软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software, 2023,34(3):1259−1276 [doi: 10.13328/j.cnki.jos.006788]
©中国科学院软件研究所版权所有.
E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel: +86-10-62562563

基于多粒度注意力网络的知识超图链接预测*

庞 俊[1,2], 刘小琪[1,2], 谷 峪[3], 王 鑫[4], 赵宇海[3], 张晓龙[1,2], 于 戈[3]

[1](武汉科技大学 计算机科学与技术学院, 湖北 武汉 430065)

[2](智能信息处理与实时工业系统湖北省重点实验室(武汉科技大学), 湖北 武汉 430065)

[3](东北大学 计算机科学与工程学院, 辽宁 沈阳 110169)

[4](天津大学 智能与计算学部, 天津 300354)

通信作者: 庞俊, E-mail: pangjun@wust.edu.cn

摘 要: 在知识图谱中进行链接预测是图谱补全的有效方法, 可以有效地改善知识图谱的数据质量. 然而, 现实生活中的关系往往是多元的, 这些包含多元关系的知识图谱可称为知识超图(knowledge hypergraph, KHG). 然而, 现有的知识超图链接预测模型忽略了多元关系的平等性(多元关系中实体不存在先后关系)与整体性(多元关系缺少一个实体则不成立). 针对以上问题, 首先提出了一种知识超图多元关系表示模型, 可以直接建模知识超图中的多元关系; 然后研究了一种基于多粒度神经网络的链接预测方法(hyperedge prediction based on multi-granular attention network, HPMG). 该模型将关系划分为多重粒度进行学习, 从不同粒度联合完成知识超图的学习和预测, 充分考虑了知识超图中不同维度多元关系的整体性. 接下来, 针对 HPMG 特征融合不充分的问题, 提出了基于多粒度注意力网络的知识超图链接预测方法 HPMG$^+$, 结合全局和局部注意力, 实现了不同特征的有区分融合, 进一步提高了模型的性能. 最后, 真实数据集上的大量实验结果验证了所提方法的效果显著地优于所有基线方法.

关键词: 知识超图; 链接预测; 多粒度; 嵌入学习; 注意力机制

中图法分类号: TP18

中文引用格式: 庞俊, 刘小琪, 谷峪, 王鑫, 赵宇海, 张晓龙, 于戈. 基于多粒度注意力网络的知识超图链接预测. 软件学报, 2023, 34(3): 1259−1276. http://www.jos.org.cn/1000-9825/6788.htm

英文引用格式: Pang J, Liu XQ, Gu Y, Wang X, Zhao YH, Zhang XL, Yu G. Knowledge Hypergraph Link Prediction Based on Multi-granular Attention Network. Ruan Jian Xue Bao/Journal of Software, 2023, 34(3): 1259−1276 (in Chinese). http://www.jos.org.cn/1000-9825/6788.htm

Knowledge Hypergraph Link Prediction Based on Multi-granular Attention Network

PANG Jun[1,2], LIU Xiao-Qi[1,2], GU Yu[3], WANG Xin[4], ZHAO Yu-Hai[3], ZHANG Xiao-Long[1,2], YU Ge[3]

[1](School of Computer Science and Technology, Wuhan University of Science and Technology, Wuhan 430065, China)

[2](Hubei Province Key Laboratory of Intelligent Information Processing and Real-time Industrial System (Wuhan University of Science and Technology), Wuhan 430065, China)

[3](School of Computer Science and Engineering, Northeastern University, Shenyang 110169, China)

[4](College of Intelligence and Computing, Tianjin University, Tianjin 300354, China)

Abstract: Link prediction in knowledge graphs is the most effective method for graph complementation, which can effectively improve the data quality of knowledge graphs. However, the relationships in real life are often multiple, thus these knowledge graphs containing multiple relationships can be called knowledge hypergraphs (KHGs). Unfortunately, the existing knowledge graph link prediction methods cannot be directly applied to knowledge hypergraphs, and the existing knowledge hypergraph link prediction models ignore the equality

* 基金项目: 国家重点研发计划(2020AAA0108503); 国家自然科学基金(62072083, 61972299)

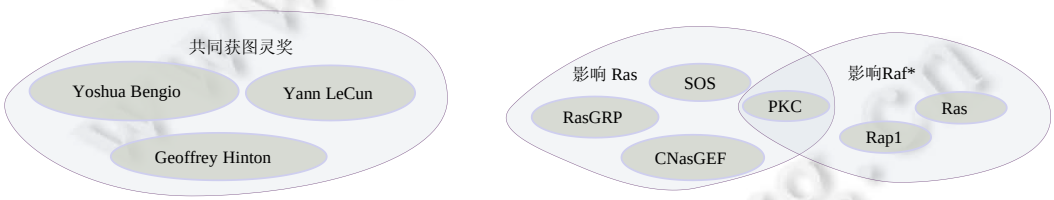
本文由“大数据治理的理论与技术”专题特约编辑杜小勇教授、杨晓春教授和童咏昕教授推荐.

收稿时间: 2022-05-15; 修改时间: 2022-07-29, 2022-09-07; 采用时间: 2022-09-23; jos 在线出版时间: 2022-10-27

(there is no sequential relationship among the elements in a multivariate relationship) and completeness (a multivariate relationship is not valid if it lacks elements) of the real-life multivariate relationships. To address these problems, a knowledge hypergraph multivariate representation model is firstly proposed, which can directly model the multivariate relationships in the knowledge hypergraph. Then, a multi-granularity neural network-based hypergraph prediction method (HPMG) is studied, which divides the relations into multiple granularities for learning and prediction from different granularities jointly. Next, to address the problem of inadequate HPMG feature fusion, HPMG$^+$ is proposed based on multi-granularity attention network for link prediction of knowledge hypergraphs, which combines all and local attention to achieve differentiated fusion of different features and further improves the performance of the model. Finally, extensive experimental results on real datasets verify that the proposed methods significantly outperform all baseline methods in terms of hyper-edge prediction.

Key words: knowledge hypergraph; link prediction; multi-granularity; embedding learning; attention mechanism

知识图谱补全，是改善知识图谱质量的核心问题. 而在知识图谱中进行链接预测，是图谱补全的有效方法[1]. 现有的知识图谱链接预测大多数仅仅考虑预测二元关系[2−7]，然而，现实生活中的关系往往是多元的. 如图 1(a)所示, Yoshua Bengio、Yann LeCun 和 Geoffrey Hinton 这 3 位科学家同时分享了 2018 年图灵奖；图 1(b)显示了人体内丝氨酸-苏氨酸激酶信号通路的蛋白质交互网络关系[8]，其中，SOS、PKC、RasGRP、CNasGEF 蛋白质共同影响了 Ras 蛋白质的含量，而 Ras，PKC，Rap1 又共同影响了 Raf*蛋白质的含量. 显然，这些多元关系中的实体具有平等性(没有主次、先后之分)和整体性(一个关系即为一个整体，缺一不可). 除此以外，在知识图谱公开数据集 Freebase 数据集中，超过 1/3 的关系都是多元关系[9]; Wikipedia 数据集中，超过 1/6 的关系都是多元关系[7]. 包含多元关系的图可以称为知识超图[10]，其中，每一个多元关系可以由超边建模. 本文主要研究知识超图链接预测问题，旨在预测超图中缺失的多元关系，以实现知识超图补全、改善图谱质量.



(a) 获奖关系　　(b) 蛋白质网络交互关系

图 1　知识超图示例

常见的知识图谱链接预测模型主要包括:

(1) TransE 方法[11]将关系建模为两个实体之间的向量加操作，然后通过评估$\|h+rt\|$来预测缺失的关系(h, t 和 r 分别表示头、尾实体和关系). 近年来, TransH[12]和 TransR[13]等 TransE 改进方法[14−18]可以获取更丰富的知识图谱结构信息，进行链接预测.

(2) 语义匹配模型[19−24]通过最小化张量匹配实体的潜在语义和向量空间表示中包含的关系，来度量关系的可信性.

(3) 高维嵌入模型基于神经网络的方法[25−28]，使用卷积和内积等运算来计算实体间的相似性，实现链接预测.

遗憾的是，以上方法都基于传统知识图谱中的二元关系进行建模，因此无法直接用来预测知识超图中缺失的多元关系.

近年来，学界也对知识超图链接预测有了一定的研究[29−37]. 现有方法的一种思路是，将多元关系转化为多个二元关系建模. 例如，HypE 算法[10]根据实体扮演的不同角色学习不同的滤波器，但是这种多元关系建模中的实体存在先后、主次之分(即非平等). 还有一种思路是直接建模多元关系，例如，NaLP 算法[34]使用 m 维的实体构建卷积网络，这种模型中虽然关系已经是多维度建模，但是没有考虑 2 到 m−1 维度的多元关系表达，因此在预测中会缺失大量少于 m 维的关系(即非整体). 而具有平等性和整体性的多元关系在现实生活中普遍存在，如图 1 所示. 学习多元关系的平等性和整体性，能丰富嵌入表示蕴含的语义信息，提高超图链接预测的

© 中国科学院软件研究所 http://www.jos.org.cn

性能.

不同于现有工作, 本文首先提出了一种基于多元关系平等性的知识超图多元关系表示模型, 可以直接建模知识超图中的多元关系. 然后研究了一种基于多粒度神经网络的知识超图链接预测方法 HPMG, 从多重粒度角度学习多元关系的整体性. HPMG 模型将关系划分为多重粒度进行学习, 能够学习到包含实体不同维度的、更丰富的多元关系信息, 从而提高知识超图链接预测的性能. 接下来提出了基于多粒度注意力网络的知识超图链接预测方法 HPMG$^+$, 进一步提高模型的性能.

本文的主要贡献如下.

1) 提出了一种知识超图多元关系表示模型, 得到一个包含多个实体的多元组来表示多元关系. 在建模时, 考虑了知识超图中多元关系的平等性.
2) 提出了一种基于多粒度神经网络的知识超图链接预测算法 HPMG, 从不同粒度联合完成知识超图的学习和预测. 充分考虑了知识超图中不同维度多元关系的整体性.
3) 提出了一种基于多粒度注意力网络的知识超图链接预测算法 HPMG$^+$, 引入一种基于注意力机制的多特征融合方法 MFF (multi-feature fusion), 对 HPMG 进行优化, 结合全局和局部注意力实现了不同特征的有区分融合, 提升了模型预测性能.
4) 大量知识超图真实数据集上的实验结果, 验证了本文提出的算法的有效性.

