

融合关系模式和类比迁移的知识图谱补全方法

宋宝燕, 刘杭生, 单晓欢, 李素, 陈泽

引用本文

宋宝燕, 刘杭生, 单晓欢, 李素, 陈泽. [融合关系模式和类比迁移的知识图谱补全方法](#)[J]. 计算机科学, 2025, 52(3): 287-294.

SONG Baoyan, LIU Hangsheng, SHAN Xiaohuan, LI Su, CHEN Ze. [Joint Relational Patterns and Analogy Transfer Knowledge Graph Completion Method](#) [J]. Computer Science, 2025, 52(3): 287-294.

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)

[自学习星型链空间自适应分配方法](#)

Self-learning Star Chain Space Adaptive Allocation Method

计算机科学, 2025, 52(3): 359-365. <https://doi.org/10.11896/jsjcx.240700140>

[基于可微知识图谱的多跳知识库问答](#)

Multi-hop Knowledge Base Question Answering Based on Differentiable Knowledge Graph

计算机科学, 2025, 52(3): 295-305. <https://doi.org/10.11896/jsjcx.240600095>

[融合知识图谱的负采样方法](#)

Negative Sampling Method for Fusing Knowledge Graph

计算机科学, 2025, 52(3): 161-168. <https://doi.org/10.11896/jsjcx.240500015>

[基于动态贝叶斯博弈的工业控制网络恶意接入检测研究](#)

Study on Malicious Access Detection in Industrial Control Networks Based on Dynamic BayesianGames

计算机科学, 2025, 52(1): 383-392. <https://doi.org/10.11896/jsjcx.231200083>

[融合情感和常识知识的对话生成模型](#)

Dialogue Generation Model Integrating Emotional and Commonsense Knowledge

计算机科学, 2025, 52(1): 307-314. <https://doi.org/10.11896/jsjcx.231100130>

融合关系模式和类比迁移的知识图谱补全方法

宋宝燕 刘杭生 单晓欢 李素 陈泽

辽宁大学信息学部 沈阳 110036

(bysong@lnu.edu.cn)

摘要 近年来,知识图谱嵌入(Knowledge Graph Embedding,KGE)作为一种主流方法在知识图谱补全任务中已取得显著效果。然而,现有KGE方法仅在数据层考虑三元组信息,忽略了不同三元组间在逻辑层存在的关系模式语义,导致现有方法仍存在一定性能缺陷。针对上述问题,提出一种融合关系模式和类比迁移的知识图谱补全方法(Fusing Relational-pattern and Analogy Transfer,RpAT)。首先,在逻辑层,根据实体关系的语义层次结构,细分为不同的关系模式;其次,在数据层,提出一种模式类比对象生成方法,该方法利用关系模式性质生成目标三元组相似类比对象,依据类比对象对缺失信息进行迁移;最后,提出一种融合了原始知识图谱嵌入模型的推理能力与类比迁移能力的综合性评分函数,以提升图谱补全性能。实验结果表明,在FB15k-237和WN18RR数据集上,相较于其他基线模型,RpAT方法的MRR值分别提升了15.5%和1.8%,验证了在知识图谱补全任务中的有效性。

关键词:知识图谱;知识图谱补全;关系模式;类比对象;类比迁移

中图分类号 TP182

Joint Relational Patterns and Analogy Transfer Knowledge Graph Completion Method

SONG Baoyan,LIU Hangsheng,SHAN Xiaohuan,LI Su and CHEN Ze

Faculty of Information,Liaoning University,Shenyang 110036,China

Abstract In recent years,knowledge graph embedding(KGE) has emerged as a mainstream approach and achieved significant results in the task of knowledge graph completion. However,existing KGE methods only consider the information of triplets at the data level,neglecting the semantic relational patterns that exist between different triplets at the logical level,leading to certain performance deficiencies in current methods. To address this issue,a knowledge graph completion method(RpAT) that integrates relational patterns and analogy transfer is proposed. Firstly,at the logical level,different relational patterns are refined according to the semantic hierarchy of entity relationships. Secondly,at the data level,a method for generating pattern analogy objects is proposed,which utilizes the properties of relational patterns to generate similar analogy objects for target triplets and transfers missing information based on these analogy objects. Finally,a comprehensive scoring function that integrates the reasoning capabilities of the original knowledge graph embedding model and the analogy transfer capabilities is proposed to enhance the performance of graph completion. Experimental results show that,compared to other baseline models,the RpAT method pimproves the MRR values by 15.5% and 1.8% on the FB15k-237 and WN18RR datasets,respectively,demonstrating its effectiveness in the task of knowledge graph completion.

Keywords Knowledge graph,Knowledge graph completion,Relational patterns,Analogy objects,Analogy transfer

到稿日期:2024-07-24 返修日期:2024-10-14

基金项目:辽宁省公共舆情与网络安全信息系统重点实验室(d252453002);辽宁省应用基础研究计划(2022JH2/101300250);教育部产学研合作协同育人项目(230701160261310);国家重点研发计划(2023YFC3304900);辽宁省教育厅高校基本科研项目(理工类)面上项目(揭榜挂帅服务地方项目)(JYTMS20230761);辽宁省自然科学基金项目博士启动项目(2023-BS-085)

This work was supported by the Liaoning Provincial Key Laboratory of Public Opinion and Network Security Information System(d252453002), Applied Basic Research Program of Liaoning Province(2022JH2/101300250),Ministry of Education University-Industry Collaborative Education Program(230701160261310),National Key R&D Program of China(2023YFC3304900),General Program of University Basic Scientific Research of Education Department of Liaoning Province(Science and Engineering)(Initiating Flagship Service for Local Projects)(JYTMS20230761) and Nature Science Foundation Program Doctoral Startup Project of Liaoning Province (2023-BS-085).

通信作者:陈泽(chenz1996@outlook.com)

1 引言

知识图谱(Knowledge Graph, KG)作为一种以图结构形式存储的知识系统,包含大量常识知识、通用规则以及其他结构化信息,这些信息通常以三元组 (h, r, t) 的形式呈现。其中,节点 h, t 代表实体,边 r 则代表实体间的关系。目前,知识图谱系统包括 Freebase^[1], WordNet^[2]等,这些方法已在词义消歧^[3]、智能推荐^[4]和信息提取^[5]等下游应用中取得显著效果。然而,由于信息抽取技术的限制,大规模的知识图谱仍面临数据不完备的挑战。

