

文章编号: 1003-0077(2026)04-0001-25

多元知识图谱链接预测研究综述

魏继耀^{1,2}, 官赛萍^{1,2}, 靳小龙^{1,2}, 郭嘉丰^{1,2}, 程学旗^{1,2}

(1. 中国科学院 计算技术研究所 智能算法安全全国重点实验室, 北京 100190;
2. 中国科学院大学 计算机科学与技术学院, 北京 100190)

摘要: 多元知识图谱(N-ary Knowledge Graph, NKG)是一种特殊的知识图谱,旨在更有效地存储现实世界中的复杂事实。与传统的知识图谱不同,NKG 允许一个事实包含两个以上的实体,简称多元事实。NKG 链接预测是为了预测 NKG 中事实的缺失元素,这对于补全 NKG 和进一步提高下游应用任务的效果至关重要。最近,这一领域受到越来越多的关注。该文首次对 NKG 链接预测研究进行了系统回顾。首先给出了 NKG 链接预测任务的定义,并对该任务的现有方法进行了分类:基于空间结构的方法、基于张量分解的方法以及基于神经网络的方法。然后,分析了每个类别中的方法的建模思路、优缺点。在方法综述的基础上,进一步整理了相关基准数据集,并对比了最先进方法在这些基准上的实验结果。接着,总结了 NKG 在生物医疗、商品推荐系统等领域的典型应用,展示了其广泛的实践价值。最后,讨论了 NKG 链接预测的未来发展方向。

关键词: 链接预测;多元知识图谱;多元事实

中图分类号: TP391

文献标识码: A

Link Prediction over N-ary Knowledge Graphs: A Survey

WEI Ji Yao^{1,2}, GUAN Sai Ping^{1,2}, JIN Xiao Long^{1,2}, GUO Jia Feng^{1,2}, CHENG Xue Qi^{1,2}

(1. State Key Laboratory of AI Safety, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China;
2. School of Computer Science and Technology, University of Chinese Academy of Sciences, Beijing 100190, China)

Abstract: N-ary Knowledge Graphs (NKGs) are a special type of knowledge graph designed to store complex real-world facts, allowing n-ary facts involving more than two entities. NKG link prediction aims to predict missing elements in n-ary facts, which is crucial for improving NKGs for downstream applications. This paper presents a systematic review of NKG link prediction research. We first define the NKG link prediction task and categorize existing approaches into three types: space structure-based methods, tensor decomposition-based methods, and neural network-based methods. Then, we analyze the modeling strategies, strengths, and limitations of methods in each category. We also summarize commonly used benchmark datasets with their construction and characteristics, and compare the performance of state-of-the-art models on these datasets. We finally highlight representative applications of NKGs in domains such as biomedical research and product recommendation systems, illustrating their practical value. We conclude by discussing future directions for NKG link prediction research.

Keywords: link prediction; N-ary knowledge graph; N-ary facts

0 引言

自 2012 年 Google 提出知识图谱(Knowledge Graph, KG)的概念以来,KG 受到了越来越多的关注^[1]。KG 存储大量事实,通常表达为三元组 (h, r, t) ,其中, h 和 t 代表头实体和尾实体, r 代表它们之间的特定关系,如爱因斯坦,获得,诺贝尔奖。多元知识图谱(N-ary

Knowledge Graph, NKG)是一种特殊类型的 KG,擅长存储复杂的事实,其中每个事实可以包含两个以上的实体,这类事实也被称为多元事实。多元事实可以表达为一个角色-值对的集合。例如,事实“爱因斯坦在苏黎世大学攻读物理学,并获得博士学位”可以表达为{人物:爱因斯坦,机构:苏黎世大学,专业:物理学,学位:博士}。相比于传统三元组,NKG 能够更全面地表达现实世界中的复杂事实。例如,在生物医学中,

收稿日期: 2025-03-12 定稿日期: 2025-07-29

基金项目: 国家自然科学基金(62441229);中国科学院先导 B 项目课题(XDB0680102)

药物-靶点-疾病等关系天然为多元事实,使用 NKG 能更完整地建模药物组合治疗的语义关系,从而提升新药发现和副作用预测的能力。多元事实在现实场景中广泛存在。例如,在知名的 Freebase 数据库中,超过三分之一的实体涉及多元事实^[2],且超过 61% 的事实类型属于多元事实类别^[3]。然而,与传统 KG 相似, NKG 在构建过程中不可避免地会出现不完整性^[4],即部分事实存在缺失元素。这种不完整性会影响下游应用任务的性能,如语义解析^[5]、信息检索^[6]、问答系统^[7]和推荐系统^[8]等。为了解决这一问题, NKG 链接预测被提出,用于预测 NKG 中事实的缺失元素,从而补全和丰富 NKG。

传统的 KG 链接预测方法仅支持以三元组形式编码事实。为了处理多元事实,这些方法通常使用虚拟实体将多元事实分解为多个三元组,例如,在 Freebase 中采用的复合值类型(Compound Value Type, CVT)实体,并分别对每个三元组进行编码^[9]。然而,相较于直接预测多元事实中的缺失元素,这种方法存在多方面的缺陷^[10]。首先,对于包含多个实体的事实预测,需要同时考虑多个三元组,而现有的三元组链接预测方法通常是独立建模单个三元组,并在每次预测时仅针对一个三元组进行推理,因此涉及多个三元组的推理变得更加困难。其次,在将多元事实分解为三元组的过程中,可能会导致结构信息的丢失^[10],从而影响链接预测的准确性。此外,引入的虚拟实体及其对应的三元组会增加模型的参数量,提高学习的难度,并可能引发潜在的不正确推理^[2]。因此,越来越多的研究者关注 NKG 链接预测。不同于传统的 KG 链接预测方法, NKG 链接预测方法直接建模多元事实,而不是将其分解为多个三元组。这类方法可以更高效地利用多元事实的完整信息,提高推理的准确性。从应用场景来看,现有方法可以分为四类,包括通用场景^[2]、时间序列场景^[11]、小样本场景^[12]和归纳场景^[13]。此外,从建模策略来看,这些方法可以分为基于空间嵌入建模的方法^[2]、基于张量分解的方法^[14]和基于神经网络的方法^[15]三类,分别从不同角度建模多元事实,提升对多元事实的表达能力和推理能力。

在 KG 链接预测领域,已有大量调研工作从不同类型的 KG 视角提供了系统性的概述,包括通用 KG 链接预测^[16-18]、时序 KG 链接预测^[19-20]、多模态 KG 链接预测^[21-22]和稀疏 KG 链接预测^[23-24]。这些调研系统梳理了各自领域的发展历程、主要方法及应用场景。然而,针对 NKG 链接预测的专门调研尚未出现,尽管该领域近年来发展迅速,已提出近 50 种方

法。这些方法采用了空间结构、张量分解、神经网络等不同建模策略,并在生物医学、金融风控、推荐系统等多个应用场景中展现出优异性能。尽管 NKG 链接预测取得了显著进展,现有综述大多集中于传统的 KG 链接预测,仅在少量研究^[25]中零散提及 NKG 链接预测,但缺乏深入分析和系统整理。这些研究的共性是不能全面总结相关方法的发展脉络、关键技术挑战及潜在改进方向,导致研究人员难以把握该领域的整体趋势。因此,本文提供了一份亟需且至关重要的 NKG 链接预测调研,为学术界和工业界提供有价值的参考。

本文其余部分组织如下:第一部分介绍 NKG 链接预测的定义、多元事实的形式化表达及现有 NKG 链接预测方法的分类;第二部分介绍现有 NKG 链接预测方法,详细阐述其设计思想、优缺点,对不同类型的方法进行对比;第三部分介绍常用的 NKG 链接预测基准数据集,并对比最先进方法在这些基准数据集上的效果;第四部分介绍 NKG 链接预测在生物医疗、商品推荐等典型应用场景中的实践价值;第五部分探讨 NKG 链接预测的未来发展方向。

1 多元知识图谱链接预测简介

本节介绍 NKG 链接预测的定义、多元事实的形式化表达,以及现有方法的分类。

1.1 基本定义

知识图谱(KG): 传统知识图谱(KG)存储现实世界中的结构化事实,每个事实通常表达为一个三元组(h, r, t),其中, h 和 t 分别表示头实体和尾实体, r 表示它们之间的关系。例如,在一个 KG 中,事实“爱因斯坦获得诺贝尔奖”可以表达为(爱因斯坦,获得,诺贝尔奖)。

多元知识图谱(NKG): 相比于 KG,多元知识图谱(NKG)存储更复杂的事实,即每个事实可能包含两个以上的实体,这类事实被称为多元事实。例如,在一个 NKG 中,多元事实“爱因斯坦曾在苏黎世大学攻读物理学,并获得博士学位。”涉及四个实体,分别是爱因斯坦、苏黎世大学、物理学和博士。因此,该事实也被称为四元(4-ary)事实。多元事实的形式化表达将在 1.2 节中详细介绍。

多元知识图谱链接预测: 多元知识图谱链接预测的目标是根据多元知识图谱中已有的事实,预测事实中缺失的元素,从而补全知识图谱。例如,对于如下

不完整的多元事实{人物: 爱因斯坦, 机构: ?, 专业: 物理学, 学位: 博士}, NKG 链接预测任务就是预测其中缺失的机构, 苏黎世大学。

1.2 多元知识图谱适用范围

NKG 是对 KG 的一种泛化与扩展, 其优势在于更准确地表示涉及多个实体的复杂事实。当事实涉及三个及以上的核心实体时, 即可优先考虑采用 NKG, 以更准确地表达这类复杂结构的事实。为了进一步判断某一领域的知识是否特别适合使用 NKG 建模, 本文提出以下三个评估维度:

(1) 多方参与性: 当一个事实涉及三个及以上彼此存在语义联系的核心实体时, 推荐采用 NKG 建模以保持事实的完整性。事实中参与实体越多, 多方参与性越强。例如, “A、B、C、D 是大学同学” 相较于 “A、B、C 是大学同学” 具有更强的多方参与性, 更推荐使用 NKG 建模, 以保持事实的完整性。

(2) 语义耦合性: 若事实内实体之间存在紧密的语义关联, 无法被合理拆解为独立的二元关系, 否则会丢失语义, 则应使用 NKG。例如, “学生 A 于 B 年

在学校 C 获得奖学金 D” 中的各要素共同构成不可分割的语义整体, 若拆解为 (学生 A, 就读, 学校 C)、(学生 A, 获得, 奖学金 D) 等多个二元事实, 将无法准确表达原始语义。

(3) 上下文依赖性: 指事实的成立依赖于时间、地点、状态等限定条件, 这些上下文信息是构成事实语义不可或缺的部分, 此时推荐采用 NKG 建模以保证事实完整的上下文信息。例如, “爱因斯坦于 1921 年在瑞士获得诺贝尔物理学奖” 中的 “1921 年” 和 “瑞士” 是关键上下文, 若在建模中被忽略, 将影响事实的准确性。因此, 这类依赖上下文的事实更适合通过 NKG 统一建模。

通过以上评估维度对事实的结构特征进行分析, 有助于在 NKG 与传统 KG 之间做出合理的建模选择, 从而提升知识表示的完整性和下游推理任务的准确性。

1.3 多元事实形式化表达

多元事实的典型形式化表达主要包括超边式表达、角色-值对式表达和超关系式表达三种。图 1 展示了这些不同形式化表达方式的示例。

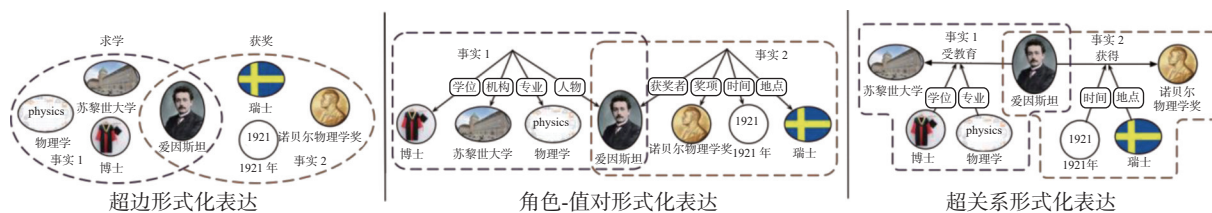


图 1 多元事实的三种形式化表达方式

超边式表达^[2]: 超边式表达将多元事实建模为超图 (Hypergraph) 中的超边, 每个超边连接一个多元事实中的所有实体, 即 $(H, e_1, \dots, e_n)$, 其中, H 表示超边, 连接事实中的每个实体; e_n 表示事实中的第 n 个实体。

例如, 图 1 中展示的事实可以表示如下:

事实 1: (求学, 爱因斯坦, 苏黎世大学, 物理学, 博士)

事实 2: (获奖, 爱因斯坦, 诺贝尔物理学奖, 1921 年, 瑞士)

在超边式表达中, 形式化后的事实中的实体是有顺序的, 且实体数量是固定的。超边中的每个位置对应一个固定的角色。该表达方式可以直接建立多元事实中多个实体之间的连接关系。当处理二元事实时, 超边式表达可以简化为传统 KG 中的三元组表示形式^[2]。然而, 其缺点在于, 所有实体的角色均被隐式地嵌入在超边中, 没有明确区分同一事实中各实体所

承担的具体角色^[10]。

角色-值对式表达^[10]: 角色-值对式表达则通过一组角色-值对表达多元事实, 即 $\{r_i: v_i\}_{i=1}^n$, 其中, r_i 表示事实中的角色, v_i 为该角色对应的实体, n 表示事实中实体的总数。

例如, 上述爱因斯坦受教育和获得诺贝尔物理学奖的事实可以分别表示如下:

事实 1: {人物: 爱因斯坦, 机构: 苏黎世大学, 专业: 物理学, 学位: 博士}

事实 2: {获奖者: 爱因斯坦, 奖项: 诺贝尔物理学奖, 时间: 1921 年, 地点: 瑞士}

在角色-值对式表示中, 一个事实内部的角色-值对是无序的, 并且可以包含任意数量的实体。这种表示方式在指定多元事实中各实体的角色方面具有较高的灵活性^[10]。然而, 它未能体现同一事实中不同实体的重要性或突出程度的差异^[26]。

超关系式表达^[26]: 超关系式表达在主三元组的基础上, 添加限定性角色-值对, 即 $(h, r, t), \{r_i: v_i\}_{i=1}^{n-2}$ 。其中, (h, r, t) 作为主三元组表示事实的核心信息, 而 $\{r_i: v_i\}_{i=1}^{n-2}$ 作为限定性信息, 用于提供补充的多元事实内容。

例如, 上述爱因斯坦受教育和获得诺贝尔奖的事实可以分别表示如下:

事实 1: ((爱因斯坦, 受教育, 苏黎世大学), {
-|专业: 物理学, 学位: 博士})

事实 2: ((爱因斯坦, 获得, 诺贝尔物理学奖), {
-|时间: 1921 年, 地点: 瑞士})

该方式在兼容传统 KG 结构的同时, 增强了对复杂多元事实的表达能力, 并能够突出事实中的核心实体, 即头实体和尾实体^[26]。但当一个事实中没有明确的主语(即头实体)或宾语(即尾实体)时, 使用超关系式表达并不合适^[15]。此外, 无论采用哪种表达方式, NKG 链接预测任务的目标始终是预测事实中缺失的元素。

这三种表达方式各有优劣, 超边式表达适用于直接建模多实体之间关系, 角色-值对式表达能更清晰地描述每个实体的具体角色, 而超关系式表达在兼容传统 KG 的同时, 增强了对重要实体的表达能力。在具体应用中, 可以根据任务需求选择适合的表达方式,