本文第 1 节简述相关工作. 第 2 节给出本文提出的多元组表示模型和问题的形式化定义. 第 3 节描述 HPMG 方法的总体框架和具体实现. 第 4 节介绍 HPMG$^+$算法. 第 5 节进行实验评估及分析. 第 6 节总结全文.

1 相关工作

本节主要介绍基于表示学习的知识超图链接预测的研究工作, 根据所用技术不同, 可分为 3 类: 距离模型方法、张量分解方法、神经网络方法. 下文将对这 3 类方法分别展开介绍.

1.1 距离模型方法

在知识图谱中使用距离模型进行链接预测的最典型方法是 TransE[11]. 它将知识图谱中的实体和二元关系都嵌入成低维向量, 认为知识图谱中的每个关系 r 是从实体 h 到尾实体 t 的平移变换, 即满足 $h+r\approx t$. 后续又有很多基于距离模型的变体被提出来, 但大都只能处理二元关系. 用距离模型进行超图链接预测的思路是: 首先, 将关系建模为多元关系中实体之间的某种转换操作; 然后, 根据实体和关系之间的联系学习嵌入表示, 进而运用到链接预测任务中. Wen 等人提出一种 m-TransH[9]方法, 将实体映射到知识超图多元关系超平面, 用映射结果的加权来定义评分函数. m-TransH 首次采用基于距离模型的方法来解决知识超图链接预测问题, 但存在不具备完全表达性的问题. Zhang 等人提出一种 RAE[29]方法, 在 m-TransH 基础上进一步改进了关联性假设, 考虑了 2 个实体同时出现在一个多元关系中的概率, 将该概率引入损失函数, 并运用全连接神经网络对模型进行训练. 但是 RAE 在建模中涉及星到团的转换, 会导致某些属性特征永久丢失.

距离模型的一个共同缺陷是, 大多数平移距离模型不具有完全表达性[38]. 因此, 它们在关系建模方面存在一定的局限性. 同时, 由于距离模型中需要构造实体 h 到尾实体 t, 将其扩展到知识超图的基本方法是增加尾实体的数量, 因此使用距离模型进行超图链接预测, 难以保证多元关系中的实体具有平等性.

1.2 张量分解方法

基于张量分解的方法将关系表示为高阶张量, 然后将高阶张量分解为多个低阶张量来学习嵌入表示. 由于此类方法在二元知识图谱中取得了较好的效果, 因此研究人员将基于张量分解的知识图谱链接预测方法推广到知识超图中. 近年来提出的 GETD[31]模型就是对 TuckER[24]模型在处理知识超图链接预测问题上的扩展, 结合 Tucker 分解[39]和 Tensor Ring 分解[40]. GETD 首先将多元关系的高阶张量表示分解为一个核张量和若干个因子张量, 为了解决核张量参数过多问题, 继续将核张量分解为多个三阶张量以降低模型复杂度. GETD 模型虽然能够完全表达所有关系类型, 且能有效减少模型的复杂性, 但其是针对固定元知识超图设计的, 不适合

© 中国科学院软件研究所 http://www.jos.org.cn

处理混合元的知识超图. HSimplE[10]和 HypE[10]是对 SimplE[38]的扩展, HSimplE 根据实体在给定关系中的位置来更新实体嵌入, HypE 根据实体在元组中的位置不同来学习不同的卷积核, 计算对应位置上实体和关系的嵌入表示. 这些模型能够有效地捕获实体和关系间的潜在交互, 但其假设实体在多元关系中扮演不同角色, 即构成多元关系的各实体是不平等的, 而所含实体平等的多元关系在现实中是普遍存在的.

如上所述, 由于张量分解方法直接将多元关系表示为高阶张量, 但是张量的大小是固定的, 因此其难以处理混合元的知识超图. 且研究者考虑了实体在多元关系中不同位置的信息差异, 针对具有不平等性的多元关系进行建模. 本文处理的是混合元知识超图, 并且考虑了多元关系中实体的平等性和整体性; 本文提出的多元关系表示模型具有平等性, 改变其中实体的位置, 多元关系保持不变.

1.3 神经网络方法

基于神经网络的模型能够学习实体之间的交互信息、知识图谱的结构信息等, 在关系建模、结构建模等方面提升了表示学习的性能. 因此, 现有大量神经网络方法被运用于知识超图链接预测任务. 根据所用技术不同, 该类方法可分为 3 类: 基于传统神经网络的模型、基于图神经网络的模型和基于超图神经网络的模型.

基于传统神经网络的模型学习多元关系内部的交互信息, 例如, NaLP[34]和 tNaLP^{+}[35]将每个多元关系表示为一组键值对(其中键就是关系, 值就是实体), 然后采用卷积和全连接神经网络对多元关系进行学习. 与 NaLP 相比, tNaLP^{+}在学习中融入了类型嵌入, 并对负采样方法进行了改进. 这两种模型对建模的关系类别均有严格的要求, 限制了模型的应用场景. HINGE[36]和 NeuInfer[37]考虑结构信息的主次, 认为仅用键值对表示多元关系会得到次优模型, 因此将多元关系表示为主三元组和一组辅助键值对. HINGE 通过卷积分别学习主三元组的关系特征表示和对应主三元组联合键值对的关系特征表示, 然后利用最小化操作融合这些相关性特征, 得到多元关系的整体特征表示, 再将其送入全连接层, 获得该关系成立的得分. NeuInfer 通过全连接神经网络分别计算主三元组的可信性得分以及主三元组和键值对的相容性得分, 利用加权和获得多元关系成立的最终得分. HINGE 和 NeuInfer 模型假设多元关系内部存在主次关系, 然而并非所有多元关系都如此, 很多多元关系包含的实体具有平等性和整体性. 基于图神经网络的模型结合图结构信息完成知识超图的建模 StarE[32]是首次且目前最优的使用图神经网络进行知识超图链接预测的方法, 采用图卷积神经网络学习目标实体的多跳领域信息, 可以有效地学习图结构信息. 它通过一个消息传递网络来获取实体和关系的嵌入, 然后将嵌入表示输入一个 Transformer 解码器, 来对多元关系评分. StarE 只能预测主三元组的实体, 且将多元关系分为主三元组和辅助键值对, 并未考虑多元关系内部的平等性和整体性. Hyper-SAGNN[41]提出了一种新的基于自注意力的图神经网络学习多元关系, 该网络可以学习节点的嵌入情况并预测具有不同元数多元关系的知识超图. 但是其忽略了关系的语义信息, 且仅能建模关系类别数量极少的知识超图, 通常仅为 3 或 4 类, 不适用于多关系类别的知识超图. 目前, 基于超图神经网络进行知识图谱嵌入学习的研究已见报道, 但完成知识超图链接预测的研究尚未见报道. H-AKRL[42]提出了一种基于超图神经网络的知识表示学习模型, 在更高层次上建模实体和属性之间的相关性. 它充分利用属性信息和结构信息之间的互补关系, 解决了知识图谱中长尾实体嵌入表示的问题. 虽然 H-AKRL 与本文提出的方法都能完成嵌入学习任务, 但本文的研究对象是知识超图而不是知识图谱. 本文首次提出了一种超图神经网络, 完成知识超图链接预测任务.

神经网络方法充分利用神经网络对非线性复杂关系的建模能力, 通过对图谱的结构特征和语义特征的学习, 实现对缺失元素的有效预测. 相比于张量分解的方法, 此类方法在处理知识超图时更加灵活多变, 可以学习混合元的知识超图信息. 相比于距离模型的方法, 此类方法具有完全表达性. 因此, 本文对该类方法展开研究. 现有神经网络的方法大都引入一组限定符(即键值对), 缺乏用限定符表示关系的粒度; 或者直接学习所有的限定符, 无法将主要信息的限定符和特定于关系的限定符区分. 本文提出的多元组表示模型直接将多元关系表示为多元组, 考虑了多元关系的平等性, 避免了因引入限定符而出现的每个限定符的主次无法区分、引入的限定符个数无法判断等问题. 且考虑到多元关系具有整体性, 提出的 HPMG 从多粒度角度学习多元关系, 充分挖掘出多元关系粗粒度和细粒度上的特征, 充分考虑了知识超图中不同维度多元关系的整体性. 此外, 引入 MFF 方法来进行特征融合, 进一步提升模型的性能.

© 中国科学院软件研究所 http://www.jos.org.cn

2 多元组表示模型和问题定义

本节给出了本文提出的多元关系多元组表示模型和知识超图链接预测的形式化定义.

定义 1(多元关系). 一个多元关系 $f=(h_1,h_2,\ldots,h_m,r)$, 其中, $h_1,h_2,\ldots,h_m$ 是实体, r 代表关系属性.

与现有相关研究采用的多元关系表示模型不同, 本文提出的上述多元组模型中实体无头尾之分, 为位置无关的一个整体. 多元关系平等性和整体性形式化定义如下所示.

- 多元关系平等性: 已知多元关系 $f_1=(h_1,\ldots,h_i,\ldots,h_j,\ldots,h_m,r)$, 若更换 f_1 中实体 h_i 和 h_j 的位置可得 $f_2=(h_1,\ldots,h_j,\ldots,h_i,\ldots,h_m,r)$, 那么等式 $f_1=f_2$ 成立.
- 多元关系整体性: 已知多元关系 $f_1=(h_1,\ldots,h_{i-1},h_i,h_{i+1},\ldots,h_m,r)$, 若 f_1 缺失任意实体 h_i 后变成 $f_2=(h_1,\ldots,h_{i-1},h_{i+1},\ldots,h_m,r)$, 那么多元关系 f_2 不成立, 即 f_2=False.

定义 2(知识超图). 一个知识超图 $KHG=\{H,F\}$由节点(实体)和超边(多元关系)组成, 其中, H 为全部的实体集合, F 为全部多元关系的集合.

定义 3(知识超图链接预测). 假设 F_{true} 表示现实世界中已知知识超图 KHG 的全部关系集合, F 表示 KHG 中已知的关系集合, 知识超图链接预测旨在满足 $\max(|F'|)$条件下查找 KHG 中缺失的多元关系集合 $F'=F_{\text{true}}-F$. $|F'|$表示集合的基数, $\max(x)$表示 x 取最大值.

根据定义 3, 预测缺失的多元关系可以转化为预测多元关系中缺失的实体. 如图 2 和图 1(b)所示, 已知两个分别影响 Ras 和 Raf*蛋白质含量的交互关系, 知识超图链接预测旨在判断是否可以加入 SOS、RasGRP、CNasGEF 实体, 使得图 2 右侧的多元关系成立.

