_{\theta}(\mathbf{y}_i | \mathbf{Z}_{\text{IB};i}) + \beta p(\mathbf{Z}_{\text{IB}} | G_i) \log \frac{p(\mathbf{Z}_{\text{IB}} | G_i)}{r(\mathbf{Z}_{\text{IB}})} \right\} \quad (8)$$

其中, $q_{\theta}(\mathbf{Y}_i | \mathbf{Z}_{\text{IB};i})$ 是真实后验 $p(\mathbf{Y} | G_{\text{IB}})$ 的变分近似, $r(\mathbf{Z}_{\text{IB}})$ 是对潜变量先验分布 $p(\mathbf{Z}_{\text{IB}})$ 的变分近似, N 为样本总数。对于式(2)中的压缩项 $I(G_{\text{IB}}; G)$,通过参数化的高斯分布对先验 $r(\mathbf{Z}_{\text{IB}})$ 和后验分布 $p(\mathbf{Z}_{\text{IB}} | G_i)$ 进行建模。

$$r(\mathbf{Z}_{\text{IB}}) = N(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) \quad (9)$$

$$p(\mathbf{Z}_{\text{IB}} | G) = \mathcal{N}(f_{\phi}^r(G_{\text{IB}}), f_{\phi}^{\Sigma}(G_{\text{IB}})) \quad (10)$$

其中, $\boldsymbol{\mu} \in R^K$, $\boldsymbol{\Sigma} \in R^{K \times K}$ 分别为由编码器 $f_{\phi}(G_{\text{IB}})$ 输出的潜变量 \mathbf{Z}_{IB} 的均值向量和对角协方差矩阵。在具体实现中,将先验均值 $\boldsymbol{\mu}_0$ 设为零向量,协方差矩阵 $\boldsymbol{\Sigma}_0$ 设为单位矩阵 \mathbf{I} 。上述两个映射函数 $f_{\phi}^r(\cdot)$ 与 $f_{\phi}^{\Sigma}(\cdot)$ 均由图卷积网络实现,下文将对这两个映射函数进行详细介绍。

为有效捕获知识图谱快照中的结构依赖,本文采用图卷积网络来学习实体和关系的嵌入表示。具体地,对于任一实体 o ,其在第 $l+1$ 层的嵌入表示通过如下方式进行更新:

$$\mathbf{h}_{o,t}^{l+1} = \sigma\left(\sum_{(s,r,o) \in \mathcal{F}_l} \frac{1}{c_o} \mathbf{W}_r^l(\psi(\mathbf{h}_{s,t}^l, \mathbf{r}_t)) + \mathbf{W}_o^l \mathbf{h}_{o,t}^l\right) \quad (11)$$

其中, $\mathbf{h}_{s,t}^l$ 和 $\mathbf{h}_{o,t}^l$ 分别表示在时刻 t 上,实体 s 和 o 在第 l 层的嵌入表示; \mathbf{W}_r^l 和 \mathbf{W}_o^l 为可学习权重参数; $\psi(\cdot)$ 表示一维卷积操作; σ 是 ReLU 激活函数; c_o 为归一化因子,其大小等于实体 o 在图中的入度大小。值得注意的是,该嵌入表示与图信息瓶颈中的表示 $\mathbf{Z}_{\text{IB}}(u)$ 一致,即 $\mathbf{Z}_{\text{IB}}(o) = \mathbf{h}_{o,t}^{l+1}$ 。此外,对于知识图谱快照中未参与任何交互的孤立实体,为确保其表示的更新,在图中为其显式添加自环边。

最后,通过池化操作对图谱中所有节点的表示进行聚合,分别获得均值映射 $f_{\phi}^r(G_{\text{IB}})$ 和协方差映射 $f_{\phi}^{\Sigma}(G_{\text{IB}})$ 。

$$(f_{\phi}^r(G_{\text{IB}}), f_{\phi}^{\Sigma}(G_{\text{IB}})) = \text{Pooling}\{\mathbf{Z}_{\text{IB}}(u), \forall u \in \mathcal{V}\} \quad (12)$$

其中,输出向量的前 K 维表示均值编码 $\boldsymbol{\mu}$,后 K 维表示协方差编码 $\boldsymbol{\Sigma}$ 。

基于此,利用重参数化技巧对最终的图表示 \mathbf{Z}_{IB} 进行采样,具体如下:

$$\mathbf{Z}_{\text{IB}} = f_{\phi}^r(G_{\text{IB}}) + f_{\phi}^{\Sigma}(G_{\text{IB}}) \odot \boldsymbol{\epsilon} \quad (13)$$

其中, $\boldsymbol{\epsilon} \in \mathcal{N}(0, \mathbf{I})$ 表示与输入无关的标准高斯噪声, \odot 表示 Hadamard 积。该表示确保从任务相关子图中采样出的 \mathbf{Z}_{IB} 同时具备压缩性和表示性。

同时,根据式(2)中的压缩项 $I(G_{\text{IB}}; G)$,可以推导出信息压缩的优化目标函数。具体而言,该压缩目标可通过最小化潜变量分布与先验分布之间的 Kullback-Leibler (KL) 散度来实现,其损失函数定义如下:

$$\mathcal{L}_{\text{GIB}} = \mathcal{D}_{\text{KL}}(p(\mathbf{Z}_{\text{IB}} | G) \| r(\mathbf{Z}_{\text{IB}})) \quad (14)$$

其中, $\mathcal{D}_{\text{KL}}(\cdot \| \cdot)$ 表示两个分布之间的 KL 散度, $p(\mathbf{Z}_{\text{IB}} | G)$ 为编码器从原始图 G 中学得的后验分布。

4.2 时间序列演化模块

时间序列演化模块旨在从实体和关系两个层面建模知识图谱快照之间的顺序性时序依赖模式。具体而言,针对给定查询 $q = (s, r, ?, t)$,采用门控循环机制,通过门控循环单元(GRU)对多个时间戳上的图谱快照进行建模,以捕捉时间演

化过程中的交互信息,并动态更新实体表示。实体表示的更新方式定义如下:

$$\mathbf{H}_t = \text{GRU}(\mathbf{H}_{t-1}, \mathbf{H}_{t-1}^{\text{GCN}}) \quad (15)$$

其中, $\mathbf{H}_t, \mathbf{H}_{t-1} \in \mathbb{R}^{|\mathcal{E}| \times d}$ 分别表示 t 时刻和 $t-1$ 时刻的 d 维实体嵌入矩阵, $\mathbf{H}_{t-1}^{\text{GCN}} \in \mathbb{R}^{|\mathcal{E}| \times d}$ 表示在 $t-1$ 时刻, 实体经过图卷积网络单步聚合后的嵌入结果。通过这一更新机制, 模型可以在保持结构信息传递能力的同时, 引入时间维度上的演化特征, 实现对实体动态表示的有效建模。

