

# 知识图谱嵌入技术研究综述<sup>\*</sup>

张天成<sup>1</sup>, 田雪<sup>1</sup>, 孙相会<sup>1</sup>, 于明鹤<sup>2</sup>, 孙艳红<sup>1</sup>, 于戈<sup>1</sup>

<sup>1</sup>(东北大学 计算机科学与工程学院, 辽宁 沈阳 110169)

<sup>2</sup>(东北大学 软件学院, 辽宁 沈阳 110169)

通信作者: 田雪, E-mail: [1901787@stu.neu.edu.cn](mailto:1901787@stu.neu.edu.cn)



**摘要:** 知识图谱 (KG) 是一种用图模型来描述知识和建模事物之间关联关系的技术. 知识图谱嵌入 (KGE) 作为一种被广泛采用的知识表示方法, 其主要思想是将知识图谱中的实体和关系嵌入到连续的向量空间中, 用来简化操作, 同时保留 KG 的固有结构. 可以使得多种下游任务受益, 例如 KG 补全和关系提取等. 首先对现有的知识图谱嵌入技术进行全面回顾, 不仅包括使用 KG 中观察到的事实进行嵌入的技术, 还包括添加时间维度的动态 KG 嵌入方法, 以及融合多源信息的 KG 嵌入技术. 对相关模型从实体嵌入、关系嵌入、评分函数等方面进行分析、对比与总结. 然后简要介绍 KG 嵌入技术在下游任务中的典型应用, 包括问答系统、推荐系统和关系提取等. 最后阐述知识图谱嵌入面临的挑战, 对未来的研究方向进行展望.

**关键词:** 知识图谱嵌入; 翻译模型; 复杂关系建模; 动态知识图谱; 关系提取

**中图法分类号:** TP182

中文引用格式: 张天成, 田雪, 孙相会, 于明鹤, 孙艳红, 于戈. 知识图谱嵌入技术研究综述. 软件学报. <http://www.jos.org.cn/1000-9825/6429.htm>

英文引用格式: Zhang TC, Tian X, Sun XH, Yu MH, Sun YH, Yu G. Overview on Knowledge Graph Embedding Technology Research. Ruan Jian Xue Bao/Journal of Software (in Chinese). <http://www.jos.org.cn/1000-9825/6429.htm>

## Overview on Knowledge Graph Embedding Technology Research

ZHANG Tian-Cheng<sup>1</sup>, TIAN Xue<sup>1</sup>, SUN Xiang-Hui<sup>1</sup>, YU Ming-He<sup>2</sup>, SUN Yan-Hong<sup>1</sup>, YU Ge<sup>1</sup>

<sup>1</sup>(School of Computer Science and Engineering, Northeastern University, Shenyang 110169, China)

<sup>2</sup>(Software College, Northeastern University, Shenyang 110169, China)

**Abstract:** Knowledge graph (KG) is a kind of technology that uses graph model to describe the relationship between knowledge and modeling things. Knowledge Graph Embedding (KGE), as a widely adopted knowledge representation method, its main idea is to embed entities and relationships in a knowledge graph into a continuous vector space, which is used to simplify operations while preserving the intrinsic structure of the KG. It can benefit a variety of downstream tasks, such as KG completion, relation extraction, etc. Firstly, the existing knowledge graph embedding technologies are comprehensively reviewed, including not only techniques using the facts observed in KG for embedding, but also dynamic KG embedding methods that add time dimensions, as well as KG embedding technologies that integrate multi-source information. The relevant models are analyzed, compared and summarized from the perspectives of entity embedding, relation embedding and scoring functions. Then, typical applications of KG embedding technologies in downstream tasks are briefly introduced, including question answering systems, recommendation systems and relationship extraction. Finally, the challenges of knowledge graph embedding are expounded, and the future research directions are prospected.

**Key words:** knowledge graph embedding (KGE); translation model; complex relationship modeling; dynamic knowledge graph; relationship extraction

知识图谱 (knowledge graph, KG) 作为人工智能的一个分支, 引起了学术界和工业界的广泛关注, 其构建与应用也得到了迅速发展. 例如 Freebase<sup>[1]</sup>, DBpedia<sup>[2]</sup>, YAGO<sup>[3]</sup>, NELL<sup>[4]</sup>, Wikidata<sup>[5]</sup> 等知识图谱已经被成功创建并应用

<sup>\*</sup> 基金项目: 国家自然科学基金 (U1811261, 61902055); 中央高校基本科研业务费 (N180716010, N2117001)

收稿时间: 2021-03-29; 修改时间: 2021-06-28; 采用时间: 2021-08-09; jos 在线出版时间: 2022-09-16

于许多现实世界应用,从语义分析<sup>[6,7]</sup>、命名实体消歧<sup>[8,9]</sup>,到信息提取<sup>[10,11]</sup>和问答系统<sup>[12,13]</sup>等.知识图谱是以现实世界的实体为节点,实体之间的关系为边的有向图.在这个图中,每个有向边连同其头实体与尾实体构成了一个三元组,即(头实体,关系,尾实体),表示头实体与尾实体通过关系进行连接.尽管知识图谱在表示结构化数据方面非常有效,但这种三元组的基本符号性质使 KG 难以操作<sup>[14]</sup>.

为了解决这个问题,近年来提出了一个新的研究方向,称为知识图谱嵌入 (knowledge graph embedding, KGE) 或知识表示学习 (knowledge representation learning, KRL),旨在将 KG 的组成部分 (包括实体和关系) 嵌入到连续的向量空间中,以在简化操作的同时保留 KG 的固有结构.与传统的表示方法相比, KGE 为 KG 中的实体和关系提供了更加密集表示,降低了其应用中的计算复杂度.此外, KGE 可以通过度量实体和关系低维嵌入的相似性来显式地捕获实体和关系之间的相似性.

尽管研究者已提出多种模型来学习 KG 中的实体和关系表示,但是目前大多数可用的技术仍然仅根据知识图谱中观察到的事实来执行嵌入任务.具体地说,给定一个 KG,首先在低维向量空间中表示实体和关系,并为每个三元组定义一个评分函数以衡量其在该空间中的合理性.然后通过最大化观察到的三元组的总合理性来学习实体和关系的嵌入.这些学习的嵌入还可以进一步用于实现各种任务,例如 KG 补全<sup>[15,16]</sup>,关系提取<sup>[10,17]</sup>,实体分类<sup>[18,19]</sup>,实体解析<sup>[18,20]</sup>等.由于在整个过程中仅要求学习的嵌入在每个单独的事实中兼容,因此对下游任务可能没有足够的预测性<sup>[21,22]</sup>.近年来,越来越多的研究者开始进一步考虑利用其他类型的信息,例如实体类型<sup>[23,24]</sup>,文本描述<sup>[25-28]</sup>,关系路径<sup>[29-31]</sup>,甚至逻辑规则<sup>[32,33]</sup>来学习更多的预测嵌入.

本文第 1 节介绍相关工作调查与基本符号定义;第 2 节对仅使用 KG 中观察到的事实进行嵌入的技术进行全面回顾,具体介绍基于距离的模型,语义匹配模型以及最新的 KGE 技术;第 3 节主要讨论了融合时间信息的动态知识图谱嵌入技术,详细介绍 t-TransE、Know-Evolve、HyTE、TDG2E 等代表性的动态 KGE 方法;第 4 节归纳了除 KG 中观察到的事实以外的结合附加信息的 KGE 技术,例如实体类别、文本描述、关系路径等.第 5 节介绍 KGE 技术在下游任务中的典型应用.第 6 节对 KGE 技术面临的挑战与未来研究方向进行讨论.最后,第 7 节对全文工作进行总结.

## 1 相关调查与符号定义

### 1.1 相关调查

先前有关知识图谱的调查论文主要集中在统计关系学习 (statistical relational learning)<sup>[34]</sup>, knowledge graph refinement<sup>[35]</sup>, 中文知识图谱构建 (Chinese knowledge graph construction)<sup>[36]</sup>, KGE<sup>[14]</sup>或 KRL<sup>[37]</sup>. Liu 等人<sup>[37]</sup>在 2016 年详细介绍了知识表示学习 (KRL) 的基本概念和主要方法,对知识表示学习面临的主要挑战、已有解决方案以及未来研究方向进行了全面总结,为后续的调查和研究奠定了坚实的基础.近年来, Lin 等人<sup>[37]</sup>以线性方式提出 KRL,着重于进行定量分析. Wang 等人<sup>[14]</sup>根据评分函数对 KRL 模型进行分类,侧重于 KRL 中使用的信息类型.

我们的调查研究以 Wang 等人<sup>[14]</sup>的调查为基础.与之不同的是,本文对基于距离的模型与语义匹配模型进行了全新角度的分类,对主流 KGE 技术进行了阐述,同时介绍了动态知识图谱嵌入方法的最新进展,并分析了相关代表模型.此外,本文讨论了结合事实以外的其他信息的嵌入技术,以及 KGE 技术的典型应用.最后,总结了 KGE 技术面临的挑战,并对其未来方向进行展望.

### 1.2 符号定义

知识图谱嵌入旨在将 KG 中的实体和关系嵌入到一个低维连续的语义空间中.为了便于说明,本小节定义几种基本符号.首先,定义知识图谱为  $G = (E, R, S)$ , 其中  $E = \{e_1, e_2, \dots, e_{|E|}\}$  是事实集合, 包含  $|E|$  种不同实体;  $R = \{r_1, r_2, \dots, r_{|R|}\}$  代表关系集合, 包含  $|R|$  种不同关系; 而  $S \subseteq E \times R \times E$  表示事实三元组集合, 一般格式为  $(h, r, t)$ , 其中  $h$  和  $t$  分别表示头、尾实体,  $r$  表示它们之间的关系.例如三元组 (BillClinton, wasPresidentOf, USA) 表示 BillClinton 和 USA 之间存在关系 wasPresidentOf. 表 1 列出了具体的符号及其描述.

表 1 基本符号定义

符号	描述	符号	描述
$G$	知识图谱	$S$	事实集合
$(h, r, t)$	事实三元组	$(\mathbf{h}, \mathbf{r}, \mathbf{t})$	嵌入三元组
$r \in R, e \in E$	关系集合与实体集合	$f_r(h, t)$	评分函数
$\sigma(\cdot), g(\cdot)$	非线性激活函数	$\mathbf{M}_r$	映射矩阵
$L$	损失函数	$\mathbb{R}^d$	$d$ 维实值空间
$\mathbb{C}^d$	$d$ 维复数空间	$\mathbb{H}^d$	$d$ 维超复数空间
$\mathbb{T}^d$	$d$ 维环面空间	$\otimes$	Hamilton乘积
$\circ$	Hadamard乘积	$R_e(\cdot)$	取复数值的实部
$\star$	循环相关	$\text{concat}(\cdot), [h, r]$	向量/矩阵连接
$\omega$	卷积滤波器	$*$	卷积操作
$[\mathbf{h}]_i$	向量 $\mathbf{h}$ 的第 $i$ 项	$[\mathbf{M}_r]_{i,j}$	矩阵 $\mathbf{M}_r$ 的第 $i$ 项

## 2 使用事实进行知识图谱嵌入

本节对仅使用事实进行知识图谱嵌入的方法采用评分函数进行划分. 评分函数用于衡量事实的合理性, 在基于能量的学习框架中也被称为能量函数. 典型类型的评分函数分为两种: 基于距离的评分函数 (如图 1(a)) 与基于相似性的评分函数 (如图 1(b)).

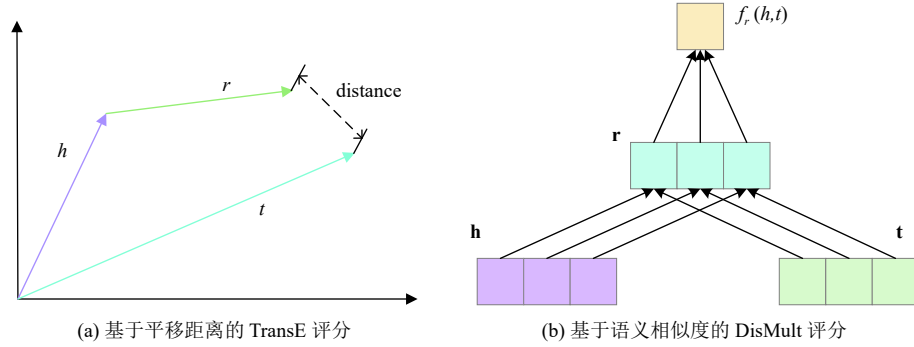


图 1 以 TransE<sup>[15]</sup>和 DistMult<sup>[38]</sup>为例说明基于距离和基于相似性匹配的评分函数

### 2.1 基于距离的模型

基于距离的模型使用基于距离的评分函数, 即通过计算实体之间的距离来衡量事实的合理性, 在这种情况下, 翻译原理  $\mathbf{h} + \mathbf{r} \approx \mathbf{t}$  被广泛使用. 也就是说, 基于距离的模型通常由关系执行翻译后, 根据两个实体之间的距离来度量一个事实的合理性. 本小节将基于距离的模型进一步细分为基本距离模型, 翻译模型和复杂关系建模.

#### 2.1.1 基本距离模型 SE

• SE: 一种直观的基于距离的方法是计算实体在关系的对应空间中的投影向量之间的距离. 结构表示 (structured embedding, SE<sup>[39]</sup>) 中的每个实体用  $d$  维向量表示, SE 为每个关系定义了两个投影矩阵  $\mathbf{M}_{r,1}$  和  $\mathbf{M}_{r,2}$ , 利用这两个投影矩阵和  $L_1$  距离学习结构嵌入为:

$$f_r(h, t) = -\|\mathbf{M}_{r,1}\mathbf{h} - \mathbf{M}_{r,2}\mathbf{t}\|_1 \quad (1)$$

该距离表明头实体  $h$  与尾实体  $t$  在关系  $r$  下的语义相关度. 然而, SE 模型对头、尾实体使用 2 个不同的矩阵进行投影, 因此, SE 模型的协同性较差, 无法精确刻画头、尾实体与关系之间语义联系的强弱.

#### 2.1.2 翻译模型

• TransE: Mikolov 等人在 2013 年提出了 Word2Vec 词表示学习模型和工具包<sup>[40,41]</sup>, 利用该模型, Mikolov 等

人发现词向量空间存在着有趣的平移不变现象. 受该现象启发, Bordes 等人于 2013 年提出了 TransE 模型<sup>[15]</sup>, 该模型将关系和实体表示为同一空间中的向量. 给定事实  $(h, r, t)$ , 关系  $r$  的向量  $\mathbf{r}$  被解释为头实体向量  $\mathbf{h}$  与尾实体向量  $\mathbf{t}$  之间的平移. 因此, 嵌入的实体  $\mathbf{h}$  和  $\mathbf{t}$  可以通过平移向量  $\mathbf{r}$  以低误差连接, 即满足:  $\mathbf{h} + \mathbf{r} \approx \mathbf{t}$ , 图 2(a) 为该方法的简洁表示. 对于每个三元组  $(h, r, t)$ , TransE 定义了如下的评分函数:

$$f_r(h, t) = -\|\mathbf{h} + \mathbf{r} - \mathbf{t}\|_{1/2} \quad (2)$$

即向量  $\mathbf{h} + \mathbf{r}$  与  $\mathbf{t}$  的  $L_1$  或  $L_2$  距离.

● UM: 非结构模型 (unstructured model, UM<sup>[42]</sup>) 是 TransE 的简单版本, 将知识图谱视为单关系图, 并设置所有  $r=0$ , 则 UM 的评分函数为:

$$f_r(h, t) = -\|\mathbf{h} - \mathbf{t}\|_2^2 \quad (3)$$

UM 通常用作其他 KGE 方法的基本基准<sup>[15,43]</sup>, 但是它不能区分不同关系.

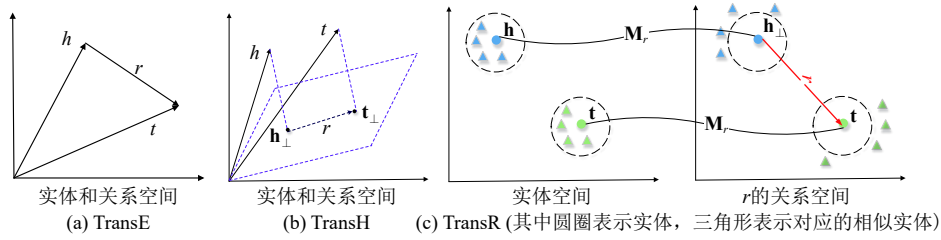


图 2 TransE, TransH, TransR 模型的基本思想说明

### 2.1.3 复杂关系建模

TransE 模型在大规模知识图谱上效果明显, 但是由于其模型简单, 导致 TransE 无法对知识库中的复杂关系建模, 这里的复杂关系定义如下. 按照知识库中关系两端连接实体的数目, 可以将关系划分为 1-1, 1-N, N-1 和 N-N 这 4 种类型, 例如 1-N 类型关系指的是一个头实体会平均对应多个尾实体. 我们将 1-N, N-1 和 N-N 称为复杂关系. 研究发现, 各种知识获取算法在处理 4 种类型关系时的性能差异较大, 例如: TransE 在处理复杂关系建模时性能降低, 这与其模型的假设密切相关.

例如: 事实“比尔·克林顿在 1993 年到 2001 年间担任美国总统”可以被抽象为如下三元组:

$$(h_i : \text{BillClinton}, r_i : \text{wasPresidentOf}, t_i : \text{USA})$$

将另一个事实“乔治·沃克·布什在 2001 年到 2009 年间担任美国总统”, 表述为如下三元组:

$$(h_j : \text{GeorgeWalkerBush}, r_j : \text{wasPresidentOf}, t_j : \text{USA})$$

显然, 上述两个三元组共享相同的尾实体与关系, 而具有不同的头实体. 此时, 如果使用 TransE 从以上两个三元组学习知识表示, TransE 会推导出  $h_i = h_j$  的错误结论. 为了解决这一问题, 近年来涌现了大量关于 TransE 的扩展模型, 接下来按照实体与关系的不同表示空间进行划分, 介绍其中的代表模型.

#### (1) Point-Wise 空间

Point-Wise 欧氏空间广泛应用于表示实体和关系, 在向量或矩阵空间中投影关系嵌入, 或者捕捉关系交互.

● TransH: TransH 模型<sup>[16]</sup>使得一个实体在涉及不同关系时具有分布式表示. 如图 2(b) 所示, TransH 将实体建模为向量, 将每个关系  $r$  建模为法向量为  $\mathbf{w}_r$  的关系特定超平面上的向量  $\mathbf{r}$  ( $\mathbf{r} \in \mathbb{R}^d$ ). 具体来说, 对于一个三元组  $(h, r, t)$ , TransH 首先将头实体向量  $\mathbf{h}$  与尾实体向量  $\mathbf{t}$  ( $\mathbf{h}, \mathbf{t} \in \mathbb{R}^d$ ) 沿法线  $\mathbf{w}_r$  ( $\mathbf{w}_r \in \mathbb{R}^d$ ) 投影到关系  $r$  对应的超平面上, 投影分别记为  $\mathbf{h}_\perp$  和  $\mathbf{t}_\perp$ , 表示如下:

$$\mathbf{h}_\perp = \mathbf{h} - \mathbf{w}_r^\top \mathbf{h} \mathbf{w}_r, \quad \mathbf{t}_\perp = \mathbf{t} - \mathbf{w}_r^\top \mathbf{t} \mathbf{w}_r \quad (4)$$

如果三元组  $(h, r, t)$  成立, 即  $\mathbf{h}_\perp + \mathbf{r} \approx \mathbf{t}_\perp$ , 假设投影在超平面上由  $\mathbf{r}$  以低误差连接, 则 TransH 的评分函数定义为:

$$f_r(h, t) = -\|\mathbf{h}_\perp + \mathbf{r} - \mathbf{t}_\perp\|_2^2 \quad (5)$$

通过引入投影到关系特定超平面的机制, TransH 使得每个实体在不同的关系中具有不同的表示形式.

● **TransR**: TransE 和 TransH 模型假定实体和关系嵌入在同一空间  $\mathbb{R}^d$  中, 但是关系和实体是完全不同的对象. 一个实体是多种属性的综合体, 而各种关系关注实体的不同属性. 因此, 某些相似的实体在实体空间中彼此接近, 而在某些特定属性上不同, 在对应的关系空间中应彼此远离. 为了解决这个问题, Lin 等人<sup>[43]</sup>提出了 TransR 方法, 该方法在不同的空间 (实体空间和关系空间) 中对实体和关系进行建模, 并在关系空间中进行翻译.

TransR 的基本思想如图 2(c) 所示, 对于每个三元组  $(h, r, t)$ , 首先将头、尾实体向量向关系  $r$  空间投影, 使得原来在实体空间中与头、尾实体相似的实体在关系  $r$  空间中被区分开. 具体来说, 对于每一个关系  $r$ , TransR 设置一个投影矩阵  $\mathbf{M}_r \in \mathbb{R}^{k \times d}$ , 将实体  $(\mathbf{h}, \mathbf{t} \in \mathbb{R}^d)$  从实体空间投影到关系  $(\mathbf{r} \in \mathbb{R}^k)$  空间. 利用投影矩阵, 实体的投影向量定义如下:

$$\mathbf{h}_\perp = \mathbf{M}_r \mathbf{h}, \mathbf{t}_\perp = \mathbf{M}_r \mathbf{t} \quad (6)$$

其中,  $\mathbf{M}_r$  是从实体空间到  $r$  的关系空间的投影矩阵. 因此, TransR 的评分函数定义为:

$$f_r(h, t) = -\|\mathbf{h}_\perp + \mathbf{r} - \mathbf{t}_\perp\|_2^2 \quad (7)$$

● **TransD**: 虽然 TransR 较 TransE 和 TransH 具有显著的改进, 但它仍然存在一些缺陷: ① 对于关系  $r$ , 头、尾实体共享相同的投影矩阵  $\mathbf{M}_r$ , 忽略了头、尾实体不同的类型和属性; ② 投影操作是实体与关系之间的交互过程, 因此, 投影矩阵仅由关系决定是不合理的; ③ 与 TransE 和 TransH 相比, 矩阵-向量乘法使 TransR 模型参数急剧增加, 因此, TransR 难以应用于大规模知识图谱.