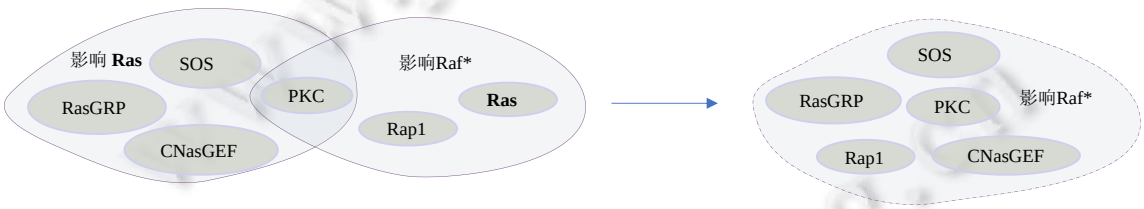


图 2 知识超图链接预测

3 HPMG 方法

多元关系具有平等性和整体性. 上文已基于平等性提出了多元组表示模型, 本节将在多元组表示模型基础上提出知识超图链接预测方法 HPMG, 该方法采用多重粒度的思想考虑多元关系的整体性. 现有知识超图链接预测方法一般只将目标元组看作一个整体, 进行嵌入表示的学习(即粗粒度). HPMG 除此之外, 还融合了各个实体组合对目标元组嵌入表示的贡献(即细粒度). 其中, i-细粒度贡献($1\leqslant i\leqslant m-1$)指从 m 个实体中任选 i 个实体获得的所有实体组合对目标元组的嵌入表示的综合贡献(详情请见第 3.3 节).

本节详细介绍了本文提出的基于多元组表示模型的知识超图链接预测方法 HPMG: 首先阐述 HPMG 方法的整体框架, 然后详细地展开介绍 HPMG 方法的粗粒度、δ-细粒度评价, 最后给出了总评分函数、损失函数以及 HPMG 方法的训练过程.

3.1 HPMG整体框架

HPMG 的基本思想是: 首先, 通过二维卷积分别学习每个关系粗粒度的超边特征和δ-细粒度的实体特征; 然后, 基于上一步获取的粗、δ-细粒度特征分别进行粗粒度评估和δ-细粒度评估; 最后, 对得到的粗、δ-细粒度评估结果进行加权求和操作, 计算出每个关系的最终评分, 完成链接预测. HPMG 方法整体框架如图 3 所示, 图 3(a)首先将多元关系 f 的实体和关系进行初始化嵌入表示. 为了更好地学习 f 内部的语义信息, 将 f 输入多头自注意力神经网络, 学习 f 内部的关联性, 得到更新后的嵌入表示(由于该操作在实际操作中较常见, 因此在图 3 中没有体现. 图 3 输入的 $h_1,h_2,\ldots,h_m,r$ 为经过多头自注意力网络更新后的嵌入表示). 然后, 将更新后

© 中国科学院软件研究所 http://www.jos.org.cn

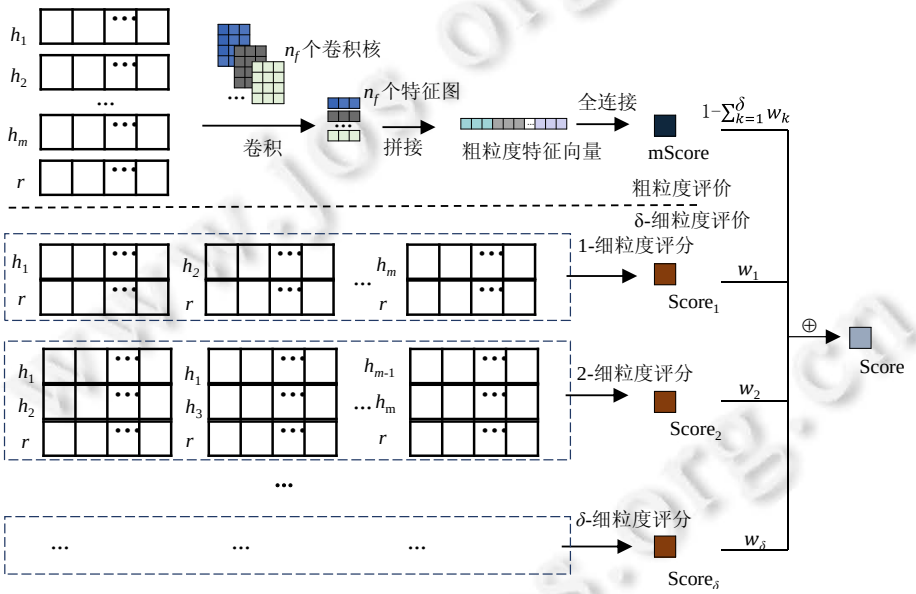
的多元关系嵌入表示分别进行粗粒度、δ-细粒度评价.

1) 粗粒度评价: 将多元关系的所有实体$(h_1,h_2,\ldots,h_m)$和关系 r 拼接后输入卷积神经网络, 得到特征表示 Fea_n. 然后, 将 Fea_n 进行拼接并输入全连接神经网络, 得到多元关系粗粒度的评分.

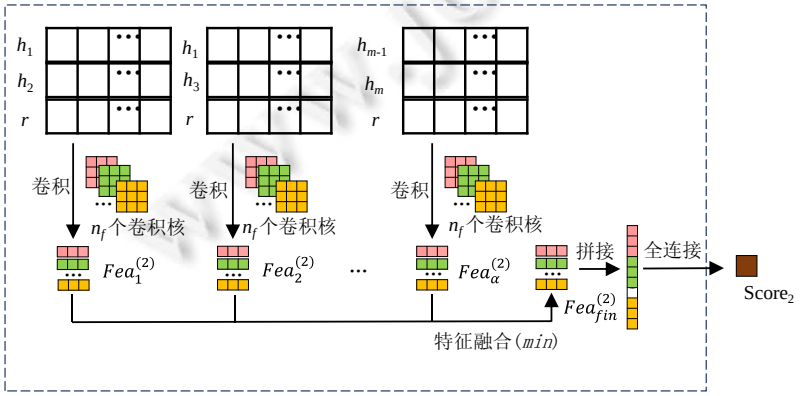
2) δ-细粒度评价: 进行 1 细粒度至δ细粒度的评分, 如图 3(b)所示.

其中, i 细粒度评分(记作 i-细粒度评分, i=1,2,…,δ)过程如下所示.

将 f 的 i 粒度组合中的每个元素分别和关系 r 进行拼接, f 的 i 粒度组合: 从多元关系 $f=(h_1,h_2,\ldots,h_m,r)$的实体集合$\{h_1,h_2,\ldots,h_m\}$中任意选出 i 个实体的所有可能结果组成的集合, 然后将这 C_m^i 组向量分别输入卷积神经网络, 得到 C_m^i 个特征图. 利用最小值(min)操作融合这 C_m^i 个特征图, 计算 i-细粒度上的特征表示 $Fea_{fin}^{(i)}$. 将 $Fea_{fin}^{(i)}$ 拼接后输入全连接神经网络, 计算多元关系的 i-细粒度的评分. 最后, 将多元关系粗粒度评分、δ-细粒度的δ个评分进行加权求和, 得到该多元关系的最终得分. 若目标关系为一元关系, 则仅采用“粗粒度”评分作为它的最终评分. 下文将分别详细介绍粗粒度评价和细粒度评价.



(a) HPMG 方法整体框架简图



(b) 2-细粒度评分具体步骤

图 3 HPMG 方法整体框架示意图

© 中国科学院软件研究所 http://www.jos.org.cn

3.2 粗粒度评价

粗粒度评价旨在学习多元关系包含的所有实体($h_1,h_2,\ldots,h_m$)与关系 r 之间的联系, 获取更丰富的超边语义特征. 关系包含的所有实体、关系之间一般存在着联系, 将其进行整体学习, 不仅可以实现所有实体之间信息的互补, 还可以使用所有实体的特征对关系信息进行更新. 具体方法伪代码如算法 1 所示.

算法 1. 粗粒度评价.

输入: 多元关系嵌入表示 $f=(h_1,h_2,\ldots,h_m,r)$.

输出: 多元关系粗粒度评分 *mScore*.

1. $F_{all}\leftarrow concat(h_1,h_2,\ldots,h_m,r)$; //计算表示多元关系实体和关系嵌入向量串联的集合 F_{all}
2. $F_{all}\leftarrow M_{self\text{-}att}(K_i,Q_i,V_i)$; //公式(2)使用多头自注意力网络更新 F_{all}
3. $Fea_{Coa}\leftarrow\sigma(BN(conv2(F_{all})))$; //计算特征矩阵 Fea_{Coa}
4. **return** $mScore\leftarrow FCN_1(Flatten(Fea_{Coa}))$. //公式(4)

为了获取含多元关系内部语义信息的关系和实体的嵌入表示, 首先对关系 f 包含的所有实体和关系进行串联操作, 得到 $F_{all}=concat(h_1,h_2,\ldots,h_m,r)$(第 1 行); 然后, 通过多头自注意力神经网络学习 F_{all} 内部语义差别并更新 F_{all}, 计算过程如公式(1)、(2)所示(第 2 行).

$$F_{all}=M_{self\text{-}att}(F_{all}W_i) \tag{1}$$

其中, $W_i=(W_i^k,W_i^q,W_i^v)$, W_i^k, W_i^q, W_i^v 表示权重矩阵, i=1,2,…,5. 矩阵 F_{all} 分别与权重矩阵相乘得到 K_i, Q_i, V_i. $M_{self\text{-}att}(\cdot)$为多头自注意力函数:

$$M_{self\text{-}att}(K_i,Q_i,V_i)=Concat\left(softmax\left(Q_1K_1^T/\sqrt{d_k}\right)V_1,\ldots,softmax\left(Q_5K_5^T/\sqrt{d_k}\right)V_5\right)W^0) \tag{2}$$

其中, W^0 为参数矩阵, d_k 表示键 K 的维数, *Concat* 为向量拼接操作, $T=m+1$.

然后, 获取多元关系 f 的粗粒度特征. 将 F_{all} 输入二维卷积神经网络, 进行标准化处理后, 使用 ReLU 非线性激活函数进行处理, 从而得到一组超关系特征矩阵 Fea_{Coa}, 计算过程如公式(3)(第 3 行).

$$Fea_{Coa}=\sigma(BN(conv2(F_{all}))) \tag{3}$$

其中, $BN(\cdot)$为 *BatchNorm2d* 函数, $\sigma(\cdot)$为非线性激活函数 ReLU, $conv2(\cdot)$表示卷积核大小为$(m+1)\times3$ 的二维卷积操作.