除了顺序性的时序依赖外, 事件演化过程还通常表现出远距离的长期依赖与周期性特征, 因此, 全局时序模式在建模时序知识图谱中事件动态方面同样具有重要作用。为此, 除了使用门控循环单元建模事件演化的顺序性时序依赖外, 本文进一步设计了基于注意力机制的 Transformer 模块, 用于捕捉图谱快照间的全局时序依赖关系, 以进一步增强模型对全局时序交互依赖的联合建模。

具体地, 在获得所有时间戳的实体表示矩阵后, 首先将这些实体矩阵表示按照时间顺序进行拼接, 以构建全局序列输入。

$$\mathbf{H}_i' = [\mathbf{H}_{t-m} : \mathbf{H}_{t-m+1} : \dots : \mathbf{H}_{t-1}] \quad (16)$$

其中, $[\cdot]$ 表示拼接操作, $\mathbf{H}_i (i \in [t-m, t-1])$ 为第 i 个图谱快照的实体表示矩阵。随后, 使用 Transformer 模块来捕捉全局时序特征^[7], 其中自注意力机制的计算过程如下:

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{Softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}}\right)\mathbf{V} \quad (17)$$

其中, $\mathbf{Q}, \mathbf{K}, \mathbf{V}$ 是 \mathbf{H}_i' 经过线性变换后获得的查询、键和值矩阵; $\text{Softmax}(\cdot)$ 表示 Softmax 操作; d_k 为每个向量的长度; $\mathbf{Q}\mathbf{K}^T$ 用于评估不同实体在各图谱快照中的关联强度, 从而加权汇聚对应的值向量 \mathbf{V} 。

为了增强模型对不同时序依赖模式的表达能力, 引入了多头注意力机制, 其定义如下:

$$\text{MHA}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \{\text{head}_0, \dots, \text{head}_{k-1}\} \quad (18)$$

$$\text{head}_i = \text{Attention}(\mathbf{Q}_i, \mathbf{K}_i, \mathbf{V}_i), i \in \{0, \dots, k-1\} \quad (19)$$

其中, head_i 为第 i 个注意力头的输出结果。经过多层感知机与 Softmax 层处理后, 可生成用于后续预测任务的实体表示向量, 从而捕捉跨快照的全局时序特征, 提升模型对复杂演化模式的建模能力。

此外, 为保证与子图序列中实体嵌入更新过程的一致性, 引入了面向关系的门控循环单元对关系表示进行动态更新。具体过程如下:

$$\mathbf{r}_t' = [\text{pooling}(\mathbf{H}_{t-1}, \mathcal{H}_{r,t}); \mathbf{r}] \quad (20)$$

$$\mathbf{R}_t = \text{GRU}(\mathbf{R}_{t-1}, \mathbf{R}_t') \quad (21)$$

其中, \mathbf{H}_{t-1} 表示 $t-1$ 时刻的实体表示矩阵, $\mathcal{H}_{r,t}$ 表示在时间步 t 下与关系 r 相关的所有实体表示, \mathbf{R}_t' 表示所有 \mathbf{r}_t' 组成的中间关系表示矩阵, \mathbf{R}_t 和 $\mathbf{R}_{t-1} \in \mathbb{R}^{|\mathcal{R}| \times d}$ 是在 t 和 $t-1$ 时刻的关系嵌入矩阵。通过上述操作, 模型有效建模了关系随时间演化的动态特征。

4.3 重复事件历史挖掘模块

重复事件历史挖掘模块旨在建模全局范围内的历史事件重复模式, 为后续解码器提供全局性的重复模式约束与先验

信息。受文献[36]的启发, 本文引入复制机制(Copy Mechanism), 以检索当前查询的潜在重复历史事件。具体而言, 对于给定的查询 $(s, r, ?, t)$, 在时间戳 t 之前的所有知识图谱快照中进行检索, 以获取当前查询的潜在候选客体。具体地, 收集了与当前查询具有相同主体 s 和相同关系 r 的所有历史四元组 (s, r, o', t') , 其中 $t' < t$, 并将这些对应的客体 o' 作为当前查询的潜在重复历史候选实体。上述过程可形式化表示为:

$$C(s, r, t) = \{o' | (s, r, o', t') \in \mathcal{G}, t' < t\} \quad (22)$$

其中, $C(s, r, t)$ 表示在时间戳 t 之前, 查询 $(s, r, ?, t)$ 检索到的潜在重复历史候选实体集合; \mathcal{G} 表示整个时序知识图谱。在时序依赖预测网络模块中, 利用该候选实体集合对模型输出的实体预测概率分布进行掩码操作。具体而言, 在后续的时间依赖预测网络模块中, 当模型计算实体预测的概率得分时, 对于出现在集合 $C(s, r, t)$ 中的实体, 适当提高其预测概率得分; 而对于不在集合中的实体, 则相应地抑制其预测概率得分。对关系预测的检索也是类似的操作。该策略实现了对历史重复依赖的显式建模, 使模型在推理过程中能够优先关注历史中已经出现过的结构模式与实体组合。通过重复事件历史挖掘模块, 模型能够有效捕捉事件发生的重复性与周期性特征, 提升对潜在重复事件的识别能力与整体预测精度。

4.4 时间依赖预测网络模块

时间依赖预测网络模块旨在对时间信息进行建模与编码, 以进一步挖掘时序事件演化过程中潜在的周期性与非周期性模式。通过引入时间感知机制, 增强模型对实体与关系随时间演变特性的建模能力。最终, 利用解码器对实体表示、关系表示与时间特征进行融合解码, 完成实体/关系预测任务。

具体地, 借鉴先前的工作^[21], 本文采用 Time2Vec 时间编码函数^[37]来同时建模周期性和非周期性的时间特征, 以增强模型对事件演化过程中复杂时间特性的表达能力。Time2Vec 通过非线性和线性组合的方式, 对时间戳进行多维映射, 其具体编码过程如下:

$$\mathcal{T}_t^p = \sin(\omega_p t + \phi_p) \quad (23)$$

$$\mathcal{T}_t^{np} = \omega_{np} t + \phi_{np} \quad (24)$$

其中, $\mathcal{T}_t^p, \mathcal{T}_t^{np} \in \mathbb{R}^{d_s}$ 分别为时间戳 t 的周期性和非周期性时间向量表示; $\omega_p, \omega_{np}, \phi_p$ 和 ϕ_{np} 为可学习参数; 正弦函数 $\sin(\cdot)$ 用于建模周期性时间变化模式; 线性项用于建模非周期性时间趋势。通过上述时间编码机制, 模型能够获得更为丰富的时间语义特征, 从而在解码阶段更准确地融合时间信息与结构信息, 提升对目标实体的预测能力。