为此, Ji 等人<sup>[44]</sup>提出改进模型 TransD, 图 3 显示了其基本思想, 每个形状表示出现在关系  $r$  的三元组中的一个实体对,  $\mathbf{M}_{rh}$  和  $\mathbf{M}_{rt}$  分别是  $\mathbf{h}$  和  $\mathbf{t}$  的投影矩阵,  $\mathbf{w}_{hi}$ ,  $\mathbf{w}_{ti}$  ( $i = 1, 2, 3$ ) 和  $\mathbf{w}_r$  是投影向量.  $\mathbf{h}_{i\perp}$  和  $\mathbf{t}_{i\perp}$  是实体的投影向量, 满足  $\mathbf{h}_{i\perp} + \mathbf{r} \approx \mathbf{t}_{i\perp}$  ( $i = 1, 2, 3$ ). TransD 为每个实体和关系定义两个向量, 一个是实体/关系表示, 另一个用于构造投影矩阵. 例如, 给定三元组  $(h, r, t)$ , 其向量为:  $\mathbf{h}, \mathbf{w}_h, \mathbf{t}, \mathbf{w}_t \in \mathbb{R}^d$  且  $\mathbf{r}, \mathbf{w}_r \in \mathbb{R}^k$ , TransD 模型进一步学习了 2 个分别将头实体与尾实体投影到关系空间的投影矩阵  $\mathbf{M}_{rh}, \mathbf{M}_{rt} \in \mathbb{R}^{k \times d}$ , 具体定义如下:

$$\mathbf{M}_{rh} = \mathbf{w}_r \mathbf{w}_h^\top + \mathbf{I}^{k \times d}, \mathbf{M}_{rt} = \mathbf{w}_r \mathbf{w}_t^\top + \mathbf{I}^{k \times d} \quad (8)$$

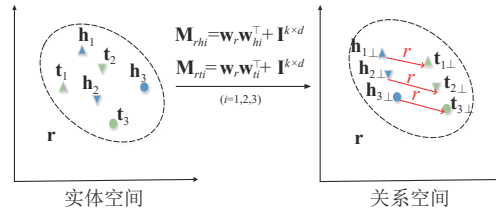


图 3 TransD 的简单说明

显然, 这里的投影矩阵  $\mathbf{M}_{rh}, \mathbf{M}_{rt}$  与实体和关系均有关, 并且利用两个投影向量构造投影矩阵解决了 TransR 面临的参数众多问题. 对于三元组  $(h, r, t)$ , TransD 的评分函数定义如下:

$$f_r(h, t) = -\|\mathbf{h} \mathbf{M}_{rh} + \mathbf{r} - \mathbf{t} \mathbf{M}_{rt}\|_2^2 \quad (9)$$

● **STransE**: Nguyen 等人通过将两个简单的关系预测模型 SE 与 TransE 进行组合, 提出了一个新的嵌入模型 STransE<sup>[45]</sup>, 该模型将每个实体表示为一个低维向量, 并通过两个矩阵和一个平移向量表示每个关系. 其评分函数定义如下:

$$f_r(h, t) = \|\mathbf{M}_{r,1} \mathbf{h} + \mathbf{r} - \mathbf{M}_{r,2} \mathbf{t}\|_{1/2} \quad (10)$$

STransE 可以看作是 TransR 模型的扩展. 它对每个关系使用两个投影矩阵, 一个用于头实体, 另一个用于尾实体, 而不是类似 TransR 对两者使用相同的矩阵.

● **TransSparse**: 前面介绍的工作 Trans(E, H, R 和 D) 都忽略了知识图谱的异质性 (异质性指知识库中一些关系连接许多实体对, 而另一些关系则不连接) 和不平衡性 (不平衡性指一个关系中头实体与尾实体的数目可能不同), 为了处理这两个问题, Ji 等人提出了 TransSparse 模型<sup>[46]</sup>, 它有 TransSparse(share) 与 TransSparse(separate) 两个版本.

为了克服异质性, TransSparse(share) 模型中投影矩阵的稀疏度由关系连接的实体对数量决定, 并且关系的两侧共享相同的投影矩阵. 具体来说, TransSparse(share) 为每个关系  $r$  设置了稀疏投影矩阵  $\mathbf{M}_r(\theta_r) \in \mathbb{R}^{k \times d}$  和平移向量



$\mathbf{r} \in \mathbb{R}^k$ ,  $N_r$  表示关系  $r$  连接的实体对数量,  $N_r$  表示它们的最大数量,  $\theta_{\min}(0 \leq \theta_{\min} \leq 1)$  是计算稀疏度的超参数. 则投影矩阵  $\mathbf{M}_r(\theta_r)$  的稀疏度  $\theta_r$  定义如下:

$$\theta_r = 1 - (1 - \theta_{\min})N_r/N_r \quad (11)$$

为了处理关系不平衡问题, Ji 等人提出了第二种模型 TransSparse(separate), 其中每个关系有两个不同的稀疏投影矩阵, 分别用于头实体与尾实体, 稀疏度由关系连接的头(尾)实体数量确定. 具体来说, 在 TransSparse(separate) 中, 为每个关系设置两个不同的稀疏矩阵  $\mathbf{M}_r^h(\theta_r^h)$  和  $\mathbf{M}_r^t(\theta_r^t)$ , 并且  $\mathbf{M}_r^h(\theta_r^h), \mathbf{M}_r^t(\theta_r^t) \in \mathbb{R}^{k \times d}$  其中下标  $r$  是关系的索引,  $h, t$  表示矩阵用于哪个实体(头或尾).  $N_r^l(l=h, t)$  表示关系  $r$  在位置  $l$  处连接不同实体的数量,  $N_r^l$  表示  $N_r^l$  中的最大数. 投影矩阵的稀疏度为:

$$\theta_r^l = 1 - (1 - \theta_{\min})N_r^l/N_r^l \quad (12)$$

投影向量定义为:

$$\mathbf{h}_\perp = \mathbf{M}_r^h(\theta_r^h)\mathbf{h}, \mathbf{t}_\perp = \mathbf{M}_r^t(\theta_r^t)\mathbf{t} \quad (13)$$

• TransM: TransM<sup>[47]</sup> 将每个事实三元组  $(h, r, t)$  与特定关系权重  $w_r$  相关联. 其核心思想是, 在学习 KG 表示时, 不同的关系可能具有不同的重要性. 相应的评分函数定义为:

$$f_r(h, t) = w_r \|\mathbf{h} + \mathbf{r} - \mathbf{t}\|_{1/2} \quad (14)$$

其中,  $\mathbf{h}, \mathbf{r}, \mathbf{t} \in \mathbb{R}^d$ , TransM 通过给复杂关系分配较低的权重, 缓解了 TransE 模型对复杂关系建模的不足. 图 4 直观地显示了 TransE 与 TransM 在建模 1-N 关系时的不同.

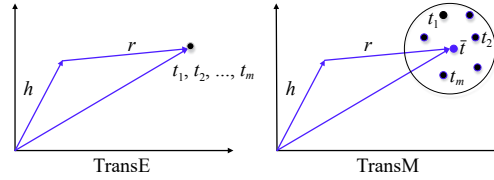


图 4 TransE 和 TransM 在建模 1-N 关系实例时的区别

• TransA: Xiao 等人<sup>[48]</sup>认为 TransE 及其之后的扩展模型存在两个主要问题: ① 复杂的关系总是导致复杂的嵌入拓扑, 而球形等势超曲面的灵活性不足以表征拓扑结构. ② 损失函数过于简化, 实体和关系向量的每个维度同等考虑, 带来大量噪声并降低了性能. 为此, Xiao 等人提出了一种基于自适应和灵活度量的嵌入方法 TransA.

TransA 利用绝对损失的自适应马氏距离取代相对不够灵活的欧几里德距离<sup>[49]</sup>. 评分函数定义为:

$$f_r(h, t) = (\|\mathbf{h} + \mathbf{r} - \mathbf{t}\|)^T \mathbf{M}_r (\|\mathbf{h} + \mathbf{r} - \mathbf{t}\|) \quad (15)$$

其中,  $(\|\mathbf{h} + \mathbf{r} - \mathbf{t}\|) = (|h_1 + r_1 - t_1|, |h_2 + r_2 - t_2|, \dots, |h_n + r_n - t_n|)$ ,  $\mathbf{h}, \mathbf{t}, \mathbf{r} \in \mathbb{R}^d$ ,  $\mathbf{M}_r \in \mathbb{R}^{d \times d}$  是与自适应度量相对应的关系特定的对称非负权重矩阵.

TransA 采用椭圆等势面代替球形等势面, 可以更好地表示由复杂关系引起的复杂嵌入拓扑. 此外, TransA 可被视为加权转换后的特征维数, 抑制了来自无关维度的噪声.

• TransF: Feng 等人<sup>[50]</sup>指出基于距离的模型采用的翻译原理太严格, 无法为复杂多样的实体和关系建模. 因此, 提出允许灵活的翻译对复杂多样的实体和关系建模, 即: 仅约束  $\mathbf{h} + \mathbf{r}$  (或  $\mathbf{t} - \mathbf{r}$ ) 的方向与  $\mathbf{t}$  (或  $\mathbf{h}$ ) 的方向相同. 其评分函数定义为:

$$f_r(h, t) = (\mathbf{h} + \mathbf{r})^T \mathbf{t} + \mathbf{h}^T (\mathbf{t} - \mathbf{r}) \quad (16)$$

其中,  $\mathbf{h}, \mathbf{r}, \mathbf{t} \in \mathbb{R}^d$ . TransF 模型遵循“灵活翻译”的原则, 因此在处理复杂关系时, 可以改善 TransE 模型的缺点.

• ITransF: 虽然 STransE 的性能优于 TransE, 但它更容易出现数据稀疏问题. Xie 等人提出了一种新的嵌入模型 ITransF<sup>[51]</sup>, 通过稀疏注意力向量学习关系与概念之间的关联, 实现了隐藏概念的发现和统计强度的传递.

ITransF 将所有概念投影矩阵堆叠为一个三维张量  $\mathbf{D} \in \mathbb{R}^{m \times n \times n}$ , 其中  $m$  是预先指定的概念投影矩阵数目,  $n$  是实体嵌入和关系嵌入的维数. ITransF 的能量函数定义为:

$$f_r(h, t) = \|\alpha_r^H \cdot \mathbf{D} \cdot \mathbf{h} + \mathbf{r} - \alpha_r^T \cdot \mathbf{D} \cdot \mathbf{t}\| \quad (17)$$

其中,  $\alpha_r^H, \alpha_r^T \in [0, 1]^m$ , 满足  $\sum_i \alpha_{r,i}^H = \sum_i \alpha_{r,i}^T = 1$  是归一化的注意力向量. 显然, 当使用  $m = 2|R|$  概念矩阵, 并将注意力向量设置为不相交的 one-hot 向量时, STTransE 可以表示为 ITransF 模型的一个特例.

• TransAt: Qian 等人<sup>[52]</sup>认为人类对关系的认知遵循一种层次化规律, 并且实体之间存在类别区分, 提出链接预测过程包含两个阶段, 在第一阶段中, 考虑通过实体的某些属性分析候选实体类别是否合理, 从合理的类别中收集候选实体; 第二阶段, 对于那些可能的三元组组合, 关注实体的细粒度属性以区分它们与特定关系的联系. 这个过程中除了学习嵌入之外还引入了学习关系相关候选对象与关系相关注意力两个任务. 而以往的模型 (如 TransH, TransR, TransSparse) 无法学习细粒度的注意力.

由此, Qian 等人<sup>[52]</sup>提出 TransAt 模型来同时学习嵌入, 关系相关候选对象和关系相关注意力. TransAt 的评分函数定义如下:

$$f_r(h, t) = P_r(\sigma(\mathbf{r}_h)\mathbf{h}) + \mathbf{r} - P_r(\sigma(\mathbf{r}_t)\mathbf{t}) \quad (18)$$

其中,  $P_r$  是仅保留与  $r$  有关维度的投影,  $\sigma$  是 Sigmoid 激活函数,  $\mathbf{r}_h, \mathbf{r}_t$  是与关系  $r$  有关的两个向量.

• TransMS: TransMS 模型<sup>[53]</sup>利用非线性函数和线性偏置向量传输多向语义, 在复杂关系的链接预测任务上取得了显著改进, 其评分函数定义为:

$$f_r(h, t) = \left\| -\tanh(\mathbf{t} \circ \mathbf{r}) \circ \mathbf{h} + \mathbf{r} - \tanh(\mathbf{h} \circ \mathbf{r}) \circ \mathbf{t} + \alpha \cdot (\mathbf{h} \circ \mathbf{t}) \right\|_{1/2} \quad (19)$$

其中, 符号  $\circ$  表示 Hadamard 乘积.

## (2) 流形和群

流形是由集合论定义为具有邻域的点的集合的拓扑空间, Point-Wise 建模是一个不适定 (从数学上讲, 不适定的代数系统通常会使解决方案不精确且不稳定) 的代数系统, 无法在大规模知识图谱中进行精确的链接预测<sup>[54]</sup>. 为了解决这些问题, Xiao 等人提出了一种基于流形的嵌入原理 (ManifoldE)<sup>[55]</sup>, 该原理可被看作是一个适定的代数系统, 它将 Point-Wise 嵌入扩展为基于流形的嵌入.

给定三元组  $(h, r, t)$ , ManifoldE 基于流形的原理  $MF(\mathbf{h}, \mathbf{r}, \mathbf{t}) \approx D_r^2$ , 当给定一个头实体和一个关系时, 尾实体位于一个高维流形中. ManifoldE 采用三元组远离流形的距离来设计评分函数:

$$f_r(h, t) = \left\| MF(\mathbf{h}, \mathbf{r}, \mathbf{t}) - D_r^2 \right\|^2 \quad (20)$$

其中,  $D_r$  是关系特定的流形参数,  $MF: \mathbb{E} \times \mathbb{L} \times \mathbb{E} \rightarrow \mathbb{R}$  是流形函数, 其中  $\mathbb{E}$  是实体集,  $\mathbb{L}$  是关系集,  $\mathbb{R}$  是实数字段.

Xiao 等人<sup>[55]</sup>介绍了基于流形嵌入的两种设置, 即 Sphere 和 Hyperplane. 在 Sphere 设置中, 使用 reproducing kernel Hilbert space (RKHS) 表示流形函数, 即:

$$MF(\mathbf{h}, \mathbf{r}, \mathbf{t}) = \|\varphi(h) + \varphi(r) - \varphi(t)\|^2 = K(h, h) + K(t, t) + K(r, r) - 2K(h, t) - 2K(r, t) + 2K(r, h) \quad (21)$$

其中,  $\varphi$  是从原始空间到 Hilbert 空间的映射,  $K$  是核函数.

另一个 Hyperplane 提出将头、尾实体嵌入两个单独超平面中, 并在它们的超平面不平时彼此相交. 因此, 将  $MF$  定义为:

$$MF(\mathbf{h}, \mathbf{r}, \mathbf{t}) = (\mathbf{h} + \mathbf{r}_h)^\top (\mathbf{t} + \mathbf{r}_t) \quad (22)$$

其中,  $\mathbf{r}_h$  和  $\mathbf{r}_t$  分别是头、尾实体的特定关系向量.

TransE 模型在计算实体距离之前, 对所有的实体和关系向量进行了正则化, 这种规范化虽然避免了向量空间的无限扩张, 但也导致了新的矛盾<sup>[56]</sup>. 为了避免正则化带来的矛盾, TorusE 模型<sup>[56]</sup>将映射空间由普通向量空间替换成了李群. 在 TransE 的嵌入模型中, 向量空间需要满足的条件有: (1) 可微的流形空间, (2) 群运算  $(+, -)$  可微且 (3) 能够定义距离函数. TorusE 模型在此基础上增加了空间紧致性的条件, 克服了 TransE 的局限性, 而且可以证明, 紧李群可以满足 TransE 遵循的优化目标和正则化条件.

Zhang 等人<sup>[56]</sup>构架了一个紧李群  $\mathbb{T}^n$  的圆环空间和圆环空间上的不同范式的距离函数  $d_{L_1}, d_{L_2}, d_{eL_2}$ , 实体和关系被表示为  $[\mathbf{h}], [\mathbf{r}], [\mathbf{t}] \in \mathbb{T}^n$ , 类似于 TransE 在  $\mathbb{R}^n$  上的优化目标  $\mathbf{h} + \mathbf{r} = \mathbf{t}$ , TorusE 在  $\mathbb{T}^n$  上构建  $[\mathbf{h}] + [\mathbf{r}] = [\mathbf{t}]$ , 并根据距离函数不同定义 3 个对应的评分函数:

$$\begin{cases} f_{L_1}(h, r, t) = 2d_{L_1}([\mathbf{h}] + [\mathbf{r}], [\mathbf{t}]) \\ f_{L_2}(h, r, t) = (2d_{L_2}([\mathbf{h}] + [\mathbf{r}], [\mathbf{t}]))^2 \\ f_{eL_2}(h, r, t) = (d_{eL_2}([\mathbf{h}] + [\mathbf{r}], [\mathbf{t}]) / 2)^2 \end{cases} \quad (23)$$

TorusE 具有比 TransE 更低的计算复杂度。

### (3) 高斯空间

前面提出的部分模型 (例如: TransE) 通过优化全局损失函数来确保 KG 中正三元组分数高于负三元组。然而, 这些模型忽略了实体和关系的 (不) 确定性。实际上, 不同的实体和关系可能包含不同的确定性, 近几年的一些研究考虑了它们的不确定性, 并将其建模为随机变量<sup>[57,58]</sup>。

● KG2E: 受高斯词嵌入的启发, 基于密度的嵌入模型 KG2E<sup>[57]</sup>引入了高斯分布来处理实体和关系的 (不) 确定性。具体来说, KG2E 将实体和关系嵌入到多维高斯分布, 每个实体和关系都由具有均值向量和协方差矩阵的高斯分布表示, 即:

$$\mathbf{h} \sim N(\mathbf{u}_h, \Sigma_h), t \sim N(\mathbf{u}_t, \Sigma_t), r \sim N(\mathbf{u}_r, \Sigma_r) \quad (24)$$

其中,  $\mathbf{u}_h, \mathbf{u}_t, \mathbf{u}_r \in \mathbb{R}^d$  是高斯分布的均值向量, 表示实体或关系在语义空间中的中心位置。  $\Sigma_h, \Sigma_t, \Sigma_r \in \mathbb{R}^{d \times d}$  是协方差矩阵, 表示该实体或关系的不确定性。

借用基于翻译的方法<sup>[15,16,43]</sup>中的概念, He 等人<sup>[57]</sup>认为从头实体到尾实体的转换结果类似于正三元组中的关系, 并使用公式  $\mathbf{h} - \mathbf{t}$  来表示, 它对应于概率分布:

$$P_e \sim N(\mathbf{u}_h - \mathbf{u}_t, \Sigma_h + \Sigma_t) \quad (25)$$

KG2E 考虑了 2 种计算相似度的方法: KL 散度和期望似然。KL 散度是一种不对称相似度, 评分函数定义如下:

$$f_r(h, t) = \int_{x \in \mathbb{R}^{4e}} N(x; \mathbf{u}_r, \Sigma_r) \log \frac{N(x; \mathbf{u}_e, \Sigma_e)}{N(x; \mathbf{u}_r, \Sigma_r)} dx \quad (26)$$

期望似然是一种对称相似度, 评分函数定义如下:

$$f_r(h, t) = \log \int_{x \in \mathbb{R}^{4e}} N(x; \mathbf{u}_e, \Sigma_e) N(x; \mathbf{u}_r, \Sigma_r) dx \quad (27)$$

● TransG: Xiao 等人<sup>[58]</sup>指出了 KGE 中的新问题——多重关系语义, 即 KG 中的一个关系可能具有与对应三元组关联的实体对所揭示的多种含义。TransG 是知识图谱嵌入的第一个生成模型, 用以解决多重关系语义问题。

TransG 模型与传统模型的对比如图 5 所示, 图 5(a) 为传统模型示例, 由于关系  $r$  所有语义均被视为是相同的, 因此现有的基于翻译的模型不能区分有效的三元组和不正确的三元组。图 5(b) 表明, 通过考虑关系的多重语义, TransG 模型可以将有效与无效的三元组区分开。

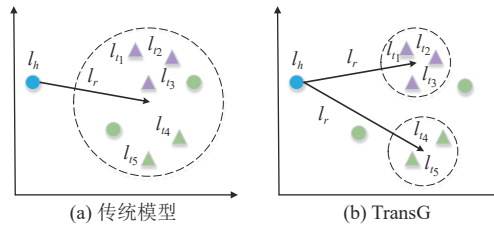


图 5 传统模型与 TransG 模型比较, 其中三角形表示正确的尾实体, 圆形表示错误的尾实体

具体来说, TransG 提出使用贝叶斯非参数无限混合嵌入<sup>[59]</sup>考虑关系的多重语义含义。对于每个实体, TransG 假定实体嵌入向量服从正态分布, 即:

$$\mathbf{h} \sim N(\mathbf{u}_h, \sigma_h^2 \mathbf{I}), \mathbf{t} \sim N(\mathbf{u}_t, \sigma_t^2 \mathbf{I}) \quad (28)$$

其中,  $\mathbf{I} \in \mathbb{R}^{d \times d}$  表示单位矩阵,  $\mathbf{u}_h, \mathbf{u}_t \sim N(0, 1)$  分别表示头、尾实体的平均嵌入向量。  $\sigma_h, \sigma_t$  分别表示头尾实体分布的方差。

TransG 认为一个关系可以具有多种语义, 应该将其表示为高斯分布的混合。因此, 将关系嵌入向量定义为:

$$\mathbf{r}_i = \mathbf{t} - \mathbf{h} \sim N(\mathbf{u}_i - \mathbf{u}_h, (\sigma_h^2 + \sigma_i^2) \mathbf{I}) \quad (29)$$

其中,  $\mathbf{r}_i$  表示关系  $r$  的第  $i$  个语义的关系嵌入向量, TransG 的评分函数定义如下:



$$f_r(h, t) = \sum_{i=1}^{M_r} \pi_r^i \exp\left(\frac{-\|h + r_i - t\|_2^2}{\sigma_h^2 + \sigma_t^2}\right) \quad (30)$$

其中,  $\pi_r^i$  是对应于关系  $r$  的第  $i$  个语义的权重因子,  $M_r$  是关系  $r$  的语义分量的数量, 由 Chinese restaurant process (CRP) 从数据中自动学习<sup>[60,61]</sup>.

### 2.1.4 模型总结

本节介绍了基于距离的评分函数的代表模型, 具体划分为基本距离模型, 翻译模型, 复杂关系建模 3 个小节进行阐述. 在复杂关系建模小节中, 按照实体与关系的不同表示空间进行细分, 表 2 对基于距离的模型进行了全面总结.