以提升多元事实推理的能力。

1.4 多元知识图谱链接预测方法分类

从应用场景的角度来看, 大多数 NKG 链接预测方法旨在通用环境下执行链接补全任务, 即在静态的稠密的 NKG 上进行推理。随着实际应用需求的多样化, 也有一些方法针对特定场景进行了优化, 如时序 NKG 链接预测、小样本 NKG 链接预测以及归纳式 NKG 链接预测。

时序 NKG 链接预测方法 主要聚焦于时间信息的建模, 适用于具有时间演化特性的数据, 能够更有效地捕捉多元事实随时间变化的动态趋势。

小样本 NKG 链接预测方法 致力于提升模型在训练样本有限(如小样本实体或关系)条件下的预测性能, 以适应低资源环境。

归纳式 NKG 链接预测方法 强调模型对训练集中未出现元素(如未见实体或关系)的推理能力, 通过利用已有知识实现更强的泛化性与迁移能力。

这些方法分别针对 NKG 链接预测中的不同关键挑战进行建模设计, 体现了对实际应用需求的深入理解与响应。基于应用场景的不同, 各类 NKG 链接预测方法的具体分类如表 1 所示。

表 1 NKG 链接预测方法应用场景分类

场景	子场景分类	方法
通用场景	—	m-TransH ^[2] , RAE ^[27] , HYPER2 ^[28] , PolygonE ^[29] , WPolygonE+ ^[30] , ShrinkE ^[31] , HYPERMONO ^[32] , m-Simple ^[3] , HypE ^[3] , GETD ^[14] , S2S ^[33] , NaLP ^[10] , t-NaLP ^[15] , NeuInfer ^[26] , HINGE ^[34] , s-HINGE ^[35] , RAM ^[36] , Enhance ^[37] , HyconvE ^[38] , HyCube ^[39] , HJE ^[4] , GRAN ^[40] , HyTransformer ^[41] , HyNT ^[42] , HIST ^[43] , NYLON ^[44] , StarE ^[45] , HAHE ^[46] , HyperFormer ^[47] , QUAD ^[48] , MSeaKG ^[49] , DHGE ^[50] , HELIOS ^[51] , HyperCL ^[52]
特殊场景	时序场景	NE-NET ^[11] , HypeTKG ^[53]
	小样本场景	HANCL ^[12] , MetaRH ^[54]
	归纳场景	QBLP ^[13] , MetaNIR ^[55]

注: 特殊应用场景可细分为时序、小样本和归纳三类: 时序场景下, 模型关注多元事实随时间演化的有效性建模; 小样本场景下, 模型旨在提升对训练样本稀缺元素的建模能力; 在归纳场景中, 模型需具备泛化到训练集中未出现元素的能力, 以应对开放世界下的预测任务。

从建模策略的角度来看, 当前的 NKG 链接预测方法主要分为三类: 基于空间嵌入建模的方法、基于张量分解的方法和基于神经网络的方法, 如表 2 所示。其中, 基于空间嵌入建模的方法通过在特定几何空间中建模多元事实中的元素(如角色、实体)之间的位置关系, 以此评估事实的可信度。根据所采用的几何空间不同, 这类方法可以进一步细分为基于欧几里得空间的方法、基于双曲空间的方法和基于复数空间的方法, 以适应不同的几何结构特征和事实建模需

求。例如, 基于双曲空间的方法适合表示具有层次性或树状结构的数据, 而基于复数空间的方法则有助于捕捉更复杂的旋转或对称关系。基于张量分解的方法则通过将所有多元事实组织成一个高阶张量(通常是三阶以上的张量), 其中每个张量单元对应一个事实的真实性。随后, 模型对该张量进行分解与参数优化, 旨在学习多元事实中的潜在模式与交互特征。由于基于张量分解的方法能够直接捕捉多实体之间的高阶交互信息, 近年来在提升链接预测准确性方面展

表 2 NKG 链接预测方法建模策略分类

方法类别	子方法分类	方法
基于空间嵌入建模的方法	基于欧几里得空间的方法	m-TransH ^[2] , RAE ^[27]
	基于双曲空间的方法	HYPER2 ^[28] , PolygonE ^[29] , WPolygonE+ ^[30]
	基于复数空间的方法	ShrinkE ^[31] , HYPERMONO ^[32]
基于张量分解的方法	—	m-Simple ^[3] , HypE ^[3] , GETD ^[14] , S2S ^[33]
基于神经网络的方法	基于FCN的方法	NaLP ^[10] , t-NaLP ^[15] , NeuInfer ^[26]
	基于CNN的方法	HINGE ^[34] , s-HINGE ^[35] , RAM ^[36] , EnhanceE ^[37] , HyconvE ^[38] , HyCubE ^[39] , HJE ^[4]
	基于GNN的方法	StarE ^[45] , HAHE ^[46] , HyperFormer ^[47] , QUAD ^[48] , MSeaKG ^[49] , DHGE ^[50] , HELIOS ^[51] , HyperCL ^[52] , NE-NET ^[11] , HypeTKG ^[53] , HANCL ^[12] , MetaRH ^[54] , QBLP ^[13] , MetaNIR ^[55]
	基于Transformer架构的方法	GRAN ^[40] , HyTransformer ^[41] , HyNT ^[42] , HIST ^[43] , NYLON ^[44]

现出良好潜力。基于神经网络的方法利用神经网络强大的表示学习能力,对多元事实中的元素及其关系进行深度编码,并基于这些编码来评估事实的有效性。该类方法已成为当前研究最为广泛和活跃的方向,并可进一步细分为:基于全连接神经网络(Full Connect Neural Networks, FCN)、基于卷积神经网络(Convolutional Neural Networks, CNN)、基于 Transformer 架构的方法以及基于图神经网络(Graph Neural Networks, GNN)的方法。不同的神经网络结构侧重于捕捉不同层次和类型的元素关联模式,例如 FCN 注重局部组合特征, CNN 擅长提取局部空间特征, Transformer 强调全局依赖关系, GNN 则专门建模复杂的结构化交互。

从事实的形式化表达方式来看,如前所述,当前的 NKG 链接预测方法主要分为三类:超边式表达、角色-值对式表达和超关系式表达。每种 NKG 链接预测

方法和其所采用的表达方式对应关系如表 3 所示。采用不同表达方式对 NKG 链接预测方法的设计具有直接影响。例如,基于超边式表达的 NKG 链接预测方法强调多实体之间整体结构关系的建模,擅长捕捉实体间复杂交互;基于角色-值对式的链接预测方法则以角色为中心,擅长捕捉不同角色对实体语义的影响;而基于超关系式的方法引入了对实体重要性的建模,能够更加准确地反映不同实体在事实中的贡献与作用差异。在现有研究中,超关系式表达应用最为广泛。这种表达方式不仅具备出色的多元事实建模能力,还能够兼容传统 KG 中的三元组事实,具有良好的灵活性与泛化性,因此在各类实际任务中表现出了良好的适应性与迁移能力。

总体而言,在具体应用场景中,研究者可根据任务类型、数据特性及推理需求,综合参考表 1~表 3 中的对比分析,灵活选择最适合的事实表达方式。

表 3 NKG 链接预测方法事实形式化表达方式分类

事实形式化表达方式	方法
超边式	m-TransH ^[2] , RAE ^[27] , m-Simple ^[3] , HypE ^[3] , GETD ^[14] , S2S ^[33] , RAM ^[36] , EnhanceE ^[37] , HyconvE ^[38] , HyCubE ^[39] , HJE ^[4]
角色-值对式	PolygonE ^[29] , NaLP ^[10] , t-NaLP ^[15] , NE-NET ^[11] , HypeTKG ^[53]
超关系式	HYPER2 ^[28] , WPolygonE+ ^[30] , HYPERMONO ^[32] , NeuInfer ^[26] , HINGE ^[34] , s-HINGE ^[35] , GRAN ^[40] , HyTransformer ^[41] , HyNT ^[42] , HIST ^[43] , NYLON ^[44] , StarE ^[45] , HAHE ^[46] , HyperFormer ^[47] , QUAD ^[48] , DHGE ^[50] , HELIOS ^[51] , HyperCL ^[52] , HANCL ^[12] , MetaRH ^[54]

1.5 NKG 链接预测方法与传统 KG 链接预测方法比较

传统 KG 与 NKG 在基本构成(实体、关系、事实)上具有一定共性,因此其链接预测方法采用空间嵌入建模、张量分解、神经网络等相似的技术路线。然而,二者在建模对象的结构和预测任务的定义上存在

显著差异,这些差异决定了 NKG 链接预测方法在结构设计和建模策略上必须进行相应的扩展和优化。具体而言,传统 KG 链接预测与 NKG 链接预测的主要差异体现在以下两个方面。

建模对象结构的差异 传统 KG 建模的是标准的

三元组事实 (h, r, t) , 而 NKG 则处理的是形式更加灵活的多元事实。例如, 对于超关系式的多元事实 $((h, r, t), \{r_i: v_i\}_{i=1}^{n-2})$, NKG 链接预测要求模型不仅能建模头实体 h 和尾实体 t 之间的关联关系 r , 还需处理限定性角色 r_i 、限定性值 v_i 之间的对应关系, 以及限定性角色-值对 $\{r_i: v_i\}_{i=1}^{n-2}$ 和主三元组 (h, r, t) 之间的相互作用等结构特征。多元事实内的这种复杂结构对模型的表达能力提出了更高的要求。

预测任务定义的差异 传统 KG 中的链接预测主要关注三元组中缺失实体的补全, 如 $(?, r, t)$ 、 $(h, r, ?)$ 或 $(h, ?, t)$; 而 NKG 中的链接预测则更加广泛和灵活, 不仅涉及实体、关系的缺失, 还包括限定性角色和值的缺失, 且常常需要同时补全多个缺失元素。这种任务设定要求模型具备更强的结构建模能力和灵活的推理机制, 以适应多种类型缺失的补全需求。

基于上述差异, 虽然 NKG 链接预测方法在技术分类上延续了传统 KG 的三大路线(空间嵌入建模、张量分解、神经网络), 但针对 NKG 的建模挑战均进行了针对性的改进。

例如, (1)基于空间嵌入建模的 NKG 链接预测方法引入角色特定的空间变换函数, 以捕捉实体在不同角色语境下的语义差异; (2)基于张量分解的 NKG 链接预测方法使用 Tucker 分解等工具处理更高维、结构更复杂的张量, 并解决嵌套结构和实体数不固定等问题; (3)基于神经网络的 NKG 链接预测方法强调对 NKG 中特有结构的建模能力, 如由事实间关联关系构成的超图结构, 或由事实内部元素构成的全连接图等。

此外, 面对更复杂的预测任务, 当前大多数 NKG 链接预测方法均支持对任意缺失元素的补全, 部分方法甚至能够同时预测多个缺失元素。例如, 基于神经网络的 HAHE 方法采用自回归编码器结合 MASK 机制, 有效支持对多元事实中多个缺失项的预测。

综上所述, 尽管 NKG 链接预测方法在技术路线上与传统 KG 方法存在相似性, 但其核心关注点和建模机制已发生根本性转变。因此 NKG 链接预测方法并非对传统 KG 方法的浅层延续, 而是在应对复杂事实与多样化预测需求的背景下, 对相关理论体系的深化与方法的创新发展。

2 多元知识图谱链接预测方法

本节介绍 NKG 链接预测方法。首先, 我们从建

模策略的角度出发, 在第 2.1~2.3 节分别介绍基于空间嵌入建模的方法、基于张量分解的方法和基于神经网络的方法, 重点分析其设计思路、优缺点。由于大多数方法针对通用场景设计, 而部分方法专为特定场景优化, 我们在第 2.4 节单独介绍面向特殊场景的 NKG 链接预测方法。最后, 在第 2.5 节对不同类型的方法进行横向对比, 以总结各类方法的特点与适用性。

2.1 基于空间嵌入建模的方法

此类方法通过定义事实中元素(如角色、实体)在特定空间中投影操作, 以建模并评估多元事实的可信度。根据建模空间的不同, 这些方法可进一步分为三类: 基于欧几里得空间的方法、基于双曲空间的方法和基于复数空间的方法。

2.1.1 基于欧几里得空间的方法

m-TransH^[2]是第一个用于 NKG 链接预测的方法。其核心动机在于, 传统 KG 建模方法通常假设所有事实都是二元的, 即每个事实仅涉及两个实体。然而, 在真实的知识库(如 Freebase)中, 存在大量多元事实, 例如“某演员在某部电影中扮演某个角色”。传统方法通常将多元事实拆解为多个三元组, 但这一过程会导致多元事实结构信息的丢失, 从而影响模型预测效果。为此, m-TransH 直接建模多元事实, 避免了三元组转换所带来的潜在结构信息损失。具体来说, 该方法使用超边形式表达多元事实, 通过一个超边连接多元事实中的所有实体。在该模型中, 每个超边类型 r 由两个正交的单位向量 $\mathbf{n}_r$ 和 $\mathbf{b}_r$ 以及一个加权函数 a_r 参数化。这些参数用于定义实体的投影空间, 并结合投影操作来衡量多元事实的合理性。具体而言, m-TransH 的事实评估函数 f_r 形式如式(1)所示。

$$f_r(t) = \left\| \sum_{\rho \in \mathcal{M}(R_r)} \mathbf{a}_r(\rho) \mathbb{P}_{\mathbf{n}_r}(t(\rho)) + \mathbf{b}_r \right\|^2 \quad (1)$$

其中, t 表示 R_r 类型超边的事实实例, $\mathcal{M}(R_r)$ 表示该事实中的角色集合, $\mathbf{a}_r(\rho)$ 是角色 ρ 相关的权重参数向量, $t(\rho)$ 表示实例 t 在角色 ρ 位置上的实体表示, $\mathbb{P}_{\mathbf{n}_r}()$ 表示投影层, $\mathbb{P}_{\mathbf{n}_r}(t(\rho))$ 表示 $t(\rho)$ 在超平面上的投影, $\mathbf{b}_r$ 是与 R_r 类型超边相关的超平面中的偏置向量。

RAE^[27]在 m-TransH 的基础上引入了多层感知机(Multilayer Perceptron, MLP), 加强了同一事实中实体之间的相似性约束, 从而提升了模型性能。

上述方法主要用于预测多元事实中单个缺失的实体, 但在实际应用中, 一个多元事实可能同时缺失

数个实体。因此, Rosso 等人^[56]提出了一个新任务, 实例重建, 即预测多元事实中的多个缺失的实体, 并开发了一个实例重建算法 RETA, 该方法通过滤掉包含低相关度实体的候选事实, 显著降低了计算复杂度。

2.1.2 基于双曲空间的方法

前述基于欧几里得空间的方法忽视了 NKG 中实体的层次结构。在 NKG 中, 实体的相关实体数量从低层次到高层次呈指数增长, 这一特性与双曲空间中的超线性长度增长相吻合, 如图 2 所示。因此, Yan 等人提出了基于双曲 Poincaré 嵌入的 HYPER2^[28]和 PolygonE^[29]方法, 通过利用双曲空间的独特性质, 捕捉 NKG 中实体之间的层次结构和复杂的关联关

系, 从而提升 NKG 链接预测的准确性。

HYPER2 采用超关系式表达, 将多元事实表达为一个核心主三元组和若干限定性角色-值对。其基于 Poincaré 球模型, 在双曲空间中初始化所有实体的嵌入, 使得层次化知识能够自然地映射到低维空间。考虑到 Poincaré 空间的计算通常较为复杂, HYPER2 通过将双曲空间中的嵌入投影到切空间(Tangent Space), 在切空间中执行信息聚合操作, 将限定性角色-值对的信息融合到头实体和尾实体的嵌入中。随后将更新后的实体嵌入投影回双曲空间, 并利用 Poincaré 距离计算头尾实体之间的关联性, 将其作为事实的可信度评分。