然而, 以上获得的周期性与非周期性时间向量表示是与具体查询事件无关的全局时间特征, 若直接用于实体/关系预测, 则难以有效地将时间模式融入事件的动态演化过程, 这可能导致模型性能下降。为此, 本文进一步对时间表示与查询语义进行融合, 以增强时间信息的任务相关性。具体而言, 使用一维卷积操作, 将所获得的周期性与非周期性时间向量与查询中的实体与关系表示进行深度融合, 使时间特征能够参与事件语义建模, 从而提升对目标实体的预测能力。具体融合过程如下:

$$m_c^n = \sum_{\tau=0}^{K-1} w_c(\tau, 0) \hat{h}_{s,t}^E(n+\tau) + w_c(\tau, 1) \hat{r}_t(n+\tau) + w_c(\tau, 2) \hat{T}_t^p(n+\tau) + w_c(\tau, 3) \hat{T}_t^{np}(n+\tau) \quad (25)$$

其中, c, K, n 分别表示卷积核的数量、核宽度和输出向量维度索引; w_c 为可学习卷积核参数; $\hat{h}_{s,t}^E, \hat{r}_t, \hat{T}_t^p$ 和 \hat{T}_t^{np} 分别表示实体嵌入、关系嵌入、周期性时间表示向量和非周期性时间表示向量。最终, 卷积输出矩阵 O_t 由所有卷积结果向量 m_c^n 拼接构成。

最后, 使用 ConvTransE^[39] 作为解码器对给定查询的候选实体/关系进行评分, 具体过程如下:

$$p^{\text{score}} = \text{softmax}(\text{ReLU}(\text{vec}(O_t) \mathbf{W}_4) \mathbf{H}_t^{E/R}) \quad (26)$$

其中, $\text{vec}(\cdot)$ 表示特征映射操作, \mathbf{W}_4 为可学习的线性变换矩阵, \mathbf{H}_t^E 表示所有实体的候选嵌入矩阵。

为进一步增强模型对复杂事件演化模式的建模能力, 本文分别计算了不同时序演化偏好下的实体/关系预测得分, 包括非重复性演化模式和重复性演化模式。针对这两类时序特征, 模型分别学习对应的预测分布, 并通过加权融合获得最终的实体/关系预测结果。具体计算式如下:

$$p^{\text{final}} = \alpha \times p^r + (1-\alpha) \times p^f \quad (27)$$

其中, p^r, p^f 分别表示重复性演化模式和非重复性演化模式下的预测得分; $\alpha \in [0, 1]$ 为可调节的加权系数, 用于平衡两种演化模式对最终预测结果的影响。

4.5 训练目标

为联合优化实体预测和关系预测任务, 本文采用交叉熵损失函数作为主要训练目标。具体地, 任务的总体损失函数定义如下:

$$\mathcal{L}_{\text{main}} = \sum_{(s,r,o,t) \in G} y_o^i \log p(o|s,r,t) + y_r^i \log p(r|s,o,t) \quad (28)$$

其中, $P(o_i | s, r, t)$ 表示模型在给定查询 $(s, r, ?, t)$ 下对实体 o 的预测得分; y_o^i 为对应的标签指示变量, 当查询事实存在于图谱中时取值为 1, 否则为 0。

此外, 为增强模型对任务相关信息的压缩能力, 本文将信息压缩任务的损失项 \mathcal{L}_{GIB} 引入训练过程中, 联合优化模型表示学习与冗余信息压缩目标。最终的整体训练目标函数定义如下:

$$\mathcal{L} = \mathcal{L}_{\text{main}} + \beta \mathcal{L}_{\text{GIB}} \quad (29)$$

其中, β 为超参数, 用于平衡实体/关系预测任务与信息压缩目标之间的权重。

5 实验与结果

本章在 4 个基准数据集上进行了实验, 以验证 GIBformer 的有效性。首先, 详细介绍实验使用的数据集、性能评估指标、对比基线模型以及模型参数设置。随后, 设计并开展全面实验, 对 GIBformer 进行深入评估, 包括模型整体性能对比、消融实验、参数敏感性分析以及案例研究。

5.1 数据集与评估指标

为了验证 GIBformer 的有效性, 本文使用了 4 个基准数据集: ICEWS14^[39], ICEWS18^[9], ICEWS05-15^[39] 和 GDELT^[40]。表 1 详细列出了各数据集的统计信息, 包括实体数量、关系类型数量、时间戳数量和时间间隔等数据。

ICEWS14: 该数据集来自综合危机预警系统 (ICEWS), 聚焦于 2014 年发生的事件。数据集包含 7128 个唯一实体, 并涵盖 230 种不同的实体关系类型。

ICEWS18: 此数据集记录了 2018 年发生的事件, 时间粒度为天。数据集包含 23033 个实体以及 256 种关系类型。

ICEWS05-15: 该数据集涵盖了从 2005 年 1 月 1 日至 2015 年 12 月 31 日期间的大量事件信息, 时间粒度为天。数据集包含 10488 个实体, 并且涵盖 251 种关系类型。

GDELT: 该数据集源自全球事件、语言和语调数据库 (GDELT), 以 15 分钟为时间单位, 记录了 2018 年 1 月 1 日至 2018 年 1 月 31 日期间的事件。该数据集包含 7691 个实体和 240 种关系类型。

为确保实验结果的公平性和可复现性, 本文遵循已有工作^[41-42]的预处理策略, 对所有数据集采用统一的数据划分方案。具体而言, 所有数据集均按照时间顺序划分为训练集 (80%)、验证集 (10%) 和测试集 (10%), 满足以下顺序关系: 时间戳(训练集) < 时间戳(验证集) < 时间戳(测试集)。该划分方式不仅保证了训练与测试数据之间的时间独立性, 而且更贴近真实世界中基于历史信息预测未来事件的应用场景, 有助于全面评估模型在时序知识图谱推理任务中的实际表现。需要特别说明的是, 本文中的训练集、验证集和测试集并未采用随机划分方式, 而是依据事件发生的时间顺序进行划分。验证集主要用于确定模型的最优参数配置和早停策略。

表 1 数据集统计

Table 1 Statistics of experimental datasets

数据集	实体数	关系数	训练集	验证集	测试集	时间间隔	时间戳数
ICEWS14	7128	230	74845	8514	7371	1 天	365
ICEWS18	23033	256	373018	45995	49545	1 天	304
ICEWS05-15	10094	251	368868	46302	46159	1 天	4017
GDELT	7691	240	1734399	238765	305241	15 分钟	2751