表 2 基于距离的模型总结

类别	模型	实体嵌入	关系嵌入	评分函数 $f_r(h, t)$
基本距离模型	SE <sup>[39]</sup>	$\mathbf{h}, \mathbf{t} \in \mathbb{R}^d$	$\mathbf{M}_{r,1}, \mathbf{M}_{r,2} \in \mathbb{R}^{d \times d}$	$-\ \mathbf{M}_{r,1}\mathbf{h} - \mathbf{M}_{r,2}\mathbf{t}\ $
翻译模型	TransE <sup>[15]</sup>	$\mathbf{h}, \mathbf{t} \in \mathbb{R}^d$	$\mathbf{r} \in \mathbb{R}^d$	$-\ \mathbf{h} + \mathbf{r} - \mathbf{t}\ _{1/2}$
	UM <sup>[42]</sup>	$\mathbf{h}, \mathbf{t} \in \mathbb{R}^d$	—	$-\ \mathbf{h} - \mathbf{t}\ _2^2$
复杂关系建模	TransH <sup>[16]</sup>	$\mathbf{h}, \mathbf{t} \in \mathbb{R}^d$	$\mathbf{r}, \mathbf{w}_r \in \mathbb{R}^d$	$-\ (\mathbf{h} - \mathbf{w}_r^T \mathbf{h} \mathbf{w}_r) + \mathbf{r} - (\mathbf{t} - \mathbf{w}_r^T \mathbf{t} \mathbf{w}_r)\ _2^2$
	TransR <sup>[43]</sup>	$\mathbf{h}, \mathbf{t} \in \mathbb{R}^d$	$\mathbf{r} \in \mathbb{R}^k, \mathbf{M}_r \in \mathbb{R}^{k \times d}$	$-\ \mathbf{M}_r \mathbf{h} + \mathbf{r} - \mathbf{M}_r \mathbf{t}\ _2^2$
	TransD <sup>[44]</sup>	$\mathbf{h}, \mathbf{t} \in \mathbb{R}^d, \mathbf{w}_h, \mathbf{w}_t \in \mathbb{R}^d$	$\mathbf{r}, \mathbf{w}_r \in \mathbb{R}^k$	$-\ \mathbf{h}(\mathbf{w}_r \mathbf{w}_h^T + \mathbf{I}) + \mathbf{r} - \mathbf{t}(\mathbf{w}_r \mathbf{w}_t^T + \mathbf{I})\ _2^2$
	STransE <sup>[45]</sup>	$\mathbf{h}, \mathbf{t} \in \mathbb{R}^d$	$\mathbf{r} \in \mathbb{R}^d$ $\mathbf{M}_{r,1}, \mathbf{M}_{r,2} \in \mathbb{R}^{d \times d}$	$\ \mathbf{M}_{r,1}\mathbf{h} + \mathbf{r} - \mathbf{M}_{r,2}\mathbf{t}\ _{1/2}$ $-\ \mathbf{M}_r(\theta_r)\mathbf{h} + \mathbf{r} - \mathbf{M}_r(\theta_r)\mathbf{t}\ _{1/2}^2$
	Point-Wise空间 TranSparse <sup>[46]</sup>	$\mathbf{h}, \mathbf{t} \in \mathbb{R}^d$	$\mathbf{r} \in \mathbb{R}^k, \mathbf{M}_r(\theta_r) \in \mathbb{R}^{k \times d},$ $\mathbf{M}_r^h(\theta_r^h), \mathbf{M}_r^t(\theta_r^t) \in \mathbb{R}^{k \times d}$	$-\ \mathbf{M}_r^h(\theta_r^h)\mathbf{h} + \mathbf{r} - \mathbf{M}_r^t(\theta_r^t)\mathbf{t}\ _{1/2}^2$
	TransM <sup>[47]</sup>	$\mathbf{h}, \mathbf{t} \in \mathbb{R}^d$	$\mathbf{r} \in \mathbb{R}^d$	$w_r \ \mathbf{h} + \mathbf{r} - \mathbf{t}\ _{1/2}$
	TransA <sup>[48]</sup>	$\mathbf{h}, \mathbf{t} \in \mathbb{R}^d$	$\mathbf{r} \in \mathbb{R}^d, \mathbf{M}_r \in \mathbb{R}^{d \times d}$	$(\mathbf{h} + \mathbf{r} - \mathbf{t})^T \mathbf{M}_r (\mathbf{h} + \mathbf{r} - \mathbf{t})$
	TransF <sup>[50]</sup>	$\mathbf{h}, \mathbf{t} \in \mathbb{R}^d$	$\mathbf{r} \in \mathbb{R}^d$	$(\mathbf{h} + \mathbf{r})^T \mathbf{t} + \mathbf{h}^T (\mathbf{t} - \mathbf{r})$
	ITransF <sup>[51]</sup>	$\mathbf{h}, \mathbf{t} \in \mathbb{R}^d$	$\mathbf{r} \in \mathbb{R}^d$	$\ \alpha_r^H \cdot \mathbf{D} \cdot \mathbf{h} + \mathbf{r} - \alpha_r^T \cdot \mathbf{D} \cdot \mathbf{t}\ _l$
	TransAt <sup>[52]</sup>	$\mathbf{h}, \mathbf{t} \in \mathbb{R}^d$	$\mathbf{r} \in \mathbb{R}^d$	$P_r(\sigma(\mathbf{r}_h)\mathbf{h}) + \mathbf{r} - P_r(\sigma(\mathbf{r}_t)\mathbf{t})$
	TransMS <sup>[53]</sup>	$\mathbf{h}, \mathbf{t} \in \mathbb{R}^{k_e}$	$\mathbf{r} \in \mathbb{R}^{k_r}, \alpha = r_{k_r+1} \in \mathbb{R}^1$	$\ -\tanh(\mathbf{t} \circ \mathbf{r}) \circ \mathbf{h} + \mathbf{r} - \tanh(\mathbf{h} \circ \mathbf{r}) \circ \mathbf{t} + \alpha \cdot (\mathbf{h} \circ \mathbf{t})\ _{1/2}$
	流形和群 ManifoldE <sup>[55]</sup>	$\mathbf{h}, \mathbf{t} \in \mathbb{R}^d$	$\mathbf{r} \in \mathbb{R}^d$	$\ M(\mathbf{h}, \mathbf{r}, \mathbf{t}) - D_r^2\ ^2$
	TorusE <sup>[56]</sup>	$[\mathbf{h}], [\mathbf{t}] \in \mathbb{T}^n$	$[\mathbf{r}] \in \mathbb{T}^n$	$\min_{(x,y) \in ([\mathbf{h}]+[\mathbf{r}]) \times [\mathbf{t}]} \ x - y\ _i$
高斯空间	KG2E <sup>[57]</sup>	$\mathbf{h} \sim N(u_h, \Sigma_h)$ $\mathbf{t} \sim N(\mathbf{u}_t, \Sigma_t)$ $\Sigma_h, \Sigma_t \in \mathbb{R}^{d \times d}$	$r \sim N(\mathbf{u}_r, \Sigma_r), u_r \in \mathbb{R}^d$ $\Sigma_r \in \mathbb{R}^{d \times d}$	$\int_{x \in \mathbb{R}^{k_e}} N(x; \mathbf{u}_r, \Sigma_r) \log \frac{N(x; \mathbf{u}_e, \sigma_e)}{N(x; \mathbf{u}_r, \Sigma_r)} dx$ $\log \int_{x \in \mathbb{R}^{k_e}} N(x; \mathbf{u}_e, \Sigma_e) N(x; \mathbf{u}_r, \Sigma_r) dx$
	TransG <sup>[58]</sup>	$\mathbf{h} \sim N(\mathbf{u}_h, \sigma_h^2 \mathbf{I})$ $\mathbf{t} \sim N(\mathbf{u}_t, \sigma_t^2 \mathbf{I})$	$\mathbf{u}_r^i \sim N(u_t - u_h, (\sigma_h^2 + \sigma_t^2) \mathbf{I})$ $r = \sum_i \pi_r^i \mathbf{u}_r^i \in \mathbb{R}^d$	$\sum_{i=1}^{M_r} \pi_r^i \exp\left(\frac{-\ h + r_i - t\ _2^2}{\sigma_h^2 + \sigma_t^2}\right)$

## 2.2 语义匹配模型

语义匹配模型利用基于相似性的评分函数, 即通过语义匹配来衡量事实的合理性. 语义匹配通常采用乘法公式  $(\mathbf{h}^T \mathbf{M}_r \approx \mathbf{t}^T)$  来变换表示空间中的头实体, 使其与尾实体相近. 本节根据实体和关系编码的不同模型结构来介绍代表性的语义匹配模型.

### 2.2.1 线性/双线性模型

线性/双线性模型(虽然这里考虑的一些模型乍看起来可能不是双线性的,但文献[62]证明了它们与双线性模型密切相关)通过将头实体投影到接近尾实体的表示空间中,将关系表述为一个线性/双线性映射,即通过应用线性运算(公式(31))/双线性运算(公式(32))编码实体和关系的相互作用。

$$g_r(\mathbf{h}, \mathbf{t}) = \mathbf{M}_r^\top \begin{pmatrix} \mathbf{h} \\ \mathbf{t} \end{pmatrix} \quad (31)$$

$$f_r(\mathbf{h}, \mathbf{t}) = \mathbf{h}^\top \mathbf{M}_r \mathbf{t} \quad (32)$$

• SME: 语义匹配能量模型 (semantic matching energy, SME)<sup>[20,42]</sup>提出寻找实体与关系之间的语义联系。在SME中,实体和关系类型共享相同的表示形式,将定义多关系图的所有符号嵌入到同一空间中(等价于删除实体和关系类型之间通常存在的概念差异),并且定义了若干投影矩阵(用来刻画实体与关系的内在联系)。SME使用神经网络架构进行语义匹配,图6表明了语义匹配能量函数的基本思想。

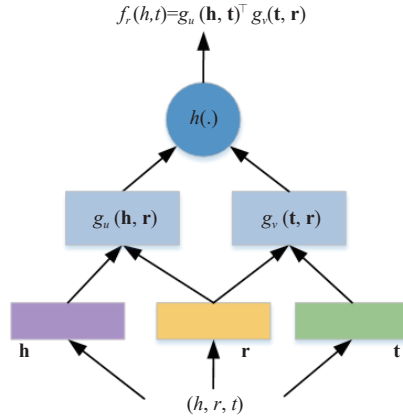


图6 语义匹配能量函数简单说明,即在语义上匹配 $(h, r)$ 和 $(r, t)$ 的实体关系对的组合

如图6所示, SME首先将输入三元组 $(h, r, t)$ 的每个符号映射到其嵌入 $\mathbf{h}, \mathbf{r}, \mathbf{t} \in \mathbb{R}^d$ , 然后关系嵌入 $\mathbf{r}$ 通过函数 $g_u(\cdot)$ 与头实体嵌入 $\mathbf{h}$ 组合, 得到 $g_u(\mathbf{h}, \mathbf{r})$ ; 并通过函数 $g_v(\cdot)$ 与尾实体嵌入 $\mathbf{t}$ 组合, 得到 $g_v(\mathbf{t}, \mathbf{r})$ . 最后, 将事实的评分函数定义为通过它们的点积匹配 $g_u(\mathbf{h}, \mathbf{r})$ 和 $g_v(\mathbf{t}, \mathbf{r})$ , 即:

$$f_r(h, t) = g_u(\mathbf{h}, \mathbf{r})^\top g_v(\mathbf{t}, \mathbf{r}) \quad (33)$$

SME为语义匹配的能量函数定义了线性形式、双线性形式两个版本, 定义如下:

SME 线性形式:

$$\begin{cases} g_u(\mathbf{h}, \mathbf{r}) = \mathbf{M}_u^1 \mathbf{h} + \mathbf{M}_u^2 \mathbf{r} + \mathbf{b}_u \\ g_v(\mathbf{t}, \mathbf{r}) = \mathbf{M}_v^1 \mathbf{t} + \mathbf{M}_v^2 \mathbf{r} + \mathbf{b}_v \end{cases} \quad (34)$$

SME 双线性形式:

$$\begin{cases} g_u(\mathbf{h}, \mathbf{r}) = (\mathbf{M}_u^1 \mathbf{h}) \circ (\mathbf{M}_u^2 \mathbf{r}) + \mathbf{b}_u \\ g_v(\mathbf{t}, \mathbf{r}) = (\mathbf{M}_v^1 \mathbf{t}) \circ (\mathbf{M}_v^2 \mathbf{r}) + \mathbf{b}_v \end{cases} \quad (35)$$

其中,  $\mathbf{M}_u^1, \mathbf{M}_u^2, \mathbf{M}_v^1, \mathbf{M}_v^2 \in \mathbb{R}^{d \times d}$  为投影矩阵;  $\mathbf{b}_u, \mathbf{b}_v \in \mathbb{R}^d$  为偏置向量, 符号 $\circ$ 表示Hadamard乘积。此外, 在文献[19]中, SME进一步扩展了它的双线性形式, 用三阶张量代替它的矩阵, 以提高其建模能力。

• LFM: 隐变量模型 (latent factor model, LFM)<sup>[62,63]</sup>采用关系特定双线性变换考虑实体和关系之间的相关性, LFM为每个三元组 $(h, r, t)$ 定义评分函数为:

$$f_r(h, t) = \mathbf{h}^\top \mathbf{M}_r \mathbf{t} \quad (36)$$

其中,  $\mathbf{M}_r \in \mathbb{R}^{d \times d}$  是关系 $r$ 对应的双线性变换矩阵。它简单有效地实现了头、尾实体的分布式表示, 较以往的模型有了很大的改进。然而, 由于LFM模型需要大量的参数来建模, 因此性能受到限制。

• **DistMult**: DistMult<sup>[38]</sup>限制  $\mathbf{M}_r$  为对角矩阵减少了 LFM 中关系参数的数量, 产生了与 TransE 模型相同数量的关系参数. 这种简化的双线性公式具有与 TransE 相同的可扩展性, 并且在链接预测任务上取得了更好的性能.

对于每个关系  $r$ , DistMult 引入一个向量嵌入  $\mathbf{r} \in \mathbb{R}^d$ , 并要求  $\mathbf{M}_r = \text{diag}(\mathbf{r})$ , 评分函数定义为:

$$f_r(h, t) = \mathbf{h}^\top \text{diag}(\mathbf{r}) \mathbf{t} \quad (37)$$

该分数仅捕获沿相同维度  $\mathbf{h}$  和  $\mathbf{t}$  分量之间的成对交互作用, 减少了每个关系的参数数量. 对于任何  $\mathbf{h}$  和  $\mathbf{t}$ , 满足  $\mathbf{h}^\top \text{diag}(\mathbf{r}) \mathbf{t} = \mathbf{t}^\top \text{diag}(\mathbf{r}) \mathbf{h}$ . 因此, 这种过度简化的模型只能处理对称关系, 该模型等效于 INDSCAL 张量分解<sup>[64]</sup>分量之间的成对交互作用, 减少了每个关系的参数数量. 对于任何  $\mathbf{h}$  和  $\mathbf{t}$ , 满足  $\mathbf{h}^\top \text{diag}(\mathbf{r}) \mathbf{t} = \mathbf{t}^\top \text{diag}(\mathbf{r}) \mathbf{h}$ . 因此, 这种过度简化的模型只能处理对称关系, 该模型等效于 INDSCAL 张量分解<sup>[64]</sup>.

• **HolE**: 为了捕获关系数据中的丰富交互并有效地进行计算, Nickel 等人提出了全息嵌入 (holographic embeddings, HolE)<sup>[65]</sup>模型, 以学习整个知识图谱的成分向量空间表示. 为了将张量积的表达能力与 TransE 的高效性和简单性相结合, HolE 使用头、尾实体向量的循环相关性表示实体对, 即使用组合运算符:  $\mathbf{a} \circ \mathbf{b} = \mathbf{a} \star \mathbf{b}$ , 其中  $\star: \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^d$  表示循环相关:

$$[\mathbf{a} \star \mathbf{b}]_k = \sum_{i=0}^{d-1} a_i b_{k+i} \bmod d \quad (38)$$

循环相关操作可以看成是张量乘法的特殊形式, 且具有不可交换性, 衡量相关性与计算效率高的优点.

通过在语义上匹配循环相关性与关系嵌入, HolE 模型的评分函数定义为:

$$f_r(h, t) = \mathbf{r}^\top (\mathbf{h} \star \mathbf{t}) \quad (39)$$

循环相关对成对交互作用进行压缩, 如图 7(b) 所示, HolE 中每个关系只需要  $O(d)$  个参数, 比 RESCAL<sup>[18,19]</sup>更有效.

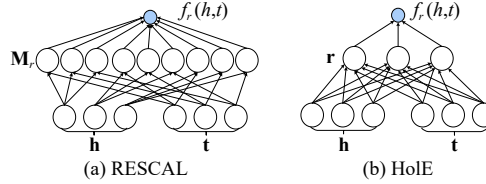


图 7 作为神经网络的 RESCAL 和 HolE, RESCAL 通过  $d^2$  个中间层表示实体对, 而 HolE 只需要  $d$  个中间层

• **ComplEx**: 由于 DistMult 模型过于简化, 只能处理对称关系, 复数嵌入 (complex embedding, ComplEx)<sup>[66]</sup>引入了复数向量空间, 使用 Hermitian 点积进行关系、头部和尾部三者的共轭式合成, 使其可以捕获对称和反对称关系, 完善了 DistMult 模型.

在 ComplEx 中, 实体和关系嵌入  $\mathbf{h}, \mathbf{r}, \mathbf{t}$  不再位于实值空间中, 而是位于复数空间中, 即:  $\mathbf{h}, \mathbf{t}, \mathbf{r} \in \mathbb{C}^d$ , 可以更好地对非对称关系建模. 形式上, ComplEx 的事实  $(h, r, t)$  的评分函数定义为:

$$f_r(h, t) = \text{Re}(\mathbf{h}^\top \text{diag}(\mathbf{r}) \bar{\mathbf{t}}) \quad (40)$$

其中,  $\bar{\mathbf{t}}$  是  $\mathbf{t}$  的共轭,  $\text{Re}(\cdot)$  表示取复数值的实部. 该评分函数不是对称的, 非对称关系的事实可以根据所涉及实体的顺序得到不同的分数.

文献 [67] 已经证明 ComplEx 在数学上等效于 HolE, 并且 ComplEx 将 HolE 归类为对嵌入施加共轭对称性的一种特殊情况. 值得注意的是, ComplEx 也可以看成是 RESCAL 的扩展, 它指定了实体和关系的复嵌入.

• **ANALOGY**: ANALOGY<sup>[68]</sup>侧重于多关系推理, 对关系数据的类比结构进行建模. ANALOGY 使用与 LFM 相同的双线性形式评分函数度量事实三元组的概率, 它的评分函数定义为:

$$f_r(h, t) = \mathbf{h}^\top \mathbf{M}_r \mathbf{t} \quad (41)$$

其中,  $\mathbf{h}, \mathbf{t} \in \mathbb{R}^d$  是头、尾实体的向量嵌入,  $\mathbf{M}_r \in \mathbb{R}^{d \times d}$  是与关系相关联的线性映射, 为了建模类比结构, 它要求关系的线性映射是正态的且相互可交换, 如下所示:

$$\begin{cases} \text{normality: } \mathbf{M}_r \mathbf{M}_r^\top = \mathbf{M}_r^\top \mathbf{M}_r, \forall r \in \mathbb{R} \\ \text{commutativity: } \mathbf{M}_r \mathbf{M}_{r'} = \mathbf{M}_{r'} \mathbf{M}_r, \forall r, r' \in \mathbb{R} \end{cases} \quad (42)$$

虽然 ANALOGY 将关系表示为矩阵,但这些矩阵可以同时分块对角化为一组稀疏的对角矩阵,并且已经证明 DistMult, HolE 和 ComplEx 方法在原则上都可以被 ANALOGY 归类为特殊情况。

• Simple: CP 分解 (canonical Polyadic decomposition)<sup>[69]</sup>是最早提出的张量分解方法,该方法为每个关系学习一个嵌入向量,并为每个实体学习两个嵌入向量,一个用于头实体嵌入,另一个用于尾实体嵌入。由于头、尾实体嵌入的学习是独立的,这导致了 CP 方法在知识图谱补全上性能较差。Simple<sup>[70]</sup>是基于 CP 的张量分量分解方法,解决了实体的两个嵌入向量之间的独立性问题。

Simple 可以被看成是一种可解释的且具有完全表达能力的双线性模型,与其他模型相比,它具有完全表达能力,同时冗余参数较少。此外,它还可以通过参数共享的方式将背景知识编码进嵌入中。Simple 引入了关系的逆,并且计算  $(h, r, t)$  和  $(t, r^{-1}, h)$  的 CP 评分的平均值,三元组  $(h, r, t)$  的评分函数定义为:

$$f_r(h, t) = \frac{1}{2}(\mathbf{h} \circ \mathbf{r} \mathbf{t} + \mathbf{t} \circ \mathbf{r}' \mathbf{h}) \quad (43)$$

其中,  $\mathbf{r}'$  是关系的逆的嵌入,符号  $\circ$  表示 Hadamard 乘积。Simple 模型在实体预测任务中性能良好,并且具有编码先验知识的能力。

### 2.2.2 矩阵分解模型

矩阵分解是获得低维向量表示的一项重要技术,在利用矩阵分解进行知识表示学习中的一个典型模型是 RESCAL<sup>[17,18]</sup>,它是一种基于三向张量分解的新型关系学习方法。RESCAL 模型将 KG 结构的建模归结为一个张量分解操作,在 RESCAL 中,KG 中的三元组  $(h, r, t)$  形成一个大的张量  $X$ ,如果三元组存在,则  $X_{hrt}=1$ ,否则为 0。张量分解旨在将  $X$  分解为实体嵌入和关系嵌入,以使  $X_{hrt}$  接近于  $\mathbf{h} \mathbf{M}_r \mathbf{t}$ 。给定事实三元组  $(h, r, t)$ ,定义其评分函数为:

$$f_r(h, t) = \mathbf{h}^T \mathbf{M}_r \mathbf{t} = \sum_{i=0}^{d-1} \sum_{j=0}^{d-1} [\mathbf{M}_r]_{ij} \cdot [\mathbf{h}]_i \cdot [\mathbf{t}]_j \quad (44)$$

其中,  $\mathbf{h}, \mathbf{t} \in \mathbb{R}^d$  分别是头、尾实体的向量表示,  $[\mathbf{h}]_i$  和  $[\mathbf{t}]_j$  分别表示向量  $\mathbf{h}, \mathbf{t}$  的第  $i, j$  项。  $\mathbf{M}_r \in \mathbb{R}^{d \times d}$  是一个与关系相关联的矩阵。该分数捕获了  $\mathbf{h}$  和  $\mathbf{t}$  的所有分量之间的成对交互作用 (见图 7(a)),  $[\mathbf{M}_r]_{ij}$  表示矩阵  $\mathbf{M}_r$  的第  $ij$  项。可以发现 RESCAL 与前面介绍的 LFM 模型相似,主要区别在于 RESCAL 优化张量  $X$  中的所有值 (包括零值),而 LFM 则集中于优化 KG 中的三元组。

TATEC<sup>[71]</sup>模型不仅对三向交互  $\mathbf{h}^T \mathbf{M}_r \mathbf{t}$  进行建模,而且还建模双向交互,例如实体与关系之间的交互。它的评分函数为:

$$f_r(h, t) = \mathbf{h}^T \mathbf{M}_r \mathbf{t} + \mathbf{h}^T \mathbf{r} + \mathbf{t}^T \mathbf{r} + \mathbf{h}^T \mathbf{D} \mathbf{t} \quad (45)$$

其中,  $\mathbf{D}$  是在所有不同关系中共享的对角矩阵。

通过引入三阶 Tucker 张量分解, TuckER<sup>[72]</sup>通过输出一个核心张量以及实体和关系的嵌入向量来学习嵌入,它的评分函数定义为:

$$f_r(h, t) = w \times_1 \mathbf{h} \times_2 \mathbf{r} \times_3 \mathbf{t} \quad (46)$$

其中,  $w \in \mathbb{R}^{d_e \times d_r \times d_e}$  是 Tucker 分解得到的核心张量,  $d_e$  和  $d_r$  分别表示实体和关系的向量维数,  $\times_n$  表示沿着第  $n$  维的张量乘法。

除了 RESCAL 模型,在知识图谱嵌入中还有其他的利用矩阵分解的研究工作。文献 [17,73] 学习头尾实体对而不是单个实体表示,形式上,它建立一个实体-关系矩阵  $Y$ ,当  $(h, r, t)$  成立时,  $Y_{hrt}$  为 1,否则为 0,利用矩阵分解将  $Y$  分解为实体对  $(h, t)$  嵌入  $\mathbf{P}$  ( $\mathbf{P} \in \mathbb{R}^d$ ) 和关系嵌入  $\mathbf{r}$  ( $\mathbf{r} \in \mathbb{R}^d$ ),事实的合理性通过  $\mathbf{P}$  与  $\mathbf{r}$  的内积衡量。类似地,文献 [74] 和文献 [75] 用两个单独的向量对头实体和关系-尾实体对建模,其中将头实体建模为向量  $\mathbf{h}$  ( $\mathbf{h} \in \mathbb{R}^d$ ),并将关系-尾实体对  $(r, t)$  建模为另一个向量  $\mathbf{P}$  ( $\mathbf{P} \in \mathbb{R}^d$ )。然而,这样的成对建模不能捕获成对的交互,并且更容易受到数据稀疏性的影响<sup>[37]</sup>。

### 2.2.3 神经网络模型

神经网络模型旨在以实体和关系的嵌入为输入,通过神经网络输出事实三元组的概率。神经网络用非线性神经激活函数和更复杂的网络结构来编码关系数据。在最近的研究中,用于编码语义匹配的神经网络取得了显著的

预测性能. 具有线性/双线性块的编码模型也可以使用神经网络进行建模, 例如: SME 模型 (具体介绍见第 3.2.1 节). 代表性的神经网络模型有 MLP<sup>[76]</sup>, SLM<sup>[77]</sup>, NTN<sup>[77]</sup>, NAM<sup>[78]</sup>等, 图 8 显示了部分相关模型的神经网络体系结构.