最后, 计算多元关系 f 的粗粒度评分. 将特征矩阵 Fea_{Coa} 按列展开后输入全连接神经网络, 计算出粗粒度评分 *mScore*, 计算过程如公式(4)所示(第 4 行).

$$mScore=FCN_1(Flatten(Fea_{Coa})) \tag{4}$$

其中, *Flatten* 表示将张量按列展开, FCN_1 表示全连接操作.

将参照文献[43]采用的算法复杂度分析方法分析本文提出的算法的复杂度. 定义 n_h 表示实体个数, n_r 表示关系个数, d 表示实体和关系空间维度大小, n_f 表示卷积核个数, K 表示卷积核大小, C_{in} 表示输入通道数目, C_{out} 表示输出特征图的大小. 算法 1 主要包括卷积和全连接两大过程. 经分析, 卷积和全连接过程时间复杂度分别为 $O(n_fK_{cor}C_{in}C_{out})$和 $O(C_{out}\times1)=O(C_{out})$. 其中, K_{cor} 表示粗粒度评价中卷积核的大小. 由于 $n_f\times K_{cor}\times C_{in}>>1$, 因此, 算法 1 的时间复杂度为 $O(n_fK_{cor}C_{in}C_{out})$. 算法 1 卷积和特征图参数个数分别为 $5\times3\times n_f=15n_f$ 和 $3\times n_f=3n_f$, 因此, 算法 1 的空间复杂度为 $O(n_f)$.

3.3 δ-细粒度评价

δ-细粒度评价旨在学习关系包含的所有实体的不同粒度组合分别与关系 r 之间的联系, 获取更丰富的实体语义特征. 多元关系包含的不同粒度的实体一般也与关系分别存在着联系. 分别对它们进行学习, 可以根据实体的特性进一步更新实体的表示. δ-细粒度学习是指对具有 m 个实体的多元关系从 1 细粒度到δ细粒度的学习($\delta=m-1$). 其中, i 细粒度的学习过程如下: 首先, 获取 i 细粒度的 C_m^i 个实体组合, 分别结合关系 r 的特征; 然后, 通过最小值操作融合这 C_m^i 个特征, 得到 i 粒度的特征表示; 最后, 通过全连接神经网络得到 i-细粒度的评分. i 细粒度的实体组合是指当细粒度为 i 时, m 个实体中的任意 i 个实体进行的组合, 其组合个数为

© 中国科学院软件研究所 http://www.jos.org.cn

C_m^i, i=1,2,…,δ. 具体方法伪代码如算法 2 所示.

算法 2. i-细粒度评价.

输入: 多元关系嵌入表示 $f=(h_1,h_2,\ldots,h_m,r)$.

输出: 多元关系 f 的 i-细粒度的评分 $Score_i$, i=1,2,…,δ.

1. $F_j^{(i)} \leftarrow concat(Com(h_j)_i, r)$; //$Com(h_j)_i$ 表示实体的不同粒度组合
2. $Fea_j^{(i)} \leftarrow \sigma(BN(conv2'(F_j^{(i)})))$; //公式(5)
3. $Fea_{fin}^{(i)} \leftarrow \min(Fea_1^{(i)},...,Fea_\alpha^{(i)})$; //公式(6)
4. **return** $Score_i \leftarrow FCN_2(Flatten(Fea_{fin}^{(i)}))$. //公式(7)

首先, 分别串联经过多头自注意力神经网络更新后的多元关系嵌入向量, 包括实体的不同粒度组合 $Com(h_j)_i$ 和关系 r, 记作 $F_j^{(i)} = concat(Com(h_j)_i, r)$. 其中, $j=1,2,...,C_m^i$, i=1,2,…,δ(第 1 行). $Com(h_j)_i$ 表示 m 元关系的 i 粒度实体组合, 例如, 当 i 为 2 时, $Com(h_j)_2=concat(h_1,h_2),concat(h_1,h_3),\ldots,concat(h_{m-1},h_m)$. 然后, 为了获取多元关系 f 实体的 i 维特征, 将 $F_j^{(i)}$ 输入一个二维卷积神经网络, 经过标准化并使用 ReLU 非线性激活函数进行处理, 从而得到实体的不同粒度组合 $Com(h_j)_i$ 基于关系 r 的特征 $Fea_j^{(i)}$ (第 2 行). 计算过程如公式(5)所示.

$$Fea_j^{(i)} \leftarrow \sigma(BN(conv2'(F_j^{(i)}))) \tag{5}$$

其中, $conv2'$为卷积核大小为$(i+1)\times 3$ 的二维卷积操作.

接着, 完成 $Fea_j^{(i)}$ 的特征融合. 使用最小值特征融合方法对上一步得到的 C_m^i 组特征分别进行特征融合(第 3 行). 计算过程如公式(6)所示.

$$Fea_{fin}^{(i)} = \min(Fea_1^{(i)},...,Fea_\alpha^{(i)}) \tag{6}$$

其中, min(·)表示取所有 $Fea_j^{(i)}$ 每个维度的最小值作为该维度的结果(j=1,2…,α), $\alpha = C_m^i$.

最后, 计算 f 的 i-细粒度评分(第 4 行). 将每个 $Fea_{fin}^{(i)}$ 按列展开后, 输入全连接神经网络, 计算出 i-细粒度评分 $Score_i$, 计算过程如公式(7)所示.

$$Score_i = FCN_2(Flatten(Fea_{fin}^{(i)})) \tag{7}$$

其中, FCN 为全连接操作. i=1,2,…,δ.

分析易知, i-细粒度评分的每个卷积的时间复杂度为 $O(K_i n_f C_{in} C_{out})$. 其中, K_i 表示 i-细粒度卷积中卷积核大小, n_f, C_{in}, C_{out} 的含义请参照第 3.2 节. 由于 i-细粒度的所有 C_m^i 次卷积同步执行, 故 i-细粒度评分的时间复杂度为 $O(K_i n_f C_{in} C_{out})$. 并行执行$\delta=m-1$ 次细粒度评分, 因此, 算法 2 的时间复杂度为 $O(K_{m-1} n_f C_{in} C_{out})$. 算法 2 的空间复杂度为 $O\left(\sum_{i=1}^{m-1} C_m^i K_i n_f\right)$.

3.4 总评分函数、损失函数和模型训练

本节介绍了本文使用的总评分函数、损失函数以及模型训练过程.

3.4.1 总评分函数

采用流行的加权和操作, 实现目标多元关系的最终评分, 其计算过程如公式(8)所示.

$$Score = \left(1-\sum_{k=1}^{\delta} w_k\right)\times mScore + \sum_{k=1}^{\delta} w_k \times Score_k \tag{8}$$

其中, 权重因子 $w_k\in(0,1)$. 为了学习 $1-m-1$ 维的特征, δ设置为 $m-1$. 如果待评估的对象是一元关系, 则只取 $mScore$ 作为最终评分. $mScore$ 和 $Score_k$ 的计算方法分别如公式(4)和公式(7)所示.

3.4.2 损失函数

HPMG 通过最小化一个 softplus 损失函数来训练本模型的参数, 其损失 Loss 函数如公式(9)所示.

$$Loss = \sum_{\omega\in\Omega} \log(1+e^{-\delta(\omega)}) + \log(1+e^{\delta(\omega')}) \tag{9}$$

其中, Ω表示多元关系的输入集. 每个正例多元关系$\omega\in\Omega$都包含一个 n 元组$(h_1,h_2,\ldots,h_m,r)$. 负例ω'通过破坏任

© 中国科学院软件研究所 http://www.jos.org.cn

意一个正例的实体 h_i 生成. $\sigma(\omega)$和$\sigma(\omega')$分别表示模型 HPMG 预测正例ω和负例ω'的得分.

3.4.3 训练过程

训练时, HPMG 通过反向传播学习模型参数, 使用具有学习速率λ的 Adam 作为优化器, 还使用非线性激活函数(ReLU)和批标准化方法来加速模型的训练过程. HPMG 方法的训练过程如算法 3 所示. 第 2–14 行不断迭代执行, 直到执行指定训练轮数 *epoch* 结束为止. 第 4–13 行对多元关系进行评分, 通过不断最小化损失进行参数训练和嵌入更新. 由于 HPMG 模型主要完成粗粒度学习和δ-细粒度学习, 因此, 根据算法 1 和算法 2 的复杂度分析, 可以计算出 HPMG 的时间复杂度和空间复杂度分别为 $O(|T_{batch}|K_{m-1}n_fC_{in}C_{out})$和 $O\left(n_fK_{cor}+\sum_{i=1}^{m-1}C_m^iK_in_f+n_hd+n_rd\right)$. 其中, $|T_{batch}|$表示训练过程中每次迭代的批量大小, 式中其他符号的含义请参照第 3.2 节. 由于 n_h 和 $n_r>>n_f$ 和 K_{cor}, 且实际情况下, m 取值一般不超过 10, 因此, 算法 2 的空间复杂度化简为 $O(n_hd+n_rd)$.

算法 3. HPMG 模型.

输入: 训练集 T, 损失函数 Loss, 训练轮数 *epoch*, 嵌入维度 d, 卷积操作的滤波器个数 n_f, 批处理大小 *batch*.

输出: 训练集中的实体嵌入表示 h 和关系嵌入表示 r.

1. 初始化关系和实体的嵌入;
2. **for** i=1,…,*epoch* **do**
3. 将 T 划分为大小为 *batch* 的批 T_{batch}^{+};
4. **for** $\forall\ T_{batch}^{+}\in T$ **do**
5. $T_{batch}^{-}\leftarrow$ 生成 T_{batch}^{+}的负样集;
6. $T_{batch}\leftarrow T_{batch}^{+}\cup T_{batch}^{-}$;
7. $Loss\leftarrow 0$;
8. **for** $\forall F\in T_{batch}$ **do** //F 表示 T_{batch} 中包含的多元组
9. $Score\leftarrow$通过公式(8)计算最终评分;
10. **end for**
11. $Loss\leftarrow$通过公式(9)计算损失;
12. 通过反向传播更新关系嵌入 r 和实体嵌入 h 以及 HPMG 中的所有参数;
13. **end for**
14. **end for**
15. **return** h 和 r

4 HPMG$^+$方法

现有知识超图链接预测方法[34–37]在获得多组特征后, 均运用求对应维度最小值的方法完成特征融合. 这种求最小值特征具有极大的随机性, 可能出现极端情况. 例如, 当某组中的某个特征表示的每个特征值恰好是本组各维特征的最小值时, 那么特征融合的结果就是该特征表示自己. 显然, 这将失去特征融合的重要意义, 对预测结果产生较大不利影响. 因此, 寻找一种能够充分、整体地学习所有特征的特征融合方法是很必要的.与现有最小值方法相比, 新方法能够更充分地融合特征, 从而融合结果不具备随机性, 不易出现极端情况. 因此, 本文提出了 HPMG 的优化算法 HPMG$^+$.