此外, 为全面评估模型性能, 本文采用时序知识图谱推理任务中常用的评估指标: 平均倒数排名 (Mean Reciprocal Rank, MRR) 和 Hits@K ($K=1, 3, 10$)。这两类指标分别从整体排序质量和前 K 名命中率两个维度衡量模型的预测效果, 具体定义如下:

$$\text{MRR} = \frac{1}{|Q|} \sum_{i=1}^{|Q|} \frac{1}{\text{rank}_i} \quad (30)$$

$$\text{Hits@K} = \frac{1}{N} \sum_{i=1}^N \mathbb{I}(\text{rank}_i \leq K), K=1, 3, 10 \quad (31)$$

其中, $|Q|$ 表示查询的总数; rank_i 表示第 i 个查询的预测结果在候选实体中的排名; $\mathbb{I}(\cdot)$ 为指示函数, 当 $\text{rank}_i \leq K$ 时取值为 1, 表示命中, 否则为 0。遵照先前的研究工作^[21], 本文使用时间感知的过滤方法来计算评估指标。

5.2 基线选择

本节将 GIBformer 与当前主流的 12 种时序知识图谱推理模型进行系统对比, 以全面评估模型在实体/关系预测任务中的有效性与优越性。参与对比的基线模型包括: R-

GCN^[6], ConvE^[43], TTransE^[44], RE-NET^[9], RE-GCN^[41], CyGNet^[36], xERTE^[24], TLogic^[10], TANGO^[45], TiGRN^[21], RETIA^[46] 和 SiMFy^[47]。下面对这 12 种代表性方法的核心思想进行简要介绍。

1) R-GCN: 该模型引入了关系图卷积网络, 通过捕捉动态关系变化来预测实体。

2) ConvE: 该模型通过多层 2D 卷积网络捕捉知识图谱中实体与关系的表示, 以有效建模图中结构信息。

3) TTransE: 该模型是 TransE^[15] 的扩展, 针对时序知识图谱推理问题, 替换了评分函数, 并通过改进的嵌入方式捕捉时间依赖关系。

4) RE-NET: 该模型利用循环神经网络对历史事件进行建模, 并结合局部图结构聚合来捕捉同一时间窗口内事件之间的连接模式。

5) RE-GCN: 该模型结合了关系图卷积网络和门控循环神经网络, 在捕捉知识图谱中的结构依赖关系的同时, 建模相邻时间事件间的时序依赖模式。

6) CyGNet: 该模型引入了一种复制机制, 通过从已知实体词汇表中预测实体的概率来生成潜在实体。

7) xERTE: 该模型通过构建推理图生成查询子图, 并使用时间关系注意力机制来引导提取最具相关性的封闭子图。

8) TLogic: 该模型通过引入逻辑规则和时间随机游走机制, 从时序知识图谱中学习时间逻辑规则, 以推理时间相关的事件关系。

9) TANGO: 该模型扩展了常微分方程的概念, 将其应用于多图卷积网络, 从而在动态多图环境中保持连续性并生成连续时间的嵌入表示。

10) TiGRN: 该模型设计了一个局部-全局时序知识图谱编码器框架, 结合周期性时间解码器进行多模式历史特征的

联合推理。

11) RETIA: 该模型通过设计关系-超关系双视角子图, 联合聚合实体和关系的邻居信息, 建模关系与实体间的时序关联约束。

12) SiMFy: 该模型基于多层感知机, 对事件序列中的结构依赖关系进行建模, 并在推理过程中结合历史事件频率信息。

5.3 实验细节

对于所有的基线模型, 本文使用其官方开源代码, 并采用原始论文中默认的超参数设置。此外, 部分基线模型的实验结果来源于先前研究^[14], 以确保对比的公平性和准确性。GIBformer 采用 Python 语言, 并基于 PyTorch 和 DGL 库 (Deep Graph Library) 实现。在所有数据集中, 实体和关系的嵌入维度 d 均统一设置为 200, 学习率设定为 0.001, 并使用 Adam 优化器进行训练。所有数据集的超参数 α 均设置为 0.3, 超参数 β 设置为 0.001, 嵌入维度 d 设为 200。图卷积网络的层数为 2, 每层的 dropout 率设为 0.2。对于 ICEWS18, ICEWS14, ICEWS05-15 和 GDELT 数据集, 最近历史长度 m 分别设置为 10, 9, 15 和 7。解码器的通道数设置为 50, 卷积核的大小设置为 4×3 。此外, 还设置了早停策略, 在所有数据集上, 将训练周期设置为 30 轮, 当验证集上的 MRR 指标连续三轮没有增长时, 结束模型训练。为减小随机种子引入的偏差, 所有实验均至少重复进行了 5 次, 并计算了相应 p 值。实验在配备 NVIDIA GeForce RTX 3090 GPU 和 Intel i7-13700KF CPU 的机器上进行。

5.4 模型总体性能比较

实体预测任务的结果如表 2 和表 3 所列, 最好结果以加粗表示, 次优结果以下划线表示; “OOM” 表示由于计算资源限制未能完成相关实验。

表 2 在 ICEWS18 和 ICEWS14 数据集上的对比实验结果(实体预测)

Table 2 Comparative experimental results on datasets ICEWS18 and ICEWS14 datasets(entity prediction)

Model	ICEWS18				ICEWS14			
	MRR	Hit@1	Hit@3	Hit@10	MRR	Hit@1	Hit@3	Hit@10
R-GCN	16.40	9.68	18.16	30.52	29.52	21.11	33.52	46.51
ConvE	24.48	15.20	27.33	43.06	31.76	22.99	36.07	49.23
TTransE	9.91	3.26	10.62	23.92	14.19	4.56	17.38	34.97
TANGO	28.46	18.98	32.41	47.06	37.73	27.91	41.83	56.29
TLogic	29.85	20.02	33.76	48.87	43.06	33.46	48.31	61.74
RE-GCN	32.62	22.39	36.79	52.68	42.00	31.63	47.20	61.65
xERTE	29.31	21.03	33.51	46.48	40.79	32.70	45.67	57.30
CyGNet	27.12	17.21	30.97	46.85	37.65	27.43	42.63	57.90
RE-NET	29.78	19.73	32.55	48.46	39.86	30.11	44.02	58.21
TiGRN	33.55	<u>23.05</u>	<u>37.94</u>	<u>54.10</u>	<u>44.04</u>	<u>33.83</u>	<u>48.95</u>	<u>63.84</u>
RETIA	32.10	21.96	36.18	51.95	43.00	32.47	48.01	63.64
SiMFy	28.65	18.99	32.45	47.62	39.54	29.56	44.56	59.18
GIBformer	34.07	23.47	38.32	54.59	44.43	34.33	49.50	64.36
p -val.	2.55×10^{-4}	4.46×10^{-4}	5.48×10^{-4}	5.32×10^{-4}	2.24×10^{-4}	6.18×10^{-4}	6.47×10^{-4}	5.19×10^{-4}