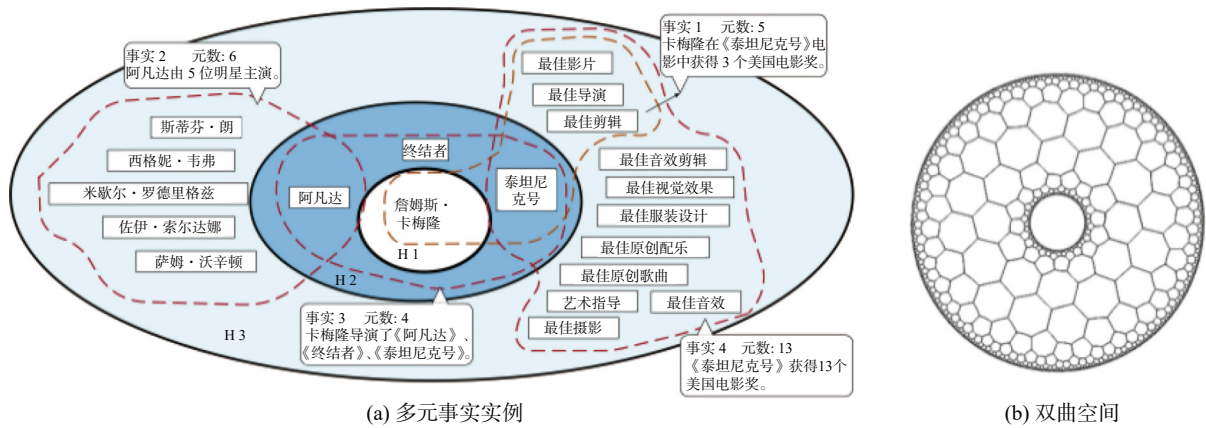


图2 Wikipedia 数据库中一些以电影导演詹姆斯·卡梅隆为中心的多元事实实例及双曲空间

为了对涉及的实体进行层次化, H1 代表顶层实体詹姆斯·卡梅隆, H2 列出了一些詹姆斯·卡梅隆的代表作(事实 1), H3 包含 H2 中电影的主演信息(事实 2)或获奖信息(事实 3 和事实 4)的实体。从 H1 到 H3, 涉及的实体数量呈指数级增长, 这与双曲空间中的超线性增长特性高度一致。而且, 事实的元数(即事实中的实体数量)越高, 其指数增长特性越明显。

与 HYPER2 不同, PolygonE 将多元事实建模为双曲空间中的陀螺多边形(gyro-polygons), 其中实体被表示为多边形的顶点, 而角色则对应于实体的平移或旋转操作。在 KG 嵌入模型中, 合理的事实通常意味着相关实体在低维空间中具有较小的距离。因此, PolygonE 通过优化基于顶点-陀螺旋心距离的事实合理性评估策略, 调整陀螺多边形的形状, 从而让相关实体更加紧密地分布在一起。这种调整增强了事实的可解释性, 同时避免了 HYPER2 方法进行多次空间投影可能带来的信息损失。

在 PolygonE 方法中, 陀螺多边形的旋心是基于所有实体等权计算得到的, 即假设多元事实中的每个实体在该事实中同等重要。然而, 在实际应用中, 实体在事实中的贡献往往不均衡。例如, 某些实体可能扮演核心角色, 而其他实体仅起辅助作用。因此, PolygonE 的等权假设可能会导致模型无法正确表达

不同实体的重要性, 从而影响对事实合理性的评估。为了解决这一问题, WPolygonE^[30]引入了一种无监督的实体权重学习方法, 根据每个实体与旋心的距离动态分配权重, 使更重要的实体在旋心计算中占据更大的比重。此外, WPolygonE 进一步优化了主要三元组对整体事实的描述能力, 约束主要三元组中心靠近完整多元事实的旋心, 利用该联系提升了模型对事实合理性的建模能力。

2.1.3 基于复数空间的方法

上述方法通常忽略了多元事实中的推理模式, 在知识表示过程中未能有效建模事实之间的逻辑关系。例如, 在 NKG 中, 存在单调性模式: 如果两个限定性角色-值对 q_i 和 q_j 满足 q_i 蕴含 q_j (即 q_i 的存在暗示了 q_j 的成立), 那么在任何事实中, 同时附加 q_i 和 q_j 时事实的真实性(真或假)和仅附加 q_i 应当保持一致。基于复数空间的方法在复数空间中建模多元事实, 擅

长于捕捉 NKG 中的推理模式,进而提升 NKG 链接预测的准确性。

ShrinkE^[31]将多元事实形式化表达为一个核心主三元组和若干限定性角色-值对。它采用几何变换的方式来处理这些多元事实,能够有效建模多元事实中的推理模式,如限定性角色-值对的单调性、蕴涵关系和互斥关系。具体来说,它通过将主三元组表达为从头实体到关系特定的盒嵌入(box embeddings)的空间映射变换,来对多元事实进行建模。每个限定性角色-值对通过“缩小”这个盒嵌入空间来缩小可能的答案集,从而实现限定性角色-值对单调性(即添加限定性角色-值对时,答案集只能变小或保持不变)。此外,限定性角色-值对之间的空间关系(如盒子的相对位置)被用来建模限定性角色-值对的蕴涵和互斥推理模式。通过这种方式,ShrinkE 能够显式地捕捉这些重要的推理模式,提高了模型的推理能力和泛化能力。

不同于 ShrinkE,为了实现限定性角色-值对单调性,HyperMono^[32]采用了锥体嵌入(cone embeddings)来表示限定性角色-值对的收缩效应。与传统的盒嵌入不同,锥体嵌入通过其有限的开口角度自然地表示一个有限的答案集及其子集。在每次添加限定性角色-值对时,锥体嵌入的角度逐渐减小,从而减少答案集的大小。相比于盒嵌入的边界偏移,锥体嵌入能更精确地捕捉到添加限定性角色-值对引起的答案集缩小。

上述方法通常采用图 1 中展示的常见的多元事实形式化表达方式,然而,少数研究则引入了嵌套事实的概念,以更灵活地表达多元事实。一个嵌套事实通常由两个原子事实组成,其中每个原子事实都是一个二元事实。例如,嵌套事实 $\langle T_1, \text{前提条件}, T_2 \rangle$ 包含两个原子事实,分别是 T_1 : (奥斯卡金像奖, 提名, 阿凡达)和 T_2 : (阿凡达, 获得, 奥斯卡金像奖)。其中 T_1 和 T_2 之间存在某种因果关系或时间先后顺序。这种表达方式相较于传统的多元事实表达方式更加灵活,能够捕捉事实间复杂的依赖关系,尤其适用于描述因果关系或条件关系。BiVE^[57]首次引入了嵌套事实的概念,并采用了复数空间的二元事实建模方法 QuatE^[58]、BiQUE^[58]对原子事实和嵌套事实进行编码。通过复数旋转和变换,该方法能够有效捕捉事实间的因果关系和其他逻辑联系。

最近, NestE^[59]在 BiVE 的基础上进行了扩展,采用了超复数嵌入,支持四维超复数空间的扩展,包括球形四元数、双曲四元数和分裂四元数等。不同类型的超复数系统具有不同的代数乘法规则,使其能够处

理更复杂的几何模式和结构,特别是在表示层次关系和复杂几何变换时。NestE 通过超复数矩阵建模嵌套关系,利用 Hamilton 乘积进行旋转,作用于原子事实矩阵。这种旋转模型能够捕捉事实之间的复杂关系,如时序依赖关系或逻辑关系。超复数矩阵在进行矩阵乘法的过程中,涉及超复数乘法,有效捕捉了原子事实之间的深层语义关系,支持复杂的逻辑模式(如霍恩规则),提升了模型对复杂推理模式的表达能力。

2.2 基于张量分解的方法

此类方法将 NKG 中的事实集合表达为高阶张量,其中每个张量单元对应一个事实的真实性。通过对该张量进行重构、分解与优化,使模型能够学习到多元事实中元素之间的潜在模式特征,以提升链接预测的准确性。

例如, m-Simple^[3]将多元事实表达为一个连接所有实体的超边。一个实体在不同事实中会表达不同的角色。例如,一个实体(如 Turing)可能在某些事实中扮演“学生”的角色,而在其他事实中则可能扮演“教授”的角色。因此, m-Simple 根据实体在事实中的位置特征来动态调整实体的表示,以便更准确地进行预测。具体来说,该方法根据实体在多元事实中的位置,利用“平移”操作调整嵌入空间中实体的嵌入,使其能够适应多元事实的结构。最终,模型将调整后的嵌入与超边嵌入结合来对多元事实进行评分。具体而言, m-Simple 的评分函数如式(2)所示。

$$\phi(r(e_i, e_j, \dots, e_k)) = \odot(r, e_i, \text{shift}(e_j, \frac{\text{len}(e_j)}{\delta}), \dots, \text{shift}(e_k, \frac{\text{len}(e_k) \cdot (\delta - 1)}{\delta}))) \quad (2)$$

其中,“ $\odot$ ”表示逐元素乘积的总和, r 表示超边 r 的嵌入, e_* 表示第 $*$ 个实体 e_* 的嵌入, $\text{shift}(a, x)$ 表示将向量 a 向左移动 x 步, $\text{len}(e)$ 表示实体嵌入向量 e 的长度, δ 是 NKG 中事实的最大元数。

与 m-Simple 不同, HypE^[3]使用卷积滤波器生成位置特定的实体嵌入,具有更强的灵活性。通过设计针对不同位置的滤波器对实体进行编码, HypE 提升了模型的表达能力。具体来说, HypE 的评分函数如式(3)所示。

$$\phi(r(e_i, e_j, \dots, e_k)) = \odot(r, f_i(e_i), \dots, f_k(e_k)) \quad (3)$$

其中, $f_*(e_*)$ 使用卷积核 f_* 生成 $*$ 位置的实体嵌入。

尽管上述基于张量分解的模型能够对多元事实进行建模,但随着涉及实体数量的增加,模型的参数

规模指数级增长, 容易导致计算复杂度激增和过拟合问题。为了解决这一问题, GETD^[14]结合了 Tucker^[60]分解和张量环(Tensor Ring, 简称 TR)分解, 有效减少了模型参数量并提升了计算效率。Tucker 分解通常用于二元事实建模, 例如, 在图 3(a)中, 通过分解核心张量 $\mathbf{W}$ 来评估二元事实的有效性。对于多元事实, $\mathbf{W}$ 需要扩展为高阶张量 $\widehat{\mathbf{W}}$, 以捕捉多元事实中超边与多个实体之间的交互。但随着嵌入维度的增加, 模型参数量呈指数级增长。因此, GETD 采用 TR 分解, 将高阶张量 $\widehat{\mathbf{W}}$ 重塑为 k 阶张量, 并进一步将其分解为 k 个低阶张量, 如图 3(b)所示, 大幅降低了计算复杂度。GETD 的评分函数如式(4)所示。

$$\phi(\mathbf{H}, \mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n) = \widehat{\mathbf{W}} \times_1 \mathbf{H} \times_2 \mathbf{e}_1 \times_3 \mathbf{e}_2 \cdots \times_{n+1} \mathbf{e}_n = \text{TR}(\mathbf{Z}_1, \dots, \mathbf{Z}_k) \times_1 \mathbf{H} \times_2 \mathbf{e}_1 \times_3 \mathbf{e}_2 \cdots \times_{n+1} \mathbf{e}_n \quad (4)$$

其中, $\widehat{\mathbf{W}}$ 表示高阶张量, $\text{TR}(\mathbf{Z}_1, \dots, \mathbf{Z}_k)$ 表示张量环分解, 将 $\widehat{\mathbf{W}}$ 分解为 k 个较低阶张量 $\{\mathbf{Z}_1, \dots, \mathbf{Z}_k\}$ 。

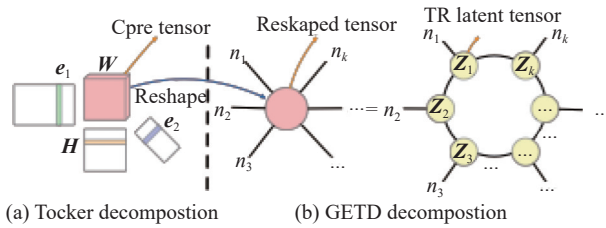


图 3 GETD 方法示意图

(a) Tucker 分解。 (b) TR 分解核心张量为多个较低阶张量, 进一步减少模型参数数量。

由于 GETD 不能同时处理不同元数的多元事实, 容易导致数据稀疏性问题。S2S^[33]通过引入嵌入共享机制来同时处理具有不同元数的多元事实数据, 有效减少数据稀疏性问题。该方法通过划分嵌入向量, 使得不同元数的事实能够共享不同维度的嵌入。此外, 张量分解模型面临的另一个挑战是过度参数化, 这会导致训练效率低下。S2S 采用神经架构搜索(NAS)算法^[61]优化核心张量, 减少冗余参数。这种方法不仅保持了模型的表达能力, 还提升了训练效率, 避免了过度参数化带来的训练开销。

2.3 基于神经网络的方法

此类方法利用神经网络对 NKG 进行编码, 然后基于编码得到特征向量评估多元事实的有效性。根据所采用的神经网络的类型不同, 这些方法大致可以分为基于 FCN、CNN、Transformer 和 GNN 四种类型。图 4 展示了每个类型中的代表性方法。

2.3.1 基于全连接网络的方法

NaLP^[10]将多元事实表达为一个角色-值对集合, 并使用 FCN 层从每一个角色-值对中提取特征。这些提取出来的特征随后被拼接并通过另一个 FCN 层处理, 最终得到事实的真实性评分。然而, NaLP 对所有角色-值对给予同等重要性, 未能充分考虑多元事实中不同实体的相对重要性和角色差异。因此, 该方法在处理具有不同重要性实体的多元事实时, 可能存在性能的局限。

为了解决这个问题, NeuInfer^[26]将多元事实形式化表达为一个主三元组, 并对任意数量的限定性角色-值对进行组合。该方法首先通过 FCN 模块分别计算主三元组的有效性评分和主三元组与限定性角色-值对之间的兼容性评分。接着, NeuInfer 将这两个评分融合, 得到事实的真实性评分, 如图 4(a)所示。NeuInfer 能够有效区分不同实体在多元事实中的作用, 并在一定程度上克服了 NaLP 的不足。

为了进一步提升 NaLP 的性能, t-NaLP^[15]提出了一种基于实体类型信息的负采样机制。具体而言, t-NaLP 通过引入实体类型信息, 选择与目标实体类型相似的实体作为负样本, 从而避免了传统负采样方法中可能导致的过于简单或不具代表性的负样本。通过这种方式, t-NaLP 增强了 NaLP 的性能, 进一步提高了模型在多元事实建模中的鲁棒性。

2.3.2 基于卷积神经网络的方法

与 NeuInfer 类似, HINGE^[34]将多元事实表达为一个核心主三元组, 并对任意数量的角色-值对进行组合。HINGE 利用 CNN 计算主三元组与每个角色-值对之间的特征向量, 并通过使用最小池化聚合这些特征, 最终通过一个 FCN 层对事实进行评分, 如图 4(b)所示。类似于 t-NaLP, s-HINGE^[35]通过结合实体类型信息选取更具代表性的负样本, 从而进一步提高了 HINGE 的性能。