知识图谱嵌入(KGE)^[6]作为一种被广泛采用的知识图谱表示方法,其核心思想是将知识图谱的实体和关系嵌入到连续的低维向量空间。为了解决知识缺失这一问题,知识图谱补全(Knowledge Graph Completion, KGC)技术应运而生。其通过整合实体和关系的语义结构,为提升知识图谱的完整性和准确性提供了有效的解决方案。KGE在低维向量空间中捕捉实体和关系的语义表示,并通过预定义的三元组评分函数进行预测,实现知识图谱补全。在知识图谱补全任务中,现有方法忽略了实体间的语义层次结构,即关系模式。例如,三元组(教师,传授,学生)是陈述“教师向学生传授知识”这一事实,实际上“教师”与“学生”间具有特定的关系模式。在不同的关系语义结构中,实体“教师”扮演着不同角色。在教学关系中,实体“教师”与“学生”构成了“教”与“学”的逆关系模式。此外,“教师”还可能具有多种身份,如“父亲”“职工”等,实体“教师”的多种身份关系语义形成了多重关系模式。尽管现有方法尝试通过融合邻域实体进行知识图谱补全,但在大规模知识图谱中,每个实体往往拥有成千上万的邻域实体,信息量大且复杂,现有方法难以有效甄别出其中的有效信息,容易受到数据噪声的干扰,难以准确捕捉关键信息,最终导致补全结果错误。为解决这一问题,本文提出了一种融合关系模式和类比迁移的知识图谱补全方法(RpAT)。首先,在逻辑层,根据 Horn 规则,将三元组细分为不同关系模式;其次,在数据层,依据已知事实,通过类比迁移相似语义信息,并提出了一种关系模式类比对象生成器,以生成目标三元组的相似属性类比对象,从而实现目标三元组缺失信息的迁移;最后,结合知识图谱嵌入模型的原始推理能力,通过插值方式与类比迁移能力相结合,实现知识图谱的准确补全。本文的主要贡献可概括为以下两个方面:

1)提出了一种新式的知识图谱补全方法 RpAT,该方法通过有效融合关系模式与类比迁移,利用已知事实生成高质量类比对象,从而显著提升了知识图谱的补全性能。

2)通过实验对比分析,验证了 RpAT 在 FB15k-237 和 WN18RR 这两个公开数据集上的卓越性能,实验指标均优于现有的典型知识图谱补全方法。

2 相关工作

知识图谱嵌入技术^[7]目前主要分为三大类:基于距离计算的方法、基于图神经网络的方法,以及嵌入额外信息的增强型方法。

基于距离计算的方法通过将实体和关系映射到连续的几何向量空间中,并利用评分函数来计算这些向量间的距离,评估实体和关系之间的相似性。其中,TransE^[8]是这类方法的典型代表方法。此外,通过在嵌入空间中定义几何关系来建模实体和关系,可以进一步提升 TransE 的性能,如 RotatE^[9], HAKE^[10], Rot-Pro^[11]和 PairRE^[12]等。尽管如此,基于距离计算的方法在处理大规模知识图谱时,往往无法捕捉到实体和关系之间更深层次的关系语义信息。基于图神经网络的方法则利用图结构来学习实体和关系之间更深层次的语义交互信息。例如,SE-GNN^[13]方法通过将关系、实体和三元组层面的语义特征整合到邻域模式中,并使用图神经网络来学习实体和关系的嵌入表示。其他基于图神经网络的方法,如 R-GCN^[14]和 CompGCN^[15]通常结合图结构中的邻居节点信息和特征来学习嵌入表示。然而,当知识图谱结构中存在噪声和稀疏性时,图神经网络模型可能会出现过拟合问题,从而影响其泛化能力。嵌入额外信息的增强型方法通过引入补充信息来提升知识图谱补全的性能。例如,CAKE^[16]方法利用常识和复杂关系生成高质量的负三元组,并结合常识与事实视图进行链接预测。PUDA^[17]和 REP^[18]在处理数据稀疏性方面表现出色,其中 PUDA 引入阳性-未标记风险估计器和对抗性数据增强策略,而 REP 则引入特定于关系的上下文函数和嵌入传播机制。尽管这些方法为解决稀疏性问题提供了有价值思路,但需要大量计算资源来训练模型,并且在处理大规模和复杂数据时可能效果不佳。SPA^[19]方法对各种关系模式进行了全面的理论分析,并基于此提出了一种融合规则置信度的评分函数,从而提升了知识图谱嵌入模型在各种关系模式上的性能。ANALOGY^[20]是首个在多关系嵌入中应用类比方法的方法,但其补全性能提升并不显著。Ankge^[21]在实体和关系建模方面需要投入大量资源来学习数据集,这无疑增加了推理时间和资源消耗。鉴于此,本文提出的方法旨在减少知识图谱补全任务在推理过程中的时间和资源消耗,同时提高 RpAT 方法的准确性,以实现更高效的知识图谱补全。

3 RpAT 方法

3.1 问题定义

知识图谱补全任务旨在预测知识图谱中缺失的数据。具体来说,给定一个知识图谱 $G = \{(h, r, t) | (h, r, t) \in \mathcal{E} \times \mathcal{R} \times \mathcal{E}\}$, 其中 $h, t \in \mathcal{E}, r \in \mathcal{R}$ 。 \mathcal{R}, \mathcal{E} 分别表示关系集和实体集, h 和 t 分别表示三元组的头、尾实体, r 表示它们之间的关系。知识图谱补全任务的目标是基于已知的事实信息,推理缺失元素。图 1 给出了 RpAT 框架图,其由 3 个模块组成:模式分类、类比迁移和链接预测。