基于这些实验结果, 有以下发现。

所有外推型时序知识图谱推理模型的性能均优于静态时序推理模型(如 R-GCN 和 ConvE)。这一结果验证了建模时间信息对于时序知识图谱推理任务的重要性, 忽视时间信息会限制模型对事件动态演化过程的建模能力。此外, 插值型时序知识图谱推理模型在所有数据集上的表现普遍较差, 主要原因在于这些模型在设计时未能有效处理未来时间戳的信

息, 所以无法捕捉跨时间的动态依赖关系。

实验结果显示, GIBformer 在 4 个基准数据集上的表现均优于其他所有外推型时序知识图谱推理模型。其优异性主要得益于, 信息瓶颈理论的引导使得模型能够专注于时序知识图谱快照中对推理任务更为关键的子图结构信息, 并有效压制那些与任务无关的冗余事件交互和噪声信息, 从而学习到更加准确且鲁棒的实体和关系表示。此外, GIB-

former 不仅能够捕捉事件演化的时序顺序模式,还使用 Transformer 架构建模了全局性的时序依赖,进一步增强

了模型对复杂事件时序演化模式的建模能力,并显著提升了模型的性能。

表 3 在 ICEWS05-15 和 GDELT 数据集上的对比实验结果(实体预测)

Table 3 Comparative experimental resultson datasets ICEWS05-15 and GDELT datasets(entity prediction)

Model	ICEWS05-15				GDELT			
	MRR	Hit@1	Hit@3	Hit@10	MRR	Hit@1	Hit@3	Hit@10
R-GCN	28.46	20.17	32.02	44.59	13.67	8.78	13.88	22.03
ConvE	32.81	23.08	37.35	52.29	19.99	12.85	20.67	33.54
TTransE	18.19	7.01	22.42	40.90	6.93	1.81	6.61	16.83
TANGO	42.62	31.94	48.53	64.09	20.17	13.10	21.57	33.97
TLogic	44.69	33.45	50.39	65.57	20.76	13.28	22.78	36.79
RE-GCN	48.03	37.33	53.90	68.51	19.69	12.46	20.93	33.81
xERTE	46.62	37.84	52.31	63.92	19.45	11.92	20.84	34.18
CyGNet	40.42	29.44	46.06	61.60	20.22	12.35	21.66	35.82
RE-NET	43.67	33.55	48.83	62.72	19.55	12.28	20.80	34.00
TiRGN	<u>49.47</u>	<u>38.56</u>	<u>55.58</u>	<u>70.71</u>	<u>21.67</u>	<u>13.63</u>	<u>23.27</u>	<u>37.60</u>
RETIA	44.37	34.02	49.64	64.42	OOM	OOM	OOM	OOM
SIMFy	41.83	31.30	47.02	62.11	19.21	11.87	20.50	33.41
GIBformer	50.03	39.12	56.08	71.27	21.97	13.92	23.55	38.08
p -val.	1.67×10^{-5}	3.28×10^{-4}	5.56×10^{-4}	6.19×10^{-4}	5.28×10^{-4}	5.16×10^{-4}	5.68×10^{-4}	2.53×10^{-4}

关系预测的结果如表 4 所列。由于部分基线模型不支持关系预测任务,因此本节仅选取支持该任务的基线模型进行对比实验。结果显示,GIBformer 在 4 个数据集上的表现均显著优于其他对比模型,展现出更强的关系建模能力。这主要是由于所提出的 GIB 卷积编码模块通过信息压缩与选择性保留机制,引导模型在实体和关系的表示中自动聚焦于对下游关系预测任务更关键的历史交互信号,从而提升了关系建模的准确性。相比之下,部分传统模型在面对复杂或高动态性的图结构时,难以识别与目标关系高度相关的上下文信息,导致性能受限。

为了在统计层面验证实验结果,对 GIBformer 以及表现最优的基线模型分别进行了 5 次重复训练,并计算了 p 值。实验结果表明,较低的 p 值进一步证实了 GIBformer 的有效性:其通过引入图信息瓶颈与 Transformer 结构,不仅能够保留与下游预测任务高度相关的关键信息,有效抑制噪声与冗余信息的干扰,还能捕捉跨图谱快照的全局时序依赖模式,从而有效地提升对缺失事件知识的预测准确性。

表 4 关系预测的结果对比(MRR)

Table 4 Results comparison of relation prediction(MRR)

Model	ICE14	ICE18	ICE05	GDELT
ConvE	38.80	37.73	37.89	18.84
ConvTransE	38.40	38.40	38.26	18.97
R-GCN	38.04	37.14	38.37	18.58
RE-GCN	39.73	39.48	38.56	19.17
TiRGN	<u>46.54</u>	<u>46.45</u>	<u>44.25</u>	<u>21.37</u>
GIBformer	47.19	46.92	45.32	22.14
p -val.	6.22×10^{-5}	1.82×10^{-4}	6.24×10^{-4}	3.61×10^{-4}

5.5 消融实验

为了验证 GIBformer 中各个关键组件的有效性,本节在 ICEWS14 数据集上进行消融实验。实验结果如表 5 所列,其中,“w/o GIB”表示去除 GIB 卷积编码模块,“w/o TF”表示去除 Transformer 模块,“w/o RH”表示去除重复事件历史挖掘模块,“w/o TE”表示去掉 Time2Vec 时间信息编码函数。

表 5 消融实验结果

Table 5 Ablation experiment results

Model	MRR	Hit@1	Hit@3	Hit@10
w/o GIB	44.12	33.76	48.79	63.65
w/o TF	44.09	33.71	49.33	63.75
w/o RH	41.26	31.15	45.99	60.93
w/o TE	43.61	32.98	48.83	63.62
GIBformer	44.43	34.33	49.50	64.36

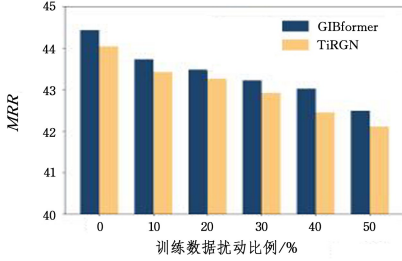
从结果中可以看出,模型设计中的每个关键组件对模型性能均有显著影响。具体而言,GIB 卷积编码模块通过在实体与关系表示学习过程中,显式压制知识图谱快照中的噪声和冗余结构依赖,使得模型能够更集中地关注与推理任务相关的子图结构信息。与去除该模块的消融实验相比,完整模型在 Hits@1 和 MRR 指标上均有明显提升,表明信息压缩机制对于提升表示学习质量和推理准确性至关重要。Transformer 模块在捕捉时序演化顺序模式的基础上,强化了快照之间的全局时序依赖建模能力。消融实验中去除 Transformer 模块后,模型对未来事件复杂演化关系的预测性能显著下降,说明该模块对于全局时序依赖的建模效果至关重要。重复事件历史挖掘模块通过检索当前查询的潜在历史重复事件,使得模型能够有效挖掘事件发生的周期性与重复性规律。相比于移除该模块的基线,保留该模块可显著提升周期性事件预测的准确率,验证了历史重复线索在增强实体预测中的重要作用。Time2Vec 时间信息编码函数通过统一建模周期性与非周期性时间信息,增强了模型对多样化时序模式的捕捉能力。消融该模块后,模型在涉及显著时间周期性特征的数据集(如 ICEWS14)上的性能下降,说明时间向量编码对于周期性时序特征理解具有重要贡献。