RAM^[36]引入了潜在空间来增强角色的语义建模。通过为每个角色引入一个潜在空间, RAM 能够捕捉到不同角色之间的语义关系。例如, 在三元事实“PlayCharacterIn”(演员、角色、电影)中, RAM 能够将“演员”和“角色”视为相似的语义角色。除了角色的潜在空间外, RAM 还为每个角色学习一个模式矩阵, 用来捕捉角色与实体之间的兼容性。此外, RAM 设计了一个多线性评分函数, 该函数结合了潜在空间和模式矩阵的表示, 通过对实体和角色的向量进行多线性乘积, 来计算该事实的合理性得分, 如式(5)所示。

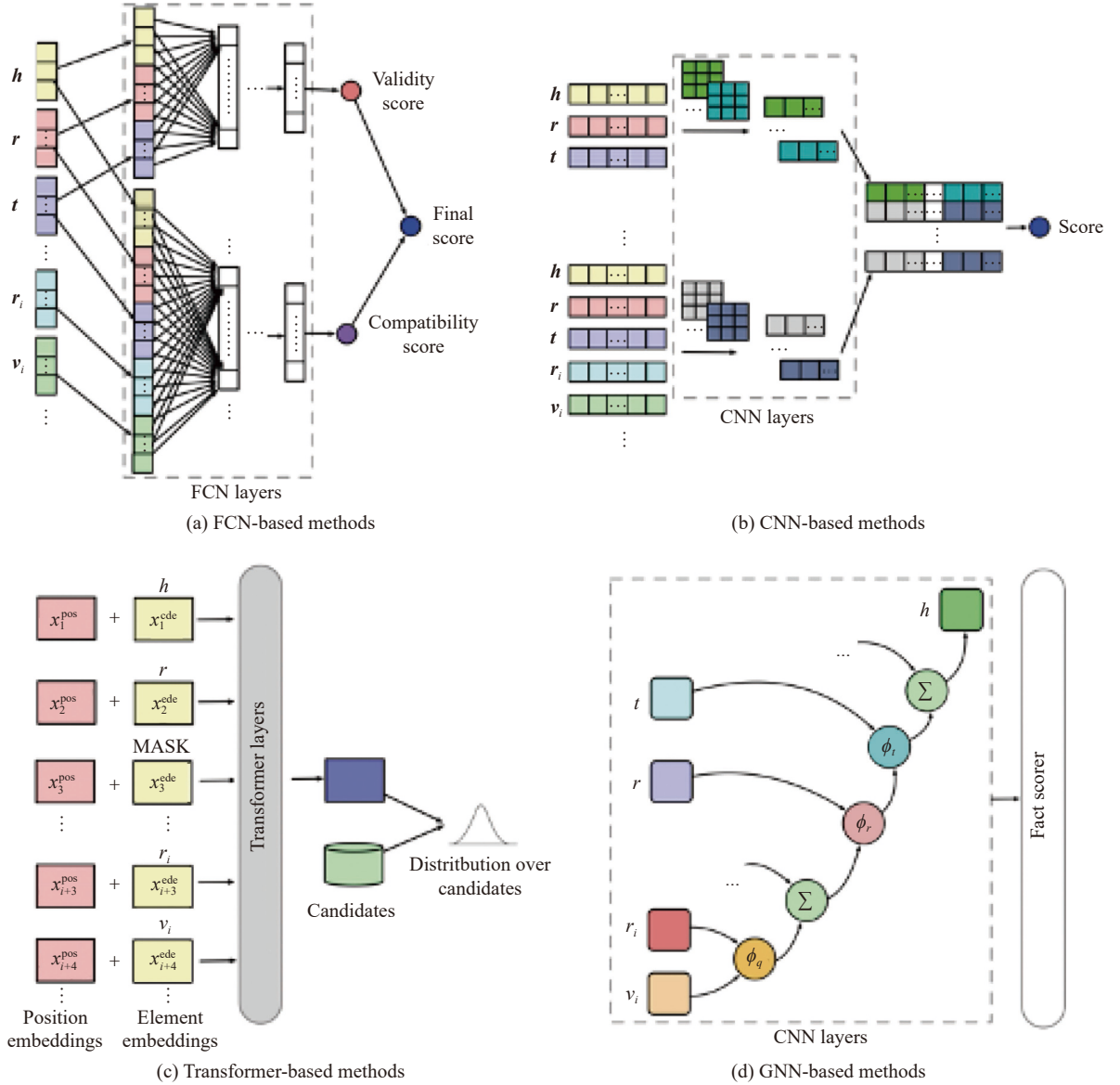


图4 不同类型的基于神经网络的方法示意图

(a) 基于 FCN 的方法 NeuInfer 通过 FCN 层从多元事实中提取特征, 然后基于这些特征评估多元事实的有效性。(b) 基于 CNN 的方法 HINGE 通过卷积操作捕获多元事实中的模式特征, 并利用这些模型特征对多元事实进行评分。(c) 基于 Transformer 的方法 HyTransformer 将多元事实编码为序列, 使用 MASK 替代需要补全的实体, 最后基于编码后的 MASK 表示预测事实中的缺失元素。(d) 基于 GNN 的方法 StarE 使用 GNN 在 NKG 上进行信息传递, 更新元素嵌入, 然后基于更新后的元素嵌入评估多元事实的有效性。

$$\phi(x) = \sum_{i=1}^{a_r} \left\langle \mathbf{u}_i^r, \mathbf{P}_i^r[1, :] \cdot \mathbf{E}_1, \dots, \mathbf{P}_i^r[a_r, :] \cdot \mathbf{E}_{a_r} \right\rangle \quad (5)$$

其中, a_r 表示 r 类型事实中包含的实体数, $\mathbf{P}_i^r$ 表示模式矩阵, $\mathbf{E}_i$ 表示事实中第 i 个实体, $\mathbf{u}_i^r$ 表示第 i 个角色的嵌入向量, $\mathbf{P}_i^r[1, :] \cdot \mathbf{E}_1$ 表示第 i 个角色与第一个实体的兼容性, 其他项依此类推。

不同于上述方法独立建模每个事实, EnhanceE^[37] 考虑实体在多元事实中的位置信息和邻居信息来更新实体的表示。位置信息通过卷积滤波器编码, 而邻居结构则通过图注意力机制建模。EnhanceE 使用

位置信息和邻居信息增强实体表示, 并将实体语义信息融合到关系表示中, 增强了关系的语义表达。

为了进一步探索 CNN 在 NKG 链接预测任务上的表示学习能力, HyconvE^[38] 使用 3D 卷积捕捉多元事实中元素之间的深层交互。它将多元事实集合转换为一个具有 $n+1$ 个通道的特征立方体, 并在 3D 空间中执行卷积操作, 以学习实体和超边之间的高阶关联特征。此外, HyconvE 结合 2D 关系感知卷积提取事实的固有模式, 利用 2D 位置感知卷积捕捉实体在不同位置上的语义信息, 从而更全面地理解多元事实的

结构特征, 增强模型的表达能力。最终, HyconvE 通过逐元素加法融合 3D 卷积和 2D 卷积两条路径的输出, 并经过全连接层映射, 衡量多元事实的可信度。

HyconvE 在模型训练中直接对真实事实进行打分, 此时模型可能会“作弊”, 直接记住训练数据, 而不是学习到元素之间的关联关系。HJE^[4]对其进行改进, 受掩码语言模型(Masked Language Model, MLM)的启发, 提出实体屏蔽机制。在训练过程中, HJE 将需要预测的实体位置替换为掩码(即不提供该实体的具体信息), 仅使用剩余实体和角色作为输入。模型只能依靠其他实体及其角色信息来推理缺失的实体, 而不会受到数据泄露的影响, 避免了依赖于数据记忆。

上述基于 CNN 的模型在提升性能的同时, 其结构变得越来越复杂, 计算开销和内存消耗显著增加, 导致训练效率较低。为了减少参数数量, 降低计算复杂度, Li 等人^[39]提出了一种增强特征交互与提取的 3D 循环卷积嵌入模型 HyCubE, 旨在在保证模型性能的同时, 提高计算效率。具体而言, 该方法通过自适应调整 3D 循环卷积核的大小, 使得多元事实中的角色与实体能够充分交互, 同时减少参数计算量。

2.3.3 基于 Transformer 的方法

HyTransformer^[41]利用 Transformer^[62]结构对多元事实进行编码。具体而言, 它通过初始化位置嵌入来表示多元事实中各个元素的类型信息, 例如实体、关系或限定性角色-值, 接着将其与元素的嵌入表示拼接后输入 Transformer 进行建模。通过这种方式, HyTransformer 能够生成缺失元素的嵌入, 进而预测候选答案概率分布, 如图 4(c)所示。然而, 该方法仅依赖 Transformer 的自注意力机制进行特征交互, 并未显式建模多元事实中不同元素之间的关联关系类型, 导致其在捕捉元素间复杂语义关系时存在一定的局限性。

为进一步增强对多元事实建模能力, GRAN^[40]采用异构图结构来表示多元事实, 其中每个节点对应多元事实中的一个元素, 并通过引入多种边类型来刻画元素之间的不同关系。该方法利用多个全连接的注意力层对异构图进行编码, 并在计算注意力权重时加入边感知偏置项, 以充分利用元素间关联关系类型信息, 从而提升模型的表达能力和推理能力, 在 NKG 链接预测任务上取得了显著的性能提升。

上述方法主要针对离散实体进行建模, 而现实世界的 NKG 常包含数值实体, 如含数值实体的限定性角色-值对 {开始时间: 1911 年} 附加于主三元组 (J. R. R. Tolkien, 受教育于, 牛津大学)。这些数值在量化信息的表达和推理中至关重要。为此, HyNT^[42]采用两

个 Transformer 模块建模含数值实体的多元事实: 上下文 Transformer 和预测 Transformer。上下文 Transformer 聚合主三元组和限定性角色-值对的嵌入表示, 以捕捉它们之间的相关性, 预测 Transformer 则帮助模型学习如何从这些嵌入表示中预测缺失的实体、关系或数值。在处理数值实体时, 上下文 Transformer 将数值实体视为一种特殊类型的实体, 为每个数值字面量分配一个嵌入向量, 预测 Transformer 则通过引入一个损失函数来预测缺失的数值字面量, 增强了模型对数值实体的表达能力。

不同于上述方法仅关注 NKG 中的结构信息, HIST^[43]通过将 HKG 中的结构信息和文本信息结合, 提高了模型对多元事实的表达能力。HIST 首次使用预训练语言模型来编码 NKG, 该方法捕获 NKG 中的文本信息, 并将其与结构信息融合, 以生成更全面的嵌入表示。这种方式使得 HIST 能够克服单纯依赖结构信息时的不足, 尤其在处理结构信息有限的实体时表现得更为出色。

最近, 研究者们还探索了如何在含噪声的 NKG 中进行鲁棒的链接预测。Yu 等人^[44]通过主动众包学习来提高多元事实的置信度评估, 增强 NKG 链接预测的鲁棒性。他们提出了一种名为 NYLON 的方法, 该方法扩展了 GRAN 来处理含噪声的 NKG 链接预测。NYLON 使用带有可学习边偏置的 Transformer 模型来计算事实置信度和元素置信度, 并基于这些置信度进行高效的选择性注释。NKG 标注者可以按置信度的升序检查多元事实的有效性, 一旦发现错误即可终止, 大幅减少标注工作量。NYLON 能够有效应对噪声干扰, 提高其在复杂和不完美数据中的表现。

2.3.4 基于图神经网络的方法

大多数基于 GNN 的方法是为通用场景设计的, 只有少数方法针对特定场景进行了定制化优化。本节讨论适用于通用场景的基于 GNN 的方法, 而特定场景的方法将在 2.4 节中讨论。

适用于通用场景的基于 GNN 的方法可以根据其实现方式进一步分为两类: 面向事实建模的方法和面向本体建模的方法。

(1) 面向事实建模的方法

这类方法通过建模 NKG 上的信息传递, 更新优化元素的嵌入表示, 然后基于更新后的嵌入表示计算多元事实的有效性。例如, StarE^[45]采用编码器-解码器架构来进行 NKG 链接预测。它将多元事实表达为一个核心主三元组和若干的限定性角色-值对。在编码过程中, StarE 首先将限定性角色-值对信息聚合到

主三元组中的关系嵌入中, 然后进行头尾实体之间的信息传递, 更新主三元组中的实体嵌入, 如图 4(d) 所示。在解码阶段, StarE 基于更新后的实体表示, 使用 ConvE^[63]、ConvKB^[64]和 Transformer 对多元事实进行评估打分。

StarE 的一个主要缺点是, 它仅将信息从限定性角色-值对聚合到头尾实体嵌入, 忽略了反向信息传递。为了解决这一问题, QUAD^[48]不仅将信息从限定性角色-值对传递到主三元组, 还通过反向信息传递, 将信息从主三元组传递到限定性角色-值对中, 从而改进了聚合过程。

Luo 等人^[46]提出, NKG 的内部结构可以表达为全局超图和局部语义序列。全局超图表示事实之间的关联关系, 而局部语义序列表示事实的内部组成结构。然而, StarE 和 QUAD 仅考虑了全局超图结构, 忽略了局部语义序列结构。因此, HAHE^[46]引入了两个注意力模块: 全局级注意力模块通过两个 GAT 层^[65]对全局超图结构进行编码, 局部级注意力模块通过异质自注意力层学习局部序列结构。此外, 与上述仅支持预测单一缺失元素的链接预测方法不同, HAHE 进一步扩展到多元素缺失的链接预测任务。具体而言, HAHE 通过在输入中使用 MASK 标记所有缺失元素, 并将标记后的多元事实送入编码器中编码, 最后对每个 MASK 位置逐一进行解码, 从而实现对任意数目缺失元素的链接高效补全。

在 NKG 上进行信息传递时, 多跳邻居聚合可能会引入噪声, 影响实体表示的准确性。为了解决这个问题, HyperFormer^[47]将局部的事实信息融入实体和关系嵌入中。它捕获来自单跳邻居的语义信息, 随后通过双向交互机制增强实体、关系和限定性角色-值对之间的互动。此外, HyperFormer 还采用 Mixture-of-Experts 策略^[66], 结合多个专家网络处理输入, 同时通过一个门控网络选择最相关的专家生成输出, 从而减少了模型的参数量。

上述方法使用固定结构和操作进行信息传递, 使得它们在处理不同形式的多元事实时不够灵活, 难以选择最合适的信息聚合函数。MSeaKG^[49]提出通过自动搜索优化神经架构, 提供了一个灵活且具有表现力的搜索空间, 包含超过八种消息传递函数, 能够同时适应 NKG 和传统 KG。

(2) 面向本体建模的方法

与上述仅对事实本身进行编码的方法不同, 面向本体建模的方法通过建模多元事实中实体的本体语义结构, 为多元事实提供更加深刻的理解。其核心目

标在于通过捕捉实体和本体层之间的复杂关联关系, 从而提升对多元事实的表达能力和推理能力。

例如, Luo 等人^[50]提出了一种基于双视角 NKG 结构的方法 DHKG, 该方法结合了事实实例视角和本体模式视角, 共同对多元事实进行编码。具体来说, DHKG 利用 GRAN 编码器和 GAT 编码器分别在这两种视角下对 NKG 进行建模。通过这种双视角设计, DHKG 显著提高了模型的性能, 超越了仅依赖事实实例视角的基准方法。