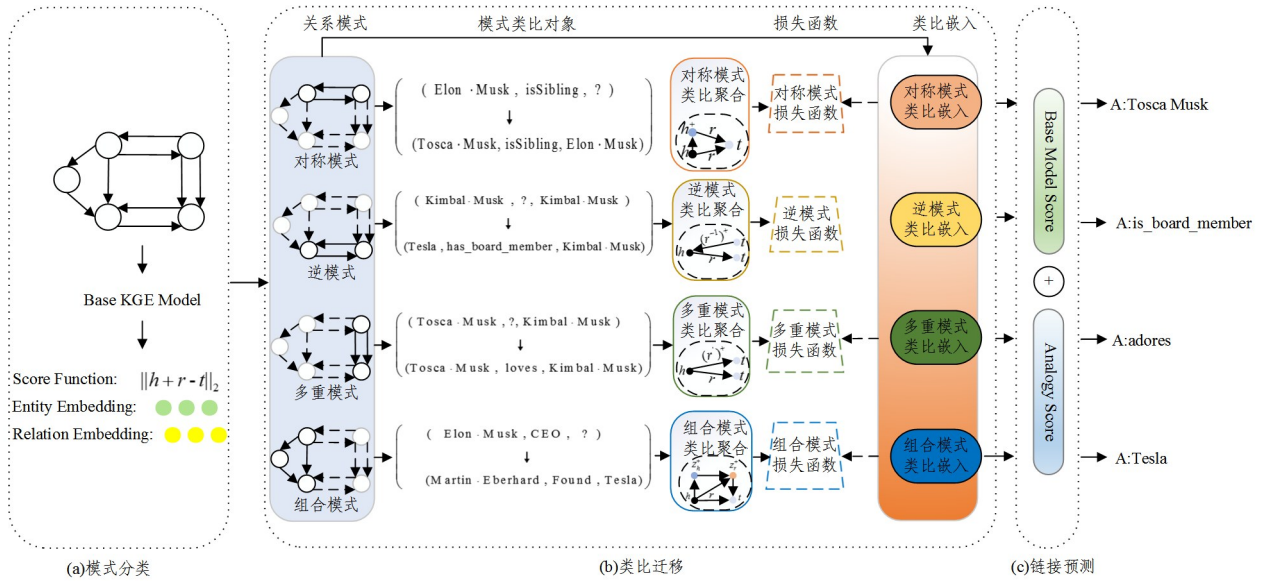


图 1 RpAT 框架图

Fig. 1 Structure of RpAT

3.2 模式分类

采用预训练模型 RotatE 将实体和关系映射到复数向量空间,将每个关系定义为从头实体到尾实体的旋转,获取实体和关系的嵌入表示, $M = \{E, R, f_{kge}, \theta\}$ 。其中 E, R 和 f_{kge} 分别是实体嵌入表、关系嵌入表和模型 M 的评分函数, θ 是其他参数的集合。 $g_{kge}(h, r)$ 是根据给定的关系

嵌入,将头实体的嵌入映射到尾实体的嵌入。具体来说,将知识图谱视作复杂的图结构,图 2 左侧显示了图谱结构,遵循关系模式规则形式^[19]定义,将知识图谱划分为关系模式。

依据现有的逻辑规则,把图 2 划分为对称模式(a)、多重模式(b)、逆模式(c)和组合模式(d)。

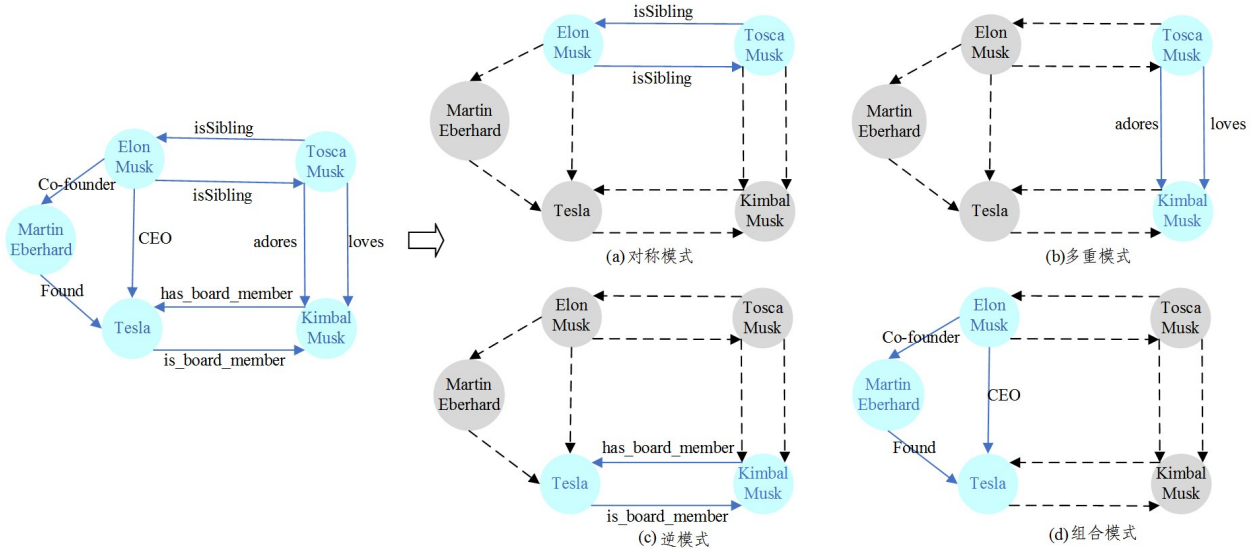


图 2 KG 模式分类

Fig. 2 KG pattern classification

3.3 类比迁移

3.3.1 类比对象生成器

在关系模式中检索与发现和目标三元组相似的实体或关系,挖掘知识图谱中关系模式与实体、关系之间隐含的内在联系,生成目标三元组类比对象。

1) 对称模式生成器。结合对称模式的特性与类比相似性,生成类比对象。例如,在对称模式下,对于给定的查询三元组 $(h, r, ?)$,首先利用知识图谱嵌入模型的评分函数

$f_{kge}(h, r, t)$,对所有候选实体进行评分,并选取评分最高的三元组作为事实三元组。由于在对称模式中,查询三元组的实体 h 与其邻接实体之间存在紧密联系,因此将已知事实实体的邻接实体替换到事实三元组的相应位置,创建新的三元组。随后,再次利用评分函数对这些新三元组进行评分,评分最高的新三元组即为最佳类比对象 (h', r, t) 。以图 2(a)为例,对于查询三元组 $(Elon \cdot Musk, isSiblings, ?)$,可以依据已知事实邻接实体 $Tosca \cdot Musk$ 替换 $Elon \cdot Musk$,形成新的三元组

(Tosca · Musk, isSibling, Elon · Musk), 并对其评分, 最终确定 Tosca · Musk 为类比对象, 从而得到查询三元组的答案是 Tosca · Musk。对称模式下的类比对象如式(1)所示:

$$\mathbf{p}_{\text{sym}} = \{\mathbf{h}_i \mid \text{Top}(\{f_{\text{kge}}(\mathbf{h}_i, \mathbf{r}, t) \mid \mathbf{h}_i \in \mathcal{E}\})_{N_s}\} \quad (1)$$

其中, $\text{Top}(\cdot)_k$ 表示在所有输入中选择前 k 个最高值的三元组, $f_{\text{kge}}(\mathbf{h}, \mathbf{r}, t)$ 是模型中预定义的评分函数, $(\mathbf{h}_i, \mathbf{r}, t)$ 是新的三元组, 作为输入。

2) 多重模式生成器。在多重模式中, 为生成类比关系对象, 需要检索实体间的不同关系。具体而言, 这要求从关系的角度进行定量分析, 以识别由关系部分引起的相似类比对象。这是因为实体间可能存在多种相似或等价的关系, 而这些关系通常共享相同的语义。如图 2(b) 所示, 实体 Tosca · Musk 和 Kimbal · Musk 间存在 adores 和 loves 两种关系, 而 adores 与 loves 通常被认为是近义词。因此, 在多重模式中, adores 的类比关系对象便是 loves。具体地, 通过分析已知事实三元组中的关系, 来推断出具有相似性质的类比关系对象 $(\mathbf{h}, \mathbf{r}', t)$ 。多重模式的类比对象定义如式(2)所示:

$$\mathbf{p}_{\text{mul}} = \{\mathbf{r}_i \mid \text{Top}(\{f_{\text{kge}}(\mathbf{h}, \mathbf{r}_i, t) \mid \mathbf{r}_i \in \mathcal{R}\})_{N_m}\} \quad (2)$$