5.6 鲁棒性分析

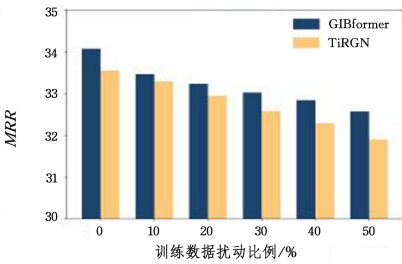
为进一步分析 GIBformer 在面对高噪声和极端稀疏数据环境中的性能表现,本节设计了在不同噪声干扰强度和不同数据稀疏度下,GIBformer 与最优基线模型 TiRGN 的性能对比实验,旨在全面评估模型在复杂数据条件下的鲁棒性。

如图 2 所示,分别在训练集中随机抽取 10%,20%,

30%, 40%, 50% 的样本并添加随机噪声, 以模拟不同强度的噪声环境。实验结果表明, 在各级噪声强度下, GIBformer 均显著优于 TiRGN。即使在极端噪声条件(50% 的训练数据被扰动)下, GIBformer 仍展现出更优的性能。这得益于所提出的 GIB 卷积编码模块在实体与关系表示学习过程中, 能够显式抑制噪声与冗余图谱信息的干扰, 同时增强对与推理密切相关的历史交互信息的关注, 从而提升了表示学习的鲁棒性。



(a) ICDWS14

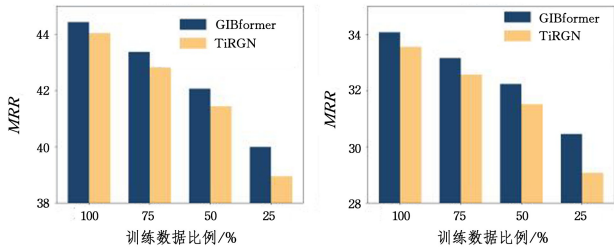


(a) ICDWS18

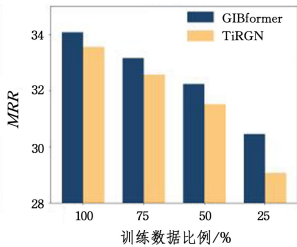
图 2 不同噪声环境下模型性能对比

Fig. 2 Comparison of model performance under different noise environments

为模拟极端数据稀疏环境, 从训练集中随机抽取 25%, 50%, 75% 和 100% 的样本用于模型训练。实验结果如图 3 所示, 在不同数据稀疏度条件下, GIBformer 的整体性能始终优于 TiRGN。这一优势主要得益于所提出的 GIB 卷积编码模块对关键结构信息的有效建模。即使在训练数据极度稀疏的情形下, 该模块仍能够成功捕获与推理任务密切相关的关键交互信号, 从而增强模型的推理精度与鲁棒性。



(a) ICDWS14



(b) ICDWS18

图 3 不同数据稀疏环境下模型性能对比

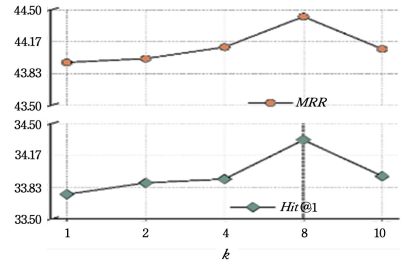
Fig. 3 Comparison of model performance under different data sparse environments

5.7 超参数分析

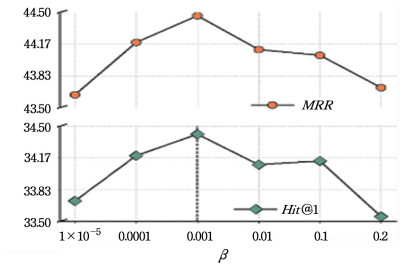
为评估关键超参数对模型性能的影响, 本节在 ICEWS14 数据集上对 Transformer 多头注意力数量 k 以及信息压缩损失项权重 β 进行了超参数敏感性分析, 实验结果如图 4 所示。

随着多头注意力头数 k 的增加, 模型能够捕获更为丰富

和全面的事件时序演化模式。当 $k=8$ 时, 模型性能达到最优, 说明此时注意力机制在建模全局时序依赖方面最为有效。然而, 进一步增大 k 值可能会引入冗余信息, 导致模型过拟合, 从而造成性能下降。对于信息压缩损失项权重 β , 较大的数值虽然可以增强模型对知识图谱快照中冗余信息的压缩能力, 但同时也可能削弱模型对实体预测任务的学习。实验结果表明, 当 $\beta=0.001$ 时, 模型在信息压缩与主任务优化之间达到较好平衡, 从而取得了最优的预测性能。



(a)



(b)

图 4 超参数敏感性实验结果

Fig. 4 Experimental results of hyperparameter sensitivity

5.8 案例研究

为深入分析模型中两个关键设计——GIB 卷积编码模块与 Transformer 模块对实体预测任务的具体影响, 本文在 ICEWS14 测试集中随机选取了 2 个查询, 并分别使用完整的 GIBformer 模型及其两个变体模型(去除 GIB 卷积编码模块、去除 Transformer 模块)进行预测, 实验结果如表 6 所列。其中, “w/o GIB” 表示移除 GIB 卷积编码模块后的变体模型, “w/o TF” 表示移除 Transformer 模块后的变体模型。分别提取了每个模型给出的 Top-5 实体预测结果, 并分析不同模型之间的表现差异。

在查询(?, Sign formal agreement, Peru) 中, GIBformer 成功预测出正确的缺失主体。在移除 GIB 卷积编码模块后, 模型明显受到来自前一时间步 $t-1$ 的图谱快照中无关实体的交互信息的影响, 尽管最终预测结果仍然正确, 但 Top-5 中其他候选实体大多为历史图谱快照中的冗余实体, 表明信息冗余和噪声信息对预测排名产生了影响。而移除 Transformer 模块后, 模型对复杂事件时序模式的建模能力明显减弱, 更偏向于事件的重复历史演化模式, 导致在历史交互频繁但与当前不相关的实体(如 Colombia) 被排在首位, 从而出现预测偏差。