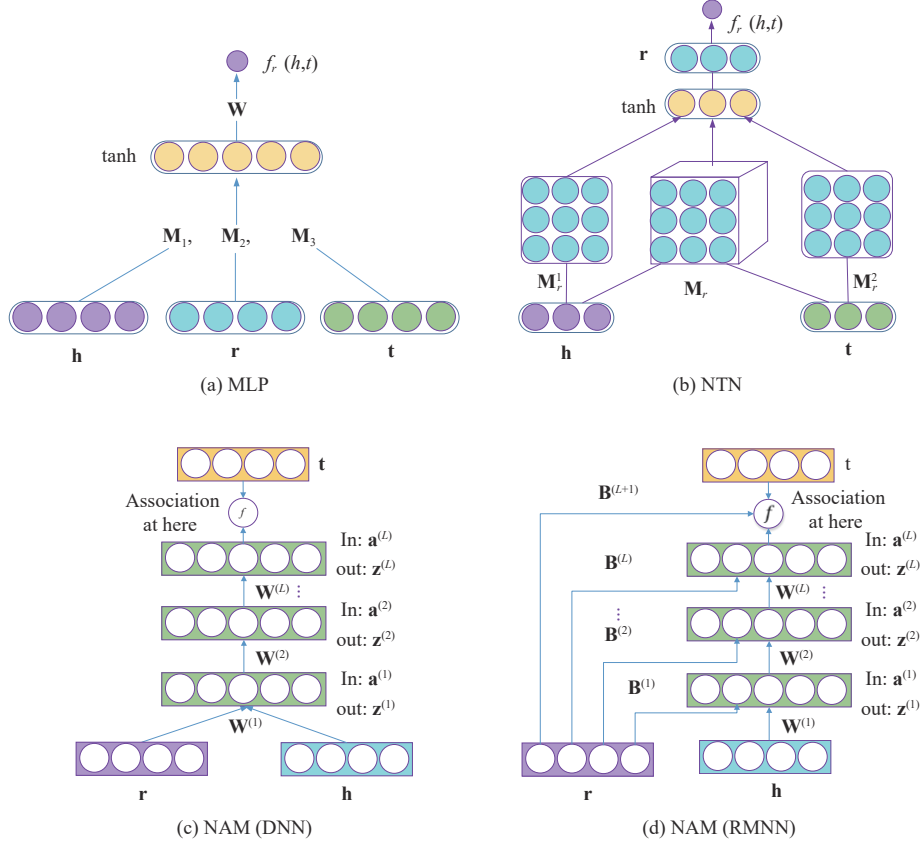


图 8 MLP, NTN, NAM (DNN) 和 NAM (RMNN) 模型的神经网络结构

- MLP: 多层感知机 (multi-layer perceptron, MLP)<sup>[76]</sup>也被称为多层人工神经网络. 它将每个关系以及实体与单个向量相关联, 使用标准的多层感知器捕获实体和关系之间的交互. MLP 模型的每个三元组  $(h, r, t)$  的评分函数定义如下:

$$f_r(h, t) = \mathbf{w}^\top \tanh(\mathbf{M}_1 \mathbf{h} + \mathbf{M}_2 \mathbf{r} + \mathbf{M}_3 \mathbf{t}) \quad (47)$$

其中,  $\mathbf{M}_1, \mathbf{M}_2, \mathbf{M}_3 \in \mathbb{R}^{d \times d}$  和  $\mathbf{w} \in \mathbb{R}^d$  是 MLP 的参数.

- SLM: 单层神经网络模型 (single layer model, SLM)<sup>[77]</sup>与 MLP 模型相似, 它尝试通过单层 MLP 神经网络的非线性操作隐式的连接实体与关系嵌入, 以减轻基本距离模型 (SE) 无法协同精确刻画实体与关系的语义联系问题. SLM 模型的评分函数定义为:

$$f_r(h, t) = \mathbf{r}^\top \tanh(\mathbf{M}_{r,1} \mathbf{h} + \mathbf{M}_{r,2} \mathbf{t}) \quad (48)$$

其中,  $\mathbf{M}_{r,1}, \mathbf{M}_{r,2} \in \mathbb{R}^{k \times d}$  是投影矩阵. 虽然 SLM 模型对基本距离模型 (SE) 进行了改进, 但是非线性操作是以一个更困难的优化问题为代价, 只提供了两个实体向量之间的微弱联系.

- NTN: 张量神经网络模型 (neural tensor network, NTN)<sup>[77]</sup>利用双线性张量在不同的维度下将头、尾实体向量联系起来, NTN 模型的每个三元组  $(h, r, t)$  的评分函数定义如下:

$$f_r(h, t) = \mathbf{r}^\top \tanh(\mathbf{h}^\top \underline{\mathbf{M}}_r \mathbf{t} + \mathbf{M}_r^1 \mathbf{h} + \mathbf{M}_r^2 \mathbf{t} + \mathbf{b}_r) \quad (49)$$

其中,  $\underline{\mathbf{M}}_r \in \mathbb{R}^{d \times d \times k}$  是一个三阶张量,  $\mathbf{M}_r^1, \mathbf{M}_r^2 \in \mathbb{R}^k \times d$  是与关系  $r$  有关的投影矩阵, 并且  $\mathbf{r}$  是关系  $r$  的向量. 可以看



出, SLM 是 NTN 将张量层数设置为 0 时的特殊情况, 同时它还可以看作是 MLP 和双线性模型的组合.

同时, 与以往的知识图谱嵌入模型不同, NTN 中的实体向量是该实体中所有单词向量的平均值. 然而, 尽管 NTN 中的张量运算可以更明确地描述实体与关系之间的复杂关系语义关联, 但 NTN 的高复杂性限制了其在大规模稀疏知识图谱上的应用.

• **NAM:** Liu 等人提出了一种基于深度神经网络的神经关联模型 (neural association model, NAM)<sup>[78]</sup>, 用于人工智能中的概率推理. 该模型利用深度神经网络中的多层非线性激活函数建模头、尾实体之间的条件概率, 文献 [78] 研究了 NAM 的两种模型结构, 一个是标准的深度神经网络 (deep neural networks, DNN), 另一个采用一种称为关系调制神经网络的结构 (relation-modulated neural nets, RMNN).

NAM-DNN 的模型结构如图 8(c) 所示. 给定事实三元组  $(h, r, t)$ , 首先在输入层中连接头实体与关系的嵌入向量, 得到  $\mathbf{z}^{(0)} = [\mathbf{h}; \mathbf{r}] \in \mathbb{R}^{2d}$ , 并将  $\mathbf{z}^{(0)}$  作为输入馈送到一个由  $L$  个整流线性隐层组成的深度神经网络, 使得:

$$\begin{cases} \mathbf{a}^{(l)} = \mathbf{W}^{(l)}\mathbf{z}^{(l-1)} + \mathbf{b}^{(l)}, (l = 1, \dots, L) \\ \mathbf{z}^{(l)} = \text{ReLU}(\mathbf{a}^{(l)}), (l = 1, \dots, L) \end{cases} \quad (50)$$

其中,  $\mathbf{W}^{(l)}$  和  $\mathbf{b}^{(l)}$  分别表示第  $l$  层的权重矩阵和偏置.

通过利用最后一个隐层输出与尾实体嵌入计算每个三元组的 Sigmoid 分数作为关联概率:

$$f_r(h, t) = \text{Sigmoid}(\mathbf{t}^\top \mathbf{z}^{(L)}) \quad (51)$$

与 NAM-DNN 不同, 如图 8(d) 所示, NAM-RMNN 将关系嵌入  $\mathbf{r}$  连接到深度神经网络中的所有隐藏层, 则有:

$$\mathbf{a}^{(l)} = \mathbf{W}^{(l)}\mathbf{z}^{(l-1)} + \mathbf{B}^{(l)}\mathbf{r}, (l = 1, \dots, L) \quad (52)$$

其中,  $\mathbf{W}^{(l)}$  和  $\mathbf{B}^{(l)}$  分别表示第  $l$  层的权重矩阵和关系特定权重矩阵. 并且 NAM-RMNN 在最上层使用以下评分函数为每个三元组计算最终得分:

$$f_r(h, t) = \text{Sigmoid}(\mathbf{t}^\top \mathbf{z}^{(L)} + \mathbf{B}^{(L+1)}\mathbf{r}) \quad (53)$$

## 2.2.4 模型总结

本节介绍了基于相似性评分函数的语义匹配模型, 并按照对实体和关系的交互进行编码的不同模型体系结构进行划分, 包括线性/双线性模型, 矩阵分解模型, 神经网络模型 3 个部分. 表 3 对语义匹配模型进行了总结.

## 2.3 最新知识图谱嵌入模型

大多数翻译模型和双线性模型是 2016 年之前提出的方法, 而最近几年研究 KGE 的方法众多. 本小节简要介绍其中的主流方法, 具体划分为卷积神经网络模型, 旋转模型, 双曲几何模型和其他模型.

### 2.3.1 卷积神经网络

卷积神经网络 (convolutional neural networks, CNNs) 在自然语言处理领域受到广泛关注, 与全连接神经网络相比, CNN 学习非线性特征, 以非常少的参数数量捕捉复杂关系. 此外, CNN 可用于学习深层表达特征. ConvE<sup>[79]</sup> 是第一个使用二维卷积层的神经链接预测模型, 其中输入实体和关系之间的交互由卷积层和全连接层建模, 图 9 显示了 ConvE 模型的体系结构. 在 ConvE 模型中, 实体和关系嵌入首先被重构和连接 (步骤①、②), 然后得到的矩阵被用作卷积层的输入 (步骤③); 将得到的特征映射张量向量化并投影到  $k$  维空间中 (步骤④), 并与所有候选尾实体嵌入匹配 (步骤⑤).

ConvE 将头实体和关系重构为二维矩阵, 通过卷积层和全连接层建模实体和关系之间的相互作用. 然后与矩阵  $\mathbf{W}$  和尾实体进行计算, 判断当前事实三元组的可信度. 形式上, ConvE 的评分函数定义如下:

$$f_r(h, t) = \sigma(\text{vec}(\sigma([\mathbf{M}_h; \mathbf{M}_r] * \omega))\mathbf{W})\mathbf{t} \quad (54)$$

其中,  $\mathbf{M}_h, \mathbf{M}_r$  分别表示头实体嵌入  $\mathbf{h}$  和关系嵌入  $\mathbf{r}$  的二维矩阵, 如果  $\mathbf{h}, \mathbf{r} \in \mathbb{R}^d$ , 则  $\mathbf{M}_h, \mathbf{M}_r \in \mathbb{R}^{d_w \times d_h}$ , 其中  $d = d_w d_h$ ,  $\text{vec}$  是将张量重构为向量的一个向量化操作.

ConvE 是用于链接预测最简单的多层卷积体系结构, 可以通过多层非线性特征学习表达语义信息, 并且该模型具有很高的参数利用率, 可以在参数分别减少了 8× 和 17× 的前提下, 得到与 DistMult<sup>[38]</sup> 和 R-GCN<sup>[80]</sup> 模型相同的性能.

表 3 语义匹配模型总结

类别	模型	实体嵌入	关系嵌入	评分函数 $f_r(h, t)$
线性/双线性模型	SME <sup>[20]</sup>	$\mathbf{h}, \mathbf{t} \in \mathbb{R}^d$	$\mathbf{r} \in \mathbb{R}^d$	$g_u(\mathbf{h}, \mathbf{r})^\top g_v(\mathbf{t}, \mathbf{r})$
	LFM <sup>[63]</sup>	$\mathbf{h}, \mathbf{t} \in \mathbb{R}^d$	$\mathbf{u}_r, \mathbf{v}_r \in \mathbb{R}^p$	$\mathbf{h}^\top \sum_{i=1}^d \alpha_i^r \mathbf{u}_i \mathbf{v}_i^\top \mathbf{t}$
	DistMult <sup>[38]</sup>	$\mathbf{h}, \mathbf{t} \in \mathbb{R}^d$	$\mathbf{r} \in \mathbb{R}^d$	$\mathbf{h}^\top \text{diag}(\mathbf{r}) \mathbf{t}$
	HolE <sup>[65]</sup>	$\mathbf{h}, \mathbf{t} \in \mathbb{R}^d$	$\mathbf{r} \in \mathbb{R}^d$	$\mathbf{r}^\top (\mathbf{h} \star \mathbf{t})$
	ComplEx <sup>[66]</sup>	$\mathbf{h}, \mathbf{t} \in \mathbb{C}^d$	$\mathbf{r} \in \mathbb{C}^d$	$\text{Re}(\mathbf{h}^\top \text{diag}(\mathbf{r}) \bar{\mathbf{t}})$
	ANALOGY <sup>[68]</sup>	$\mathbf{h}, \mathbf{t} \in \mathbb{R}^d$	$\mathbf{M}_r \in \mathbb{R}^{d \times d}$	$\mathbf{h}^\top \mathbf{M}_r \mathbf{t}$
	SimpleE <sup>[69]</sup>	$\mathbf{h}, \mathbf{t} \in \mathbb{R}^d$	$\mathbf{r}, \mathbf{r}' \in \mathbb{R}^d$	$\frac{1}{2}(\mathbf{h} \circ \mathbf{r} \mathbf{t} + \mathbf{t} \circ \mathbf{r}' \mathbf{t})$
矩阵分解模型	RESCAL <sup>[18]</sup>	$\mathbf{h}, \mathbf{t} \in \mathbb{R}^d$	$\mathbf{M}_r \in \mathbb{R}^{d \times d}$	$\mathbf{h}^\top \mathbf{M}_r \mathbf{t}$
	TATEC <sup>[71]</sup>	$\mathbf{h}, \mathbf{t} \in \mathbb{R}^d$	$\mathbf{r} \in \mathbb{R}^d, \mathbf{M}_r \in \mathbb{R}^{d \times d}$	$\mathbf{h}^\top \mathbf{M}_r \mathbf{t} + \mathbf{h}^\top \mathbf{r} + \mathbf{t}^\top \mathbf{r} + \mathbf{h}^\top \mathbf{D} \mathbf{t}$
	Tucker <sup>[72]</sup>	$\mathbf{h}, \mathbf{t} \in \mathbb{R}_e^d$	$\mathbf{r} \in \mathbb{R}_r^d$	$w \times_1 \mathbf{h} \times_2 \mathbf{r} \times_3 \mathbf{t}$
神经网络模型	MLP <sup>[76]</sup>	$\mathbf{h}, \mathbf{t} \in \mathbb{R}^d$	$\mathbf{r} \in \mathbb{R}^d$	$w^\top \tanh(\mathbf{M}_1 \mathbf{h} + \mathbf{M}_2 \mathbf{r} + \mathbf{M}_3 \mathbf{t})$
	SLM <sup>[77]</sup>	$\mathbf{h}, \mathbf{t} \in \mathbb{R}^d$	$\mathbf{r} \in \mathbb{R}^k, \mathbf{M}_r^1, \mathbf{M}_r^2 \in \mathbb{R}^{k \times d}$	$\mathbf{r}^\top \tanh(\mathbf{M}_{r,1} \mathbf{h} + \mathbf{M}_{r,2} \mathbf{t})$
	NTN <sup>[77]</sup>	$\mathbf{h}, \mathbf{t} \in \mathbb{R}^d$	$\mathbf{r}, b_r \in \mathbb{R}^k, \mathbf{M}_r \in \mathbb{R}^{d \times d \times k}, \mathbf{M}_r^1, \mathbf{M}_r^2 \in \mathbb{R}^{k \times d}$	$\mathbf{r}^\top \tanh(\mathbf{h}^\top \mathbf{M}_r \mathbf{t} + \mathbf{M}_r^1 \mathbf{h} + \mathbf{M}_r^2 \mathbf{t} + b_r)$
	NAM <sup>[78]</sup>	$\mathbf{h}, \mathbf{t} \in \mathbb{R}^d$	$\mathbf{r} \in \mathbb{R}^d$	$\text{Sigmoid}(\mathbf{t}^\top \mathbf{z}^{(L)} + \mathbf{B}^{(L+1)} \mathbf{r})$

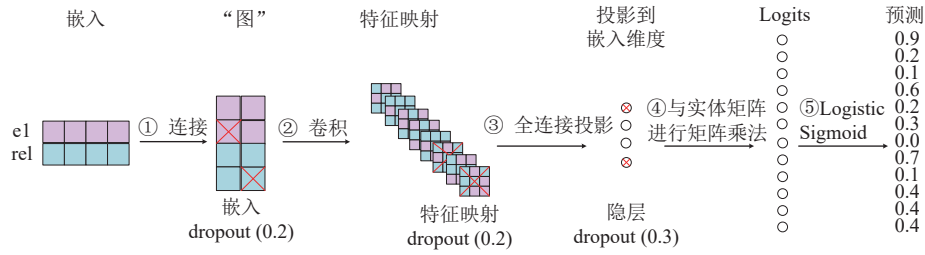


图 9 ConvE 模型体系结构

然而, Nguyen 等人<sup>[81]</sup>认为 ConvE 只考虑了头实体向量  $\mathbf{h}$  或关系向量  $\mathbf{r}$  中不同维度条目之间的局部关系, 没有考虑嵌入三元组  $(\mathbf{h}, \mathbf{r}, \mathbf{t})$  中相同维度条目之间的全局关系, 忽略了过渡特征. 针对这个问题, Nguyen 等人提出用于知识库补全的实体和关系嵌入模型 ConvKB<sup>[81]</sup>. ConvKB 采用 CNN 编码实体与关系的级联, 而不需要重构, 与捕捉局部关系的 ConvE 相比, ConvKB 保留了过渡特征.

图 10 显示了 ConvKB 的计算过程 (嵌入大小  $k=4$ , 卷积核数目  $\tau=3$ , 激活函数  $g=\text{ReLU}$ ). 在 ConvKB 中, 每个实体或关系都与唯一的  $k$  维嵌入相关联. 对于每个三元组  $(h, r, t)$ , 对应的  $k$  维嵌入三元组  $(\mathbf{h}, \mathbf{r}, \mathbf{t})$  表示为一个  $k \times 3$  的输入矩阵, 将该输入矩阵馈送到卷积层, 在卷积层中使用对应  $1 \times 3$  形状的不同过滤器提取嵌入三元组的相同维度条目之间的全局关系. 这些过滤器在输入矩阵的每一行上重复操作, 以产生不同的特征映射. 设  $\omega$  和  $\tau$  分别表示过滤器的集合和过滤器的数目, 即:  $\tau = |\omega|$ , 则得到  $\tau$  个特征映射. 将  $\tau$  个特征映射连接成单个特征向量, 通过点积计算该向量与权重向量  $\mathbf{w}$  ( $\mathbf{w} \in \mathbb{R}^{\tau \times 1}$ ) 以给出三元组  $(h, r, t)$  的得分. ConvKB 模型的评分函数定义如下:

$$f_r(h, t) = \text{concat}(g([\mathbf{h}, \mathbf{r}, \mathbf{t}] * \omega)) \cdot \mathbf{w} \quad (55)$$

ConvKB 模型可以看作是 TransE 进一步建模全局关系的扩展.

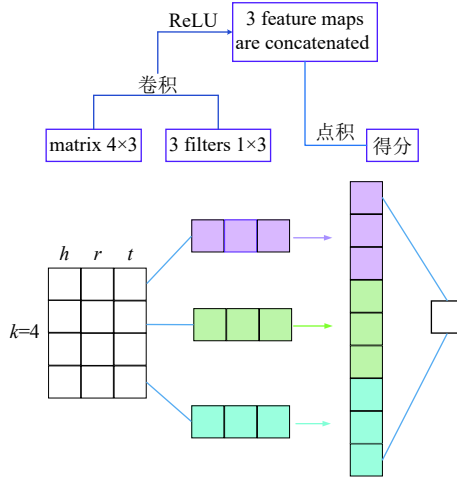


图 10 ConvKB 涉及的计算过程

• HypER: HypER<sup>[82]</sup>利用超网络<sup>[83]</sup>生成一维关系特定的卷积过滤器, 以实现知识图谱中不同关系之间的多任务知识共享, 同时超网络体系结构能够在非线性表现和要学习的参数数量之间进行权衡, 因此该模型也简化了二维 ConvE 模型引入的实体和关系嵌入之间的交互. 此外, HypER 模型使用超网络生成的关系特定卷积过滤器对嵌入的头实体的每个维度进行卷积. 因此, 该模型具有更强的表达能力.