3) 逆模式生成器。如表 1 所列, 不同方法对逆关系的数学定义各不相同。与多重模式类似, 在生成类比关系对象时, 逆模式关注的焦点是关系本身, 而不涉及特定实体。对于给定三元组, 逆模式生成器通过检索和分析, 构造逆模式的类比关系对象 $(t, \mathbf{r}^{-1}, \mathbf{h})$ 。如图 2(c) 所示, 在 Tesla 公司与董事会成员 Kimbal · Musk 之间的关系中, 能够观察到逆模式的属性关系。从公司和董事会成员的不同视角分析这种关系, 可以发现存在两种事实表述: (1) “Tesla 公司拥有董事会成员 Kimbal · Musk”, 即三元组 (Tesla, has_board_member, Kimbal · Musk); (2) “Kimbal · Musk 是作为 Tesla 公司的董事会成员之一”, 即三元组 (Kimbal · Musk, is_board_member, Tesla)。这些表述体现了逆模式在关系分析中的重要作用。因此逆模式类比对象定义如式(3)所示:

$$\mathbf{p}_{\text{inv}} = \{\mathbf{r}_i^{-1} \mid \text{Top}(\{f_{\text{kge}}(t, \mathbf{r}_i^{-1}, \mathbf{h}) \mid \mathbf{r}_i^{-1} \in \mathcal{R}\})_{N_i}\} \quad (3)$$

表 1 基线模型的逆关系表示

Table 1 Inverse relation representation of baseline models

Model	Inverse Relation
TransE	$\mathbf{r}^{-1} = -\mathbf{r}$
RotatE	$\mathbf{r}^{-1} = \bar{\mathbf{r}}$
HAKE	$\mathbf{r}_m^{-1} = -\mathbf{r}_m \quad \mathbf{r}_p^{-1} = \bar{\mathbf{r}}$
PairRE	$\mathbf{r}^{H^{-1}} = 1/\mathbf{r}^H \quad \mathbf{r}^{T^{-1}} = 1/\mathbf{r}^T$

4) 组合模式生成器。结合对称模式和多重模式的检索机制所生成的类比对象被用于组合模式的类比迁移。以图 2(d) 为例, 在寻找查询三元组 (Elon · Musk, CEO, ?) 的类比对象时, 从已知事实中得知, Martin · Eberhard 是 Tesla 的创始人。同时, Martin · Eberhard 和 Elon · Musk 都是联合创始人。因此, 基于上述信息, (Elon · Musk, CEO) 的类比对象为 (Martin · Eberhard, Founder)。这样的类比迁移可以帮助我们更好地理解 and 扩展知识图谱中的关系模式。实际上, 在大多数知识图谱中, 组合模式的类比对象候选

对的数量极大。为减少检索候选对的成本, 受到局部性原理^[17]的启发, 通常分别选择具有高三元组分数的 m 个实体和 n 个关系, 分别进行两两配对。因此, 组合模式的类比对象如式(4)所示:

$$\mathbf{p}_{\text{comp}} = \{(\mathbf{h}_i, \mathbf{r}_i) \mid \text{Top}(\{f_{\text{kge}}(\mathbf{h}_i, \mathbf{r}_i, t) \mid \mathbf{h}_i \in \mathbf{E}_m, \mathbf{r}_i \in \mathbf{R}_n\})_{N_c}\} \quad (4)$$

3.3.2 模式类比对象聚合

为更好地捕捉这些模式类比对象的特征并提升其表示能力, 采用加权方式将这些类比对象聚合嵌入。对称模式类比对象聚合表示、多重模式类比对象聚合表示和逆模式类比对象聚合嵌入表示分别如式(5)、式(6)、式(7)所示。

$$\mathbf{sym}^+ = \sum_{\mathbf{h}_i \in \mathbf{E}_m} \mathbf{h}_i S(f_{\text{kge}}(\mathbf{h}_i, \mathbf{r}, t)) \quad (5)$$

$$\mathbf{mul}^+ = \sum_{\mathbf{r}_i \in \mathbf{R}_n} \mathbf{r}_i S(f_{\text{kge}}(\mathbf{h}, \mathbf{r}_i, t)) \quad (6)$$

$$\mathbf{inv}^+ = \sum_{\mathbf{r}_i^{-1} \in \mathbf{R}_n} \mathbf{r}_i^{-1} S(f_{\text{kge}}(t, \mathbf{r}_i^{-1}, \mathbf{h})) \quad (7)$$

其中, $S(\cdot)$ 表示 softmax 函数, 将实数向量转换为概率分布, 使得每个元素的取值范围在 0~1 之间, 并且所有元素的和等于 1。组合模式 \mathbf{comp}^+ 首先分别计算 $\mathbf{comp}_{\text{sym}}^+$ 和 $\mathbf{comp}_{\text{mul}}^+$ 的聚合嵌入, 通过 $g_{\text{kge}}(\mathbf{h}, \mathbf{r})$ 得到组合模式 \mathbf{comp}^+ 的聚合嵌入, 如式(10)所示:

$$\mathbf{comp}_{\text{sym}}^+ = \sum_{\mathbf{h}_i \in \mathbf{E}_m} \mathbf{h}_i S(f_{\text{kge}}(\mathbf{h}_i, \mathbf{r}, t)) \quad (8)$$

$$\mathbf{comp}_{\text{mul}}^+ = \sum_{\mathbf{r}_i \in \mathbf{R}_n} \mathbf{r}_i S(f_{\text{kge}}(\mathbf{h}, \mathbf{r}_i, t)) \quad (9)$$

$$\mathbf{comp}^+ = g_{\text{kge}}(\mathbf{comp}_{\text{sym}}^+, \mathbf{comp}_{\text{mul}}^+) \quad (10)$$

3.3.3 类比嵌入

通过对知识图谱嵌入模型进行训练, 得到了预训练模型参数表示 $\mathbf{M} = \{\mathbf{E}, \mathbf{R}, f_{\text{kge}}, \boldsymbol{\theta}\}$, 其中 \mathbf{E}, \mathbf{R} 和 f_{kge} 分别是实体嵌入表、关系嵌入表和模型 \mathbf{M} 的评分函数, $\boldsymbol{\theta}$ 是其他参数的集合。通过类比嵌入, RpAT 方法增强了模型 \mathbf{M} 在类比迁移方面的能力。利用模型的实体和关系嵌入来训练每种模式的类比嵌入, 从而获得类比对象的结果。这些类比对象不仅用于对结果的分析, 也作为链接预测任务的输出。在多重模式中, 类比嵌入 \mathbf{mul}_a 通过模型的关系嵌入向量 $\mathbf{V}_r^R \in \mathbf{R}^d$, 将三元组 $(\mathbf{h}, \mathbf{r}, t)$ 的关系 \mathbf{r} 的初始嵌入表示映射到多重模式的类比嵌入表示中, 如式(11)所示:

$$\mathbf{mul}_a = \frac{1}{2} (\mathbf{W}_1 (\mathbf{V}_r^R + \mathbf{r}) + \mathbf{W}_2 (\mathbf{V}_r^R \circ \mathbf{r})) \quad (11)$$