在预测查询(Peru, Demand change in institutions, ?) 时, 完整的 GIBformer 准确预测出正确答案 Ecuador。移除 GIB 卷积编码模块后, 模型容易受到近期历史知识图谱快照中无

关冗余信息和噪声的干扰,导致正确实体仅排第 5 位,且排名 Top-4 的候选实体多为与当前事件无关的噪声实体。移除 Transformer 模块后,模型对远距离周期性等复杂时序模式的建模能力进一步下降,过于关注近期频繁交互的实体,最终导致预测结果偏离。

综上所述,GIB 卷积编码模块能够有效压制历史图谱快

照中的冗余结构信息与噪声信息,有助于模型提取关键历史子图结构,并关注与实体预测任务高度相关的事件交互信息。同时,Transformer 模块所建模的快照间全局时序特征显著提升了模型捕捉远距离周期性等复杂事件演化模式的能力,使模型能够跳出局部时序顺序依赖的限制,更全面地理解复杂事件的演化时序规律,从而提高预测准确性。

表 6 案例研究实验结果

Table 6 Case study experiment results

T-4	T-3	T-2	T-1	T	Answer	Ours	w/o GIB	w/o Former
(Peru, Express intention to meet, Bolivia) (Peru, Call for diplomatic cooperation, Bolivia) (Peru, Express intention to meet, Colombia) (Colombia, Requirement, Peru) (Peru, Make a visit, Ecuador) (Peru, Requirement, Ecuador)	(Peru, Express intent to meet or negotiate, Bolivia) (Colombia, Participate in negotiation, Peru) (Peru, Participate in negotiation, Colombia) (Colombia, Demand change in institutions, Peru) (Peru, Express intent to meet or negotiate, Ecuador) (Bolivia, Engage in negotiation, Uruguay)	(Bolivia, Host a visit, Peru) (Peru, Sign formal agreement, Colombia) (Peru, Make statement, Colombia) (Colombia, Sign formal agreement, Peru) (Uruguay, Engage in negotiation, Venezuela) (Peru, Consult, Chile)	(Peru, Engage in negotiation, Bolivia) (Bolivia, Host a visit, Peru) (Peru, Make optimistic comment, Bolivia) (Peru, Consult, Ecuador) (Uruguay, Engage in negotiation, Venezuela) (Venezuela, Express intent to meet or negotiate, Peru)	(?, Sign formal agreement, Peru)	Bolivia	Bolivia Colombia Ecuador Chile Uruguay	Bolivia Venezuela Uruguay Brazil Honduras	Colombia Bolivia Ecuador Chile Venezuela
				(Peru, Demand change in institutions, ?)	Ecuador Bolivia Colombia Venezuela Chile	Bolivia Venezuela Nicaragua Ecuador Brazil	Bolivia Venezuela Ecuador	Bolivia Chile Uruguay Venezuela Ecuador

结束语 本文提出了一种融合图信息瓶颈理论与 Transformer 的时序知识图谱推理框架 GIBformer。该框架首次将信息瓶颈理论引入时序知识图谱推理任务中,通过对图谱快照中的冗余结构信息与噪声交互信息进行有效压缩,从而增强了与实体预测任务高度相关的子图结构表示能力。此外,本文进一步设计了全局-局部联合建模机制,利用 Transformer 架构捕捉知识图谱中复杂事件的全局时序演化依赖关系,显著提升了模型对复杂动态演化模式的建模能力。在多个主流基准数据集上的实验结果充分验证了本文方法的有效性。GIBformer 在建模实体与关系的动态演变方面展现出优异性能,具有较好的应用潜力,适用于金融风控、推荐系统、智能问答等时序推理场景。然而,本文方法在面对极端稀疏或高噪声数据时推理精度仍受限,且在大规模知识图谱上的计算开销较高。未来的研究将进一步引入稀疏注意力机制、对抗训练等技术,进一步提升模型的鲁棒性与扩展性。

参考文献

[1] CHEN J, CHEN J, YU Z. Incorporating structured common-sense knowledge in story completion[C]// Proceedings of the AAAI Conference on Artificial Intelligence, 2019: 6244-6251.

[2] WU S, SUN F, ZHANG W, et al. Graph neural networks in recommender systems: a survey[J]. ACM Computing Surveys, 2022, 55(5): 1-37.

[3] CHEN B, SHEN H, CHENG Z, et al. Unveiling cross-modal consistency: Taming inter-and intra-modal noise for robust multi-modal knowledge graph completion[J]. Information Processing & Management, 2026, 63(2): 104472.

[4] CHEN B, YANG Y, CHENG Z, et al. Semantic duality in hypergraphs: Uncertainty-aware bipolar evidence aggregation for temporal knowledge graph reasoning[J]. Expert Systems with Ap-

plications, 2026, 299: 130162.

[5] CHEN M Q, ZHANG C L, YU H. Temporal knowledge graph reasoning model incorporating cross-time commonality features [J]. Journal of Chongqing University of Posts and Telecommunications (Natural Science Edition), 2025, 37(5): 708-716.

[6] SCHLICHTKRULL M, KIPF T N, BLOEM P, et al. Modeling relational data with graph convolutional networks[C]// The Semantic Web: 15th International Conference. Springer, 2018: 593-607.

[7] LIANG K, MENG L, LIU M, et al. A survey of knowledge graph reasoning on graph types: Static, dynamic, and multi-modal[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024, 46(12): 9456-9478.

[8] CHEN B, WU J, LIU X, et al. Enhancing Temporal Knowledge Graph for Future Event Prediction with Long-Term Dense Graph[J]. Tsinghua Science and Technology, 2026, 31(1): 621-638.

[9] JIN W, QU M, JIN X, et al. Recurrent Event Network: Autoregressive Structure Inference over Temporal Knowledge Graphs [C]// Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2020: 6669-6683.

[10] LIU Y, MA Y, HILDEBRANDT M, et al. Tlogic: Temporal logical rules for explainable link forecasting on temporal knowledge graphs[C]// Proceedings of the AAAI Conference on Artificial Intelligence, 2022: 4120-4127.

[11] TISHBY N, PEREIRA F C, BIALEK W. The information bottleneck method[EB/OL]. https://www.princeton.edu/~wbialek/our_papers/tishby+al_99.pdf.

[12] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]// Proceedings of the 31st International Conference on Neural Information Processing Systems, 2017: 6000-6010.