HypER 的可视化如图 11 所示, 头实体嵌入  $\mathbf{h}$  与由超网络  $\mathbf{H}$  根据关系嵌入  $\mathbf{r}$  创建的滤波器  $\mathbf{F}_r$  进行卷积, 将获得的特征映射  $\mathbf{M}_r$  通过权重矩阵  $\mathbf{W}$  和非线性函数  $f$  映射到  $d_e$  维空间中, 并利用内积操作与每个尾实体向量  $\mathbf{t} \in \mathbf{T}$  组合, 为每个三元组提供分数, 最后应用 Sigmoid 函数预测分数. HypER 的关系特定评分函数定义为:

$$f_r(h, t) = f(\text{vec}(\mathbf{h} * \mathbf{F}_r) \mathbf{W}) \mathbf{t} = f(\text{vec}(\mathbf{h} * \text{vec}^{-1}(\mathbf{rH})) \mathbf{W}) \mathbf{t} \quad (56)$$

其中,  $\text{vec}^{-1}$  表示将向量重塑为矩阵的操作, 非线性函数  $f$  为 ReLU, 超网络  $\mathbf{H} \in \mathbb{R}^{d_r \times l_f n_f}$ , 其中  $l_f$  表示滤波器长度,  $n_f$  为每个关系的滤波器数量. 关系嵌入  $\mathbf{r} \in \mathbb{R}^{d_r}$ , 滤波器  $\mathbf{F}_r = \text{vec}^{-1}(\mathbf{rH}) \in \mathbb{R}^{l_f n_f}$ . 特征映射  $\mathbf{M}_r \in \mathbb{R}^{l_m n_f}$ , 其中特征映射长度  $l_m = d_e - l_f + 1$ .

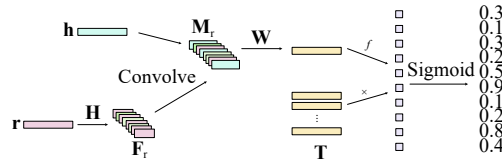


图 11 HypER 模型结构的可视化

HypER 是第一个通过将关系特定过滤器与实体嵌入卷积以非线性地结合实体与关系嵌入的模型, 在链接预测任务上取得了良好的性能.

### 2.3.2 旋转模型

现实世界中的 KG 通常是不完整的, 因此, 预测缺少的链接是知识图谱面临的一个首要问题. 为了预测缺失链接, 从观察到的事实中找到建模和推断对称/反对称, 反转和合成模式的方法是非常重要的.

实际上, 许多现有的方法一直试图隐式或显式的建模上述关系模式, 但现有模型通常仅能捕获关系模式的一部分. 例如: TransE 模型将关系表示为翻译, 旨在隐式地建模反转和合成模式, 但不能对对称关系进行建模; DistMult 模型对头实体、关系和尾实体之间的三向交互进行建模, 旨在对对称模式进行建模, 但其不能有效的建模反对称关系; ComplEx 通过引入复杂的嵌入扩展 DisMult, 以更好地对反对称关系进行建模, 但它不能有效地推断合成模式.

针对现有模型无法同时对上述 3 种模式进行建模这一问题, 受欧拉公式启发, Sun 等人提出了一种新的知识图谱嵌入方法 RotatE<sup>[84]</sup>, 该方法能够同时对反转, 对称/反对称和合成等关系模式进行建模和推理.

RotatE 模型将实体和关系映射到复数向量空间, 并将每个关系定义为从头实体到尾实体的旋转. 即给定三元组  $(h, r, t)$ , 期望  $\mathbf{t} = \mathbf{h} \circ \mathbf{r}$ , 其中  $\mathbf{t}, \mathbf{h}, \mathbf{r} \in \mathbb{C}^d$  是嵌入, 模长  $|r_i| = 1$ . 因此, 对于复数空间中的每个维度, 期望:  $t_i = h_i r_i$ , 其中  $h_i, r_i, t_i \in \mathbb{C}$

按照上述定义, 对于每个三元组, 将 RotatE 的评分函数定义如下:

$$f_r(h, t) = \|\mathbf{h} \circ \mathbf{r} - \mathbf{t}\| \quad (57)$$

通过将每个关系定义为复数向量空间中的旋转, RotatE 可以对上述 3 种类型的关系模式进行建模和推理, 并且由于 RotatE 模型在时间和内存上都保持线性, 因此易于扩展到大型知识图谱. 此外, 2018 年提出的 TorusE 模型可以看做是 RotatE 的特例, 其中嵌入模长设置为固定值, 而 RotatE 定义在整个复数空间上, 具有更强的表示能力.

QuatE<sup>[85]</sup>扩展了复值空间, 引入更有表现力的超复数表示来建模实体和关系, 在提供几何解释的同时满足了对称/反对称, 反转和合成等关系模式的建模需求. 具体来说, QuatE 利用四元数嵌入表示实体和关系, 每个四元数嵌入是超复数空间  $\mathbb{H}$  中的一个向量, 它具有 3 个虚分量  $\mathbf{i}, \mathbf{j}, \mathbf{k}$ , 则四元数可以表示为  $Q = a + b\mathbf{i} + c\mathbf{j} + d\mathbf{k}$ . QuatE 的评分函数定义为:

$$f_r(h, t) = \mathbf{h} \otimes \frac{\mathbf{r}}{|\mathbf{r}|} \cdot \mathbf{t} \quad (58)$$

其中,  $h, r, t \in \mathbb{H}^d$ , 表示内积. QuatE 利用四元数表示的优势, 实现了头、尾实体之间丰富且富有表现力的语义匹配. 与 RotatE 只有一个旋转平面不同, QuatE 具有两个旋转平面. 而且, 与欧拉角相比, 四元数可以避免万向节锁问题, 同时, 四元数也比旋转矩阵更有效且数值稳定.

除了 RotatE 和 QuatE 利用复数空间解决关系模式的建模外, DihEdral 模型<sup>[86]</sup>利用群论知识解决上述关系模式问题. Xu 等人提出的 DihEdral 模型, 旨在利用二面体群表示建模知识图谱中的关系, 这是第一次尝试在 KGE 中使用有限非交换群解决关系模式问题的方法.

对于一个  $k$  边 ( $k \in \mathbb{Z}^+$ ) 多边形, 对应的二面体群表示为  $\mathbb{D}_k$ , 其中  $\mathbb{D}_k$  由  $2k$  个元素组成, 包括  $K$  个旋转操作  $\mathbf{O}_k$  和  $K$  个反射操作  $\mathbf{F}_k$ . 需要注意的是, 当  $k$  可以被 4 整除时, 旋转矩阵  $\mathbf{O}_k^{(K/4)}$  和  $\mathbf{O}_k^{(3K/4)}$  是反对称的, 反射矩阵  $\mathbf{F}_k^{(m)}$  和旋转矩阵  $\mathbf{O}_k^{(0)}$ ,  $\mathbf{O}_k^{(K/2)}$  是对称的,  $\mathbb{D}_4$  的表示如图 12 所示, 每个子图表示对左上角 ( $\mathbf{O}_4^{(0)}$  上方) 的 ACL 正方形应用相应操作后的结果, 上面一行对应旋转操作, 下面一行对应反射操作.

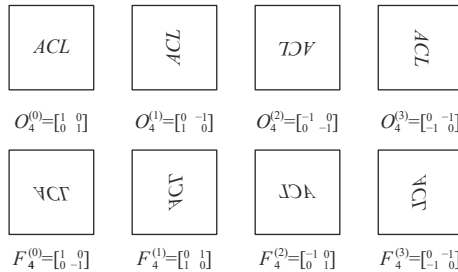


图 12  $\mathbb{D}_4$  中的元素

DihEdral 模型使用  $\mathbb{D}_k$  中的群元素建模关系. 关系矩阵采用块对角形式  $\mathbf{R} = \text{diag}(\mathbf{R}^{(1)}, \mathbf{R}^{(2)}, \dots, \mathbf{R}^{(L)})$ , 其中  $\mathbf{R}^{(l)} \in \mathbb{D}_k$  且  $l \in 1, 2, \dots, L$ . 对应的嵌入向量  $\mathbf{h} \in \mathbb{R}^{2L}$  和  $\mathbf{t} \in \mathbb{R}^{2L}$  的形式分别为  $[\mathbf{h}^{(1)}, \mathbf{h}^{(2)}, \dots, \mathbf{h}^{(L)}]$  和  $[\mathbf{t}^{(1)}, \mathbf{t}^{(2)}, \dots, \mathbf{t}^{(L)}]$ , 其中  $\mathbf{h}^{(l)}, \mathbf{t}^{(l)} \in \mathbb{R}^2$ . 因此, 双线性形式的三元组  $(h, r, t)$  的分数可以写成  $L$  个分量的和, 即:

$$\mathbf{h}^\top \mathbf{R} \mathbf{t} = \sum_{l=1}^L \mathbf{h}^{(l)\top} \mathbf{R}^{(l)} \mathbf{t}^{(l)} \quad (59)$$

其中, 每个分量  $\mathbf{R}^{(l)}$  都是一个二面体群元素的表示矩阵.

### 2.3.3 双曲几何模型

虽然嵌入方法已经在许多应用中被证明是成功的, 但它们存在一个普遍的局限性, 即对复杂模式的建模能力

本质上受到嵌入空间维数的限制. 目前还没有一种方法能够在不丢失信息的情况下计算大型图结构数据的嵌入.

Adcock 等人<sup>[87]</sup>的实证分析表明, 许多现实世界网络表现出潜在的树状结构. 为了利用这种结构性学习更有效的表示方法, Nickel 等人<sup>[88]</sup>提出在双曲空间 (具有常负曲率的空间) 中计算嵌入, 这是因为双曲空间天然适合于建模层次结构, 或者也可以认为, 双曲空间是树的连续版本, 相比欧式空间 (零曲率空间) 可以更准确、更简洁的表示分层数据. Nickel 等人<sup>[88]</sup>提出了基于双曲空间 Poincare ball 模型的 Poincare 方法, 非常适合于基于梯度的优化.

Nickel 等人<sup>[88]</sup>将符号数据嵌入到双曲空间  $\mathbb{H}$  中, 设  $B^d = \{\mathbf{x} \in \mathbb{R}^d \mid \|\mathbf{x}\| < 1\}$  是开放的  $d$  维单位球, 其中  $\|\cdot\|$  表示欧几里德范数, 双曲空间的 Poincare ball 模型对应于黎曼流形  $(B^d, g_x)$ , 即:

$$g_x = \left( \frac{2}{1 - \|\mathbf{x}\|^2} \right)^2 g^E \quad (60)$$

其中,  $\mathbf{x} \in B^d$  且  $g^E$  表示欧几里德度量张量.

Poincare 模型能够学习符号数据的层次表示, 设  $D = \{(h, t)\}$  是观察到的名词对之间的上位关系集合, 然后学习  $D$  中所有符号嵌入, 使得相关对象在嵌入空间中接近. 则 Poincare 模型的评分函数定义为:

$$f_r(h, t) = \sum_{(h, t) \in D} \log \frac{e^{-d(h, t)}}{\sum_{t' \in N(h)} e^{-d(h, t')}} \quad (61)$$

其中,  $N(h) = \{t' \mid (h, t') \in D\} \cup \{t\}$  是  $h$  的负例集.

实际上, 多关系知识图谱往往表现出多个层次结构, 为了解决这一问题, 提出了 MuRP 模型<sup>[89]</sup>. MuRP 相对于 Poincare 模型而言更加完善, 它是一种在双曲空间的 Poincare ball 模型中嵌入分层多关系数据的方法, 通过 Möbius 矩阵-向量乘法和 Möbius 加法学习关系特定参数来转换实体嵌入.

一组实体可以在不同的关系下形成不同的层次结构, 而一个理想的嵌入模型应该同时捕获所有层次结构, 受 Word2Vec 词嵌入中类比结构启发, MuRP 定义了多关系图嵌入的评分函数:

$$f_r(h, t) = -d(\mathbf{h}^{(r)}, \mathbf{t}^{(r)})^2 + b_h + b_t = -d(\mathbf{R}\mathbf{h}, \mathbf{t} + \mathbf{r})^2 + b_h + b_t \quad (62)$$

其中,  $d: \mathcal{E} \times \mathcal{R} \times \mathcal{E} \rightarrow \mathbb{R}^+$  是距离函数,  $b_h, b_t \in \mathbb{R}$  分别表示头、尾实体  $h$  和  $t$  的标量偏差.  $\mathbf{R} \in \mathbb{R}^{d \times d}$  是对角关系矩阵.  $\mathbf{h}^{(r)} = \mathbf{R}\mathbf{h}$  和  $\mathbf{t}^{(r)} = \mathbf{t} + \mathbf{r}$  表示在应用相应的关系特定变换后的头部和尾部实体嵌入.

将该评分函数与双曲几何相结合, MuRP 模型的评分函数定义为:

$$f_r(h, t) = -d_{\mathbb{B}}(\mathbf{h}_h^{(r)}, \mathbf{h}_t^{(r)})^2 + b_h + b_t = -d_{\mathbb{B}}(\exp_0^c(\mathbf{R} \log_0^c(\mathbf{h}_h)), \mathbf{h}_t \oplus_c \mathbf{r}_h)^2 + b_h + b_t \quad (63)$$

其中,  $\mathbf{h}_h, \mathbf{h}_t \in \mathbb{B}_c^d$  分别表示头、尾实体  $h$  和  $t$  的双曲嵌入,  $\log$  表示的是对数映射,  $0$  表示的是庞加莱球的  $0$  点,  $c$  为参数,  $\mathbf{r}_h \in \mathbb{B}_c^d$  是关系  $r$  的双曲平移向量.  $\mathbf{h}_h^{(r)} \in \mathbb{B}_c^d$  是通过 Möbius 矩阵-向量乘法得到的关系调整后的头实体嵌入, 而关系调整的尾实体嵌入  $\mathbf{h}_t^{(r)} \in \mathbb{B}_c^d$  是通过将关系向量  $\mathbf{r}_h$  与尾实体嵌入  $\mathbf{h}_t$  进行 Möbius 加法得到的,  $\oplus_c$  表示 Möbius 加法操作.

关系矩阵是对角的, 因此 MuRP 的参数数量随实体和关系的数量线性增加, 使其可以用于大型知识图谱.

#### 2.3.4 其他模型

除了利用卷积神经网络对 KGE 进行建模的方法, 最近也有研究将胶囊网络应用于解决 KGE 问题. 例如: CapsE 模型<sup>[90]</sup>, 该模型探索了一种用于对关系三元组建模的胶囊网络. 总体来说, CapsE 在 ConvKB 卷积提取特征映射后加入两个胶囊层, 在第一层中, 构造  $k$  个胶囊, 其中来自所有特征映射相同维度的条目被封装到一个相应的胶囊中, 每个胶囊可以捕捉嵌入三元组中相应维度条目之间的许多特征. 这些特征被概括并输入到第二层的胶囊中, 该胶囊产生一个向量输出, 其长度用作三元组的分数. CapsE 的评分函数定义如下:

$$f_r(h, t) = \|\text{capsnet}(g([\mathbf{h}, \mathbf{r}, \mathbf{t}] * \omega))\| \quad (64)$$

其中,  $\text{capsnet}$  表示胶囊网络运算. CapsE 是首个考虑将胶囊网络用于知识图谱补全的方法.

此外, CrossE 模型<sup>[91]</sup>进一步考虑了实体与关系之间的双向影响, 显式地建模交叉交互. Cross 模型主要包括 4



个步骤: (1) 为头实体  $h$  生成交互嵌入  $\mathbf{h}_I$ ; (2) 生成关系  $r$  的交互嵌入  $\mathbf{r}_I$ ; (3) 结合交互嵌入  $\mathbf{h}_I, \mathbf{r}_I$ ; (4) 比较组合嵌入与尾实体嵌入  $\mathbf{t}$  的相似性. CrossE 的评分函数定义为:

$$f_r(h, t) = \sigma(\tanh(\mathbf{c}_r \circ \mathbf{h} + \mathbf{c}_r \circ \mathbf{h} \circ \mathbf{r} + \mathbf{b}) \mathbf{t}^\top) \quad (65)$$

其中,  $\mathbf{c}_r \circ \mathbf{h} = \mathbf{h}_I, \mathbf{c}_r \circ \mathbf{h} \circ \mathbf{r} = \mathbf{r}_I$ , 且  $\mathbf{b} \in \mathbb{R}^{1 \times d}$  是全局偏置向量. CrossE 模型在复杂数据集上获得的实验结果, 展示了在 KGE 中建模交叉交互的有效性.

### 2.3.5 模型总结

本节归纳总结了最新的知识图谱嵌入方法, 由卷积神经网络模型, 旋转模型, 双曲几何模型和其他模型 4 个部分组成, 表 4 对相关的最新知识图谱嵌入研究方法进行了总结.

表 4 最新的知识图谱嵌入模型总结

类别	模型	实体嵌入	关系嵌入	评分函数 $f_r(h, t)$
卷积神经网络	ConvE <sup>[79]</sup>	$\mathbf{M}_h \in \mathbb{R}^{d_w \times d_h}, \mathbf{t} \in \mathbb{R}^d$	$\mathbf{M}_r \in \mathbb{R}^{d_w \times d_h}$	$\sigma(\text{vec}(\sigma([\mathbf{M}_h; \mathbf{M}_r] * \omega)) \mathbf{W}) \mathbf{t}$
	ConvKB <sup>[81]</sup>	$\mathbf{h}, \mathbf{t} \in \mathbb{R}^d$	$\mathbf{r} \in \mathbb{R}^d$	$\text{concat}(g([\mathbf{h}, \mathbf{r}, \mathbf{t}] * \omega)) \mathbf{w}$
	HypER <sup>[82]</sup>	$\mathbf{h}, \mathbf{t} \in \mathbb{R}^d$	$\mathbf{r} \in \mathbb{R}^{d_r}$	$f(\text{vec}(\mathbf{h} * \text{vec}^{-1}(\mathbf{r} \mathbf{H})) \mathbf{W}) \mathbf{t}$
旋转模型	RotatE <sup>[84]</sup>	$\mathbf{h}, \mathbf{t} \in \mathbb{C}^d$	$\mathbf{r} \in \mathbb{C}^d$	$\ \mathbf{h} \circ \mathbf{r} - \mathbf{t}\ $
	QuatE <sup>[85]</sup>	$\mathbf{h}, \mathbf{t} \in \mathbb{H}^d$	$\mathbf{r} \in \mathbb{H}^d$	$\mathbf{h} \otimes \frac{\mathbf{r}}{ \mathbf{r} } \cdot \mathbf{t}$
	DihEdral <sup>[86]</sup>	$\mathbf{h}^{(1)}, \mathbf{t}^{(1)} \in \mathbb{R}^2$	$\mathbf{R}^{(l)} \in \mathbb{D}_k$	$\sum_{l=1}^L \mathbf{h}^{(l)\top} \mathbf{R}^{(l)} \mathbf{t}^{(l)}$
双曲几何模型	Poincare <sup>[88]</sup>	$\mathbf{h}, \mathbf{t} \in \mathcal{B}^d$	—	$\sum_{(h,t) \in D} \log \frac{e^{-d(h,t)}}{\sum_{t' \in N(h)} e^{-d(h,t')}}$
	MuRP <sup>[89]</sup>	$\mathbf{h}_h, \mathbf{h}_t \in \mathbb{B}_c^d$	$\mathbf{r}_h \in \mathbb{B}_c^d$	$-d_{\mathbb{B}}(\exp_0^c(\mathbf{R} \log_0^c(\mathbf{h}_h)), \mathbf{h}_t \oplus_c \mathbf{r}_h)^2 + b_h + b_t$
其他模型	CapsE <sup>[90]</sup>	$\mathbf{h}, \mathbf{t} \in \mathbb{R}^d$	$\mathbf{r} \in \mathbb{R}^{d_r}$	$\ \text{capsnet}(g([\mathbf{h}, \mathbf{r}, \mathbf{t}] * \omega))\ $
	CrossE <sup>[91]</sup>	$\mathbf{h}, \mathbf{t} \in \mathbb{R}^d$	$\mathbf{r} \in \mathbb{R}^{d_r}$	$\sigma(\tanh(\mathbf{c}_r \circ \mathbf{h} + \mathbf{c}_r \circ \mathbf{h} \circ \mathbf{r} + \mathbf{b}) \mathbf{t}^\top)$

## 2.4 小 结

第 2 节主要介绍了 3 大类知识图谱嵌入方法, 即: 基于距离的模型, 语义匹配模型与最新的 KGE 模型, 并结合已有的研究成果对其进行了分析. 根据上述分析结果, 表 5 从类别、方法、提出年份及优缺点 4 方面对这几类知识图谱嵌入方法的部分代表模型进行对比.

## 3 动态知识图谱嵌入

当前 KGE 的研究主要集中于静态知识图谱, 其中事实不会随时间发生变化, 例如: TransE, TransH, TransR, RESCAL 等. 但是, 在实际应用中, 知识图谱通常是动态的, 例如 Twitter 中的社交知识图, DBLP 中的引文知识图等, 其中事实随时间演变, 仅在特定时间段内有效. 以往的静态 KGE 方法完全忽略了时间信息, 这使得静态 KGE 方法无法在这些实际场景中工作. 因此, 有必要设计一种用于动态知识图谱嵌入的方法.

### 3.1 模型介绍

t-TransE<sup>[92]</sup>是第一个考虑将时间信息用于 KGE 的方法, 使用时间顺序约束建模中时间敏感关系之间的转换, 并强制嵌入在时间上保持一致. t-TransE 认为时间敏感事实的发生时间可以表示事实和关系的时间敏感关系的特定时间顺序. 因此, t-TransE 是一个结合时间顺序信息而提出来的时间感知链接预测模型.

具体来说,  $t$ -TransE 在事实三元组中添加时间维度, 表示为四元组:  $(e_i, r, e_j, t_r)$ , 其中  $t_r$  表示事实的发生时间. 为了结合成对时间事实之间的时序信息, 它假设先前时间敏感关系向量可以通过时间转换演化为后续时间敏感关系向量. 例如, 有两个共享同一头实体的时间事实:  $(e_i, r_i, e_j, t_1)$  和  $(e_i, r_j, e_k, t_2)$ , 且时序约束为  $t_1 < t_2$ , 那么可以假设: 先前关系  $r_i$  经过时间转换后应接近后续关系  $r_j$ , 即  $r_i \mathbf{M} \approx r_j$ , 矩阵  $\mathbf{M}$  可以捕捉关系之间的时间顺序信息.