本节主要介绍了 HPMG$^+$算法. HPMG 采用最小值特征融合方法, 将每一维度的最小值作为该维度的最终融合特征. 该方法虽然简单, 但存在极大的随机性, 会导致特征融合不够充分, 从而影响最终的预测性能. 不同特征一般对学习任务产生不同的贡献, 如特征融合时考虑该特性, 可以使特征有区分地融合, 从而提升知识超图链接预测的性能. 注意力机制是一种常用于描述不同要素的贡献不同的机制, 因此, 本文引入了一种基于注意力机制的多特征融合方法 MFF, 完成 HPMG 模型中 m 个实体的细粒度特征融合任务, 得到 HPMG$^+$算法. 为了阐述 MFF 如何完成细粒度特征融合任务, 图 4 列举了 HPMG$^+$的 2-细粒度评分流程. 其他细粒度评

© 中国科学院软件研究所 http://www.jos.org.cn

分过程类推. 其中, 虚线框部分表示 2-细粒度评分的 MFF 特征融合. MFF 方法采用注意力机制, 实现特征的有区分融合, 能够有效提升 HPMG 的性能. 由于 MFF 使用了注意力模块, 因此下文在介绍 MFF 方法之前, 先详细描述所涉及的注意力模块.

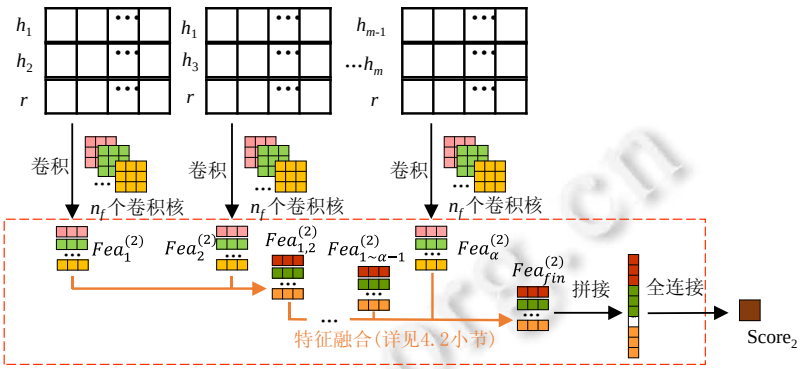


图 4 HPMG$^+$算法的 2-细粒度评分示意图

4.1 注意力模块

本节介绍注意力模块 ATT (attention module), 其示意图如图 5 所示. 首先, 将输入特征通过复制扩充为通道数为 C 的特征表示; 然后, 通过全局注意力分支和局部注意力分支分别提取输入特征的全局特征注意力和局部注意力, 将上一步得到的全局注意力和局部注意力进行广播求和操作, 并输入 Sigmoid 函数计算出注意力权重; 接着, 将上一步计算出的注意力权重和原始输入特征 X 相乘后, 通过对通道求平均的操作得到特征图 X'.

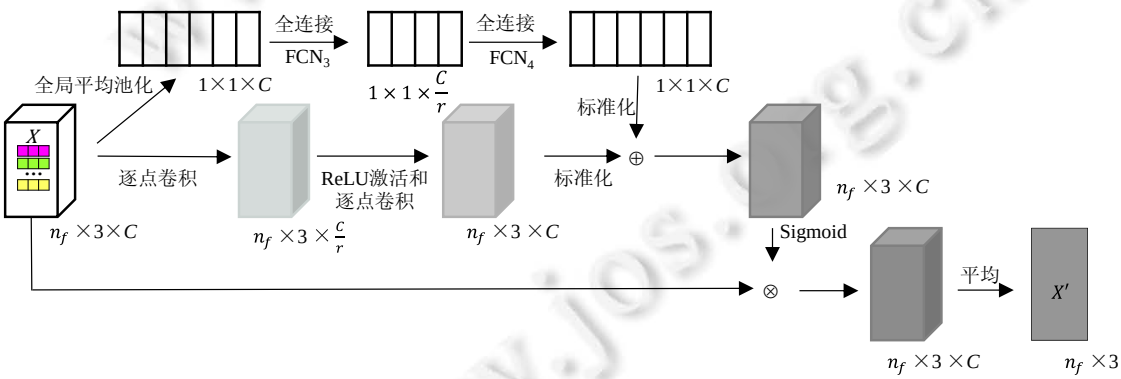


图 5 注意力模块 ATT 的示意图

全局注意力特征提取过程如公式(10)所示, 通过全局平均池化和全连接操作获得特征的全局注意力. 首先, 使用全局平均池化操作(GAP)将原始输入特征压缩成一维参数, 该一维参数获得了原始特征的全局视野, 感受区域更广; 然后, 采用第 1 个全连接层 FCN_3 对重要性进行预测, 得到重要性参数; 最后, 用 ReLU 激活, 并输入全连接层 FCN_4 以恢复原始特征的维度. 最终得到的 X_g 包含了原始特征全局的重要性:

$$X_g=B[FCN_4[\sigma(FCN_3(GAP(X)))]] \tag{10}$$

其中, GAP 为全局平均池化操作, FCN_3 和 FCN_4 为全连接操作, σ为 ReLU 函数.

局部注意力特征提取过程如公式(11)所示, 通过点卷积(PWConv)来获取特征的局部注意力. 将特征 X 输入两层卷积神经网络(两层中间使用 ReLU 进行激活). 先采用核大小为 $C/r\times C\times1\times1$ 的 $Conv_1$ 减少通道数, 然后通过核大小为 $C\times C/r\times1\times1$ 的 $PWConv_2$ 卷积恢复通道数. 最终得到的 X_l 可以保留、突出显示特征的细微细节:

$$X_l=B[PWConv_2[\sigma(B(PWConv_1(X)))]] \tag{11}$$

© 中国科学院软件研究所 http://www.jos.org.cn

最后, 将上述的全局注意力和局部注意力通过广播求和后, 经过 Sigmoid 函数激活得到权重值, 利用该权重值对输入特征 X 做注意力操作后, 对各通道特征进行求平均值操作, 得到仅有一个通道的输出 X', 如公式(12)所示.

$$X'=F_{att}(X)=avg(X\otimes\delta(X_l\oplus X_g)) \tag{12}$$

其中, δ是 Sigmoid 函数, $\oplus$为广播求和操作, $\otimes$为对应元素相乘, avg 为求各个通道的平均.

4.2 MFF方法

本节主要介绍 i-细粒度的 C_m^i 个特征的特征融合方法 MFF. MFF 的主要思想是: 首先, 采用第 4.1 节描述的注意力模块获取任意两个输入特征的注意力特征图; 然后, 根据注意力特征图对这两个特征进行加权融合; 以上两个特征的融合结果与第 3 个特征进行类似的融合. 反复如此, 能够实现多个特征的融合.

下文以 2-细粒度为例, 描述 MFF 方法, 其示意图如图 6 所示. 其中, $Fea_i^{(2)}$ 表示特征 $(i=1,\ldots,C_m^2)$, $Fea_{1,2}^{(2)}$ 表示 $Fea_1^{(2)}$ 和 $Fea_2^{(2)}$ 的融合特征. 下文首先介绍两个特征融合的方法[44], 然后推广到多个特征的融合.

不失一般性, 设已知两个特征 $Fea_1^{(2)}$ 和 $Fea_2^{(2)}$, 首先将这两个特征求和得到初始特征集; 然后, 将上一步集成的结果输入注意力模块 ATT, 得到融合 $Fea_1^{(2)}$ 和 $Fea_2^{(2)}$ 的特征图; 接着, 利用该特征图分别对 $Fea_1^{(2)}$ 和 $Fea_2^{(2)}$ 加权融合, 最终得到融合 $Fea_1^{(2)}$ 和 $Fea_2^{(2)}$ 的特征表示 $Fea_{1,2}^{(2)}$. 其计算过程如公式(13)所示.

$$Fea_{1,2}^{(2)}=F_f(Fea_1^{(2)},Fea_2^{(2)})=F_{att}(Fea_1^{(2)}+Fea_2^{(2)})\otimes Fea_1^{(2)}+(1-F_{att}(Fea_1^{(2)}+Fea_2^{(2)}))\otimes Fea_2^{(2)} \tag{13}$$

若有 $\alpha=C_m^i$ 个特征需要融合, 先将任意两个特征进行融合, 然后将上一步的融合结果与第 3 个特征进行融合. 如此循环操作, 直至融合完第α个特征为止. 其计算过程如公式(14)所示.

$$Fea_{fin}^{(i)}=MFF(Fea_1^{(i)},\ldots,Fea_\alpha^{(i)})=F_f(F_f(Fea_1^{(i)},\ldots,Fea_{\alpha-1}^{(i)}),Fea_\alpha^{(i)}) \tag{14}$$

分析易知, MFF 时间复杂度和空间复杂度分别为 $O\left(\frac{C_m^iC^2}{r}\right)$ 和 $O\left(\frac{C_m^iC^2}{r}\right)$.

由于 $\frac{C_m^iC^2}{r}<<|T_{batch}|K_{m-1}n_fC_{in}C_{out}$, 因此, HPMG$^+$的时间复杂度和空间复杂度分别为 $O(|T_{batch}|n_fC_{in}C_{out})$和 $O(n_hd+n_rd)$. 式中各符号含义请参照第 3.4.3 节.

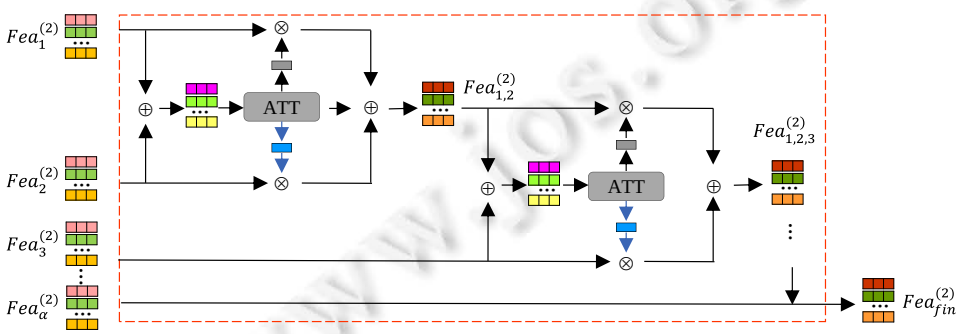


图 6 MFF 方法

5 实 验

本节主要在真实数据集上验证了本文提出的 HPMG 方法和 HPMG$^+$的链接预测能力. 实验分为 3 组: 1) 在两个知识超图真实数据集上, 通过与基线模型比较, 验证了本文提出的 HPMG 模型以及 HPMG$^+$模型链接预测的优越性; 2) 通过消融实验, 分析了 HPMG$^+$的粗粒度评价模块和细粒度评价模块对模型性能的影响; 3) 分析不同参数设置对 HPMG$^+$方法性能的影响.