然而, DHKG 未能考虑到实体在不同事实中可能关联多种本体类型, 影响了模型的预测精度。为此, HELIOS^[51]采用两层 GAT: 首先使用一个 GAT 对实体的本体类型集建模, 生成静态特征; 然后使用另一个 GAT 动态调整本体类型在当前事实中的权重。最后, HELIOS 利用自注意力机制捕捉模式元组内的关联性, 以优化本体类型预测。

尽管 DHKG 和 HELIOS 在一定程度上改进了实体本体信息的建模, 它们依然忽略了实体的本体特征通常是层次化的, 而非平面结构。因此, Lu 等人^[52]提出了一种对比学习框架 HyperCL, 该框架通过引入关系感知图注意力网络, 能够有效捕捉本体的层次结构, 将其纳入实体表示中。

此外, Chen 等人^[67]提出了一种结合逻辑规则与嵌入模型的可解释框架 HyperMLN。该方法不局限于上述三种 NKG 链接预测建模策略, 可与任意 NKG 链接预测方法结合, 以增强模型的可解释性。HyperMLN 通过一组逻辑规则发现新事实, 并引入权重来表示其置信度。随后 HyperMLN 利用随机场模型^[68]学习多元事实之间的依赖关系, 使用变分期望最大化 (Expectation-Maximization, EM) 算法进行训练, 优化逻辑规则集合。HyperMLN 通过 E 步和 M 步交替优化, 其中 E 步通过嵌入模型生成的分布来更新模型隐含变量, 而 M 步则通过优化逻辑规则的权重来更新模型参数。

2.4 特殊场景下的方法

少数基于 GNN 的 NKG 链接预测方法被设计用于特定场景, 包括时序、小样本和归纳场景。

2.4.1 时序场景下的方法

多元事实常包含时间信息。然而, 大多数现有 NKG 链接预测方法忽略了时间维度, 未能有效处理与时间相关的事实变化, 导致它们在时序任务中表现较差。为了解决这一问题, NE-Net^[11]提出了一种基于 GNN 的实体-角色编码器, 用于捕捉实体演变过程中

的精确表示。通过建模固定时间下实体与角色间关系,并结合与推理元素相关的特定解码器,NE-Net 能够处理时序 NKG 链接预测任务。最近,HyperTKG^[53]在编码时序 NKG 时引入了注意力机制,有效捕捉不同限定性角色-值对对事实推理的不同贡献。同时 HyperTKG 解决了时序性和非时序性事实联合建模的问题。考虑到真实世界不仅包含随时间变化的事实,还包括时间不变的事实,如“官方语言”或“子类”关系等,HyperTKG 将其与时序事实共同建模,它们的有效性会在整个时间范围内保留下来。

2.4.2 小样本场景下的方法

在许多实际应用中,NKG 中的关系只存在有限的支持实例,而现有方法通常假设关系具有大量的支持样本,因此在处理数据量较少的关系时,模型性能往往较差。为了解决这一问题,Zhang 等人^[12]提出了一种在小样本场景 NKG 链接预测方法 HANCL。该方法首先通过邻居编码器编码实体邻居信息,并将其融入实体表示中增强实体表示。随后基于增强后的实体嵌入生成小样本关系表示,基于注意力机制聚合不同支持实例中的小样本关系表示,得到整体的小样本关系嵌入。最后 HANCL 采用对比学习的思想,通过计算查询事实与支持实例中关系表示的相似性,评估查询事实的有效性。与 HANCL 不同,MetaRH^[54]采用元学习机制解决小样本链接预测任务。首先,MetaRH 通过 GNN 聚合邻居信息,生成初步的小样本关系表示,然后基于支持实例优化这些表示。最后利用优化后的小样本关系表示,对查询事实中的缺失实体进行预测。通过这种方式,MetaRH 有效地提升了模型对

小样本关系的泛化能力。

2.4.3 归纳场景下的方法

在现实场景中,新元素往往会在训练阶段之后出现,而传统方法通常无法有效处理这些未见过的元素。为此,Ali 等人^[13]提出了一种归纳式链接预测方法 QBLP,该方法利用与未见元素相关的额外支持样例和文本描述,为这些新元素生成嵌入,从而进行归纳场景下的 NKG 链接预测。同一时期,Wei 等人^[55]结合元学习技术,提出了一种自适应的嵌入生成方法 MetaNIR,能够为未见实体和角色生成有效的嵌入表示。该方法通过从现有的 NKG 中采样生成多个任务(每个任务包括支持集和查询集),来模拟归纳场景。通过在多个任务上进行训练,MetaNIR 学习到如何基于支持集为查询集事实中的未见元素生成嵌入表示,进一步提高了模型对新元素的泛化能力。

2.5 不同 NKG 方法比较

为了直观地比较上述每类方法特点,表 4 展示了它们的建模思想、优缺点。基于空间嵌入建模的方法具有较高的计算效率和较小的参数量,适用于编码大规模 NKG。然而,由于其建模思想较为简单,这类方法存在模型表达能力不足的缺点,通常无法达到理想的效果。基于张量分解的方法具有较强的表达能力,能够充分建模 NKG,但需要更多的参数,限制了其在大规模 NKG 中的应用。基于神经网络的方法则具有更强的特征学习能力,能够处理元素间的复杂关系,但其可解释性较弱,导致其在一些任务中的透明度较差。

表 4 不同类型的方法对比

方法类型	建模思想	优点	不足
基于空间嵌入建模的方法	该方法通过空间位置变换建模多元事实中元素之间的关系,确保其嵌入向量在嵌入空间中满足特定几何约束	具有较低的时间复杂度、较少的模型参数和较快的训练效率,能够处理大规模数据集且计算开销小	模型表达能力有限,无法捕捉实体与角色之间的复杂交互,通常导致预测结果较差
基于张量分解的方法	该方法将 NKG 表达为高阶张量,通过对该张量进行分解,使模型能够学习到 NKG 中的潜在模式和结构	拥有强大的表达能力,能够捕捉 NKG 中元素间的复杂关系与交互,模型效果较好	时间复杂度较高,尤其在处理高元数的多元事实时,计算量非常大
基于神经网络的方法	该方法使用神经网络编码多元事实,从中提取模式特征,以预测 NKG 中事实的缺失元素	能有效提取多元事实中元素的复杂特征,通常具有较高的预测准确性,模型效果最好	训练过程中需要大量数据和较长时间,且预测结果缺乏可解释性

3 NKG 链接预测实验

本节介绍 NKG 链接预测的基准数据集、常用评测指标,以及现有方法在这些基准上的最佳性能表现。

3.1 基线数据集

NKG 链接预测的基准数据集可分为两类:通用场景基准数据集和特殊场景基准数据集。特殊场景进一步划分为三类:时序场景、小样本场景和归纳场

景。下面将分别介绍这些基准数据集。

3.1.1 通用场景

JF17K^[2]、WikiPeople^[10]和 WD50K^[45]是通用场景下 NKG 链接预测任务中最常用的基准数据集,广泛用于评估模型对多元事实建模的能力。其中,JF17K 数据集来源于 Freebase^[9],而 WikiPeople 和 WD50K 来

源于 Wikidata^[69]。表 5 展示了这些数据集的基本统计信息,其中,#E、#R 分别表示数据集中实体、角色的数量,#Arity 表示多元事实的最大元数,即一个多元事实内最多包含实体数, N 表示多元事实的比例,Train、Valid 和 Test 分别表示训练集、验证集和测试集。以下将详细介绍这些基准数据集。

表 5 通用场景下 NKG 链接预测基准数据集

数据集	#E	#R	#Arity	$N/\%$	#Train	#Valid	#Test
JF17K ^[2]	28 645	501	2~6	45.9	76 379	0	24 568
WikiPeople ^[10]	47 765	193	2~9	11.6	305 725	38 223	38 281
WD50K ^[45]	47 155	531	2~67	13.6	166 435	23 913	46 159

注:表中统计信息均来自原始论文。

JF17K 数据集由 Wen 等人^[2]基于 Freebase 构建。他们首先下载了完整的 Freebase 数据,并剔除了涉及字符串、环境类型和数值的三元组。接着,从剩余的事实中恢复多元事实,并对出现次数少于 5 次的实体进行了过滤,以确保每个实体至少参与 5 个事实。这一过程确保了数据集中实体的有效性。然而,JF17K 数据集仍然存在两个主要问题:①测试集中 44.5% 的多元事实的主三元组在训练集中也有出现,这可能导致数据泄露,使得模型无法有效评估其泛化能力;②数据集缺乏验证集,导致模型的调优和泛化评估受到限制,进一步影响模型性能的优化。尽管如此,JF17K 数据集仍在促进 NKG 链接预测方法的发展和评估中发挥了重要作用。

WikiPeople 由 Guan 等人^[10]从 Wikidata 中抽取与人类相关的事实构建而成。他们首先过滤掉了包含图片或缺乏实际意义的实体的事实,并筛选出出现次数超过 30 次的实体,最终将这些实体的相关事实纳入数据集。同时,WikiPeople 的构建,避免了 JF17K 数据集中存在的两个主要问题,即数据泄露和缺乏验证集。然而,WikiPeople 数据集中仅包含 11.6% 的多元事实,这使得其中事实的复杂程度相对较为简单。并且,这些多元事实中的大多数涉及的是文字型实体(如数值、日期和时间字符串)。在传统 KG 中,这类文字型实体通常会被过滤掉,去除后,WikiPeople 数据集中的多元事实比例将不足 3%。

WD50K 由 Galkin 等人^[45]以 FB15K-237 中的实体作为种子集合,从 Wikidata 中提取以这些种子实体为头实体的事实构建得到。为了避免 WikiPeople 数据集中文字信息过多的问题,他们特别过滤掉了文字型实体,以提升数据集的质量和实用性。此外,他们还

移除了出现次数少于 2 的实体,以提高数据集的可靠性。为了进一步减少数据泄露的风险,他们还特别去除了训练集或验证集中与测试集主三元组相同的事实。这一系列改进措施使得 WD50K 数据集成为一个相对更可靠的 NKG 链接预测数据集。

此外,还有一些 NKG 数据集,尽管使用频率较低,但在特定领域内仍然具有较高的研究价值。例如,FB-AUTO^[3]数据集专注于汽车领域,提取自 Freebase,涵盖了大量与汽车相关的多元事实。该数据集的构建对于汽车领域的知识推理和多元事实建模有着独特的意义。而 MedCKG^[70]则是首个专注于医学领域的 NKG 数据集,基于中国医科大学提供的临床数据构建。MedCKG 数据集为医学领域的知识推理任务提供了支持,对于医学 KG 的构建、推理以及临床决策的辅助具有重要的应用价值。

3.1.2 时序场景

Hou 等人^[11]构建了两个专门用于时序场景下 NKG 链接预测任务的数据集: NWIKI 和 NICE。NWIKI 数据集源自 Wikidata,包含大量多元事实,为时序 NKG 链接预测提供丰富的基础数据。为了确保模型能够有效训练,他们在数据集构建过程中筛选了高频实体,并过滤掉了低频实体。与 NWIKI 数据集不同,NICE 数据集来源于 ICEWS^①,其中时序信息更加显著,特别适用于需要建模随时间动态变化的事实的任务。这两个数据集在时序 NKG 领域提供了重要的实验基础,促进了对基于时序的多元事实建模方法的研究。与 Hou 等人同期,Ding 等人^[53]以 Wikidata11k^[71]和 YAGO1830^[72]作为基础数据集,搜索其中事实

① <https://dataverse.harvard.edu/dataverse/icews>

Wikidata 中的限定性角色-值对,从而扩展原始传统时序 KG 为时序 NKG,得到 Wiki-hy 和 YAGO-hy 数据集。表 6 展示了 NWIKI、NICE、Wiki-hy 和 YAGO-hy

数据集的统计信息,其中 Timestamps 表示时间戳,记录了各个事实发生的具体时间,Time Interval 表示事实之间的最小时间间隔。

表 6 时序场景下 NKG 链接预测基线数据集。表中统计信息均来自原始论文。

数据集	#E	# Timestamps	Time Interval	N/%	#Train	#Valid	#Test
NWIKI ^[11]	17 481	205	1年	81.9	108 397	14 370	15 591
NICE ^[11]	10 860	4 017	24小时	97.5	368 868	5 268	46 159
Wiki-hy ^[53]	11 140	507	1年	9.5	111 252	13 900	13 926
YAGO-hy ^[53]	10 026	188	1年	6.9	51 193	10 973	10 977

3.1.3 小样本场景

Zhang 等人^[12]在小样本场景下构建了三个 NKG 链接预测数据集: WikiAnimals、WikiCompanies 和 WD50K-Few。这些数据集专注于模拟现实场景下稀缺关系的学习挑战,以评估模型在有限数据条件下的泛化能力。其中, WikiAnimals 和 WikiCompanies 分别从 Wikidata 中抽取包含动物和公司相关实体的事实得到,而 WD50K-Few 则从 WD50K 数据集中筛选部分数据构建而成。它们为研究小样本 NKG 链接预测提供了重要的实验基准。与 Zhang 等人同期, Wei 等人^[54]进一步构建了 F-WikiPeople、F-JF17K 和 F-WD50K

数据集,分别基于 WikiPeople、JF17K 和 WD50K 进行小样本场景扩展得到。这些数据集涵盖了不同的知识领域,并在数据分布和多元事实结构上各具特色,进一步丰富了小样本 NKG 链接预测任务的实验基础。表 7 汇总了这些小样本 NKG 链接预测基准数据集的关键统计信息。其中,“#E-q”和“#R-q”分别表示限定性角色-值对中的实体和角色,反映了数据集中多元事实的复杂性,而“#Tasks”列则表示小样本关系链接预测任务的数量。这些数据集为探索在小样本场景的 NKG 链接预测表示学习方法提供了坚实的实验基础。

表 7 小样本场景下 NKG 链接预测基线数据集

数据集	#E	#R	#E-q	#R-q	N/%	#Tasks	#Facts
WikiAnimals ^[12]	2 925 278	167	396 739	48	49.7	49	5 964 839
WikiCompanies ^[12]	30 781	164	27 511	127	18.1	125	1 128 040
WD50K-Few ^[12]	47 156	532	5 460	45	13.6	126	236 507
F-WikiPeople ^[54]	40 529	237	4 663	75	9.0	30	319 140
F-JF17K ^[54]	19 721	480	4 928	127	47.6	52	91 572
F-WD50K ^[54]	43 802	697	10 242	85	13.1	118	379 653

注:表中统计信息均来自原始论文。

3.1.4 归纳场景

Ali 等人^[13]基于 WD50K 构建了一系列专门用于归纳式 NKG 链接预测的数据集,其中包括多个不同设定的数据集,以验证模型在归纳场景下的泛化能力。本节重点介绍其中具有代表性的 WD20K(25)、WD20K(100) V1 和 WD20K(100) V2,其他数据集的详细信息可参考原文。在归纳场景下,这些数据集中的实体通常包含文本描述或附加的推理图(推理图含有未见实体相关的支持实例),以辅助在测试阶段生成未见实体的表示。其中, WD20K(25) 仅提供文本描述作为辅助信息,不提供含未见实体的推理

图,模型必须依赖文本特征完成归纳式链接预测任务。相比之下, WD20K(100) V1 和 WD20K(100) V2 除了文本描述外,还额外提供推理图,以帮助模型利用结构信息推理未见实体的表示。此外, WD20K(100) V1 的训练数据规模较 WD20K(100) V2 更大,使得模型能够在训练阶段学习到更丰富的特征。数据集名称中括号内的数字表示该数据集中包含多元事实的比例,反映了不同数据集在事实多样性上的差异。Wei 等人^[55]进一步构建了 JF-Ext、WIKI-Ext 和 WD-Ext 数据集,分别基于 JF17K、WikiPeople 和 WD50K 进行扩展,为归纳式 NKG 链接预测提供了更多高质量的