- [13] CHEN B, XIAO C, ZHOU F. Natural evolution-based dual-level aggregation for temporal knowledge graph reasoning[C]// Findings of the association for computational linguistics: EMNLP 2024. 2024;9274-9284.
- [14] LEBLAY J, CHEKOL M W. Deriving validity time in knowledge graph[C]// Companion Proceedings of the Web Conference 2018. 2018;1771-1776.
- [15] BORDES A, USUNIER N, GARCIA-DURAN A, et al. Translating embeddings for modeling multi-relational data[C]// Proceedings of the 27th International Conference on Neural Information Processing Systems. 2013;2785-2795.
- [16] CHEN K, WANG Y, LI Y, et al. Rotateqvs: Representing temporal information as rotations in quaternion vector space for temporal knowledge graph completion[J]. arXiv:2203.07993, 2022.
- [17] DASGUPTA S S, RAY S N, TALUKDAR P. HYTE: Hyperplane-based temporally aware knowledge graph embedding[C]// Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. 2018;2001-2011.
- [18] WANG J, WANG B, GAO J, et al. QDN: A quadruplet distributor network for temporal knowledge graph completion[J]. IEEE Transactions on Neural Networks and Learning Systems, 2024, 35(10):14018-14030.
- [19] LI Z, JIN X, LI W, et al. Temporal knowledge graph reasoning based on evolutionary representation learning[C]// Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval. 2021;408-417.
- [20] DENG S, RANGWALA H, NING Y. Dynamic knowledge graph based multi-event forecasting[C]// Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2020;1585-1595.
- [21] LI Y, SUN S, ZHAO J. TiRGN: Time-Guided Recurrent Graph Network with Local-Global Historical Patterns for Temporal Knowledge Graph Reasoning[C]// IJCAI. 2022;2152-2158.
- [22] YANG L, WU F, ZHENG Z, et al. Heterogeneous Graph Information Bottleneck[C]// IJCAI. 2021;1638-1645.
- [23] XU Y, OU J, XU H, et al. Temporal knowledge graph reasoning with historical contrastive learning[C]// Proceedings of the AAAI Conference on Artificial Intelligence. 2023;4765-4773.
- [24] HAN Z, CHEN P, MA Y, et al. Explainable subgraph reasoning for forecasting on temporal knowledge graphs[C]// International Conference on Learning Representations. 2020.
- [25] TISHBY N, ZASLAVSKY N. Deep learning and the information bottleneck principle[C]// 2015 IEEE Information Theory Workshop (ITW). IEEE, 2015;1-5.
- [26] SUN Q, LI J, PENG H, et al. Graph structure learning with variational information bottleneck[C]// Proceedings of the AAAI Conference on Artificial Intelligence. 2022;4165-4174.
- [27] YU J, XU T, RONG Y, et al. Recognizing predictive substructures with subgraph information bottleneck[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 46(3):1650-1663.
- [28] WU T, REN H, LI P, et al. Graph information bottleneck[J]. Advances in Neural Information Processing Systems, 2020, 33:20437-20448.
- [29] YUAN H, SUN Q, FU X, et al. Dynamic graph information bottleneck[C]// Proceedings of the ACM Web Conference 2024. 2024;469-480.
- [30] HU Z, GUTIÉRREZ-BASULTO V, XIANG Z, et al. Transformer-based entity typing in knowledge graphs[J]. arXiv:2210.11151, 2022.
- [31] XU Z, YE P, CHEN H, et al. Ruleformer: Context-aware rule mining over knowledge graph[C]// Proceedings of the 29th International Conference on Computational Linguistics. 2022;2551-2560.
- [32] HAN Z, LIAO R, LIU B, et al. Enhanced temporal knowledge embeddings with contextualized language representations[C]// ICLR. 2022.
- [33] SUN H, GENG S, ZHONG J, et al. Graph hawkes transformer for extrapolated reasoning on temporal knowledge graphs[C]// Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing. 2022;7481-7493.
- [34] GAO Y, HE Y, KAN Z, et al. Learning joint structural and temporal contextualized knowledge embeddings for temporal knowledge graph completion[C]// Findings of the Association for Computational Linguistics: ACL 2023. 2023;417-430.
- [35] JANG E, GU S, POOLE B. Categorical reparameterization with gumbel-softmax[J]. arXiv:1611.01144, 2016.
- [36] ZHU C, CHEN M, FAN C, et al. Learning from history: Modeling temporal knowledge graphs with sequential copy-generation networks[C]// Proceedings of the AAAI Conference on Artificial Intelligence. 2021;4732-4740.
- [37] KAZEMI S M, GOEL R, EGHBALI S, et al. Time2vec: Learning a vector representation of time[J]. arXiv:1907.05321, 2019.
- [38] SHANG C, TANG Y, HUANG J, et al. End-to-end structure-aware convolutional networks for knowledge base completion[C]// Proceedings of the AAAI Conference on Artificial Intelligence. 2019;3060-3067.
- [39] GARCIA-DURAN A, DUMANČIĆ S, NIEPERT M. Learning Sequence Encoders for Temporal Knowledge Graph Completion[C]// Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. 2018;4816-4821.
- [40] LEETARU K, SCHRODT P A. Gdelt: Global data on events, location, and tone, 1979-2012[C]// ISA Annual Convention. Citeseer, 2013;1-49.
- [41] LI Z, JIN X, LI W, et al. Temporal knowledge graph reasoning based on evolutionary representation learning[C]// Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval. 2021;408-417.
- [42] LI Z, GUAN S, JIN X, et al. Complex Evolutional Pattern Learning for Temporal Knowledge Graph Reasoning[C]// Pro-

ceedings of the 60th Annual Meeting of the Association for Computational Linguistics, 2022;290-296.

[43] DETTMERS T, MINERVINI P, STENETORP P, et al. Convolutional 2d knowledge graph embeddings[C]//Proceedings of the AAAI Conference on Artificial Intelligence, 2018.

[44] JIANG T, LIU T, GE T, et al. Towards time-aware knowledge graph completion[C]//Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics, 2016: 1715-1724.

[45] HAN Z, DING Z, MA Y, et al. Learning neural ordinary equations for forecasting future links on temporal knowledge graphs [C]//Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, 2021;8352-8364.

[46] LIU K, ZHAO F, XU G, et al. RETIA: relation-entity twin-interact aggregation for temporal knowledge graph extrapolation [C]//2023 IEEE 39th International Conference on Data Engineering(ICDE). IEEE, 2023;1761-1774.

[47] LIU Z, TAN L, LI M, et al. SiMFy: A Simple Yet Effective Ap-

proach for Temporal Knowledge Graph Reasoning[C]// Findings of the Association for Computational Linguistics:EMNLP 2023, 2023;3825-3836.



XIN Yichen, born in 1981, Ph.D candidate. His main research interests include data mining and knowledge graph.



LI Ye, born in 1993, Ph.D, associate researcher. His main research interests include deep learning and spatio-temporal data mining.

(责任编辑:李亚辉)