© 中国科学院软件研究所 http://www.jos.org.cn

5.1 数据集

本文实验采用了文献[34–37]使用的所有两个常用的知识超图真实公开数据集 JF17K 和 WiKiPeople.

(1) JF17K[9]是一个基于 Freebase 开发的超关系公开数据集.

(2) WiKiPeople[33]是一个从 WiKidata 过滤而来, 专注于人类类型实体的超关系数据集.

上述数据集均包含多元关系, 其统计信息如表 1 所示. $|E|$与$|R|$分别表示实体与关系的数量, *Fact*(*train*)和 *Fact*(*test*)分别表示训练集与验证集包含元组的数量, *Arity* 表示数据集包含的多元关系的元数.

表 1 数据集的统计信息

Dataset	$\|E\|$	$\|R\|$	*Fact*(*train*)	*Fact*(*valid*)	*Fact*(*test*)	*Arity*
JF17K	28 645	322	76 379	–	24 568	2–6
WikiPeople	34 839	375	205 725	38 223	38 281	2–7

5.2 评价指标

本文采用流行的平均倒数排名 MRR 和命中率 Hits@k 进行知识超图超边链接预测模型的性能评估. 这两个评价指标均与测试集中正例在所有负例中的排名有关. 该排名计算方法如下所示: 已知一个多元关系的集合 F, 设 f 为测试集 F_{test} 中任意一个正例, 对于 f 中任意位置上的一个实体, 用其他所有非此元素的实体替代, 得到该关系的一组负例; 然后, 从这些负例中删除已被 F 包含的正例, 得到候选样本; 接着, 将上一步得到的候选样本和原始正例 f 输入预测模型获取评分, 并根据评分进行排名, 从而获取正例 f 的名次.

Hits@k 表示排名前 k 的所有正例样本的比率, 即测试集中排名进入前 k 的正例关系的数目除以测试集所有关系的数目. 其计算公式如公式(15)所示. Hits@k 取值范围为[0,1], 值越大, 表明推理算法效果越好.

$$Hit@k=\frac{\sum_{f\in F_{test}} rank(f)\leqslant n?1:0}{count(F_{test})} \tag{15}$$

其中, $rank(f)\leqslant n?1:0$ 为判别排名是否进入前 n 的条件表达式: 若成立, 则结果为 1; 否则为 0. $Count(F_{test})$表示测试集包含的关系的总数.

平均倒数排名(MRR)计算每个正例名次的倒数的平均值, 其计算公式如公式(16)所示. MRR 取值范围为[0,1], 值越大, 效果越好.

$$MRR=\frac{\sum_{f\in F_{test}} \frac{1}{rank(f)}}{count(F_{test})} \tag{16}$$

本文取 Hits@1, Hits@10 和 MRR 作为性能评价指标, 评价本文提出的模型的有效性.

5.3 基线模型

本文采用目前知识超图链接预测的各类最优方法作为基线模型, 包括基于平移的 *m*-TransH、RAE 方法, 基于神经网络的 tNaLP$^+$、HINGE、NeuInfer 方法. 由于基于张量分解的方法只能对固定元关系进行链接预测, 因此未考虑此类方法. 本文均采用基线模型公开的最好结果进行实验对比.

(1) *m*-TransH[9]是一种从传统知识图谱链接预测方法 TransH[12]推广而来的知识超图链接预测方法.

(2) RAE[29]是一种在 *m*-TransH 基础上进一步改进关联性假设得到的知识超图链接预测方法.

(3) tNaLP$^+$[35]是由 NALP[34]改进而来, 它将多元关系表示为一组键值对, 并在 NALP 基础上引入了类型兼容性评分, 还进一步改进了负采样技术. 由于 tNaLP$^+$性能优于 NALP, 因此只将 tNaLP$^+$方法作为基线模型.

(4) HINGE[36]和 NeuInfer[37]均将多元关系表示为主三元组和辅助键值对, 并通过神经网络单独学习主三元组和辅助键值对.

(5) StarE[32]将图卷积神经网络推广到知识超图, 通过一个消息传递网络来获取实体和关系的嵌入.

使用 PyTorch 实现 HPMG 和 HPMG$^+$方法, 相关参数设置如下: 卷积神经网络中滤波器数量设置为 400;

© 中国科学院软件研究所 http://www.jos.org.cn

批处理大小设置为 128; 学习率设置为 0.000 05; 嵌入大小设置为 100; δ设置为 m–1, m 表示多元组的元数.

5.4 链接预测性能比较

本节分别在 JF17K 和 WiKiPeople 数据集上对比了本文提出的 HPMG、HPMG$^+$和基线模型的 Hits@1、Hits@10、MRR. 实验结果见表 2, 其中, 加粗的数据为每项评价指标的最优结果.

表 2 HPMG、HPMG$^+$和基线方法在 JF17K 和 WiKiPeople 数据集上的 Hits@1、Hits@10、MRR 对比

Method	JF17K			WikiPeople		
	Hits@1	Hits@10	MRR	Hits@1	Hits@10	MRR
m-TransH	0.206	0.463	0.206	0.063	0.301	0.063
RAE	0.215	0.467	0.215	0.059	0.306	0.059
tNaLP+	0.370	0.598	0.449	0.269	0.473	0.339
HINGE	0.361	0.624	0.449	0.415	0.585	0.476
NeuInfer	0.436	0.675	0.517	0.282	0.467	0.350
STARE	0.496	0.725	0.574	0.398	**0.648**	0.491
HPMG	0.521	0.693	0.573	0.417	0.628	0.501
HPMG$^+$	**0.528**	**0.741**	**0.592**	**0.429**	0.641	**0.502**

表 2 表明, 在 JF17K 和 WiKiPeople 数据集上, HPMG 方法的 Hits@1, Hits@10 和 MRR 均明显优于除 STARE 以外的所有基线方法. 其中, 对于 JF17K 数据集, Hits@1, Hits@10 和 MRR 分别提高了 8.5%–31.5%, 1.8%–23%和 5.6%–36.7%; 对于 WiKiPeople 数据集, Hits@1, Hits@10 和 MRR 分别提高了 13.5%–35.8%, 16.1%–32.7%和 14.9%–44.2%. 这是因为本文提出的 HPMG 不但采用了考虑了多元关系平等性的多元组模型表示多元关系, 对初始嵌入进行了预处理, 还考虑了知识超图的整体性, 从多粒度的角度学习关系, 从而提升了知识超图的链接预测性能. 并且, HINGE 和 NeuInfer 模型结构过于复杂, 未对多元关系进行预处理, 使得多元关系嵌入表示包含的语义信息较少; tNaLP$^+$将多元关系表示为辅助键值, 损失了很多结构化信息; *m*-TransH 和 RAE 均扩展传统知识图谱的方法而来, 其在知识超图上的表现力和适用度不强.

与 STARE 相比, HPMG 取得了具有竞争力的结果: 在 JF17K 数据集上, HPMG 的 Hits@1 提升了 2.5%, MRR 和 Hits@10 分别降低了 0.1%和 3.2%; 在 WiKiPeople 数据集上, HPMG 的 Hits@1 和 MRR 分别提升了 1.9%和 1%, Hits@10 降低了 2%. STARE 在某些方面稍优于本文提出的 HPMG, 这是因为 STARE 方法通过学习多元关系中的辅助信息补充了三元组的语义信息, 且其公开的实验结果只对主三元组中的实体进行了预测, 预测范围相对较小. HPMG 采用了本文提出的多元组模型表示多元关系, 能够预测多元组中缺失的所有实体, 其预测范围比 STARE 更广. 综上所述, 2 个真实数据集上的大量实验结果表明, 本文提出的 HPMG 方法在绝大多数情况下的评价指标均优于所有的基线模型, 在其他情况下也取得了具有竞争力的实验结果. 这验证了本文提出的多元组表示模型和 HPMG 方法具有比现有表示模型以及链接预测方法更优的链接预测效果.

表 2 还表明, 在 JF17K 和 WiKiPeople 数据集上, HPMG$^+$方法的各项评价指标不但优于 HPMG, 还在绝大多数情况下优于所有基线模型. 具体而言, 与 HPMG 相比, 在 JF17K 数据集上, HPMG$^+$的 Hits@1, Hits@10 和 MRR 分别提高了 0.7%, 4.8%和 1.9%; 在 WiKiPeople 数据集上, HPMG$^+$的 Hits@1, Hits@10 和 MRR 分别提高了 1.2%, 1.3%和 0.1%. 与所有基线模型相比, 在 JF17K 数据集上, HPMG$^+$的 Hits@1, Hits@10 和 MRR 分别提高了 3.2%–32.2%, 1.6%–27.8%和 1.8%–38.6%. WiKiPeople 数据集上的对比结果如下所示: HPMG$^+$的 Hits@1, MRR 分别提高了 3.1%–37%, 1.1%–44.3%; 其 Hits@10 虽然比 STARE 低 0.7%, 但比其他基线模型提高了 5.6%–34%, 具有较强的竞争力. 上述实验结果验证了 MFF 方法特征融合程度比最小值方法更充分.

5.5 消融实验

为了验证粗粒度评价模块和细粒度评价模块对模型性能的影响, 本节设计了 HPMG$^+$-Coarse 和 HPMG$^+$-Fine 模型: HPMG$^+$-Coarse 是仅保留了粗粒度评价模块、去掉了细粒度评价模块的 HPMG$^+$模型, HPMG$^+$-Fine 是仅保留了细粒度评价模块、去掉了粗粒度评价模块的 HPMG$^+$模型. HPMG$^+$, HPMG$^+$-Coarse 和 HPMG$^+$-Fine 模型在 JF17K 和 WiKiPeople 数据集上的 Hits@1、Hits@10 和 MRR 的实验结果见表 3.