测试基准。这些数据集不仅涵盖了不同领域的知识,还在数据规模、多元事实比例和实体信息丰富度上有所差异,为后续研究提供了更加全面的实验基

准。表 8 总结了这些归纳式 NKG 链接预测基准数据集的统计信息。其中,“Inference”列表示测试时所使用的推理图。

表 8 归纳场景下 NKG 链接预测基线数据集

数据集	Train		Validation		Test		Inference	
	#Facts	N/%	#Facts	N/%	#Facts	N/%	#Facts	N/%
WD20K(25) ^[13]	39 819	30.0	4 252	25.0	3 453	22.0	0	0
WD20K(100) V1 ^[13]	7 785	100.0	295	100.0	364	100.0	2 667	100.0
WD20K(100) V2 ^[13]	4 146	100.0	538	100.0	678	100.0	4 274	100.0
JF-Ext ^[55]	3 305	54.0	1 061	30.0	1 283	21.0	5 012	28.1
WIKI-Ext ^[55]	3 905	2.1	6 480	2.5	4 733	2.9	4 880	6.6
WD-Ext ^[55]	5 112	1.0	2 610	3.6	3 053	2.0	3 382	6.1

注:表中统计信息均来自原始论文。

3.2 评测指标

在评测过程中,需要计算每个候选项的得分并进行排序。正确答案的排名越靠前,模型性能越优秀。在 NKG 链接预测任务中,MRR 和 Hits@K 是最常用的评测指标,用于衡量模型在不同方面的链接预测能力。

3.2.1 平均倒数排名

平均倒数排名(Mean Reciprocal Rank, MRR)主要用于评估正确答案在预测结果中的排名质量,反映模型的整体链接预测效果,计算如式(6)所示。

$$\text{MRR} = \frac{1}{|Q|} \sum_{q \in Q} \frac{1}{\text{rank}_q} \quad (6)$$

其中, Q 表示查询集合, rank_q 为正确答案在排序列表中的排名。

MRR 的取值范围在 0 到 1 之间,值越大表示模型的预测性能越优秀。

3.2.2 Hits@K

Hits@K 计算所有查询中正确答案排名在 K 以内的比例,用于衡量模型在不同精度水平下的性能,满足特定应用需求,如式(7)所示。

$$\text{Hits@K} = \frac{\left| \{q \in Q: \text{rank}_q \leq K\} \right|}{|Q|} \quad (7)$$

其中, Hits@K 的取值范围同样在 0 到 1 之间,数值越大表明模型越能将正确答案排在前列。常见的 K 值包括 1、3、5 和 10,分别对应不同应用场景下对预测精度的要求。

3.3 实验结果

本节对比现有方法在上述基准数据集上的最新

实验结果。我们分别在通用、时序、小样本和归纳四类场景下,比较代表性方法的性能,以评估当前模型的能力。

3.3.1 通用场景

表 9 展示了 JF17K、WikiPeople 和 WD50K 这三个最常用于通用场景的数据集上,不同 NKG 链接预测方法的实验结果。最佳结果以加粗表示,次优结果以下划线标出。对比方法包括基于空间嵌入建模的方法 m-TransH^[2]和 RAE^[27],基于张量分解的方法 HypE^[3]和 S2S^[33],以及基于神经网络的方法 HINGE^[34]、NeuInfer^[26]、GRAN^[40]和 HAHE^[46]。实验结果表明,基于神经网络的方法整体表现最佳,其中基于 GNN 的 HAHE 在所有数据集上均取得最优性能,显著优于基于 FCN 的 NeuInfer、基于 CNN 的 HINGE 和基于 Transformer 的 GRAN。这表明,GNN 结构在多元事实编码方面具备更强的建模能力,能够更有效地学习元素间的复杂关联关系。此外,基于空间嵌入建模的方法(m-TransH、RAE)表现最差,说明简单的投影变换难以建模多元事实。而基于张量分解的方法(HypE、S2S)虽比空间结构方法略优,但仍显著落后于基于神经网络的方法,表明其在处理复杂的多元事实时存在一定局限性。整体来看,GNN 的引入显著提升了多元事实的建模能力,相较于传统的空间结构和张量分解方法,在 NKG 链接预测任务中展现出更强的表示学习能力。

3.3.2 时序场景

NE-Net^[11]和 HypeTKG^[53]是专门针对时序场景的 NKG 链接预测的方法。表 10 展示了它们与一些代表性基线模型在时序数据集 NWIKI 和 Wiki-hy 上的实验结果。CEN^[73]、TiGRN^[74]和 DE-Simple^[75]是代表性

表 9 通用场景下 NKG 链接预测实验结果

类型	方法	JF17K		WikiPeople		WD50K	
		MRR	Hits@1	MRR	Hits@1	MRR	Hits@1
基于空间结构的方法	m-TransH ^[2]	0.102	0.069	—	—	—	—
	RAE ^[27]	0.310	0.219	0.172	0.102	—	—
基于张量分解的方法	HypE ^[3]	0.494	0.408	0.292	0.162	—	—
	S2S ^[33]	0.528	0.457	0.372	0.277	—	—
基于神经网络的方法	HINGE ^[34]	0.473	0.397	0.333	0.259	—	—
	NeuInfer ^[26]	0.517	0.436	0.350	0.282	0.232	0.164
	GRAN ^[40]	<u>0.656</u>	<u>0.582</u>	<u>0.479</u>	<u>0.410</u>	<u>0.309</u>	<u>0.240</u>
	HAHE ^[46]	0.668	0.597	0.495	0.420	0.402	0.327

注：“—”表示该方法在原论文中未在相应数据集上进行测试。HypE、S2S结果来自文献[33]，其余方法结果来自文献[46]。

表 10 时序场景下 NKG 链接预测实验结果

方法	NWIKI			Wiki-hy		
	MRR	Hits@1	Hits@10	MRR	Hits@1	Hits@10
HINGE ^[34]	0.217	0.191	0.259	0.543	0.497	0.694
HypE ^[3]	0.252	0.249	0.257	<u>0.624</u>	<u>0.604</u>	<u>0.658</u>
CEN ^[73]	0.406	0.302	0.610	—	—	—
DE-Simple ^[75]	0.138	0.108	0.191	0.351	0.218	0.640
TiGRN ^[74]	<u>0.611</u>	<u>0.506</u>	0.811	—	—	—
NE-Net ^[11]	0.720	0.668	<u>0.802</u>	—	—	—
HypeTKG ^[53]	—	—	—	0.693	0.642	0.792

注：NWIKI、Wiki-hy数据集结果分别来自文献[11,53]。

的时序二元事实链接预测方法，能够有效建模 KG 中的时间信息，但它们未考虑多元事实中的限定性角色-值对信息。HINGE^[34]和 HypE^[3]是 NKG 链接预测领域的代表性方法，但未利用时序 NKG 中的时间信息。实验结果表明，NE-Net 和 HypeTKG 在两个数据集上分别取得了最佳性能，显著超越了时序二元事实建模方法和非时序多元事实建模方法。例如，NE-Net 在 NWIKI 数据集上的 MRR 达到 0.720，远高于其他方

法，进一步验证了其在时序 NKG 链接预测任务中的有效性。这一结果表明，在时序 NKG 链接预测任务中，同时建模限定性角色-值对信息和时间演化信息能够显著提升模型的预测能力。

3.3.3 小样本场景

HANCL^[12]和 MetaRH^[54]是专门针对小样本场景的 NKG 链接预测的方法。表 11 展示了它们在小样本数据集 WikiAnimals 和 F-WD50K 上的实验结果。

表 11 小样本场景 NKG 链接预测实验结果

方法	WikiAnimals			F-WD50K		
	MRR	Hits@1	Hits@10	MRR	Hits@1	Hits@10
StarE ^[45]	0.265	<u>0.233</u>	0.215	0.102	0.057	0.177
GRAN ^[40]	0.253	0.199	0.221	<u>0.126</u>	<u>0.077</u>	0.222
FSRL ^[76]	0.236	0.201	0.230	—	—	—
FAAN ^[77]	<u>0.270</u>	0.225	<u>0.246</u>	0.116	0.059	<u>0.226</u>
MetaR ^[78]	—	—	—	0.108	0.064	0.183
HANCL ^[12]	0.318	0.288	0.258	—	—	—
MetaRH ^[54]	—	—	—	0.192	0.109	0.340

注：WikiAnimals、F-WD50K数据集结果分别来自文献[12,54]。

FSRL^[76]、FAAN^[77]和 MetaR^[78]是代表性的小样本二元事实链接预测方法,能够在有限的训练数据下进行关系表示学习,但未考虑多元事实中的限定性角色-值对信息。实验结果表明, HANCL 和 MetaRH 在两个数据集上均取得了最佳性能,显著优于其他方法。例如, HANCL 在 WikiAnimals 数据集上的 MRR 达到 0.318,相比其他方法均有明显提升。这一结果表明,在小样本 NKG 链接预测任务中,仅依赖传统的小样本学习方法难以充分挖掘多元事实中的知识,结合限定性角色-值对信息可以有效增强模型的推理能力,显著提升预测性能。

3.3.4 归纳场景

QBLP^[13]和 MetaNIR^[55]是专门针对归纳式 NKG 链接预测的方法。表 12 展示了它们在 WD20K(100)

V1、WD20K(100) V2 和 WD-Ext 数据集上的实验结果。BLP^[79]是传统 KG 领域代表性的归纳式链接预测方法,它通过编码文本描述生成未见实体的嵌入,而 StarE^[45]和 CompGCN^[80]则基于推理图上的邻居信息生成未见实体的嵌入。BLP 和 CompGCN 忽略了多元事实中的限定性角色-值对信息。实验结果表明, QBLP、StarE 和 MetaNIR 方法取得了较好的结果,验证了利用多元事实中的限定性角色-值对信息对于归纳式 NKG 链接预测任务是必要的。此外,在 WD20K(100) V1 数据集上, StarE 表现优于 QBLP,而在 WD20K(100) V2 数据集上, QBLP 表现优于 StarE。这表明,在训练数据量较大时,基于推理图生成未见实体表示更具有优势,而在训练数据量较小的情况下,依赖文本描述的方法更为有效。

表 12 归纳式场景 NKG 链接预测实验结果

方法	WD20K(100) V1			WD20K(100) V2			WD-Ext		
	MRR	Hits@1	Hits@10	MRR	Hits@1	Hits@10	MRR	Hits@1	Hits@10
BLP ^[79]	0.057	0.019	0.123	0.039	0.014	0.092	—	—	—
CompGCN ^[80]	0.104	<u>0.057</u>	0.183	0.025	0.007	0.053	—	—	—
StarE ^[45]	0.112	0.061	<u>0.212</u>	<u>0.049</u>	<u>0.019</u>	<u>0.110</u>	<u>0.079</u>	<u>0.021</u>	<u>0.131</u>
QBLP ^[13]	<u>0.107</u>	0.039	0.245	0.066	0.034	0.120	—	—	—
MetaNIR ^[55]	—	—	—	—	—	—	0.582	0.433	0.901

注: WD20K(100) V1、WD20K(100) V2数据集结果来自文献^[13], WD-Ext数据集结果来自文献^[13]。

4 NKG 链接预测的应用

由于具备表达多实体间复杂语义关系的能力, NKG 已广泛应用于生物医学、商品推荐系统、金融科技等领域的知识建模与推理任务中。本节选取上述三个典型场景,介绍 NKG 链接预测的具体应用与价值。

4.1 生物医学领域

在生物医学领域,知识通常以复杂的多实体间关联关系的形式存在,例如“某药物通过某靶点治疗某疾病”或“某基因突变在某人群中导致某种疾病”。在此类事实中,实体之间语义关联紧密,采用 NKG 表达能够通过明确的角色分配(如药物、靶点、疾病、人群)更全面地刻画事实的语义结构,从而提升知识表示的准确性与完整性。例如,在肿瘤学知识库中,大量信息涉及基因、基因组变异与癌症治疗手段之间的关系^[81-82]。Peng 等人^[83]将文本构建为图结构,并在其

上进行信息传播以学习实体表示,从而识别潜在的多元事实以挖掘生物医学知识。基于生物医学 NKG,链接预测任务可进一步挖掘潜在有效的药物组合与适应症之间的关联,助力药物重定位与组合疗法的发现。例如, Wang 等人^[70]首次将 NKG 应用于医疗领域,构建了基于中国医科大学临床数据的 MedCKG 数据集,并通过 NKG 链接预测辅助生成个性化药物治疗方案。此外, NKG 也可以被用于医学材料的自动化检索。EDEL^[84]在肿瘤学数据集^[85-86]和蛋白质翻译数据集^[87]上实现了对可用于补全不完整多元事实查询的相关文献的自动检索。

4.2 商品推荐系统

在推荐系统中,用户行为常常受到时间、地理位置、设备类型、行为类型等多种上下文因素影响,例如,“一个用户在某地某时间使用某设备浏览某商品”。传统 KG 通常采用(用户,购买,物品)这类三元组建模,难以捕捉上述丰富的上下文语义。而 NKG 可通过建模多元事实,例如,超边形式的事实(用户购

物事实, 用户, 物品, 行为类型, 时间, 地理位置), 能够实现对用户行为的精细建模与语义解释。NKG 上的链接预测任务可用于挖掘用户在特定上下文中的潜在兴趣, 从而提升推荐结果的准确性。例如, 基于 NKG 的推荐方法, SDK^[88]通过整体建模多元事实, 有效增强了用户与商品的表示能力, 尤其在冷启动和数据稀疏场景下展现出更优的泛化能力, 这得益于其对多维信息的融合建模与深度推理能力。HeLLM^[89]将超边形式的多元事实结构建模与大语言模型 (Large Language Model, LLM) 的语义建模深度融合, 构建了一个面向多模态推荐的框架, 显著提升了 LLM 在用户偏好建模方面的时效性、结构性与语义表达能力。HKGNN^[90]则通过建模用户相关的超关系形式多元事实, 保留了更丰富的语义信息, 有效缓解了数据稀疏问题; 同时引入自注意力机制对用户访问序列进行建模, 实现了更加个性化的推荐效果。

4.3 金融科技领域

金融科技领域中的事实往往涉及多个关键要素, 例如, “一家银行在某一时间向某客户发放贷款, 用于特定项目”, 或 “某机构在某市场投资某资产并面临特定风险”。传统 KG 难以有效表达这类复杂多元事实, 而 NKG 能准确建模事实中的各要素及其对应角色 (如贷款人、借款人、用途、风险程度等), 更全面地揭示金融实体之间的交互关系。基于构建的结构化金融事实, NKG 链接预测可支持风险评估、相似案例检索等智能分析任务, 从而显著提升金融数据的自动化分析能力。例如, Hou 等人^[11]基于真实金融交易数据构建了金融领域的 NKG, 通过链接预测识别潜在高风险贷款事实, 有效提升了风险预警的准确率; Engel^[91]则通过建模多元金融事实, 整合国际贸易数据库, 支持复杂查询与分析, 增强了全球供应链数据的整合与弹性分析能力。

5 未来展望

尽管 NKG 链接预测取得了显著进展, 但作为一个新兴研究领域, 仍面临诸多挑战。未来的研究可以从以下几个方向深入探索。

5.1 基于大语言模型的 NKG 链接预测

LLM 在 KG 链接预测任务中表现出了卓越的能力。KG-BERT^[92]将三元组事实转化为文本序列, 用其对 BERT^[93]进行微调, 有效地执行了链接预测任务。