同理, 逆模式的类比嵌入 \mathbf{inv}_a 通过预训练模型 \mathbf{M} 获得基础的关系投影向量 $\mathbf{V}_{r^{-1}}^R \in \mathbf{R}^d$, \mathbf{r}^{-1} 逆关系的计算如表 1 所列, 逆模式的类比嵌入如式(12)所示。

$$\mathbf{inv}_a = \frac{1}{2} (\mathbf{W}_1 ((\mathbf{V}_{r^{-1}}^R + \mathbf{r}) + \mathbf{W}_2 (\mathbf{V}_{r^{-1}}^R \circ \mathbf{r})) \quad (12)$$

因为实体间通常存在多种不同的对称关系, 所以需要通过对模型 \mathbf{M} 的关系向量 $\mathbf{V}_r^R \in \mathbf{R}^d$ 和实体向量 $\mathbf{V}_h^E \in \mathbf{R}^d$, 将三元组 $(\mathbf{h}, \mathbf{r}, t)$ 中的实体初始嵌入 \mathbf{h} 和关系初始嵌入 \mathbf{r} 映射到对称模式的类比嵌入。对称模式的类比嵌入 \mathbf{sym}_a 如式(13)所示:

$$\mathbf{sym}_a = \frac{1}{2} ((\mathbf{W}_1 (\mathbf{V}_h^E \circ \mathbf{r}) + \mathbf{W}_2 (\mathbf{V}_r^R \circ \mathbf{h}))) \quad (13)$$

其中, $\mathbf{V}_h^E \in \mathbf{R}^d$ 是实体的投影向量, \mathbf{d}_e 是实体的隐藏维度, $\mathbf{W}_1, \mathbf{W}_2$ 是可学习的权重矩阵。组合模式类比迁移 \mathbf{comp}_a 的类比

函数结合了对称和多重模式的类比嵌入,如式(14)所示:

$$\mathit{comp}_a = g_{\text{kge}}(\mathit{sym}_a, \mathit{mul}_a) \quad (14)$$

3.3.4 损失函数

使用 \mathbf{o}^+ 来表示不同模式类比对象的聚合嵌入,而 \mathbf{o}^a 则用于表示特定模式的类比嵌入。设计损失函数的目标是减小 \mathbf{o}^+ 和 \mathbf{o}^a 在向量空间中的距离,即减小它们之间的差异性,增加它们之间的相似度。这样的策略有助于提升模型在理解和使用知识图谱中类比关系方面的能力。损失函数如式(15)所示:

$$\begin{aligned} \mathcal{L}(\mathbf{o}^+, (\mathbf{h}, \mathbf{r}, \mathbf{t})) = & \log \sigma(\alpha (\|\mathbf{o}^+ - \mathbf{o}^a\|_2 - f_{\text{kge}}(\mathbf{h}, \mathbf{r}, \mathbf{t}))) + \\ & \gamma \cdot \frac{1}{N} \sum_{i=1}^N \left(1 - \frac{\|\mathbf{o}_i^+ \cdot \mathbf{o}_i^a\|}{\|\mathbf{o}_i^+\| \cdot \|\mathbf{o}_i^a\|} \right) \end{aligned} \quad (15)$$

其中 α, γ 是训练的超参数, σ 是 *sigmoid* 激活函数, $\|\cdot\|_p$ 是 L_p 范式。对于不同的三元组而言,特定的关系模式的类比迁移推理能力并不相同,因此要为每种关系模式的损失函数添加关系模式权重参数。最终训练目标损失函数如式(16)所示:

$$\begin{aligned} \min \text{loss} = & \sum_{(\mathbf{h}, \mathbf{r}, \mathbf{t}) \in \mathcal{F}} (\beta_{\text{sym}} \mathcal{L}(\mathit{sym}_a, (\mathbf{h}', \mathbf{r}, \mathbf{t})) + \\ & \beta_{\text{inv}} \mathcal{L}(\mathit{inv}_a, (\mathbf{t}, \mathbf{r}^{-1}, \mathbf{h})) + \beta_{\text{mul}} \mathcal{L}(\mathit{mul}_a, (\mathbf{h}, \mathbf{r}', \\ & \mathbf{t})) + \beta_{\text{comp}} \mathcal{L}(\mathit{comp}_a, (\mathbf{h}', \mathbf{r}', \mathbf{t}))) \end{aligned} \quad (16)$$

当涉及不同关系模式类比迁移的推理能力时,引入 $\beta_{\text{sym}}, \beta_{\text{inv}}, \beta_{\text{mul}}$ 和 β_{comp} 来调整梯度下降,其计算类似于 *softmax* 函数。不同关系模式的权重参数如式(17)所示:

$$\begin{aligned} \beta_{\text{sym}} = & e^{f_{\text{kge}}(\mathit{sym}^+, \mathbf{r}, \mathbf{t})} / \mathcal{T}, \beta_{\text{inv}} = e^{f_{\text{kge}}(\mathit{inv}^+, \mathbf{t}, \mathbf{h})} / \mathcal{T} \\ \beta_{\text{mul}} = & e^{f_{\text{kge}}(\mathit{mul}^+, \mathbf{h}, \mathbf{r}, \mathbf{t})} / \mathcal{T} \\ \beta_{\text{comp}} = & e^{f_{\text{kge}}(\mathit{comp}^+, \mathbf{h}, \mathbf{r}, \mathbf{t})} / \mathcal{T} \\ \mathcal{T} = & e^{f_{\text{kge}}(\mathit{sym}^+, \mathbf{r}, \mathbf{t})} + e^{f_{\text{kge}}(\mathit{inv}^+, \mathbf{t}, \mathbf{h})} + \\ & e^{f_{\text{kge}}(\mathit{mul}^+, \mathbf{h}, \mathbf{r}, \mathbf{t})} + e^{f_{\text{kge}}(\mathit{comp}^+, \mathbf{h}, \mathbf{r}, \mathbf{t})} \end{aligned} \quad (17)$$