© 中国科学院软件研究所 http://www.jos.org.cn

表 3 消融实验结果

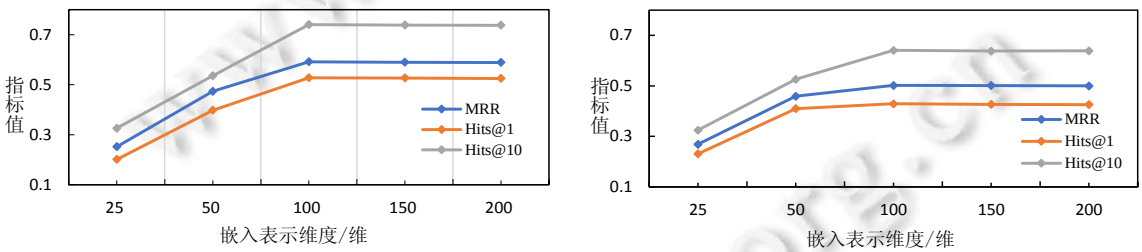
Method	JF17K			WikiPeople		
	Hits@10	Hits@1	MRR	Hits@10	Hits@1	MRR
HPMG$^+$-Coarse	0.626	0.353	0.445	0.557	0.292	0.382
HPMG$^+$-Fine	0.544	0.279	0.369	0.519	0.254	0.345
HPMG$^+$	**0.741**	**0.528**	**0.592**	**0.641**	**0.429**	**0.502**

由表 3 可知，HPMG$^+$在两个真实数据集上的各项评价指标均优于 HPMG$^+$-Coarse 和 HPMG$^+$-Fine. 例如，在 JF17K 数据集上，HPMG$^+$的 MRR 分别比 HPMG$^+$-Coarse 和 HPMG$^+$-Fine 提升了 14.7%和 22.3%. 结合表 3 的实验结果可知，与 HPMG、基线模型相比，HPMG$^+$在绝大多数情况下均取得了较优的链接预测性能，而 HPMG$^+$-Coarse 和 HPMG$^+$-Fine 则优势不明显，甚至处于劣势. 上述实验结果表明，粗粒度评价模块和细粒度评价模块虽然单独作用不明显，但它们共同、相互的作用，能使 HPMG$^+$的性优于其他所有对比模型. 因为粗粒度评价模块和细粒度评价模块分别提取了多元关系粗粒度的超边特征和细粒度的实体特征，这两部分特征具有互补的特性. 表3还显示，粗粒度评价模块对实验结果的影响比细粒度评价模块更大. 因为超边包含了比实体更丰富的语义信息，使粗粒度评价模块提取的超边特征优于细粒度评价模块提取的实体特征.

5.6 参数敏感性分析

本节在 JF17K 和 WiKiPeople 数据集上分别研究了如下超参数对模型性能的影响：嵌入维度、卷积滤波器数量. 假设嵌入维度∈{25,50,100,150,200}，滤波器数量∈{100,200,300,400,500,600}. 为公平起见，在研究当前超参数时，其余超参数与第 5.3 节实验的设置相同.

嵌入维度对各项评价指标影响的实验结果如图 7 所示.

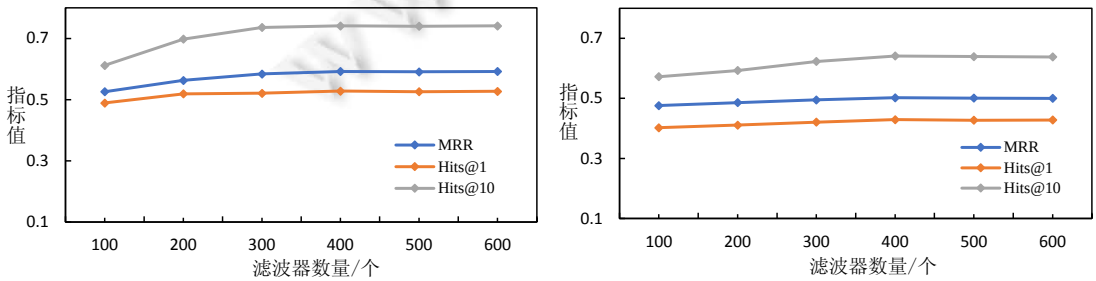


(a) JF17K (b) WiKiPeople

图 7 嵌入表示维度对模型性能的影响

图 7(a)表明，在 JF17K 数据集上，HPMG$^+$方法的各项评价指标随维度的增加呈亚线性增长，并在嵌入维度达到 100 时开始趋于平稳. 图 7(b)表明：在 WiKiPeople 数据集上，HPMG$^+$方法的各项评价指标先随维度的增加呈亚线性增长；当嵌入维度达到 100 以后，随着维度的增加，各项评价指标趋于稳定，并呈现稍稍下降的趋势. 综上，当嵌入维度超过 100 时，HPMG$^+$模型趋于稳定.

卷积滤波器数量对各项评价指标影响的实验结果如图 8 所示.



(a) JF17K (b) WiKiPeople

图 8 滤波器数量对模型性能的影响

图 8(a)和图 8(b)显示, JF17K 数据集和 WiKiPeople 数据集均有类似的实验结果: 随着滤波器数量的增加, HPMG$^+$方法各项评价指标先不断增加; 当滤波器数量为 400 时, HPMG$^+$方法各项评价指标基本最优; 超过 400 后, 滤波器数量继续增加, HPMG$^+$方法各项评价指标基本不变. 综上, 当滤波器数量超过 400, HPMG$^+$模型处于稳定状态.

6 总 结

本文提出了一种基于多粒度注意力网络的神经网络方法, 解决知识超图链接预测问题.

- 首先, 针对现有多元关系表示模型忽略了多元关系平等性和整体性的问题, 本文提出了一种知识超图多元关系表示模型, 在对知识超图多元关系的直接建模中考虑了知识超图中的多元关系的平等性.
- 然后, 基于本文提出的多元组表示模型, 提出了一种基于多粒度神经网络的知识超图链接预测方法 HPMG. HPMG 将关系划分为多重粒度进行学习和预测, 充分考虑了知识超图中不同维度多元关系的整体性.
- 接下来, 针对 HPMG 特征融合不充分的问题, 提出了一种基于多粒度注意力网络的 HPMG$^+$方法, 结合全局和局部注意力实现了不同特征的有区分融合, 进一步提高了模型的性能.

两个知识超图真实数据集上的大量实验结果表明, 本文提出的 HPMG 和 HPMG$^+$方法的链接预测性能显著地超过所有基线模型.

接下来将进一步探索不同元数关系的 1-细粒度–δ-细粒度评分中的最优组合, 以继续提升模型的性能. 本文提出的方法学习了 n 元关系 1–δ细粒度所有评分(δ=n–1), 虽提高了链接预测性能, 但也增加了时间开销. 因此, 下一步将研究不同元数的关系需要学习的最优细粒度评分组合, 旨在提高链接预测性能的同时降低时间开销. 例如, 当关系元数为 3 时, 仅需考虑 1-细粒度评分或 2-细粒度评分; 当关系元数为 5 时, 可忽略 1-细粒度和 3 细粒度评分, 只学习 2-细粒度和 4-细粒度评分.

References:

[1] Yu CM, Zhang ZG, Kong LG. A comparative study of knowledge graph representation models for link prediction. Data Analysis and Knowledge Discovery, 2021, 5(11): 29–44 (in Chinese with English abstract). [doi: 10.11925/infotech.2096-3467.2021.0491]

[2] Zhang Y, Meng XF. InterTris: Specific domain knowledge graph representation learning by interaction among triple elements. Chinese Journal of Computers, 2021, 44(8): 1535–1548 (in Chinese with English abstract). [doi: 10.11897/SP.J.1016.2021.01535]

[3] Wang MH, Qiu LL, Wang XL. A survey on knowledge graph embeddings for link prediction. Symmetry, 2021, 13(3): Article No.485.

[4] Shu ST, Li S, Hao XH, Zhang LP. Knowledge graph embedding technology: A review. Journal of Frontiers of Computer Science and Technology, 2021, 15(11): 2048–2062 (in Chinese with English abstract). [doi: 10.3778/j.issn.1673-9418.2103086]

[5] Wang X, Chen WX, Yang YJ, Zhang XW, Feng ZY. Research on knowledge graph partitioning algorithms. Chinese Journal of Computers, 2021, 44(1): 235–260 (in Chinese with English abstract). [doi: 10.11897/SP.J.1016.2021.00235]

[6] Guan SP, Jin XL, Jia YT, Wang YZ, Cheng XQ. Knowledge reasoning over knowledge graph: A survey. Ruan Jian Xue Bao/Journal of Software, 2018, 29(10): 2966–2994 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/5551.htm [doi: 10.13328/j.cnki.jos.005551]

[7] Tian L, Zhang JC, Zhang JH, Zhou WT, Zhou X. Knowledge graph survey: Representation, construction, reasoning and knowledge hypergraph theory. Journal of Computer Applications, 2021, 41(8): 2161–2186 (in Chinese with English abstract). 2161–2186. [doi: 10.11772/j.issn.1001-9081.2021040662]

[8] Sun Y, Liu WZ, Liu T, Feng X, Yang N, Zhou HF. Signaling pathway of MAPK/ERK in cell proliferation, differentiation, migration, senescence and apoptosis. Journal of Receptors and Signal Transduction, 2015, 35(6): 600–604.

[9] Wen JF, Li JX, Mao YY, Chen SN, Zhang RC. On the representation and embedding of knowledge bases beyond binary relations. In: Proc. of the25th Int'l Joint Conf. on Artificial Intelligence. San Francisco: Morgan Kaufmann Publishers, 2016. 1300–1307.

© 中国科学院软件研究所 http://www.jos.org.cn

[10] Fatemi B, Taslakian P, Vazquez D, Poole D. Knowledge hypergraphs: Prediction beyond binary relations. In: Proc. of the 29th Int'l Joint Conf. on Artificial Intelligence. San Francisco: Morgan Kaufmann, 2020. 2191−2197.

[11] Bordes A, Usunier N, Garcia-Duran A, Weston J, Yakhnenko O. Translating embeddings for modeling multi-relational data. In: Proc. of the 27th Annual Conf. on Neural Information Processing Systems. MIT, 2013. 2787−2795.

[12] Wang Z, Zhang JW, Feng JL, Chen Z. Knowledge graph embedding by translating on hyperplanes. In: Proc. of the 28th AAAI Conf. on Artificial Intelligence. Palo Alto: AAAI, 2014. 1112−1119.

[13] Lin YK, Liu ZY, Sun MS, Liu Y, Zhu X. Learning entity and relation embeddings for knowledge graph completion. In: Proc. of the 29th AAAI Conf. on Artificial Intelligence. Palo Alto: AAAI, 2015. 2181−2187.