表 5 部分 KGE 模型对比

类别	方法	提出年份	优缺点
基于距离的模型	TransE <sup>[15]</sup>	2013	优点: 第一个基于翻译的方法; 易于训练, 参数较少, 易于扩展到大型数据库; 缺点: 不能很好处理复杂关系
	TransH <sup>[16]</sup>	2014	优点: 能很好处理一对多、多对多、多对一的复杂关系; 缺点: 无法区分两个语义相近的实体
	TransR <sup>[43]</sup>	2015	优点: 不仅考虑关系的多样性, 而且考虑实体; 缺点: 计算复杂, 忽略头、尾实体不同的类型和属性
	TransD <sup>[44]</sup>	2015	优点: 计算简单、参数少; 缺点: 并非所有新事实都可以从存在的情况中推论得出
	STransE <sup>[45]</sup>	2016	优点: 对每个关系使用两个投影矩阵, 性能优于TransE; 缺点: 容易出现数据稀疏问题
	TranSparse <sup>[46]</sup>	2016	优点: 能很好处理异质性和不平衡性; 缺点: 只考虑了稀疏模式, 有一定的局限性
	TorusE <sup>[56]</sup>	2018	优点: 第一个正式讨论TransE正则化问题的模型, 可扩展到大型KG; 缺点: 模型不够通用
	KG2E <sup>[57]</sup>	2015	优点: 能对KG中实体和关系的不确定性进行建模; 缺点: 在N-N关系中效果较差
语义匹配模型	TransG <sup>[58]</sup>	2016	优点: 具有较高的实体区分度, 考虑KGE中的多重关系语义问题; 缺点: 采用边训练边聚类的方法, 比较繁琐
	LFM <sup>[64]</sup>	2012	优点: 采用关系特定双线性变换考虑实体和关系之间的相关性; 缺点: 需要大量参数进行建模
	DistMult <sup>[38]</sup>	2015	优点: 限制关系特定双线性变换矩阵为对角矩阵, 减少关系参数数量; 缺点: 模型过度简化, 只能处理对称关系
	SLM <sup>[77]</sup>	2013	优点: 减轻基本距离模型无法协同精确刻画实体与语义联系问题; 缺点: 具有更困难的优化问题
最新的KGE模型	NTN <sup>[77]</sup>	2013	优点: 明确地描述实体与关系之间的复杂语义关联; 缺点: 高复杂性限制了其在大规模稀疏知识图谱上的应用
	ConvE <sup>[79]</sup>	2018	优点: 通过多层非线性特征学习表达语义信息, 具有很高的参数利用率; 缺点: 没有考虑嵌入三元组中相同维度条目之间的全局关系, 忽略了过渡特征
	RotatE <sup>[84]</sup>	2019	优点: 能够同时对反转, 合成和对称/反对称等关系模式建模和推理; 缺点: 只有一个旋转平面, 且欧拉角中具有万向节锁问题
	Poincare <sup>[88]</sup>	2017	优点: 在双曲空间中计算嵌入, 适合基于梯度的优化; 缺点: 多关系知识图往往表现出多个层次结构, 模型不够完善

为了考虑事实发生时间,  $t$ -TransE 将时间顺序约束转化为一个优化问题, 并将时间顺序得分函数定义如下:

$$g(r_i, r_j) = \|\mathbf{r}_i \mathbf{M} - \mathbf{r}_j\|_1 \quad (66)$$

其中,  $\mathbf{M} \in \mathbb{R}^{n \times n}$  是成对时间顺序关系对  $(r_i, r_j)$  之间的转换矩阵, 当关系对按时间顺序排列时, 期望该分数较低, 否则, 该分数较高.

然而,  $t$ -TransE 不是直接将时间整合到学习的嵌入中, 而是首先学习关系之间的时间顺序. 然后在 KGE 阶段将这些关系顺序合并为约束, 因此,  $t$ -TransE 学习到的嵌入不是显式时间感知的.

Know-Evolve<sup>[93]</sup>使用双线性嵌入学习方法对 KG 元素的非线性时间演化进行建模. 然而, 它将域限制为本质上非常密集的基于事件交互类型的数据集. 文献 [94] 对关系与时间的交互作用进行建模, 研究了将时间嵌入向量与

关系嵌入向量相结合的各种方法, 例如连接, 求和或点积运算. 文献 [95] 将时间戳视为一个 (从 0 到 9) 数字序列, 并使用 LSTM 编码关系向量和时间数字.

Dasgupta 等人从 TransH<sup>[16]</sup>方法中获得启发, 在 2018 年提出了 HyTE 方法<sup>[96]</sup>. HyTE 是一个基于超平面的时间感知知识图谱嵌入方法, 该方法将每个时间戳与相应的超平面相关联, 将时间显式的合并到实体关系空间中. 因此, 它不仅能够利用时间指导来进行 KG 推理, 还可以预测缺少时间注释的关系事实的时间范围.

HyTE 首先为  $(h, r, t)$  形式的三元组组成的静态知识图谱添加单独的时间维度. 考虑四元组  $(h, r, t, [\tau_s, \tau_e])$ , 其中  $\tau_s$  和  $\tau_e$  表示三元组  $(h, r, t)$  有效的开始和结束时间. 因此, 给定时间戳, 可以将时间范围内的输入 KG 分割为多个静态子图, 这些静态图由在各个时间步中有效的三元组组成.

HyTE 将时间表示为超平面, 对于 KG 中的  $T$  个时间步, 可以得到  $T$  个不同的超平面, 分别由法向量  $w_{t1}, w_{t2}, \dots, w_{tT}$  表示. 将在时间  $\tau$  投影到时间特定的超平面  $w_\tau$  上, 分别得到投影向量  $P_\tau(\mathbf{h})$ ,  $P_\tau(\mathbf{t})$ ,  $P_\tau(\mathbf{r})$ .

与 transH 类似, HyTE 期望在时间  $\tau$  处有效的三元组具有如下的映射约束关系:  $P_\tau(\mathbf{h}) + P_\tau(\mathbf{r}) \approx P_\tau(\mathbf{t})$ . HyTE 的评分函数定义为:

$$f_\tau(h, r, t) = \|P_\tau(\mathbf{h}) + P_\tau(\mathbf{r}) - P_\tau(\mathbf{t})\|_{1/2} \quad (67)$$

为了在保持动态 KG 固有结构的同时, 在学习的嵌入中结合时间信息. 一种通常的做法是将动态 KG 分割成多个静态子 KG, 每个子 KG 对应于特定的时间段, 然后分别在这些段上学习嵌入 (HyTE, Flexible Translation 就采用了这种做法). 这类模型虽然在嵌入过程中考虑了 KG 的时间信息, 但不能显式地建模动态 KG 的演化过程. 最近, Tang 等人提出了一个结合时间演化的可感知时间间隔的动态知识图谱嵌入方法 (TDG2E)<sup>[97]</sup>, TDG2E 是一种鲁棒的动态知识图谱嵌入方法, 将时间信息直接编码到学习的嵌入中.

TDG2E 既保留了当前子 KG 的结构信息, 又能同时保留动态 KG 的演化模式. 它首先将时间范围内的输入 KG 分割为多个静态子 KG, 每个子 KG 对应于一个时间单元, 并将每个子 KG 的实体和关系投影到时间感知超平面中. 然后利用基于 GRU 的模型处理动态知识图谱嵌入的学习过程中涉及到的子 KG 之间的依赖关系. 此外, 考虑累积的结构信息直接导致连续的结构, 进一步引入辅助损失, 通过利用先前的结构信息 (即 GRU 的隐藏状态) 监督下一个子 KG 的学习过程, 其公式如下所示:

$$L_{\text{aux}}(\mathbf{W}, \mathbf{R}_s, \mathbf{R}_r, \mathbf{R}_p) = \sum_{\tau=1}^{T-1} \|\mathbf{p}_\tau - \mathbf{w}_{\tau+1}\|_2 \quad (68)$$

辅助损失  $L_{\text{aux}}$  利用隐状态  $\mathbf{p}_\tau$  监督超平面  $\mathbf{w}_{\tau+1}$  的学习, 引入辅助损失可以帮助每个超平面, 在保留当前子 KG 结构信息的同时保留动态知识图的演化模式. 当 GRU 处理大量的子 KG 时, 辅助损失减小了反向传播的难度.

为了进一步解决动态 KG 的时间不平衡问题, TDG2E 在 GRU 中设计了一个时间间隔门 (类似于更新门), 引入相邻子 KG 之间的时间间隔以更有效地建模动态 KG 的演化模式. 如图 13 是提出改编的 GRU 图, 标记为红色的是添加的部分, 红色圆圈表示时间间隔门  $T_\tau$ .

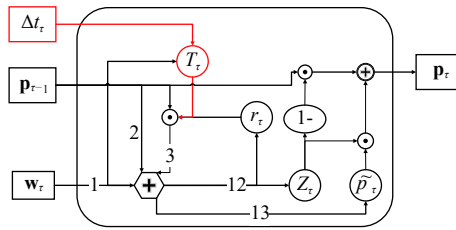


图 13 TDG2E 提出的改编 GRU 的图

时间间隔门  $T_\tau$  的计算公式为:

$$T_\tau = \sigma(\mathbf{R}_T \mathbf{w}_\tau + \sigma(\Delta t_\tau \mathbf{R}_t) + \mathbf{b}_t) \quad (69)$$

其中,  $\mathbf{R}_T \in \mathbb{R}^{d \times d}$ ,  $\mathbf{R}_t \in \mathbb{R}^{d \times d}$  是权重矩阵,  $\mathbf{b}_t$  是偏差,  $\Delta t_\tau$  表示第  $(\tau + 1)$  个时间段和第  $\tau$  个时间段之间的时间间隔, 定

义如下:

$$\Delta t_{\tau} = t_s^{\tau+1} - t_s^{\tau} \quad (70)$$

其中,  $t_s^{\tau}$  是第  $\tau$  个时间段的开始时间.

TDG2E 方法与其他现有的静态/动态嵌入方法相比, 它不仅可以保留当前子 KG 的结构信息, 而且可以保留动态 KG 的演化模式. 通过引入时间间隔, TDG2E 可以更有效地对动态 KG 的演化模式进行建模.

### 3.2 小结

本节介绍了典型的动态知识图谱嵌入方法, 其中分析了 t-TransE, Know-Evolve, HyTE, TDG2E 等相关模型, 表 6 将 TDG2E 模型与其他动态 KGE 方法进行对比, 直观地展示了 TDG2E 方法的优越性. 然而, 从大量文献中可以得出结论: 现有的大多数知识图谱嵌入方法仍然关注于静态知识图谱, 忽略了知识图谱中时间范围信息的可用性与重要性. 实际上, 在表示学习过程中合并时间信息可能会产生更好的 KG 嵌入, 时间感知的知识图谱嵌入研究仍然是一个有待于进一步探索的领域.

表 6 TDG2E 模型与其他动态 KGE 方法对比

比较方法	存在的问题	TDG2E 优势	TDG2E 采用的解决办法
HyTE, Flexible Translation	独立学习不同子KG, 不能显式建模动态KG演化过程	同时保留当前子KG的结构信息与动态KG的时间演化模式	利用基于GRU的模型捕获动态KG中相邻子KG的依赖关系; 引入辅助损失, 利用先前的结构信息监督后续超平面的学习过程
HyTE, Flexible Translation, t-TransE	没有完全解决KG中时间戳分布不平衡的问题	解决了动态KG面临的时间不平衡问题	在GRU中设计时间间隔门, 引入相邻子KG之间的时间间隔

## 4 融合多源信息的知识图谱嵌入

多源信息提供了知识图谱中三元组事实以外的信息, 能够帮助构建更加精准的知识表示, 仅使用事实进行知识图谱嵌入的方法忽略了蕴含在多源信息中的丰富知识, 例如: 实体类别信息、文本描述信息、关系路径等. 充分利用这些多源信息对于降低实体与关系之间的模糊程度, 进而提高推理预测的准确度至关重要.

### 4.1 实体类别

实体类别包含实体结构化的先验知识, 这些先验知识是由人工构建的, 能够在知识图谱三元组的结构信息上提供准确的辅助信息, 加深模型对三元组的理解.

Guo 等人提出 (semantically smooth embedding, SSE)<sup>[23]</sup>模型, 认为属于同一语义类型的实体在嵌入空间中距离应该比较接近. SSE 利用流行学算法来约束这种平滑性假设, 将两种算法的约束条件加到最大间隔方法里作为整个模型损失函数的正则化项, 从而达到约束嵌入空间语义平滑的作用.

SSE 模型默认所有实体只有一个类别, 然而, 现实世界中的实体不仅有多多个类别, 而且类别间可能具有层次关系. 实体类型可以作为不同关系的头实体与尾实体的约束, 例如关系 DirectorOf 的头实体的类型应该是人, 尾实体的类型应该是电影作品. Xie 等人<sup>[24]</sup>提出一种融合实体层次类型信息的模型 (type-embodied knowledge representation learning, TKRL), 引入具有层次结构的实体类别信息以及与关系之间的约束信息. TKRL 可以看做是带有实体层次类别信息的 TransR 模型.

Jin 等人<sup>[98]</sup>提出 TEKRL 模型, 引入注意力机制来捕获实体类别和三元组之间的潜在联系, 自动地学习实体的不同类别对某种特定关系的不同重要程度, 解决了其他模型在使用实体类别信息时需要引入额外规则的问题.

### 4.2 文本描述

知识图谱中很多实体都带有描述信息, 这些信息能够作为知识图谱中结构化信息的辅助, 帮助模型学习更精准的知识表示. 知识库的构建资源也往往从文本中获取, 因此实体描述文本能天然地与知识空间进行交互. 那些仅仅基于知识图结构化信息知识表示模型无法处理不在知识图中的实体, 而联合文本嵌入的方法可以做到互补,

使得模型可以学习到那些在文本中出现而不在知识图中的实体。

Wang 等人<sup>[25]</sup>首先提出联合知识图谱和实体描述文本的知识表示学习模型,该模型基于 TransE<sup>[15]</sup>和 Skip-gram 模型<sup>[40]</sup>的基本思想,利用实体名称或者维基百科锚文本作为对齐原则。在实际场景中,由于实体名称歧义性较大,利用实体名称对齐的原则会打乱文本原有的语义空间,而利用维基百科锚文本对齐的原则过于依赖特定的数据源。为了解决以上问题,Zhong 等人<sup>[26]</sup>提出利用实体描述文本作为对齐原则。类似的,Zhang 等人<sup>[99]</sup>也尝试使用实体名称和实体描述文本中词向量的均值作为实体的文本表示。

为了利用整个文本的语序语义信息,Xie 等人<sup>[27]</sup>提出一种融合实体描述的知识表示模型 (description-embodied knowledge representation learning, DKRL),在 TransE 模型的基础上融合实体描述的文本信息,为每一个实体设置两种知识表示。然而 DKRL 是一种弱关联建模,在融合实体基于结构的表示和基于文本的表示时没有足够的交互过程。Xiao 等人提出 SSP 模型<sup>[100]</sup>,将三元组的嵌入表示投影到语义子空间中,在语义子空间上学习实体的两种表示,与 DKRL 不同的是 SSP 采用主题模型建模实体的文本表示。相关的其他模型还有 TEKE<sup>[28]</sup>、ATEKE<sup>[101]</sup>等。

### 4.3 逻辑规则

逻辑规则 (这里所说的逻辑规则主要指一阶 Horn 子句,例如:  $\forall x, y: HasWife(x, y) \Rightarrow HasSpouse(x, y)$ , 表明任何有两个 *HasWife* 关系相连的实体都有 *HasSpouse* 关系) 包含丰富的背景信息。目前, AMIE<sup>[102]</sup>, AMIE+<sup>[103]</sup>, RLvLR<sup>[104]</sup>等规则挖掘方法可以自动从 KG 中提取逻辑规则。

Guo 等人将三元组看做原子,提出了 KALE<sup>[32]</sup>方法。给定一个逻辑规则, KALE 利用实体集中的实体初始化逻辑规则,并采用 t-norm 模糊逻辑连接原子三元组,将复合公式的真值定义为其成分真值的组合。通过这种方式, KALE 以一个统一的结构来表示三元组和规则。

由于硬规则依赖于手工设计与验证,而某些文本信息可以提取出来作为软规则。基于这个思想,Guo 等人在 2017 年提出了 (rule-guided embedding, RUGE)<sup>[105]</sup>。该模型采用迭代交互计算的形式,首先利用软规则和学习到的知识表示预测无标签三元组的标签,然后利用预测的软标签和 KG 中已有标签的三元组重新完善知识表示。

### 4.4 关系路径

关系路径是指两个实体之间的多步关系,而不仅限于两个实体之间直接相连的关系。多步关系包含了两个实体之间丰富的语义关系,有助于多步推理。

Lin 等人<sup>[29]</sup>在 TransE 模型的基础上将两个实体之间的多步关系路径看做两个实体之间相连的关系,提出了 PTransE 模型。Guo 等人<sup>[30]</sup>提出了另一种融合多步关系路径的知识表示学习模型,利用关系路径构建新的三元组,并且对 TransE 模型和 RESCAL 模型进行了扩展。

Niu 等人<sup>[106]</sup>认为目前基于关系路径的表示学习模型仅利用路径上关系或实体表示的数值计算结果作为关系路径的表示,如果某些关系或实体表示不准确就会造成误差传播,并且这种方式缺乏可解释性。因此,提出一种联合路径和规则的知识表示学习模型 RPJE,利用 Horn 规则组合多步关系路径,并且关联组合后的路径与关系之间的语义信息以获得更加准确的路径表示。

### 4.5 其他信息

除了可以融入上面介绍的信息源外,还有一些其他的信息也可以融入到知识图谱的语义空间中,用于增强知识的语义表示。例如 Xie 等人<sup>[107]</sup>尝试融合实体图像信息学习实体跨模态的知识表示,提出了 (image-embodied knowledge representation learning, IKRL)。该模型为每一个实体设置两种知识表示,并采用 AlexNet<sup>[108]</sup>网络作为图像的特征提取器。

尽管融合多步关系路径的模型引入了其他的实体和关系,但是绝大部分模型只是将关系路径看做是两个实体之间的新关系,知识图谱的图结构信息并没有被完全利用。Feng 等人<sup>[109]</sup>提出一种图感知的表示学习模型 (graph aware knowledge embedding, GAKE),由于不同的实体和关系对于一个实体的影响不同,该模型利用注意力机制赋予不同关系和实体不同的权重。文献<sup>[109]</sup>表明这种考虑图结构的知识表示模型在元组分类和链接预测任务上达



到了很好的效果.

目前有一些嵌入模型融合上述两种或多种信息来丰富实体和关系的语义信息,如 Du 等人<sup>[110]</sup>提出一种融合实体描述信息及实体类型信息的表示学习方法;Xing 等人<sup>[111]</sup>提出 MKRL 模型,同时融合实体描述、层次类型和文本关系 3 种类型的信息.

#### 4.6 小 结

本节主要介绍了在基于 KG 自身固有结构信息的基础上融合其他信息源的知识图谱嵌入模型,信息源大致分为语义信息(包括实体类型、文本描述)、逻辑规则以及关系路径等.近年来,随着 KGE 技术的不断发展,也有研究考虑将图像信息,关系层次等信息融入 KG 语义空间中以增强知识的语义表示.

### 5 知识图谱嵌入的应用

近年来,知识驱动的应用在信息检索和问答等领域取得了巨大成功,这些应用有望帮助准确深入地了解用户需求,并给出适当响应.知识图谱嵌入方法的核心思想是将每个实体、关系表示为一个低维向量,而学习到的实体、关系嵌入可以受益于多种下游任务.在本节中,我们将介绍 KGE 的典型应用.

#### 5.1 基于知识图谱嵌入的问答

随着大规模知识图谱的兴起,基于知识图谱的问答(QA)成为重要的研究方向,引起了人们的广泛关注.现实世界的领域中通常包含数百万到数十亿个事实,其庞大的数据量和复杂的数据结构使得用户很难访问其中有价值的知识.为了解决这个问题,提出了基于知识图谱的问答(QA-KG).

QA-KG 旨在利用知识图谱中的事实来回答自然语言问题.可以帮助普通用户在不知道 KG 数据结构的情况下,高效地访问 KG 中对自己有价值的知识.然而,由于涉及到语义分析<sup>[112]</sup>和实体链接<sup>[113,114]</sup>等多个具有挑战性的子问题,QA-KG 的问题还远未得到解决.近年来,随着 KGE 在不同的实际应用中表现出的有效性,人们开始探索其在解决 QA-KG 问题中的一些潜在作用.

Bordes 等人<sup>[115]</sup>基于训练问题和问题释义学习单词,关系和实体的低维表示,以便将新问题和候选事实投影到同一空间中进行比较. Yang 等人<sup>[116,117]</sup>利用问题和潜在事实的逻辑性质,将问题和候选答案投影到统一的低维空间中.还有一些基于深度学习的模型<sup>[118-122]</sup>通过将问题中的单词输入神经网络来实现这种投影.

值得注意的是,最近, Huang 等人<sup>[123]</sup>提出了一个简单有效的基于知识图谱嵌入的问答框架(KEQA),旨在解决简单问题,即 QA-KG 中最常见的问题类型. KEQA 不是直接推断问题的头实体和谓词,而是在 KGE 空间中联合恢复自然语言问题的头实体,关系和尾实体表示来回答问题.最后,基于知识图谱子集(FB2M、FB5M<sup>[123]</sup>)和问答数据集 SimpleQuestions<sup>[115]</sup>进行实验,通过与 7 个最新提出的 QA-KG 算法进行对比,KEQA 凭借在简单问题上获得 20.3% 的准确性改进获得了优于所有基线的性能<sup>[124]</sup>.此外,为了验证在使用不同的 KGE 算法时 KEQA 的通用性,分别使用 TransE<sup>[15]</sup>、TransH<sup>[16]</sup>、TransR<sup>[43]</sup>执行知识图谱嵌入,实验结果表明 KGE 算法显著提高了 KEQA 的性能,与 KEQA\_noEmbed(没有使用 KGE 算法)相比,KEQA 基于 TransE 时实现了 3.1% 的改进,并且 KEQA 在使用不同的 KGE 算法时性能相近,证明了 KEQA 的通用性,此外,即使不使用 KGE,KEQA 仍然可以获得与最先进的 QA-KG 方法相当的性能,验证了 KEQA 的健壮性.

#### 5.2 推荐系统

在过去的几年中,利用知识图谱的推荐系统已被证明与最先进的协作过滤系统具有竞争力,能有效地解决新项目和数据稀疏性等问题<sup>[124-128]</sup>.最近,KGE 的流行促进了利用 KGE 捕获实体语义进行推荐这一研究热点,使用 KGE 已被证明对推荐系统有效.