3.4 链接预测

本文在构建 RpAT 方法的评分函数时,借鉴了 KNN-LM^[22]的方法,采用了插值技术,这种技术巧妙地将基础模型的推理能力与类比迁移的推理能力结合起来。其定义如式(18)所示。这种插值策略不仅保留了基础模型的原始推理能力,还增强了模型对知识图谱深层次语义关系的理解,从而提高了链接预测的准确性和效率。

$$\begin{aligned} \text{score}(\mathbf{r}, \mathbf{t}) = & f_{\text{kge}}(\mathbf{h}, \mathbf{r}, \mathbf{t}) + \lambda_p R p_\tau (\alpha_{\text{sym}} f_{\text{kge}}(\mathit{sym}_a, \mathbf{r}, \mathbf{t}) + \\ & \alpha_{\text{mul}} f_{\text{kge}}(\mathbf{h}, \mathit{mul}_a, \mathbf{t}) + \alpha_{\text{inv}} f_{\text{kge}}(\mathbf{t}, \mathit{inv}_a, \mathbf{h}) + \\ & \alpha_{\text{comp}} f_{\text{kge}}(\mathit{sym}_a, \mathit{mul}_a, \mathbf{t})) \end{aligned} \quad (18)$$

其中, τ 表示 p 关系模式的逻辑规则集合; λ_p 表示 p 关系模式的超参数;而 $R p_\tau$ 为规则 τ 的平均置信度,用于衡量修改后的分数的可信度; $\alpha_{\text{sym}}, \alpha_{\text{inv}}, \alpha_{\text{mul}}, \alpha_{\text{comp}}$ 分别是对称模式、逆模式、多重模式和组合模式的参数权重。

4 实验结果与分析

4.1 数据集与评价指标

本文采用知识图谱补全任务中广泛使用的 FB15k-237 和 WN18RR 两个数据集进行对比实验。FB15k-237 是从 Freebase 知识图谱中提取的一个子集,它包含 14 541 个实体和 237 个关系类型;WN18RR 是从 WordNet 知识图谱中提取的

一个子集,它包含 40 943 个实体和 11 个关系类型。数据集详细信息如表 2 所列。

表 2 WN18RR 和 FB15k-237 数据集统计信息
Table 2 Statistics of WN18RR and FB15k-237 datasets

Dataset	# ent	# rel	# triples(train/valid/test)
WN18RR	40 943	11	86 835/3 034/3 134
FB15k-237	14 541	237	272 115/17 535/20 466

在评价指标方面,本文采用 3 种广泛使用的评估指标,分别为平均倒数排名 (Mean Reciprocal Rank, MRR) 和排名不大于 k 的比例 (Hit@ k)。MRR 和 (Hit@ k) 值越大证明方法效果越好。

4.2 对比方法

1) 基于距离的方法通过评分函数可以衡量实体之间的相似性或关联性。经典的方法有: TransE^[7], ANALOGY^[19], RotatE^[8], HAKE^[9], Rot-Pro^[10], PairRE^[11]。

2) 基于图神经网络的方法是利用神经网络直接在图结构上进行学习和推理。现有的方法有: R-GCN^[14], CompGCN^[15], SE-GNN^[13]。

3) 基于增强型的方法通过将知识图谱嵌入模型与附加信息相结合,增强学习嵌入的表达能力。现有的方法有: CAKE^[16], SPA^[19], PUDA^[17], REP^[18], Ankgc^[21]。

4.3 参数设置

在实验中,利用 $PCA=0.8$ 和 $HC=0.5$ 的组合来使用 AMIE3^[23] 挖掘逻辑规则把数据集的三元组划分成为 4 种关系模式。框架 RpAT 超参数选择设置如下: 对称模式、逆模式和多重模式的超参数在集合 $\{\pm 1, \pm 2, \pm 3, \pm 4, \pm 5, \pm 10, \pm 50, \pm 100\}$ 内调整,而组合模式的超参数在集合 $\{\pm 1 \times 10^{-5}, \pm 1 \times 10^{-4}, \pm 1 \times 10^{-3}, \pm 1 \times 10^{-2}, \pm 0.01, \pm 0.02, \pm 0.05, \pm 0.1, \pm 0.2, \pm 0.5, \pm 1, \pm 2, \pm 5\}$ 内调整。同时,使用网格搜索法来选择方法最优的超参数。搜索 4 种模式的类比对象的数目 N_s, N_i, N_m 和 $N_c \in \{1, 3, 5, 10, 20\}$, 4 种模式的基本权重 $\alpha_{\text{sym}}, \alpha_{\text{inv}}, \alpha_{\text{mul}}$ 和 $\alpha_{\text{comp}} \in \{0.01, 0.05, 0.1, 0.2, 0.3\}$ 。损失函数权重 α 和 γ 设为 1 和 10。

4.4 链接预测实验

为了验证本文提出的方法 RpAT 的有效性,我们将其与现有的先进方法进行了比较。实验结果如表 3 所列。表 3 中的数据清晰地展示了 RpAT 在两个数据集上的卓越表现,其在所有评估指标上均实现了显著的提升。这一结果表明,通过结合知识图谱中的逻辑规则, RpAT 能够更有效地捕捉实体关系间的语义关联性,从而提高了知识图谱嵌入任务的性能和链接预测的准确性。与 PUDA 和 REP 这两种增强型知识图谱嵌入模型相比, RpAT 在 FB15k-237 和 WN18RR 数据集上大多数评估指标均显示出优势。例如,在 FB15k-237 数据集上, RpAT 的 MRR 指标分别比 PUDA 和 REP 高出 6.2% 和 10.7%; 在 WN18RR 数据集上,其 MRR 指标分别提升了 4.8% 和 3.4%。这一性能提升归功于 RpAT 基于逻辑规则划分关系模式并采用类比方法进行知识图谱补全的策略,这证明了逻辑规则划分和模式类比方法在知识图谱嵌入中的必要性。此外,为了验证 RpAT 与广泛的知识图谱嵌入模型的兼容性,选择了 4 个基于距离的知识图谱嵌入模型:

TransE, RotatE, HAKE, PairRE 作为基础模型并进行比较。表 4 中的数据显示, RpAT 在所有指标上都显著优于这些基础模型, 特别是在 FB15k-237 数据集上, RpAT 的 MRR 指标提高了约 10%。这表明 RpAT 与广泛的知识图谱嵌入模型兼容性显著。值得注意的是, 基于 HAKE 的 RpAT 在 FB15k-237 数据集上实现了更加显著的提升, 这可能要归功于 HAKE 通过实体的深度分层来建模不同层次的层次结构,

从而促进了类比对象的迁移。我们还观察到, 与 WN18RR 数据集相比, RpAT 在 FB15k-237 数据集上的提升更为显著。我们推测导致这一差异的原因是 FB15k-237 具有更丰富的关系模式, 从而为逻辑规则的应用和模式类比的迁移提供了更多的机会。这些实验结果综合表明, RpAT 不仅在提高知识图谱嵌入质量方面具有显著优势, 而且在兼容性和准确率方面也表现出色。