[14] Fan M, Zhou Q, Chang E, Zheng TF. Transition-based knowledge graph embedding with relational mapping properties. In: Proc. of the 28th Pacific Asia Conf. on Language, Information and Computation. Bangkok: Department of Linguistics, Chulalongkorn University, 2014. 328−337.

[15] Ji GL, He SZ, Xu LH, Liu K, Zhao J. Knowledge graph embedding via dynamic mapping matrix. In: Proc. of the 53rd Annual Meeting of the Association for Computational Linguistics. Stroudsburg: Association for Computational Linguistics, 2015. 687−696.

[16] Jia YT, Wang YZ, Lin HL, Jin XL, Cheng XQ. Locally adaptive translation for knowledge graph embedding. In: Proc. of the 30th AAAI Conf. on Artificial Intelligence. Palo Alto: AAAI, 2016. 992−998.

[17] Ji GL, Liu K, He SZ, Zhao J. Knowledge graph completion with adaptive sparse transfer matrix. In: Proc. of the 30th AAAI Conf. on Artificial Intelligence. Palo Alto: AAAI, 2016. 985−991.

[18] Xiao H, Huang ML, Zhu XY. From one point to a manifold: knowledge graph embedding for precise link prediction. In: Proc. of the 25th Int'l Joint Conf. on Artificial Intelligence. San Francisco: Morgan Kaufmann, 2016. 1315−1321.

[19] Nickel M, Tresp V, Kriegel HP. A three-way model for collective learning on multi-relational data. In: Proc. of the 28th Int'l Conf. on Machine Learning. New York: Association for Computing Machinery, 2011. 809−816.

[20] Nickel M, Rosasco L, Poggio TA. Holographic embeddings of knowledge graphs. In: Proc. of the 30th AAAI Conf. on Artificial Intelligence. Palo Alto: AAAI, 2016. 1955−1961.

[21] Liu HX, Wu YX, Yang YM. Analogical inference for multi-relational embeddings. In: Proc. of the 34th Int'l Conf. on Machine Learning. New York: Association for Computing Machinery, 2017. 2168−2178.

[22] Yang BS, Yih WT, He XD, Gao JF, Deng L. Embedding entities and relations for learning and inference in knowledge bases. In: Proc. of the Int'l Conf. on Learning Representations. San Diego, 2015.

[23] Trouillon T, Welbl J, Riedel S, Gaussier É, Bouchard G. Complex embeddings for simple link prediction. In: Proc. of the 33nd Int'l Conf. on Machine Learning. New York: Association for Computing Machinery, 2016. 2071−2080.

[24] Balaevi I, Allen C, Hospedales TM. TuckER: Tensor factorization for knowledge graph completion. In: Proc. of the Conf. on Empirical Methods in Natural Language Processing and the 9th Int'l Joint Conf. on Natural Language Processing. Stroudsburg: Association for Computational Linguistics, 2019. 5184−5193.

[25] Jiang XT, Wang Q, Wang B. Adaptive convolution for multi-relational learning. In: Proc. of the Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Stroudsburg: Association for Computational Linguistics, 2019. 978−987.

[26] Nathani D, Chauhan J, Sharma C, Kaul M. Learning attention-based embeddings for relation prediction in knowledge graphs. In: Proc. of the 57th Conf. of the Association for Computational Linguistics. Stroudsburg: Association for Computational Linguistics, 2019. 4710−4723.

[27] Nguyen DQ, Nguyen TD, Dat QN, Phung DQ. A novel embedding model for knowledge base completion based on convolutional neural network. In: Proc. of the Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Stroudsburg: Association for Computational Linguistics, 2018. 327−333.

[28] Dettmers T, Minervini P, Stenetorp P, Riedel S. Convolutional 2D knowledge graph embeddings. In: Proc. of the 32nd AAAI Conf. on Artificial Intelligence. Palo Alto: AAAI, 2018. 1811−1818.

[29] Zhang RC, Li JP, Mei JJ, Mao YY. Scalable instance reconstruction in knowledge bases via relatedness affiliated embedding. In: Proc. of the 27th Int'l World Wide Web Conf. New York: Association for Computing Machinery, 2018. 1185−1194.

© 中国科学院软件研究所 http://www.jos.org.cn

[30] Wang PY, Duan L, Guo ZS, Jiang WP, Zhang YD. Knowledge hypergraph link prediction model based on tensor decomposition. Journal of Computer Research and Development, 2021, 58(8): 1599−1611 (in Chinese with English abstract). [doi: 10.7544/issn 1000-1239.2021.20210315]

[31] Liu Y, Yao QM, Li Y. Generalizing tensor decomposition for *n*-ary relational knowledge bases. In: Proc. of the 29th Int'l World Wide Web Conf. New York: Association for Computing Machinery, 2020. 1104−1114.

[32] Galkin M, Trivedi P, Maheshwari G, Usbeck R, Lehmann J. Message passing for hyper-relational knowledge graphs. In: Proc. of the Conf. on Empirical Methods in Natural Language Processing. Stroudsburg: Association for Computational Linguistics, 2020. 7346−7359.

[33] Liu Y, Yao QM, Li Y. Role-aware modeling for *n*-ary relational knowledge bases. In: Proc. of the 30th Int'l World Wide Web Conf. New York: Association for Computing Machinery, 2021. 2660−2671.

[34] Guan SP, Jin XL, Wang YZ, Cheng XQ. Link prediction on *n*-ary relational data. In: Proc. of the 28th Int'l World Wide Web Conf. New York: Association for Computing Machinery, 2019. 583−593.

[35] Guan SP, Jin XL, Guo JF, Wang YZ, Cheng XQ. Link prediction on *n*-ary relational data based on relatedness evaluation. IEEE Trans. on Knowledge and Data Engineering, 2021, 35(1): 672−685.

[36] Rosso P, Yang DQ, Cudré-Mauroux P. Beyond triplets: Hyper-relational knowledge graph embedding for link prediction. In: Proc. of the 29th Int'l World Wide Web Conf. New York: Association for Computing Machinery, 2020. 1885−1896.

[37] Guan SP, Jin XL, Guo JF, Wang YZ, Cheng XQ. Neuinfer: Knowledge inference on *n*-ary facts. In: Proc. of the 58th Annual Meeting of the Association for Computational Linguistics. Stroudsburg: Association for Computational Linguistics, 2020. 6141−6151.

[38] Kazemi SM, Poole D. Simple embedding for link prediction in knowledge graphs. In: Proc. of the 32th Annual Conf. on Neural Information Processing Systems. MIT, 2018. 4289−4300.

[39] Tucker L. Some mathematical notes on three-mode factor analysis. Psychometrika, 1966, 31(3): 279−311.

[40] Zhao QB, Zhou GX, Xie SL, Zhang LQ, Cichocki A. Tensor ring decomposition. arXiv:1606.05535, 2016.

[41] Zhang RC, Zou YS, Ma J, Hyper-SAGNN: A self-attention based graph neural network for hypergraphs. In: Proc. of the Int'l Conf. on Learning Representations. 2020.

[42] Xu YW, Zhang HJ, Cheng K, Liao XL, Zhang ZX, Li YB. Knowledge graph embedding with entity attributes using hyper graph neural networks. Intelligent Data Analysis, 2022, 26(4): 959−975.

[43] Yan BC, Wang CK, Guo GY, Lou YK. TinyGNN: Learning efficient graph neural networks. In: Proc. of the 26th ACMSIGKDD Int'l Conf. on Knowledge Discovery & Data Mining. 2020. 1848−1856.

[44] Dai YM, Gieseke F, Oehmcke S, Wu YQ, Barnard K. Attentional feature fusion. In: Proc. of the IEEE/CVF Winter Conf. on Applications of Computer Vision. 2021. 3560−3569.

附中文参考文献:

[1] 余传明, 张贞港, 孔令格. 面向链接预测的知识图谱表示模型对比研究. 数据分析与知识发现, 2021, 5(11): 29−44. [doi: 10.11925/infotech.2096-3467.2021.0491]

[2] 张祎, 孟小峰. InterTris: 三元交互的领域知识图谱表示学习. 计算机学报, 2021, 44(8): 1535−1548. [doi: 10.11897/SP.J.1016.2021.01535]

[4] 舒世泰, 李松, 郝晓红, 张丽平. 知识图谱嵌入技术研究进展. 计算机科学与探索, 2021, 15(11): 2048−2062. [doi: 10.3778/j.issn.1673-9418.2103086]

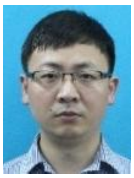
[5] 王鑫, 陈蔚雪, 杨雅君, 张小旺, 冯志勇. 知识图谱划分算法研究综述. 计算机学报, 2021, 44(1): 235−260. [doi: 10.11897/SP.J.1016.2021.00235]

[6] 官赛萍, 靳小龙, 贾岩涛, 王元卓, 程学旗. 面向知识图谱的知识推理研究进展. 软件学报, 2018, 29(10): 2966−2994. http://www.jos.org.cn/1000-9825/5551.htm [doi: 10.13328/j.cnki.jos.005551]

[7] 田玲, 张谨川, 张晋豪, 周望涛, 周雪. 知识图谱综述——表示、构建、推理与知识超图理论. 计算机应用, 2021, 41(8): 2161−2186. [doi: 10.11772/j.issn.1001-9081.2021040662]

© 中国科学院软件研究所 http://www.jos.org.cn

[30] 王培妍，段磊，郭正山，蒋为鹏，张译丹．基于张量分解的知识超图链接预测模型．计算机研究与发展，2021，58(8): 1599–1611. [doi: 10.7544/issn1000-1239.2021.20210315]



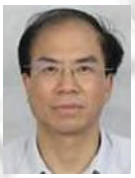
庞俊(1983－)，男，博士，副教授，CCF 专业会员，主要研究领域为图数据管理，图挖掘，大数据管理和分析．

刘小琪(1998－)，女，硕士生，CCF 学生会员，主要研究领域为知识图谱．

谷峪(1981－)，男，博士，教授，博士生导师，CCF 高级会员，主要研究领域为图、空间数据管理，大数据分析．

王鑫(1981－)，男，博士，教授，博士生导师，CCF 杰出会员，主要研究领域为知识图谱数据管理，图数据库，大规模知识处理．

赵宇海(1975－)，男，博士，教授，博士生导师，CCF 高级会员，主要研究领域为数据库，数据挖掘，机器学习，软件工程，生物信息学．



张晓龙(1963－)，男，博士，教授，博士生导师，主要研究领域为机器学习，图像处理．

于戈(1962－)，男，博士，教授，博士生导师，CCF 会士，主要研究领域为数据管理理论与技术，分布与并行系统．

www.jos.org.cn