KG-LLaMA^[94]通过微调 LLaMA^[95]从 KG 中学习结构化知识, 取得了更好效果。KICGPT^[96]首先使用传统链接预测模型对候选答案进行粗排, 然后使用 LLM 对高概率的候选项进行重排序, 提升了传统模型的性能, 缓解了 LLM 的幻觉问题。HIST^[43]是当前唯一将 LLM 应用于 NKG 链接预测的工作, 它使用 GNN 提取 NKG 中的结构信息, 然后通过结构化软提示有效地将文本和结构信息结合, 输入到 LLM, 对候选事实进行评分。LLM 仍未在 NKG 链接预测任务中得到广泛应用, 主要面临以下挑战:

(1) 如何将结构化的多元事实转换为适合 LLM 处理的输入格式: NKG 中的多元事实由多个实体和角色组成, 这些信息需要转化成 LLM 能够理解的文本格式。然而, 直接将结构化数据转换为文本序列可能丧失元素间复杂的语义关联关系。如何设计一种合适的转换机制, 以确保 LLM 能够有效地理解和处理这些多元事实, 成为了一个关键挑战。

(2) 如何设计有效的训练任务, 使 LLM 能够学习并记忆多元事实: 由于 NKG 中包含大量的事实和复杂的关系, 仅依赖传统的文本序列训练方式可能无法让 LLM 有效地捕捉到 NKG 中的复杂结构信息。因此, 开发能够引导 LLM 学习多元事实并记忆这些信息的训练任务显得尤为重要。

(3) 如何解决 LLM 的输入长度限制, 以便高效处理候选答案: LLM 在处理大规模输入时通常存在输入长度的限制, 这对于处理包含大量候选答案的 NKG 链接预测任务尤其棘手。如何设计有效的策略来突破这一限制, 处理大规模候选答案, 是提升 LLM 在 NKG 领域应用的关键。

5.2 特殊场景下的 NKG 链接预测

大多数 NKG 链接预测方法主要针对通用场景, 针对特殊情况的研究相对有限。关于特殊场景——时序、小样本和归纳——的研究仍处于初期阶段, 进一步探索存在丰富的机会。例如, NE-NET 和 HypeTKG 方法能够有效捕捉实体在 NKG 中的全局结构特征, 并在时序场景中预测未来事实, 但其未能充分利用多元事实内的局部结构信息。QBLP 方法在归纳场景下能够对未见实体进行推理, 但仍未解决对未见角色的建模问题。小样本方法如 HANCL 和 MetaRH 依赖大量小样本任务进行训练, 而在实际应用中, 获取足够的小样本任务训练数据往往较为困难。

此外, 现实世界中的 NKG 往往具有动态性, 新事实不断引入, 现有方法需要微调模型或重新训练整个

数据集,这可能导致知识遗忘或者消耗大量计算资源。因此,为不断增长的 NKG 开发链接预测方法,在高效适应新事实的同时保持已有知识,也是未来研究的关键挑战之一。

5.3 可解释的 NKG 链接预测

据我们所知,HyperMLN^[67]是目前唯一可解释的 NKG 链接预测方法。该方法利用随机场模型^[68]学习多元事实之间的依赖关系,通过挖掘自反(inverse)、对称(symmetric)和子关系(subrelation)三类一阶逻辑规则来解释预测结果。然而,HyperMLN 未能刻画多元事实中常见的组合(composition)规则。未来研究可以着重探索更丰富的逻辑规则学习方法,以提升模型的可解释性,并增强其在实际应用中的可信度。

5.4 交叉学科启发的 NKG 链接预测

NKG 链接预测本质上是一个涉及多实体、多角色的结构化语义建模问题,其研究发展可以从多个交叉学科中汲取理论支持与方法启发。例如,离散数学研究中的异构图、超图及高阶结构理论,为 NKG 中多元事实的结构建模提供了坚实的理论基础。Luo 等人^[46]提出将 NKG 表示为同时包含全局超图结构与局部语义序列的复合图,并从两个层面对其进行建模,体现了图理论对 NKG 表示范式的指导作用。线性代数及多线性代数提供了高维张量运算与分解技术,如 Tucker 分解、TR 分解等,为多元事实的高效建模提供了方法支撑。Liu 等人^[14]基于张量分解方法对 NKG 进行建模,并引入 TR 分解以提升建模效率。几何学启发了将几何嵌入(如盒状嵌入、球面嵌入)用于准确建模多元事实限定性角色-值对的单调性等几何特性。Xiong 等人^[31]则利用盒状几何嵌入对多元事实建模,将每个限定性角色-值对建模为对盒嵌入空间的逐步“收缩”,从而有效捕捉限定性结构中的单调性结构。统计学有望引入因果模型(如贝叶斯网络、潜在变量模型)分析事实之间的因果关系,推动 NKG 从“相关性”建模向“因果性”建模转变,提升推理的准确性与可解释性。与离散数学、代数学与统计学等交叉学科的融合正在为 NKG 链接预测开辟新的研究方向。未来,借助更多的相关学科理论工具,有望进一步提升 NKG 链接预测模型的表达能力和可解释性。

6 总结

NKG 链接预测已成为一个重要的研究方向。本

文首次对该领域的相关工作进行了系统性梳理。首先,我们介绍了 NKG 及其链接预测任务的定义,并从应用场景和建模策略两个维度对现有方法进行了分类。随后,我们深入分析了各类方法的特点,归纳了其优势与局限。在总结现有方法后,我们进一步介绍了 NKG 链接预测的基准数据集,包括数据集的来源、构建方式及当前主流方法在这些数据集上的实验表现。接着,我们回顾了 NKG 链接预测在生物医学、商品推荐、科技金融等实际场景中的典型应用,展现了其在复杂事实建模与智能推理任务中的潜在价值。最后,我们探讨了 NKG 链接预测的未来发展方向,包括 LLM 的应用、特殊场景下的挑战、模型可解释性的提升、交叉学科的启发改进。希望本综述能够为后续研究提供有价值的参考,并推动该领域的进一步发展。

参考文献

- [1] SUCHANEK F M, KASNECI G, WEIKUM G. YAGO: A core of semantic knowledge unifying WordNet and wikipedia [J]. *Semantic Web*, 2007, 1(1): 17-33.
- [2] WEN J, LI J, MAO Y, et al. On the representation and embedding of knowledge bases beyond binary relations[C]// *Proceedings of the 25th International Joint Conference on Artificial Intelligence*. 2016: 1300-1307.
- [3] FATEMI B, TASLAKIAN P, VAZQUEZ D, et al. Knowledge hypergraphs: Prediction beyond binary relations[C]// *Proceedings of the 29th International Joint Conference on Artificial Intelligence Main track*. 2020: 2191-2197.
- [4] LI Z, WANG C, WANG X, et al. HJE: Joint convolutional representation learning for knowledge hypergraph completion[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2024, 36(8): 3879-3892.
- [5] HECK L, HAKKANI-TÜR D, TUR G. Leveraging knowledge graphs for web-scale unsupervised semantic parsing [C]// *Proceedings of the Interspeech. ISCA*, 2013: 1594-1598.
- [6] RUDNIK C, EHRHART T, FERRET O, et al. Searching news articles using an event knowledge graph leveraged by wikidata[C]// *Proceedings of the Companion Proceedings of the World Wide Web Conference*. San Francisco, USA: ACM, 2019: 1232-1239.
- [7] SOUZA COSTA T, GOTTSCHALK S, DEMIDOVA E. Event-QA: A dataset for event-centric question answering over knowledge graphs[C]// *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*.

- Virtual Event Ireland: ACM, 2020: 3157-3164.
- [8] LIANG S, SHAO J, ZHANG J, et al. Graph-based non-sampling for knowledge graph enhanced recommendation [J]. IEEE Transactions on Knowledge and Data Engineering, 2023, 35(9): 9462-9475.
- [9] BOLLACKER K, EVANS C, PARITOSH P, et al. Free-base: A collaboratively created graph database for structuring human knowledge[C]//Proceedings of the ACM SIGMOD International Conference on Management of Data - SIGMOD '08. Vancouver, Canada: ACM Press, 2008: 1247-1250.
- [10] GUAN S, JIN X, WANG Y, et al. Link prediction on N-ary relational data[C]//Proceedings of the World Wide Web Conference, San Francisco, CA, USA: ACM Press, 2019: 583-593.
- [11] HOU Z, JIN X, LI Z, et al. Temporal knowledge graph reasoning based on n-tuple modeling[C]//Proceedings of the Association for Computational Linguistics, Singapore: Association for Computational Linguistics, 2023: 1090-1100.
- [12] ZHANG X, LIANG X, ZHENG X, et al. When true becomes false: Few-shot link prediction beyond binary relations through mining false positive entities[C]//Proceedings of the 30th ACM International Conference on Multimedia. Lisboa Portugal: ACM, 2022: 4063-4071.
- [13] ALI M, BERRENDORF M, GALKIN M, et al. Improving inductive link prediction using hyper-relational facts (Extended Abstract)[C]//Proceedings of the 20th International Semantic Web Conference, 2021: 74-92.
- [14] LIU Y, YAO Q, LI Y. Generalizing tensor decomposition for N-ary relational knowledge bases[C]//Proceedings of the Web Conference. Taipei, Taiwan: ACM, 2020: 1104-1114.
- [15] VINAY S NAVALE, CHETAN JANE, ANVITA B, et al. Link prediction on N-ary relational data based on relatedness evaluation[J]. International Journal for Research in Applied Science and Engineering Technology, 2023, 11(8): 1597-1609.
- [16] ZHANG Z, CUI P, ZHU W. Deep learning on graphs: A survey[J]. IEEE Transactions on Knowledge and Data Engineering, 2022, 34(1): 249-270.
- [17] DAI Y, WANG S, XIONG N N, et al. A survey on knowledge graph embedding: Approaches, applications and benchmarks[J/OL]. Electronics, 2020, 9(5): 750.
- [18] NGUYEN D Q. A survey of embedding models of entities and relationships for knowledge graph completion[C]//Proceedings of the Graph-based Methods for Natural Language Processing (TextGraphs). Barcelona, Spain (Online): Association for Computational Linguistics, 2020: 1-14.
- [19] CAI B, XIANG Y, GAO L, et al. Temporal knowledge graph completion: A survey[C]//Proceedings of the 32nd International Joint Conference on Artificial Intelligence, 2023: 6545-6553.
- [20] WANG J, WANG B, QIU M, et al. A Survey on temporal knowledge graph completion: Taxonomy, progress, and prospects[J]. arXiv preprint arXiv: 2308.02457, 2023.
- [21] LIANG K, MENG L, LIU M, et al. A survey of knowledge graph reasoning on graph types: Static, dynamic, and multimodal[J]. arXiv preprint arXiv: 2212.05767, 2023.
- [22] ZHU X, LI Z, WANG X, et al. Multi-modal knowledge graph construction and application: A survey[J]. IEEE Transactions on Knowledge and Data Engineering, 2022, 36(2): 715-735.
- [23] HUBERT N, MONNIN P, PAULHEIM H. Beyond transduction: A survey on inductive, few shot, and zero shot link prediction in knowledge graph[J]. arXiv preprint arXiv: 2312.04997, 2023.
- [24] MA H, WANG D Z. A survey on few-shot knowledge graph completion with structural and commonsense knowledge [J]. arXiv preprint arXiv: 2301.01172, 2023.
- [25] SHEN T, ZHANG F, CHENG J. A comprehensive overview of knowledge graph completion[J]. Knowledge-Based Systems, 2022, 255: 109597.
- [26] GUAN S, JIN X, GUO J, et al. NeuInfer: Knowledge inference on N-ary facts[C/OL]//Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Online: Association for Computational Linguistics, 2020: 6141-6151[2021-10-30]. <https://www.aclweb.org/anthology/2020.acl-main.546>.
- [27] ZHANG R, LI J, MEI J, et al. Scalable instance reconstruction in knowledge bases via relatedness affiliated embedding[C]//Proceedings of the World Wide Web Conference on World Wide Web, Lyon, France: ACM Press, 2018: 1185-1194.
- [28] YAN S, ZHANG Z, SUN Z, et al. HYPER2- Hyperbolic embedding for hyper-relational link prediction[J]. Neucomputing, 2022: 440-451.
- [29] YAN S, ZHANG Z, SUN X, et al. PolygonE: Modeling n-ary relational data as gyro-polygons in hyperbolic space [J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2022, 36(4): 4308-4317.
- [30] YAN S, ZHANG Z, XU G, et al. Modeling N-ary relational data as gyro-polygons with learnable gyro-centroid [J]. Knowledge-Based Systems, 2022: 109164.

- [31] XIONG B, NAYYER M, PAN S, et al. Shrinking embeddings for hyper-relational knowledge graphs[C]//Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics, 2023: 13306-13320.
- [32] HU Z, GUTIÉRREZ, BASULTO V, XIANG Z, et al. HyperMono: A monotonicity-aware approach to hyper-relational knowledge representation [J]. arXiv preprint arXiv: 2404.09848, 2024.
- [33] DI S, YAO Q, CHEN L. Searching to sparsify tensor decomposition for N-ary relational data[C]//Proceedings of the Web Conference. Ljubljana Slovenia: ACM, 2021: 4043-4054.
- [34] ROSSO P, YANG D, CUDRÉ-MAUROUX P. Beyond triplets: Hyper-relational knowledge graph embedding for link prediction[C]//Proceedings of the Web Conference. Taipei Taiwan: ACM, 2020: 1885-1896.
- [35] LU Y, YANG D, WANG P, et al. Schema-aware hyper-relational knowledge graph embeddings for link prediction[J]. IEEE Transactions on Knowledge and Data Engineering, 2024: 2614-2628.
- [36] LIU Y, YAO Q, LI Y. Role-aware modeling for N-ary relational knowledge bases[C]//Proceedings of the Web Conference. Ljubljana Slovenia: ACM, 2021: 2660-2671.
- [37] WANG C, LI Z, WANG X, et al. EnhanceE: Enhanced entity and relation embedding for knowledge hypergraph link prediction[C]//Proceedings of the ACM Web Conference. Austin TX USA: ACM, 2023: 115-118.
- [38] WANG C, WANG X, LI Z, et al. HyConvE: A novel embedding model for knowledge hypergraph link prediction with convolutional neural networks[C]//Proceedings of the ACM Web Conference. Austin TX USA: ACM, 2023: 188-198.
- [39] LI Z, WANG X, LI J, et al. HyCubE: Efficient knowledge hypergraph 3D circular convolutional embedding IEEE Transactions on Knowledge and Data Engineering[J]. 2025, 37(4): 1902-1914.
- [40] WANG Z, HUI B, ZHOU X, et al. Improvement on message passing of hyper-relational knowledge graph [C]//Proceedings of the 5th International Conference on E-Business, Information Management and Computer Science. Hong Kong: ACM, 2022: 7-11.
- [41] YU D, YANG Y. Improving hyper-relational knowledge graph completion[J]. arXiv preprint arXiv: 2104.08167, 2021.
- [42] CHUNG C, LEE J, WHANG J J. Representation learning on hyper-relational and numeric knowledge graphs with transformers[C]//Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2023.
- [43] LI L, WANG H, LI J, et al. Integrating structure and text for enhancing hyper-relational knowledge graph representation via structure soft prompt tuning[C]//Proceedings of the 33rd ACM International Conference on Information and Knowledge Management. Boise ID USA: ACM, 2024: 1226-1234.
- [44] YU W, YANG J, YANG D. Robust link prediction over noisy hyper-relational knowledge graphs via active learning [C]//Proceedings of the ACM on Web Conference 2024. Singapore Singapore: ACM, 2024: 2282-2293.
- [45] GALKIN M, TRIVEDI P, MAHESHWARI G, et al. Message passing for hyper-relational knowledge graphs [C]//Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, 2020: 7346-7359.
- [46] LUO H, E H, YANG Y, et al. HAHE: Hierarchical attention for hyper-relational knowledge graphs in global and local level[C]//Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2023: 8095-8107.
- [47] DING K, LIANG A J, PERROZI B, et al. HyperFormer: Learning expressive sparse feature representations via hyper-graph transformer[C]//Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2023: 2062-2066.
- [48] SHOMER H, JIN W, LI J, et al. Learning representations for hyper-relational knowledge graphs[C]//Proceedings of the ACM on Web Conference 2025, 2025: 3288-3299.
- [49] DI S, CHEN L. Message function search for knowledge graph embedding[C]//Proceedings of the ACM Web Conference. Austin TX USA: ACM, 2023: 2633-2644.
- [50] LUO H, E H, TAN L, et al. DHGE: Dual-view hyper-relational knowledge graph embedding for link prediction and entity typing[C]//Proceedings of the 37th AAAI Conference on Artificial Intelligence and 35th Conference on Innovative Applications of Artificial Intelligence and 13th Symposium on Educational Advances in Artificial Intelligence, 2023: 6467-6474.
- [51] LU Y, DENG B, YU W, et al. HELIOS: Hyper-relational schema modeling from knowledge graphs[C]//Proceedings of the 31st ACM International Conference on Multimedia, 2023: 4053-4064.
- [52] LU Y, YU W, JING X, et al. HyperCL: A contrastive learning framework for hyper-relational knowledge graph embedding with hierarchical ontology[C]//Proceedings of the Association for Computational Linguistics, 2024: 2918-2929.
- [53] DING Z, WU J, WU J, et al. Temporal Fact Reasoning