表 3 RpAT 的实验结果

Table 3 Experimental results of RpAT

method	FB15k-237				WN18RR			
	MRR	Hit@1	Hit@3	Hit@10	MRR	Hit@1	Hit@3	Hit@10
TransE ^[7]	0.317	0.223	0.352	0.504	0.224	0.022	0.390	0.520
ANALOGY ^[19]	0.256	0.165	0.290	0.436	0.405	0.363	0.429	0.474
RotatE ^[8]	0.336	0.244	0.370	0.524	0.473	0.428	0.491	0.564
HAKE ^[9]	0.349	0.252	0.385	0.545	0.496	0.452	0.513	0.580
Rot-Pro ^[10]	0.344	0.246	0.383	0.540	0.457	0.397	0.482	0.577
PairRE ^[11]	0.348	0.254	0.384	0.539	0.455	0.413	0.469	0.539
R-GCN ^[13]	0.249	0.151	0.264	0.417	—	—	—	—
CompGCN ^[14]	0.355	0.264	0.384	0.535	0.479	0.443	0.494	0.546
SE-GNN ^[12]	0.365	0.271	0.399	0.549	0.484	0.446	0.509	0.572
CAKE ^[15]	0.321	0.226	0.355	0.515	—	—	—	—
SPA ^[18]	0.338	0.261	0.366	0.525	—	—	—	—
PUDA ^[16]	0.369	0.268	0.408	0.578	0.481	0.436	0.498	0.582
REP ^[17]	0.354	0.262	0.388	0.540	0.488	0.439	0.505	0.588
Ankge ^[20]	<u>0.385</u>	<u>0.288</u>	<u>0.428</u>	<u>0.572</u>	<u>0.500</u>	<u>0.454</u>	<u>0.515</u>	0.587
Ours	0.392	0.291	0.452	<u>0.572</u>	0.505	0.454	0.516	0.595

表 4 RpAT 与其他模型的兼容性

Table 4 Compatibility of RpAT with other models

method	FB15k-237				WN18RR			
	MRR	Hit@1	Hit@3	Hit@10	MRR	Hit@1	Hit@3	Hit@10
TransE	0.317	0.223	0.352	0.504	0.224	0.022	0.390	0.520
RpAT-TransE	0.355	0.255	0.379	0.536	0.238	0.033	0.401	0.529
RotatE	0.336	0.244	0.370	0.524	0.473	0.428	0.491	0.564
RpAT-RotatE	0.371	0.279	0.412	0.552	0.482	0.431	0.500	0.579
HAKE	0.349	0.252	0.385	0.545	0.496	0.452	0.513	0.580
RpAT-HAKE	0.392	0.291	0.452	0.572	0.505	0.454	0.516	0.595
PairRE	0.348	0.254	0.384	0.539	0.455	0.413	0.469	0.539
RpAT-PairRE	0.381	0.282	0.421	0.565	0.467	0.415	0.485	0.561

4.5 消融实验

为了准确评估 RpAT 模型中各个模块的不同补全推理能力, 我们在 FB15k-237 和 WN18RR 这两个数据集上进行了一系列消融研究。表 5 列出了对比成果。消融实验分为 3 个变体: 1) 移除了由逻辑规则构建的关系模式 (w/o relation-patterns), 仅保留了类比对象; 2) 去除了不同模式类比对象 (w/o analogy object), 仅保留了逻辑规则定义的关系模式; 3) 移除了所有模块 (w/o all), 仅依赖基础模型的预测功能。实验数据清晰地表明, 完整的 RpAT 方法在两个数据集上的所有评估指标均展现出了卓越的性能。例如, 在 FB15k-237 数据集上, MRR 指标分别提升了 5.3%, 6.2%, 12.3%, 而在 WN18RR 数据集上, MRR 指标分别提升了 0.4%, 1.6%, 1.8%。同样地, Hits@10 指标在 FB15k-237 数据集上分别提升了 1.9%, 3.1%, 4.7%, 在 WN18RR 数据集上分别提升了 0.2%, 2.2%, 2.6%。这些结果有力地证明了逻辑规则划

分的关系模式与类比迁移方法相结合能够显著提升基础模型的性能, 并突显了 RpAT 每个模块之间的相辅相成。

表 5 消融实验

Table 5 Ablation experiment

Models	FB15k-237		WN18RR	
	MRR	Hit@10	MRR	Hit@10
RpAT	0.392	0.572	0.505	0.595
w/o relation-pattern	0.372	0.561	0.503	0.594
w/o analogy object	0.369	0.554	0.497	0.582
w/o all	0.349	0.545	0.496	0.580

4.6 对比实验

对比实验中, 我们将本文提出的方法 RpAT 与两种当前最先进的方法 Ankge 和 SPA 在 FB15k-237 数据集上的训练时间和补全任务预测性能指标 MRR 进行了比较。图 3 给出了这 3 种方法在不同训练时间下的 MRR 值。从图中可以明显看出, 随着训练时间的增加, RpAT 不仅在 MRR 值上呈现

出稳定的趋势,而且这3种方法在推理时间、MRR指标的性能上都有着显著的差距。这表明RpAT在处理效率上优于其他两种方法。这种效率的提高得益于RpAT中采用的逻辑层对实体语义结构的精细划分,这一技术能够在不损害知识图谱补全任务质量的前提下,加快数据处理的速度。在MRR值的表现上,RpAT同样展现出了卓越的性能。在整个训练时间范围内,RpAT始终保持着最高的MRR值,这充分证明了其在准确性上超越了当前现有的方法。例如,当3种方法的训练时间同时达到30h,RpAT实现了0.392的MRR值,而Ankge和SPA的MRR值分别仅为0.381和0.338。这一对比凸显了RpAT方法在利用关系模式的特性方面所展现的显著优势。通过RpAT方法,能够实现高效的类比迁移,从而显著提升训练效率并增强预测的准确性。

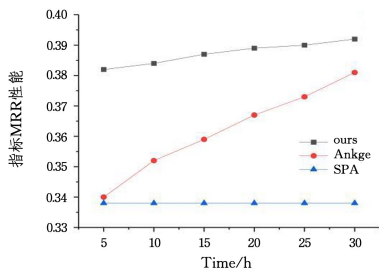


图3 对比方法训练

Fig. 3 Contrastive method training

结束语 本文提出了一种融合关系模式和类比迁移的知识图谱补全方法(RpAT)。该方法在知识图谱逻辑层根据实体和关系语义层次结构细分为三元组的关系模式,依据关系模式生成相似的类比对象进行迁移,从而避免发生实体和关系出现在不同关系语义结构中的问题。此外,依据已知的事实信息类比迁移,有效减少了推理时间和资源消耗。相比之前的基线模型,可以捕获更准确的语义嵌入。实验结果表明,RpAT方法优于现有的方法,提升了补全任务的性能。RpAT的关键在于如何准确地生成模式类比对象。未来,可以考虑根据知识图谱的特点,自适应地选择合适的模式类比对象进行迁移。

参考文献

[1] BOLLACKER K, EVANS C, PARITOSH P, et al. Freebase: a collaboratively created graph database for structuring human knowledge[C]//Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data, 2008:1247-1250.