Zhang 等人提出使用 TransR 的协作知识图嵌入(collaborative knowledge base embedding, CKE)<sup>[129]</sup>,以学习结合视觉和文本嵌入的项目结构表示.深度知识感知网络(deep knowledge-aware network, DKN)<sup>[130]</sup>利用 TransD 学习实体嵌入,并通过将它们与词嵌入相结合来设计 CNN 框架,用于新闻推荐.但是,由于需要提前学习实体嵌入,DKN 不能以端到端的方式进行训练.为了实现端到端的训练, MKR(multi-task feature learning approach for

knowledge graph)<sup>[131]</sup>通过共享潜在特征和建模高阶项-实体交互,将多任务知识图谱表示和推荐关联起来。Ai等人<sup>[132]</sup>通过 TransE 方法学习用户和项目嵌入,并基于投影空间中的用户-项目相似度评分进行推荐。文献[133]为优惠推荐任务提出了一个神经分解(neural factorization, NF)模型,以 KG 的形式对可用数据进行建模,并使用 TransE 学习实体和关系的嵌入。

最近,Sha 等人提出了一种新颖的注意力知识图谱嵌入(attentive knowledge graph embedding, AKGE)框架<sup>[134]</sup>,以更好地利用 KG 进行有效推荐。该框架以交互特定的方式充分利用了 KG 的语义和拓扑,为推荐结果提供了可解释性。此外,Ni 等人描述了一种用于 Wikipedia 的基于嵌入的实体推荐框架<sup>[135]</sup>,该框架将 Wikipedia 组织成一系列彼此重叠的图,从它们的拓扑结构和内容中学习互补的实体表示,并将其与轻量级的学习方法相结合,以推荐 Wikipedia 上的相关实体。通过使用 Wikipedia 作为框架的输入,两个实体推荐数据集(一个由 Yahoo!内部编辑团队创建,另一个从谷歌的搜索结果页面中抓取)作为基础事实,进行离线和在线评估,证明了所产生的嵌入和推荐在质量和用户参与度方面表现良好。

### 5.3 关系提取

关系提取(relation extraction, RE)是信息提取中的一项重要任务,旨在根据两个给定实体的上下文来提取它们之间的关系。由于 RE 具有提取文本信息的能力,并使许多自然语言处理应用受益(例如:信息检索,对话生成,问答等),因此受到很多研究者的青睐。

常规的监督模型已经在关系提取任务中得到深入研究,但是,它们的性能在很大程度上依赖于训练数据的规模和质量。为了构建大规模数据,Mintz 等人<sup>[136]</sup>提出了一种新颖的远程监督(distant supervision, DS)机制,通过将现有知识图谱与文本对齐来自动标记训练实例。DS 使 RE 模型能够在大规模的训练语料库上工作,因此远程监督的 RE 模型<sup>[137-139]</sup>已经成为从纯文本中提取新事实的主流方法。但是,这些方法仅在知识获取中使用纯文本中的信息,而忽略了 KG 结构所包含的丰富信息。

受 KG 丰富的知识启发,很多研究工作在 KG 的指导下扩展了 DS 模型。Weston 等人<sup>[140]</sup>提出将 TransE 与现有的远程监督的 RE 模型相结合以提取新的事实,并且获得了较大改进。此外,Han 等人<sup>[141]</sup>提出了一种针对 KRL 和 RE 的联合表示学习框架,文献[37]证实了现有的 KRL 模型可以有效增强远程监督的 RE 模型。最近,Han 等人<sup>[142]</sup>提出了一个通用的联合表示学习框架,用于知识图谱补全(knowledge graph completion, KGC)和从文本中提取关系(relation extraction, RE)两个任务,该框架适用于非严格对齐的数据。此外,Lei 等人<sup>[143]</sup>提出了一种具有双向知识提炼的神经关系提取框架,以协同使用不同的信息源,减轻了远程监督关系提取中的噪声标签问题。但是,这些工作忽略了关系之间的丰富关联。Zhang 等人<sup>[144]</sup>提出 KG 中的关系符合三层层次关系结构(hierarchical relation structure, HRS),并扩展了现有的 KGE 模型:TransE, TransH 和 DistMult,以利用 HRS 的信息学习知识表示。Zhang 等人在 FB15k<sup>[15]</sup>、FB15k237<sup>[145]</sup>、FB13<sup>[77]</sup>、WN18<sup>[15]</sup>和 WN11<sup>[77]</sup>数据集上进行了链接预测和三元组分类任务的实验评估,结果表明,相比于原始模型以及其他基线模型 TransE、TransH、DistMult,扩展模型(TransE-HRS、TransH-HRS、DistMult-HRS)始终获得最佳性能,验证了模型的有效性,同时也证明了考虑关系结构对于 KG 补全非常有效。

### 5.4 小结

本节介绍了 KGE 技术在现实世界中的典型应用,然而,除了上述应用外,也有研究者致力于将知识图谱嵌入技术编码到其他任务中,例如对话系统<sup>[146,147]</sup>、实体消歧<sup>[148,149]</sup>、实体分类<sup>[150-152]</sup>、实体对齐<sup>[153-156]</sup>、知识图谱补全<sup>[43,157]</sup>等。

## 6 挑战与展望

目前,KGE 作为处理大型知识图谱的一种方便有效的工具,被广泛探索并应用于多种知识驱动型任务,极大地提高了任务的性能,同时也存在许多可能的有待探索的领域。在本小节中,我们将讨论 KGE 面临的挑战及其未来研究方向。

## 6.1 面临的挑战

### 6.1.1 探索 KG 的内部和外部信息

KG 中的实体和关系具有复杂的特性和丰富的信息,而这些信息尚未得到充分考虑. 本小节将讨论为增强 KGE 方法的性能而需要进一步探索的内部和外部信息.

- 知识类型: 不同的 KGE 方法在处理 1-1, 1-N, N-1 和 N-N 关系时具有不同的性能, 这表明针对不同类型的知识或关系需要设计不同的 KGE 框架. 然而, 现有的 KGE 方法简单地将所有关系分为 1-1, 1-N, N-1 和 N-N 关系, 不能有效地描述知识的特征. 根据知识的认知和计算特性, 现有知识可分为以下几种类型: (1) 表示实体之间从属关系 (如 has part). (2) 表示实体属性信息 (如 nationality). (3) 表示实体之间的相互关系 (如 friend of). 这些不同类型的关系应该采用不同的方式建模.

- 多语言嵌入: 文献 [40] 观察到不同语言的向量空间之间对应概念的几何排列具有很强的相似性, 并提出两个向量空间之间的跨语言映射在技术上是可行的. 多语言 KG 对于知识共享具有重要意义, 并且在跨语言信息检索, 机器翻译, 问答等领域发挥着重要作用. 然而, 现有的关于多语言 KG 嵌入的研究很少, 因此多语言 KGE 的研究是一项有待解决的有意义但又具有挑战性的工作.

- 多源信息学习: 随着网络技术的快速发展, 如今的互联网不仅包含页面和超链接, 音频、图片和视频等多源信息也越来越多地出现在网络上. 因此, 如何高效地利用从文本到视频的多源信息已成为 KGE 中的一个关键且具有挑战性的问题. 现有的利用多源信息的方法尚处于初步阶段, 诸如社交网络之类的其他形式的多源信息仍然独立于知识图谱表示的构建, 因此还有待进一步研究.

- One-shot/Zero-shot 学习: 近年来, One-shot/Zero-shot 学习在单词表示, 情感分类, 机器翻译等各个领域蓬勃发展. One-shot/Zero-shot 学习的目的是从一个只有少量实例的类或一个从未见过的类的实例中学习, 在知识图谱表示中, 一个实际的问题是低频实体和关系的学习比高频实体和关系的学习更差. 然而, 借助实体和关系的多语言和多模态表示, 低频实体和关系的表示可以在一定程度上得到改善. 此外, 有必要设计新的 KGE 框架, 使其更适用于低频实体和关系的表示学习.

### 6.1.2 知识应用的复杂性

KG 在各种应用中发挥着重要的作用, 例如 Web 搜索, 知识推理和问答. 但是, 由于现实世界中知识应用的复杂性, 难以高效地利用 KG. 在本小节中, 将讨论在实际应用中使用 KG 时遇到的问题.

- KG 质量低: 知识应用的主要挑战之一是大型 KG 本身的质量问题. Freebase, DBpedia, Yago, Wikidata 等典型的 KG 通常是从互联网上的大量纯文本中自动获取知识来获取事实三元组. 由于缺乏人工标注, 这些 KG 遭受噪音和矛盾的问题. 当涉及到实际应用时, 这些噪音和矛盾将导致错误传播. 因此, 如何自动检测现有 KG 中的矛盾或错误已成为将 KG 的信息纳入实际应用中的重要问题.

- KG 体积过大: 现有的 KG 过于繁琐, 无法有效地部署在实际应用中. 此外, 由于 KG 的体积过大, 现有的一些方法由于计算复杂度的问题也并不实用. 因此, 有必要在现有的方法上进行改进.

- KG 不断变化: 随着时间推移, 不断有新的知识产生. 现有的 KGE 方法由于其优化目标与 KG 中的所有事实三元组相关, 因此每次 KG 发生变化时都需要从头开始重新学习模型. 如果在实际应用中使用 KG, 那么它既费时又不实用. 因此, 设计一种可以进行在线学习并逐步更新模型参数的 KGE 框架对 KG 的应用至关重要.

## 6.2 未来方向

### 6.2.1 统一框架

一些知识图谱表明学习模型已经被证明是等价的. 例如, 文献 [67] 证明 HolE 和 ComplEx 在数学上等价于具有某些约束的链接预测. ANALOGY<sup>[68]</sup>提供了包括 DistMult, ComplEx 和 HolE 在内的几种代表性模型的统一视图. Wang 等人<sup>[158]</sup>探讨了几种双线性模型之间的联系. Chandrhas 等人<sup>[159]</sup>探索了加法和乘法 KGE 模型的几何理解. 大多数工作使用不同的模型来描述知识获取和关系提取. 然而, 以类似于图网络统一框架的方式进行的统一研究<sup>[157]</sup>是弥合研究差距的一种有价值的方法.

### 6.2.2 可解释性

知识表示的可解释性是知识获取和实际应用中的关键问题. 现有方法已为可解释性作出了初步努力. ITransF<sup>[51]</sup>采用稀疏向量进行知识迁移, 通过注意力可视化进行解释. CrossE<sup>[91]</sup>利用基于嵌入的路径搜索生成对链接预测的解释, 探索了知识图谱的解释方案. 然而, 这些神经模型在透明度和可解释性方面受到了限制, 一些方法结合逻辑规则来提高互操作性, 从而将黑盒神经模型与符号推理相结合. 因此, 应该进一步研究可解释性并提高预测知识的可靠性.

### 6.2.3 可扩展性

在大规模知识图谱中, 可扩展性非常重要. 几种嵌入方法利用简化来降低计算代价, 例如, 通过循环相关运算简化张量积<sup>[65]</sup>. 但是, 这些方法仍然难以扩展到数以百万计的实体和关系中. 最近的神经逻辑模型<sup>[160,161]</sup>中的规则是由简单的蛮力搜索产生的, 这使得它们在大规模知识图上表现不足. ExpressGNN<sup>[162]</sup>试图使用 NeuralLP<sup>[163]</sup>进行有效的规则归纳. 但是, 要处理繁琐的深层架构和不断增长的知识图还需要进一步完善.

### 6.2.4 自动构建

当前的 KG 高度依赖于人工构建, 这是劳动密集且昂贵的. 知识图谱在不同认知智能领域的广泛应用需要从大规模的非结构化内容中自动构建知识图谱. 最近的研究主要是在现有知识图的监督下进行半自动构建. 面对多模态、异构性和大规模应用, 自动构建仍然是未来亟待解决的重要问题.

## 7 总 结

知识图谱作为一种语义网络拥有极强的表达能力和建模灵活性, 可以对现实世界中的实体、概念、属性以及它们之间的关系进行建模. 随着最近出现的知识表示学习、知识获取方法和各种知识图谱应用, 知识图谱引起了越来越多的研究关注. 知识图谱嵌入旨在将实体和关系嵌入到连续向量空间中, 在各种面向实体的任务中得到了重要应用. 本文围绕知识图谱嵌入技术的研究现状, 通过回顾仅使用事实进行知识图谱嵌入的方法、添加时间维度的动态 KGE 方法以及融合多源信息的 KGE 技术介绍了现有的知识图谱嵌入技术. 并简要讨论了 KGE 技术在下游任务中的实际应用. 最后总结了知识图谱嵌入领域所面临的挑战, 并对其未来的方向做出展望. 我们进行这项调查的目的是对当前 KGE 的代表性研究工作进行总结, 并且希望这一探索可以为 KGE 的未来研究提供帮助.

## References:

- [1] Bollacker KD, Evans C, Paritosh P, Sturge T, Taylor J. Freebase: A collaboratively created graph database for structuring human knowledge. In: Proc. of the 2008 ACM SIGMOD Int'l Conf. on Management of Data. Vancouver: Association for Computing Machinery, 2008. 1247–1250. [doi: 10.1145/1376616.1376746]
- [2] Auer S, Bizer C, Kobilarov G, Lehmann J, Cyganiak R, Ives Z. DBpedia: A nucleus for a web of open data. In: Proc. of the 6th Int'l Semantic Web Conf. and the 2nd Asian Semantic Web Conf. on the Semantic Web. Busan: Springer, 2007. 722–735. [doi: 10.1007/978-3-540-76298-0\_52]
- [3] Suchanek FM, Kasneci G, Weikum G. Yago: A core of semantic knowledge. In: Proc. of the 16th Int'l Conf. on World Wide Web. Banff: Association for Computing Machinery, 2007. 697–706. [doi: 10.1145/1242572.1242667]
- [4] Carlson A, Betteridge J, Kisiel B, Settles B, Hruschka ER, Mitchell TM. Toward an architecture for never-ending language learning. In: Proc. of the 24th AAAI Conf. on Artificial Intelligence. Atlanta: AAAI Press, 2010. 1306–1313.
- [5] Vrandečić D, Krötzsch M. Wikidata: A free collaborative knowledgebase. Communications of the ACM, 2014, 57(10): 78–85. [doi: 10.1145/2629489]
- [6] Berant J, Chou A, Frostig R, Liang P. Semantic parsing on freebase from question-answer Pairs. In: Proc. of the 2013 Conf. on Empirical Methods in Natural Language Processing (EMNLP). Seattle: Association for Computational Linguistics, 2013. 1533–1544.
- [7] Heck LP, Hakkani-Tür D, Tür G. Leveraging knowledge graphs for web-scale unsupervised semantic parsing. In: Proc. of the 14th Annual Conf. of the Int'l Speech Communication Association (ISCA). Lyon: International Speech Communication Association, 2013. 1594–1598.
- [8] Damjanovic D, Bontcheva K. Named entity disambiguation using linked data. In: Proc. of the 9th Extended Semantic Web Conf. New York: Association for Computing Machinery, 2012. 231–240.



- [9] Zheng ZC, Si XC, Li FT, Chang EY, Zhu XY. Entity disambiguation with freebase. In: Proc. of the 2012 IEEE/WIC/ACM Int'l Conf. on Web Intelligence and Intelligent Agent Technology. Macao: IEEE Computer Society, 2012. 82–89. [doi: [10.1109/WI-IAT.2012.26](https://doi.org/10.1109/WI-IAT.2012.26)]
- [10] Hoffmann R, Zhang CL, Ling X, Zettlemoyer L, Weld DS. Knowledge-based weak supervision for information extraction of overlapping relations. In: Proc. of the 49th Annual Meeting of the Association for Computational Linguistics: Human. Portland: Association for Computational Linguistics (ACL), 2011. 541–550.
- [11] Daiber J, Jakob M, Hokamp C, Mendes PN. Improving efficiency and accuracy in multilingual entity extraction. In: Proc. of the 9th Int'l Conf. on Semantic Systems. Graz: Association for Computing Machinery, 2013. 121–124. [doi: [10.1145/2506182.2506198](https://doi.org/10.1145/2506182.2506198)]
- [12] Bordes A, Weston J, Usunier N. Open question answering with weakly supervised embedding models. In: Proc. of the 2014 European Conf. on Machine Learning and Knowledge Discovery in Databases. Nancy: Springer, 2014. 165–180. [doi: [10.1007/978-3-662-44848-9\\_11](https://doi.org/10.1007/978-3-662-44848-9_11)]
- [13] Bordes A, Chopra S, Weston J. Question answering with subgraph embeddings. In: Proc. of the 2014 Conf. on Empirical Methods in Natural Language Processing (EMNLP). Doha: Association for Computational Linguistics (ACL), 2014. 615–620. [doi: [10.3115/v1/D14-1067](https://doi.org/10.3115/v1/D14-1067)]
- [14] Wang Q, Mao ZD, Wang B, Guo L. Knowledge graph embedding: A survey of approaches and applications. IEEE Trans. on Knowledge and Data Engineering, 2017, 29(12): 2724–2743. [doi: [10.1109/TKDE.2017.2754499](https://doi.org/10.1109/TKDE.2017.2754499)]
- [15] Bordes A, Usunier N, García-Durán A, Weston J, Yakhnenko O. Translating embeddings for modeling multi-relational data. In: Proc. of the 26th Int'l Conf. on Neural Information Processing Systems (NIPS). Lake Tahoe: Curran Associates Inc., 2013. 2787–2795.
- [16] 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 (AAAI). Québec City: AAAI Press, 2014. 1112–1119.
- [17] Riedel S, Yao LM, McCallum A, Marlin BM. Relation extraction with matrix factorization and universal schemas. In: Proc. of the 2013 Conf. of the North American Chapter of the Association for Computational Linguistics (NAACL). Atlanta: Association for Computational Linguistics, 2013. 74–84.
- [18] 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 (ICML). Bellevue: Omnipress, 2011. 809–816.
- [19] Nickel M, Tresp V, Kriegel HP. Factorizing YAGO: Scalable machine learning for linked data. In: Proc. of the 21st Int'l Conf. on World Wide Web (WWW). Lyon: Association for Computing Machinery, 2012. 271–280. [doi: [10.1145/2187836.2187874](https://doi.org/10.1145/2187836.2187874)]
- [20] Bordes A, Glorot X, Weston J, Bengio Y. A semantic matching energy function for learning with multi-relational data: Application to word-sense disambiguation. Machine Learning, 2014, 94(2): 233–259. [doi: [10.1007/s10994-013-5363-6](https://doi.org/10.1007/s10994-013-5363-6)]
- [21] Wang Q, Wang B, Guo L. Knowledge base completion using embeddings and rules. In: Proc. of the 24th Int'l Conf. on Artificial Intelligence (IJCAI). Buenos Aires: AAAI Press, 2015. 1859–1866.
- [22] Wei ZY, Zhao J, Liu K, Qi ZY, Sun ZY, Tian GH. Large-scale knowledge base completion: Inferring via grounding network sampling over selected instances. In: Proc. of the 24th ACM Int'l Conf. on Information and Knowledge Management (CIKM). Melbourne: Association for Computing Machinery, 2015. 1331–1340. [doi: [10.1145/2806416.2806513](https://doi.org/10.1145/2806416.2806513)]
- [23] Guo S, Wang Q, Wang B, Wang LH, Guo L. Semantically smooth knowledge graph embedding. In: Proc. of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th Int'l Joint Conf. on Natural Language Processing. Beijing: Association for Computational Linguistics (ACL), 2015. 84–94. [doi: [10.3115/v1/P15-1009](https://doi.org/10.3115/v1/P15-1009)]
- [24] Xie RB, Liu ZY, Sun MS. Representation learning of knowledge graphs with hierarchical types. In: Proc. of the 25th Int'l Joint Conf. on Artificial Intelligence (IJCAI). New York: IJCAI/AAAI Press, 2016. 2965–2971.
- [25] Wang Z, Zhang JW, Feng JL, Chen Z. Knowledge graph and text jointly embedding. In: Proc. of the 2014 Conf. on Empirical Methods in Natural Language Processing (EMNLP). Doha: Association for Computational Linguistics (ACL), 2014. 1591–1601. [doi: [10.3115/v1/D14-1167](https://doi.org/10.3115/v1/D14-1167)]
- [26] Zhong HP, Zhang JW, Wang Z, Wan H, Chen Z. Aligning knowledge and text embeddings by entity descriptions. In: Proc. of the 2015 Conf. on Empirical Methods in Natural Language Processing. Lisbon: Association for Computational Linguistics (ACL), 2015. 267–272. [doi: [10.18653/v1/D15-1031](https://doi.org/10.18653/v1/D15-1031)]
- [27] Xie RB, Liu ZY, Jia J, Luan HB, Sun MS. Representation learning of knowledge graphs with entity descriptions. In: Proc. of the 30th AAAI Conf. on Artificial Intelligence (AAAI). Phoenix: AAAI Press, 2016. 2659–2665.
- [28] Wang ZG, Li JZ. Text-enhanced representation learning for knowledge graph. In: Proc. of the 25th Int'l Joint Conf. on Artificial Intelligence (IJCAI). New York: AAAI Press, 2016. 1293–1299.
- [29] Lin YK, Liu ZY, Luan HB, Sun MS, Rao SW, Liu S. Modeling relation paths for representation learning of knowledge bases. In: Proc. of the 2015 Conf. on Empirical Methods in Natural Language Processing (EMNLP). Lisbon: Association for Computational Linguistics