- over Hyper-Relational Knowledge Graphs[C]//Proceedings of the Association for Computational Linguistics, 2024: 355-373.
- [54] WEI J, GUAN S, JIN X, et al. Few-shot link prediction on hyper-relational facts[C]//Proceedings of the Joint International Conference on Computational Linguistics, Language Resources and Evaluation. Torino, Italia: ELRA and ICCL, 2024: 7196-7207.
- [55] WEI J, GUAN S, JIN X, et al. Inductive link prediction in N-ary knowledge graphs[C]//Proceedings of the 31st International Conference on Computational Linguistics. Abu Dhabi, UAE: Association for Computational Linguistics, 2025: 8885-8896.
- [56] ROSSO P, YANG D, OSTAPUK N, et al. RETA: A schema-aware, end-to-end solution for instance completion in knowledge graphs[C]//Proceedings of the Web Conference. Ljubljana Slovenia: ACM, 2021: 845-856.
- [57] CHUNG C, WHANG J J. Learning representations of bi-level knowledge graphs for reasoning beyond link prediction[J]. arXiv preprint arXiv: 2302.02601, 2023.
- [58] GUO J, KOK S. BiQUE: Biquaternionic embeddings of knowledge graphs[C]//Proceedings of the Conference on Empirical Methods in Natural Language Processing. Online and Punta Cana, Dominican Republic: Association for Computational Linguistics, 2021: 8338-8351.
- [59] XIONG B, NAYYERI M, LUO L, et al. NestE: Modeling nested relational structures for knowledge graph reasoning[J]. arXiv preprint arXiv: 2312.09219, 2023.
- [60] BALAŽEVIĆ I, ALLEN C, HOSPEDALES T M. Tucker: Tensor factorization for knowledge graph completion[C]//Proceedings of the Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing. 2019: 5184-5193.
- [61] DONG X, GABRILOVICH E, HEITZ G, et al. Knowledge vault: A web-scale approach to probabilistic knowledge fusion[C]//Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York New York USA: ACM, 2014: 601-610.
- [62] 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.
- [63] DETTMERS T, MINERVINI P, STENETORP P, et al. Convolutional 2D knowledge graph embeddings[C]//Proceedings of the 32nd AAAI Conference on Artificial Intelligence and 30th Innovative Applications of Artificial Intelligence Conference and 8th AAAI Symposium on Educational Advances in Artificial Intelligence, 2018: 1811-1818.
- [64] NGUYEN D Q, NGUYEN T D, NGUYEN D Q, et al. A novel embedding model for knowledge base completion based on convolutional neural network[C]//Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, New Orleans, Louisiana: Association for Computational Linguistics, 2018: 327-333.
- [65] VELIC-KOVIC P, CUCURULL G, CASANOVA A, et al. Graph attention networks[C]//Proceedings of the ICLR 2018, 2018.
- [66] SHAZEER N, MIRHOSEINI A, MAZIARZ K, et al. Outrageously large neural networks: The sparsely-gated mixture-of-experts layer[J]. arXiv preprint arXiv: 1701.06538, 2017.
- [67] CHEN Z, WANG X, WANG C, et al. Explainable link prediction in knowledge hypergraphs[C]//Proceedings of the 31st ACM International Conference on Information & Knowledge Management. Atlanta GA USA: ACM, 2022: 262-271.
- [68] RICHARDSON M, DOMINGOS P. Markov logic networks [J]. Machine Learning, 2006, 62(1-2): 107-136.
- [69] VRANDEČIĆ D, KRÖTZSCH M. Wikidata: A free collaborative knowledgebase[J]. Communications of the ACM, 2014, 57(10): 78-85.
- [70] WANG W, ZHANG X, ZHANG J, et al. MHRE: Multivariate link prediction method for medical hyper-relational facts[J]. Applied Intelligence, 2024,54:1311-1334.
- [71] JUNG J, JUNG J, KANG U. Learning to walk across time for interpretable temporal knowledge graph completion[C]//Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining. Virtual Event Singapore: ACM, 2021: 786-795.
- [72] HAN Z, CHEN P, MA Y, et al. Explainable subgraph reasoning for forecasting on temporal knowledge graphs [C]//Proceedings of the International Conference on Learning Representations, 2021.
- [73] LI Z, GUAN S, JIN X, et al. Complex evolutionary pattern learning for temporal knowledge graph reasoning[C]//Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). 2022: 290-296.
- [74] LI Y, SUN S, ZHAO J. TiRGN: Time-guided recurrent graph network with local-global historical patterns for temporal knowledge graph reasoning[C]//Proceedings of the 31st

- International Joint Conference on Artificial Intelligence. Vienna, Austria: International Joint Conferences on Artificial Intelligence Organization, 2022: 2152-2158.
- [75] GOEL R, KAZEMI S M, BRUBAKER M, et al. Diachronic embedding for temporal knowledge graph completion[C]//Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(04): 3988-3995.
- [76] ZHANG C, YAO H, HUANG C, et al. Few-shot knowledge graph completion[J]. arXiv preprint arXiv: 1911.11298, 2019.
- [77] SHENG J, GUO S, CHEN Z, et al. Adaptive attentional network for few-shot knowledge graph completion[C]//Proceedings of the Conference on Empirical Methods in Natural Language Processing. Online: Association for Computational Linguistics, 2020: 1681-1691.
- [78] CHEN M, ZHANG W, ZHANG W, et al. Meta relational learning for few-shot link prediction in knowledge graphs[C]//Proceedings of the Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing. Hong Kong, China: Association for Computational Linguistics, 2019: 4216-4225.
- [79] DAZA D, COCHEZ M, GROTH P. Inductive entity representations from text via link prediction[C]//Proceedings of the Web Conference. Ljubljana Slovenia: ACM, 2021: 798-808.
- [80] VASHISHTH S, SANYAL S, NITIN V, et al. Composition-based multi-relational graph convolutional networks[J]. arXiv preprint arXiv: 1911.03082, 2019.
- [81] LEVER J, JONES M R, DANOS A M, et al. Text-mining clinically relevant cancer biomarkers for curation into the CIViC database[J/OL]. Genome Medicine, 2019, 11(1): 78.
- [82] LEE K, LEE S, PARK S, et al. BRONCO: Biomedical entity relation oncology corpus for extracting gene-variant-disease-drug relations[J/OL]. Database, 2016, 2016[2025-06-01]. <https://academic.oup.com/database/article/doi/10.1093/database/baw043/2630320>.
- [83] PENG N, POON H, QUIRK C, et al. Cross-sentence n-ary relation extraction with graph LSTMs[J]. Transactions of the Association for Computational Linguistics, 2017, 5: 101-115.
- [84] WANG X D, LESER U. Enhancing document retrieval for curating n-ary relations in knowledge bases[J]. arXiv preprint arXiv: 2504.10613, 2025.
- [85] CHAKRAVARTY D, GAO J, PHILLIPS S, et al. OncoKB: A precision oncology knowledge base[J/OL]. JCO Precision Oncology, 2017(1): 1-16.
- [86] GRIFFITH M, SPIES N C, KRYSIK K, et al. CIViC is a community knowledgebase for expert crowd sourcing the clinical interpretation of variants in cancer[J]. Nature Genetics, 2017, 49(2): 170-174.
- [87] The UniProt consortium. UniProt: A worldwide hub of protein knowledge[J]. Nucleic Acids Research, 2019, 47(D1): D506-D515.
- [88] LIU Y, XUAN H, LI B, et al. Self-supervised dynamic hypergraph recommendation based on hyper-relational knowledge graph[J]. arXiv preprint arXiv: 2308.07752, 2023.
- [89] GUO X, ZHANG T, WANG Y, et al. Multi-modal hypergraph enhanced LLM learning for recommendation[J]. ACM Transactions on Multimedia Computing, Communications and Applications, 2025, 21(10): 1-23.
- [90] ZHANG J, LI Y, ZOU R, et al. Hyper-relational knowledge graph neural network for next POI recommendation[J]. World Wide Web, 2024, 27(4): 46.
- [91] ENGEL F, VANIN M, KRDZAVAC N. Anticipate risk with the value and trade flows knowledge graph[C]//Proceedings of the 3rd International Workshop on Linked Data-driven Resilience Research, 2024.
- [92] YAO L, MAO C, LUO Y. KG-BERT: BERT for knowledge graph completion[J]. arXiv preprint arXiv: 1909.03193, 2019.
- [93] DEVLIN J, CHANG M W, LEE K, et al. BERT: Pre-training of deep bidirectional transformers for language understanding[C]//Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Minneapolis, Minnesota: Association for Computational Linguistics, 2019: 4171-4186.
- [94] YAO L, PENG J, MAO C, et al. Exploring large language models for knowledge graph completion[C]//Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing, 2025: 20970-20984.
- [95] TOUVRON H, LAVRIL T, IZACARD G, et al. LLaMA: Open and efficient foundation language models[J]. arXiv preprint arXiv: 2302.13971, 2023.
- [96] WEI Y, HUANG Q, ZHANG Y, et al. KICGPT: Large language model with knowledge in context for knowledge graph completion[C]//Proceedings of the Association for Computational Linguistics: EMNLP. Singapore: Association for Computational Linguistics, 2023: 8667-8683.



隗继耀(1998—), 共同第一作者, 博士, 主要研究领域为知识图谱。

E-mail: weijiyao20z@ict.ac.cn



官赛萍(1991—), 共同第一作者, 博士, 副研究员, 主要研究领域为知识图谱、数据价值评估与提升。

E-mail: guansaiping@ict.ac.cn



靳小龙(1976—), 通信作者, 博士, 研究员, 主要研究领域为大数据知识工程、知识计算、知识图谱。

E-mail: jinxiaolong@ict.ac.cn

会议通知|第六届中国情感计算大会(CCAC2026)

将于7月10日-12日在武汉举行

情感计算作为人工智能领域的核心前沿方向,正深度驱动认知心理与共情交互、脑电接口、舆情与社会治理、智能人机交互、大模型情感理解等领域的创新突破。历经五届匠心深耕,中国情感计算大会(CCAC)已成长为国内情感计算领域规格最高、影响力最广、产学研融合最深的标杆性学术交流平台。

今日,我们正式官宣:第六届中国情感计算大会(CCAC2026)定于2026年7月10日—7月12日在湖北武汉举办,由武汉纺织大学与武汉大学联合承办。江城聚贤,智启新章。我们诚挚邀约海内外情感计算领域顶尖专家、青年学者、行业先锋与企业代表齐聚武汉,共话学科前沿,共研技术突破,共促成果落地,共绘情感计算发展新蓝图。

CCAC 历届精彩回顾

自2021年创办以来,中国情感计算大会始终坚守“交流前沿成果、凝聚行业力量、推动技术创新”的初心,足迹跨越南北五城,每一站都留下了学术碰撞、思想激荡的精彩印记,见证着中国情感计算从起步探索到蓬勃发展的完整历程。

2021·北京·清华大学

首届CCAC于北京盛大启幕,作为领域开篇盛会,汇聚海内外顶尖学者专家,正式开启中文信息学会框架下情感计算领域系统化、专业化学术交流的新纪元,高起点擘画领域发展蓝图,为后续大会树立标杆。

2022·苏州·苏州大学

大会落地江南苏州,首次增设前沿技术讲习班、青年学生论坛等特色板块,搭建起“基础研究—学术交流—人才培养”一体化平台,让情感计算理念广泛传播,为青年学者成长护航。

2023·西安·西安交通大学

CCAC走进古都西安,伴随专委会正式去“筹”,大会规模与影响力再创新高,近四百名代表齐聚一堂。大会新增企业论坛、技术评测研讨等板块,推动情感计算从实验室走向产业应用,实现产学研深度融通。

2024·南昌·江西财经大学与江西师范大学

大会以“让大模型更懂情感”为核心主题,聚焦大模型时代情感计算新机遇、新挑战,众多知名专家围绕情感理解、生成与对齐深入研讨,为新时代领域发展锚定方向。

2025·成都·西华大学

CCAC亮相天府成都,设置特邀报告、青年科学家论坛、前沿趋势专场、多模态专题研讨等丰富环节,联动京东科技、科大讯飞、百度等头部企业,全面展示文本、语音、视觉等多模态情感计算最新成果,打造集学术交流、技术展示、人才引进于一体的综合性平台。

五届盛会,薪火相传,CCAC始终聚焦多模态情感认知、智能交互与应用落地,从单一学术研讨,逐步进阶为“学术+产业+人才”深度融合的高端平台,已然成为国内情感计算成果首发、技术对接、产学研协同创新的核心枢纽,持续助推领域高质量发展。

2026年,CCAC首次落户武汉,这座兼具人文底蕴与科创活力的中部重镇,正依托人工智能、大数据、智能制造等产业,全速推进科技创新建设。武汉纺织大学与武汉大学强强联手,凭借两校在计算机科学、人工智能、多媒体信息处理等领域的深厚学科积淀,全力打造一届高标准、有特色、重实效的顶级学术盛会。

本届大会将延续往届优质议程,汇聚海内外顶尖学者带来重磅特邀报告,开设青年科学家、前沿趋势、企业应用等特色论坛,搭配技术评测、海报交流等环节,同时结合武汉区位优势打造专属议题,全方位展现领域前沿成果与技术突破。江城七月,荷风送香,目前大会筹备工作已全面启动,征稿、注册等相关事宜将陆续公布,敬请持续关注官方通知。

我们诚挚邀请海内外情感计算领域专家学者、青年学子、产业精英,2026年7月10日—12日,相聚武汉,共赴情感计算学术盛宴,携手探索人机共情新边界,共同谱写中国情感计算事业的崭新篇章!

(中国中文信息学会)