[2] MILLER G A. WordNet: a lexical database for English [J]. Communications of the ACM, 1995, 38(11): 39-41.

[3] YANG L, CHEN H, LI Z, et al. Give us the facts: Enhancing large language models with knowledge graphs for fact-aware language modeling[J]. IEEE Transactions on Knowledge and Data Engineering, 2024, 36(7): 3091-3110.

[4] XUN T Y, LIU X H, ZHAO W D. Knowledge Graph and User Interest Based Recommendation Algorithm[J]. Chinese Journal of Computer Science, 2024, 51(2): 55-62.

[5] QU X, GU Y, XIA Q, et al. A survey on arabic named entity

recognition: Past, recent advances, and future trends[J]. IEEE Transactions on Knowledge and Data Engineering, 2023, 36(3): 943-959.

[6] CAO J, FANG J, MENG Z, et al. Knowledge graph embedding: A survey from the perspective of representation spaces[J]. ACM Computing Surveys, 2024, 56(6): 1-42.

[7] ZHANG T C, SUN X H, SUN X H, et al. Overview on Knowledge Graph Embedding Technology Research [J]. Journal of Software, 2023, 34(1): 277-311.

[8] BORDES A, USUNIER N, GARCIA-DURAN A, et al. Translating embeddings for modeling multi-relational data[C]//Advances in Neural Information Processing Systems. Cambridge, MA: MIT Press, 2013: 2787-2795.

[9] SUN Z, DENG Z H, NIE J Y, et al. Rotate: Knowledge graph embedding by relational rotation in complex space[C]//ICLR, 2019.

[10] ZHANG Z, CAI J, ZHANG Y, et al. Learning hierarchy-aware knowledge graph embeddings for link prediction[C]//Proceedings of the AAAI Conference on Artificial Intelligence, 2020: 3065-3072.

[11] SONG T, LUO J, HUANG L. Rot-pro: Modeling transitivity by projection in knowledge graph embedding[J]. Advances in Neural Information Processing Systems, 2021, 34: 24695-24706.

[12] CHAO L, HE J, WANG T, et al. Pairre: Knowledge graph embeddings via paired relation vectors[C]//Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, 2020: 4360-4369.

[13] LI R, CAO Y, ZHU Q, et al. How does knowledge graph embedding extrapolate to unseen data: a semantic evidence view[C]//Proceedings of the AAAI Conference on Artificial Intelligence, 2022: 5781-5791.

[14] SCHLICHTKRULL M, KIPF T N, BLOEM P, et al. Modeling relational data with graph convolutional networks[C]//The Semantic Web: 15th International Conference, ESWC 2018, Heraklion, Crete, Greece, June 3-7, 2018, Proceedings 15. Springer International Publishing, 2018: 593-607.

[15] VASHISHTH S, SANYAL S, NITIN V, et al. Composition-based multi-relational graph convolutional networks [C]//ICLR, 2020.

[16] NIU G, LI B, ZHANG Y, et al. CAKE: A scalable common-sense-aware framework for multi-view knowledge graph completion[C]//Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics, 2022: 2867-2877.

[17] TANG Z, PEI S, ZHANG Z, et al. Positive-unlabeled learning with adversarial data augmentation for knowledge graph completion[C]//Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence Main Track, 2022: 2248-2254.

[18] WANG H, DAI S, SU W, et al. Simple and Effective Relation-based Embedding Propagation for Knowledge Representation Learning[C]//Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence Main Track, 2022:

2755-2761.

- [19] JIN L, YAO Z, CHEN M, et al. A Comprehensive Study on Knowledge Graph Embedding over Relational Patterns Based on Rule Learning[C] // International Semantic Web Conference. Cham: Springer Nature Switzerland, 2023: 290-308.
- [20] LIU H, WU Y, YANG Y. Analogical inference for multi-relational embeddings[C] // International Conference on Machine Learning. PMLR, 2017: 2168-2178.
- [21] YAO Z, ZHANG W, CHEN M, et al. Analogical inference enhanced knowledge graph embedding[C] // Proceedings of the AAAI Conference on Artificial Intelligence. 2023: 4801-4808.
- [22] KHANDELWAL U, LEVY O, JURAFSKY D, et al. Generalization through memorization: Nearest neighbor language models [C] // ICLR. 2020.
- [23] LAJUS J, GALÁRRAGA L, SUCHANEK F. Fast and exact rule mining with AMIE 3[C] // The Semantic Web; 17th International Conference, ESWC 2020, Heraklion, Crete, Greece, May

31-June 4, 2020, Proceedings 17. Springer International Publishing, 2020: 36-52.



SONG Baoyan, born in 1965, Ph.D., professor. Her main research interests include data stream processing and graph data processing.



CHEN Ze, born in 1996, Ph.D candidate. His main research interests include graph data processing and natural language processing.

(责任编辑:杨雪敏)

CCF 和 ACM SIGCOMM 在京会谈

CCF 和 ACM SIGCOMM(美国计算机协会数据通信专业组)于 2025 年 2 月 15 日在北京进行会谈。ACM 数据通信专业组主席(SIGCOMM Chair)、美国 UIUC 大学教授 Matthew Caesar, 理事会成员兼教育总监(Education Director)、厦门大学教授向乔, SIGCOMM 驻中国大使(SIGCOMM Ambassador in China)、清华大学教授李丹, 以及 CCF 理事长孙凝晖、秘书长唐卫清、国际合作工作委员会主任陈熙霖、网络与数据通信专业委员会主任李克秋和互联网专业委员会副主任徐明伟参与此次会议。

双方就各自组织历史、发展历程和现状进行了简单介绍, Matthew Caesar 提出 SIGCOMM 会议汇集了来自学术界和工业界的中国出版物, 对中国人才的创新表示欣赏, 未来希望通过举办学会活动、夏令营合作、国际网络项目、奖励等促进双边学者沟通及技术发展。CCF 和 ACM SIGCOMM 参会代表一致认为彼此之间未来可就通讯领域技术及学术进行深度交流, 促进学会发展及技术进步。



从左到右:李丹、李克秋、陈熙霖、唐卫清、Mathew Caesar、徐明伟、向乔

附:

ACM(Association for Computing Machinery), 美国计算机协会, 于 1947 年成立。专注于计算机科学和技术, 促进技术发展和创新, 支持计算机领域的多样性、公平和包容性等。

CCF 与 ACM 是兄弟学会, 长期以来开展了一系列友好合作, 共同服务 CCF 会员。ACM 每年协办 CNCC, 并为 CNCC 邀请国际资深讲者。双方在深入了解和相互信任的基础上, 共同评选和颁发 CCF 青年科技奖(CCF-ACM Award for AI), 并完全按照国际惯例评奖。

据 CCF 微信公众号