- (ACL), 2015. 705–714. [doi: [10.18653/v1/D15-1082](https://doi.org/10.18653/v1/D15-1082)]
- [30] Guu K, Miller J, Liang P. Traversing knowledge graphs in vector space. In: Proc. of the 2015 Conf. on Empirical Methods in Natural Language Processing (EMNLP). Lisbon: Association for Computational Linguistics (ACL), 2015. 318–327. [doi: [10.18653/v1/D15-1038](https://doi.org/10.18653/v1/D15-1038)]
  - [31] Toutanova K, Lin V, Yih WT, Poon H, Quirk C. Compositional learning of embeddings for relation paths in knowledge base and text. In: Proc. of the 54th Annual Meeting of the Association for Computational Linguistics (ACL). Berlin: Association for Computational Linguistics (ACL), 2016. 1434–1444. [doi: [10.18653/v1/P16-1136](https://doi.org/10.18653/v1/P16-1136)]
  - [32] Guo S, Wang Q, Wang LH, Wang B, Guo L. Jointly embedding knowledge graphs and logical rules. In: Proc. of the 2016 Conf. on Empirical Methods in Natural Language Processing (EMNLP). Austin: Association for Computational Linguistics (ACL), 2016. 192–202. [doi: [10.18653/v1/D16-1019](https://doi.org/10.18653/v1/D16-1019)]
  - [33] Rocktäschel T, Singh S, Riedel S. Injecting logical background knowledge into embeddings for relation extraction. In: Proc. of the 2015 Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL). Denver: Association for Computational Linguistics (ACL), 2015. 1119–1129. [doi: [10.3115/v1/N15-1118](https://doi.org/10.3115/v1/N15-1118)]
  - [34] Nickel M, Murphy K, Tresp V, Gabrilovich E. A review of relational machine learning for knowledge graphs. Proc. of the IEEE, 2016, 104(1): 11–33. [doi: [10.1109/JPROC.2015.2483592](https://doi.org/10.1109/JPROC.2015.2483592)]
  - [35] Paulheim H. Knowledge graph refinement: A survey of approaches and evaluation methods. Semantic Web, 2017, 8(3): 489–508.
  - [36] Wu TX, Qi GL, Li C, Wang M. A survey of techniques for constructing Chinese knowledge graphs and their applications. Sustainability, 2018, 10(9): 3245. [doi: [10.3390/su10093245](https://doi.org/10.3390/su10093245)]
  - [37] Lin YK, Han X, Xie RB, Liu ZY, Sun MS. Knowledge representation learning: A quantitative review. arXiv: 1812.10901, 2018.
  - [38] 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 3rd Int'l Conf. on Learning Representations (ICLR). San Diego: International Conference on Learning Representations, 2015.
  - [39] Bordes A, Weston J, Collobert R, Bengio Y. Learning structured embeddings of knowledge bases. In: Proc. of the 25th AAAI Conf. on Artificial Intelligence (AAAI). San Francisco: AAAI Press, 2011. 301–306.
  - [40] Mikolov T, Sutskever I, Chen K, Corrado G, Dean J. Distributed representations of words and phrases and their compositionality. In: Proc. of the 26th Int'l Conf. on Neural Information Processing Systems (NIPS). Lake Tahoe: Curran Associates Inc, 2013. 3111–3119.
  - [41] Mikolov T, Chen K, Corrado G, Dean J. Efficient estimation of word representations in vector space. In: Proc. of the 1st Int'l Conf. on Learning Representations (ICLR). Scottsdale: Int'l Conference on Learning Representations, 2013.
  - [42] Bordes A, Glorot X, Weston J, Bengio Y. Joint learning of words and meaning representations for open-text semantic parsing. In: Proc. of the 15th Int'l Conf. on Artificial Intelligence and Statistics (AISTATS). La Palma: JMLR, 2012. 127–135.
  - [43] 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 (AAAI). Austin: AAAI Press, 2015. 2181–2187.
  - [44] 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 and the 7th Int'l Joint Conf. on Natural Language Processing. Beijing: Association for Computational Linguistics (ACL), 2015. 687–696. [doi: [10.3115/v1/P15-1067](https://doi.org/10.3115/v1/P15-1067)]
  - [45] Nguyen DQ, Sirts K, Qu LZ, Johnson M. STransE: A novel embedding model of entities and relationships in knowledge bases. In: Proc. of the 2016 Conf. of the North American Chapter of the Association for Computational Linguistics (NAACL). San Diego: Association for Computational Linguistics (ACL), 2016. 460–466. [doi: [10.18653/v1/N16-1054](https://doi.org/10.18653/v1/N16-1054)]
  - [46] 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 (AAAI). Phoenix: AAAI Press, 2016. 985–991.
  - [47] 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 Computing (PACLIC). Phuket: Department of Linguistics, Chulalongkorn University, 2014. 328–337.
  - [48] Xiao H, Huang ML, Hao Y, Zhu XY. TransA: An adaptive approach for knowledge graph embedding. arXiv: 1509.05490, 2015.
  - [49] Wang F, Sun JM. Survey on distance metric learning and dimensionality reduction in data mining. Data Mining and Knowledge Discovery, 2015, 29(2): 534–564. [doi: [10.1007/s10618-014-0356-z](https://doi.org/10.1007/s10618-014-0356-z)]
  - [50] Feng J, Huang ML, Wang MD, Zhou MT, Hao Y, Zhu XY. Knowledge graph embedding by flexible translation. In: Proc. of the 15th Int'l Conf. on Principles of Knowledge Representation and Reasoning (KR). Cape Town: AAAI Press, 2016. 557–560.
  - [51] Xie QZ, Ma XZ, Dai ZH, Hovy E. An interpretable knowledge transfer model for knowledge base completion. In: Proc. of the 55th Annual Meeting of the Association for Computational Linguistics (ACL). Vancouver: Association for Computational Linguistics (ACL), 2017. 950–962. [doi: [10.18653/v1/P17-1088](https://doi.org/10.18653/v1/P17-1088)]

- [52] Qian W, Fu C, Zhu Y, Cai D, He XF. Translating embeddings for knowledge graph completion with relation attention mechanism. In: Proc. of the 27th Int'l Joint Conf. on Artificial Intelligence (IJCAI). Stockholm: IJCAI.org, 2018. 4286–4292. [doi: [10.24963/ijcai.2018/596](https://doi.org/10.24963/ijcai.2018/596)]
- [53] Yang SH, Tian JD, Zhang HL, Yan JC, He H, Jin YH. TransMS: Knowledge graph embedding for complex relations by multidirectional semantics. In: Proc. of the 28th Int'l Joint Conf. on Artificial Intelligence (IJCAI). Macao: IJCAI.org, 2019. 1935–1942. [doi: [10.24963/ijcai.2019/268](https://doi.org/10.24963/ijcai.2019/268)]
- [54] Ji SX, Pan SR, Cambria E, Marttinen P, Yu P. A survey on knowledge graphs: Representation, acquisition, and applications. IEEE Trans. on Neural Networks and Learning Systems, 2022, 33(2): 494–514. [doi: [10.1109/TNNLS.2021.3070843](https://doi.org/10.1109/TNNLS.2021.3070843)]
- [55] 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. New York: AAAI Press, 2016. 1315–1321.
- [56] Ebisu T, Ichise R. TorusE: Knowledge graph embedding on a lie group. In: Proc. of the 32nd AAAI Conf. on Artificial Intelligence (AAAI). New Orleans: AAAI Press, 2018. 1819–1826.
- [57] He SZ, Liu K, Ji GL, Zhao J. Learning to represent knowledge graphs with Gaussian embedding. In: Proc. of the 24th ACM Int'l Conf. on Information and Knowledge Management (CIKM). Melbourne: Association for Computing Machinery, 2015. 623–632. [doi: [10.1145/2806416.2806502](https://doi.org/10.1145/2806416.2806502)]
- [58] Xiao H, Huang ML, Zhu XY. TransG: A generative model for knowledge graph embedding. In: Proc. of the 54th Annual Meeting of the Association for Computational Linguistics. Berlin: Association for Computational Linguistics (ACL), 2016. 2316–2325. [doi: [10.18653/v1/p16-1219](https://doi.org/10.18653/v1/p16-1219)]
- [59] Griffiths TL, Ghahramani Z. The Indian buffet process: An introduction and review. The Journal of Machine Learning Research, 2011, 12: 1185–1224.
- [60] Blei DM, Griffiths TL, Jordan MI. The nested Chinese restaurant process and bayesian nonparametric inference of topic hierarchies. Journal of the ACM, 2010, 57(2): 7. [doi: [10.1145/1667053.1667056](https://doi.org/10.1145/1667053.1667056)]
- [61] Aldous DJ. Exchangeability and related topics. In: Hennequin PL, ed. École d'Été de Probabilités de Saint-Flour XIII—1983. Berlin: Springer, 1985. 1–198. [doi: [10.1007/BFb0099421](https://doi.org/10.1007/BFb0099421)]
- [62] Sutskever I, Salakhutdinov R, Tenenbaum JB. Modelling relational data using bayesian clustered tensor factorization. In: Proc. of the 22nd Int'l Conf. on Neural Information Processing Systems (NIPS). Vancouver: Curran Associates Inc., 2009. 1821–1828.
- [63] Jenatton R, Le Roux N, Bordes A, Obozinski G. A latent factor model for highly multi-relational data. In: Proc. of the 25th Int'l Conf. on Neural Information Processing Systems (NIPS). Lake Tahoe: Curran Associates Inc, 2012. 3167–3175.
- [64] Carroll JD, Chang JJ. Analysis of individual differences in multidimensional scaling via an n-way generalization of “Eckart-Young” decomposition. Psychometrika, 1970, 35(3): 283–319. [doi: [10.1007/BF02310791](https://doi.org/10.1007/BF02310791)]
- [65] Nickel M, Rosasco L, Poggio TA. Holographic embeddings of knowledge graphs. In: Proc. of the 30th AAAI Conf. on Artificial Intelligence (AAAI). Phoenix: AAAI Press, 2016. 1955–1961.
- [66] Trouillon T, Welbl J, Riedel S, Gaussier É, Bouchard G. Complex embeddings for simple link prediction. In: Proc. of the 33rd Int'l Conf. on Machine Learning (ICML). New York: JMLR, 2016. 2071–2080.
- [67] Hayashi K, Shimbo M. On the equivalence of holographic and complex embeddings for link prediction. In: Proc. of the 55th Annual Meeting of the Association for Computational Linguistics (ACL). Vancouver: Association for Computational Linguistics (ACL), 2017. 554–559. [doi: [10.18653/v1/P17-2088](https://doi.org/10.18653/v1/P17-2088)]
- [68] Liu HX, Wu YX, Yang YM. Analogical inference for multi-relational embeddings. In: Proc. of the 34th Int'l Conf. on Machine Learning (ICML). Sydney: PMLR, 2017. 2168–2178.
- [69] Hitchcock FL. The expression of a tensor or a polyadic as a sum of products. Journal of Mathematics and Physics, 1927, 6(1-4): 164–189. [doi: [10.1002/sapm192761164](https://doi.org/10.1002/sapm192761164)]
- [70] Kazemi SM, Poole D. Simple embedding for link prediction in knowledge graphs. In: Proc. of the 32nd Int'l Conf. on Neural Information Processing Systems (NIPS). Montréal: Curran Associates Inc., 2018. 4289–4300.
- [71] García-Durán A, Bordes A, Usunier N. Effective blending of two and three-way interactions for modeling multi-relational data. In: Proc. of the 2014 European Conf. on Machine Learning and Knowledge Discovery in Databases. Nancy: Springer, 2014. 434–449. [doi: [10.1007/978-3-662-44848-9\\_28](https://doi.org/10.1007/978-3-662-44848-9_28)]
- [72] Balazevic I, Allen C, Hospedales T. TuckER: Tensor factorization for knowledge graph completion. In: Proc. of the 2019 Conf. on Empirical Methods in Natural Language Processing and the 9th Int'l Joint Conf. on Natural Language Processing (EMNLP). Hong Kong: Association for Computational Linguistics (ACL), 2019. 5185–5194. [doi: [10.18653/v1/D19-1522](https://doi.org/10.18653/v1/D19-1522)]
- [73] Fan M, Zhao DL, Zhou Q, Liu ZY, Zheng TF, Chang EY. Distant supervision for relation extraction with matrix completion. In: Proc. of

- the 52nd Annual Meeting of the Association for Computational Linguistics (ACL). Baltimore: Association for Computational Linguistics (ACL), 2014. 839–849. [doi: [10.3115/v1/P14-1079](https://doi.org/10.3115/v1/P14-1079)]
- [74] Tresp V, Huang Y, Bundschuh M, Rettinger A. Materializing and querying learned knowledge. In: Proc. of the 1st ESWC Workshop on Inductive Reasoning and Machine Learning on the Semantic Web. Heraklion: CEUR-WS, 2009.
- [75] Huang Y, Tresp V, Nickel M, Rettinger A, Kriegel HP. A scalable approach for statistical learning in semantic graphs. *Semantic Web*, 2014, 5(1): 5–22. [doi: [10.3233/SW-130100](https://doi.org/10.3233/SW-130100)]
- [76] Dong X, Gabrilovich E, Heitz G, Horn W, Lao N, Murphy K, Strohmann T, Sun SH, Zhang W. Knowledge vault: A Web-scale approach to probabilistic knowledge fusion. In: Proc. of the 20th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining (KDD). New York: Association for Computing Machinery, 2014. 601–610. [doi: [10.1145/2623330.2623623](https://doi.org/10.1145/2623330.2623623)]
- [77] Socher R, Chen DQ, Manning CD, Ng AY. Reasoning with neural tensor networks for knowledge base completion. In: Proc. of the 26th Int'l Conf. on Neural Information Processing Systems (NIPS). Lake Tahoe: Curran Associates Inc, 2013. 926–934.
- [78] Liu Q, Jiang H, Evdokimov A, Ling ZH, Zhu XD, Wei S, Hu Y. Probabilistic reasoning via deep learning: Neural association models. arXiv: 1603.07704, 2016.
- [79] Dettmers T, Minervini P, Stenetorp P, Riedel S. Convolutional 2D knowledge graph embeddings. In: Proc. of the 32nd AAAI Conf. on Artificial Intelligence (AAAI). New Orleans: AAAI Press, 2018. 1811–1818.
- [80] Schlichtkrull M, Kipf TN, Bloem P, van den Berg R, Titov I, Welling M. Modeling relational data with graph convolutional networks. In: Proc. of the 15th Int'l Conf. on Semantic Web. Heraklion: Springer International Publishing, 2018. 593–607. [doi: [10.1007/978-3-319-93417-4\\_38](https://doi.org/10.1007/978-3-319-93417-4_38)]
- [81] Nguyen DQ, Nguyen TD, Nguyen DQ, Phung D. A novel embedding model for knowledge base completion based on convolutional neural network. In: Proc. of the 2018 Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL). New Orleans: Association for Computational Linguistics (ACL), 2018. 327–333. [doi: [10.18653/v1/N18-2053](https://doi.org/10.18653/v1/N18-2053)]
- [82] Balažević I, Allen C, Hospedales TM. Hypernetwork knowledge graph embeddings. In: Proc. of the 28th Int'l Conf. on Artificial Neural Networks and Machine Learning (ICANN). Munich: Springer International Publishing, 2019. 553–565. [doi: [10.1007/978-3-030-30493-5\\_52](https://doi.org/10.1007/978-3-030-30493-5_52)]
- [83] Ha D, Dai AM, Le QV. HyperNetworks. In: Proc. of the 5th Int'l Conf. on Learning Representations (ICLR). Toulon: OpenReview.net, 2017.
- [84] Sun ZQ, Deng ZH, Nie JY, Tang J. RotatE: Knowledge graph embedding by relational rotation in complex space. In: Proc. of the 7th Int'l Conf. on Learning Representations (ICLR). New Orleans: OpenReview.net, 2019.
- [85] Zhang S, Tay Y, Yao LN, Liu Q. Quaternion knowledge graph embeddings. In: Proc. of the 33rd Int'l Conf. on Neural Information Processing Systems (NIPS). Vancouver: Curran Associates Inc., 2019. 246.
- [86] Xu CR, Li RJ. Relation embedding with dihedral group in knowledge graph. In: Proc. of the 57th Annual Meeting of the Association for Computational Linguistics (ACL). Florence: Association for Computational Linguistics (ACL), 2019. 263–272. [doi: [10.18653/v1/P19-1026](https://doi.org/10.18653/v1/P19-1026)]
- [87] Adcock AB, Sullivan BD, Mahoney MW. Tree-like structure in large social and information networks. In: Proc. of the 13th IEEE Int'l Conf. on Data Mining. Dallas: IEEE Computer Society, 2013. 1–10. [doi: [10.1109/ICDM.2013.77](https://doi.org/10.1109/ICDM.2013.77)]
- [88] Nickel M, Kiela D. Poincaré embeddings for learning hierarchical representations. In: Proc. of the 31st Int'l Conf. on Neural Information Processing Systems (NIPS). Long Beach: Curran Associates Inc., 2017. 6341–6350.
- [89] Balažević I, Allen C, Hospedales T. Multi-relational poincaré graph embeddings. In: Proc. of the 33rd Int'l Conf. on Neural Information Processing Systems. Vancouver: Curran Associates Inc., 2019. 401.
- [90] Nguyen DQ, Vu T, Nguyen TD, Nguyen DQ, Phung D. A capsule network-based embedding model for knowledge graph completion and search personalization. In: Proc. of the 2019 Conf. of the North American Chapter of the Association for Computational Linguistics (NAACL). Minneapolis: Association for Computational Linguistics (ACL), 2019. 2180–2189. [doi: [10.18653/v1/N19-1226](https://doi.org/10.18653/v1/N19-1226)]
- [91] Zhang W, Paudel B, Zhang W, Bernstein A, Chen HJ. Interaction embeddings for prediction and explanation in knowledge graphs. In: Proc. of the 12th ACM Int'l Conf. on Web Search and Data Mining (WSDM). Melbourne: Association for Computing Machinery, 2019. 96–104. [doi: [10.1145/3289600.3291014](https://doi.org/10.1145/3289600.3291014)]
- [92] Jiang TS, Liu TY, Ge T, Sha L, Li SJ, Chang BB, Sui ZF. Encoding temporal information for time-aware link prediction. In: Proc. of the 2016 Conf. on Empirical Methods in Natural Language Processing (EMNLP). Austin: Association for Computational Linguistics (ACL), 2016. 2350–2354. [doi: [10.18653/v1/D16-1260](https://doi.org/10.18653/v1/D16-1260)]
- [93] Trivedi R, Dai HJ, Wang YC, Song L. Know-evolve: Deep temporal reasoning for dynamic knowledge graphs. In: Proc. of the 34th Int'l

- Conf. on Machine Learning (ICML). Sydney: JMLR, 2017. 3462–3471.
- [94] Leblay J, Chekol MW. Deriving validity time in knowledge graph. In: Proc. of 2018 the Web Conf. Lyon: Association for Computing Machinery, 2018. 1771–1776. [doi: [10.1145/3184558.3191639](https://doi.org/10.1145/3184558.3191639)]
  - [95] García-Durán A, Dumančić S, Niepert M. Learning sequence encoders for temporal knowledge graph completion. In: Proc. of the 2018 Conf. on Empirical Methods in Natural Language Processing (EMNLP). Brussels: Association for Computational Linguistics (ACL), 2018. 4816–4821. [doi: [10.18653/v1/D18-1516](https://doi.org/10.18653/v1/D18-1516)]
  - [96] Dasgupta SS, Ray SN, Talukdar PP. HyTE: Hyperplane-based temporally aware knowledge graph embedding. In: Proc. of the 2018 Conf. on Empirical Methods in Natural Language Processing (EMNLP). Brussels: Association for Computational Linguistics (ACL), 2018. 2001–2011. [doi: [10.18653/v1/D18-1225](https://doi.org/10.18653/v1/D18-1225)]
  - [97] Tang XL, Yuan R, Li QY, Wang TY, Yang HZ, Cai YD, Song HJ. Timespan-aware dynamic knowledge graph embedding by incorporating temporal evolution. IEEE Access, 2020, 8: 6849–6860. [doi: [10.1109/ACCESS.2020.2964028](https://doi.org/10.1109/ACCESS.2020.2964028)]
  - [98] Jin J, Wan HY, Lin YF. Knowledge graph representation learning fused with entity category information. Computer Engineering, 2021, 47(4): 77–83 (in Chinese with English abstract). [doi: [10.19678/j.issn.1000-3428.0057353](https://doi.org/10.19678/j.issn.1000-3428.0057353)]
  - [99] Zhang DX, Yuan B, Wang D, Liu R. Joint semantic relevance learning with text data and graph knowledge. In: Proc. of the 3rd Workshop on Continuous Vector Space Models and their Compositionality. Beijing: ACL, 2015. 32–40. [doi: [10.18653/v1/W15-4004](https://doi.org/10.18653/v1/W15-4004)]
  - [100] Xiao H, Huang ML, Meng L, Zhu XY. SSP: Semantic space projection for knowledge graph embedding with text descriptions. In: Proc. of the 31st AAAI Conf. on Artificial Intelligence (AAAI). San Francisco: AAAI Press, 2017. 3104–3110.
  - [101] An B, Chen B, Han XP, Sun L. Accurate text-enhanced knowledge graph representation learning. In: Proc. of the 2018 Conf. of the North American Chapter of the Association for Computational Linguistics (NAACL). New Orleans: Association for Computational Linguistics (ACL), 2018. 745–755. [doi: [10.18653/v1/N18-1068](https://doi.org/10.18653/v1/N18-1068)]
  - [102] Galárraga LA, Teflioudi C, Hose K, Suchanek FM. AMIE: Association rule mining under incomplete evidence in ontological knowledge bases. In: Proc. of the 22nd Int'l Conf. on World Wide Web (WWW). New York: Association for Computing Machinery, 2013. 413–422. [doi: [10.1145/2488388.2488425](https://doi.org/10.1145/2488388.2488425)]
  - [103] Galárraga L, Teflioudi C, Hose K, Suchanek FM. Fast rule mining in ontological knowledge bases with AMIE+. The VLDB Journal, 2015, 24(6): 707–730. [doi: [10.1007/s00778-015-0394-1](https://doi.org/10.1007/s00778-015-0394-1)]
  - [104] Omran PG, Wang KW, Wang Z. An embedding-based approach to rule learning in knowledge graphs. IEEE Trans. on Knowledge and Data Engineering, 2021, 33(4): 1348–1359. [doi: [10.1109/TKDE.2019.2941685](https://doi.org/10.1109/TKDE.2019.2941685)]
  - [105] Guo S, Wang Q, Wang LH, Wang B, Guo L. Knowledge graph embedding with iterative guidance from soft rules. In: Proc. of the 32nd AAAI Conf. on Artificial Intelligence (AAAI). New Orleans: AAAI Press, 2018. 4816–4823.
  - [106] Niu GL, Zhang YF, Li B, Cui P, Liu S, Li JY, Zhang XW. Rule-guided compositional representation learning on knowledge graphs. In: Proc. of the 34th AAAI Conf. on Artificial Intelligence (AAAI). New York: AAAI Press, 2020. 2950–2958. [doi: [10.1609/aaai.v34i03.5687](https://doi.org/10.1609/aaai.v34i03.5687)]
  - [107] Xie RB, Liu ZY, Luan HB, Sun MS. Image-embodied knowledge representation learning. In: Proc. of the 26th Int'l Joint Conf. on Artificial Intelligence (IJCAI). Melbourne: IJCAI.org, 2017. 3140–3146. [doi: [10.24963/ijcai.2017/438](https://doi.org/10.24963/ijcai.2017/438)]
  - [108] Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. In: Proc. of the 25th Int'l Conf. on Neural Information Processing Systems (NIPS). Lake Tahoe: Curran Associates Inc., 2012. 1097–1105.
  - [109] Feng J, Huang ML, Yang Y, Zhu XY. GAKE: Graph aware knowledge embedding. In: Proc. of the 26th Int'l Conf. on Computational Linguistics (COLING). Osaka: The COLING 2016 Organizing Committee, 2016. 641–651.
  - [110] Du WQ, Li BC, Wang R. Representation learning of knowledge graph integrating entity description and entity type. Journal of Chinese Information Processing, 2020, 34(7): 50–59 (in Chinese with English abstract). [doi: [10.3969/j.issn.1003-0077.2020.07.005](https://doi.org/10.3969/j.issn.1003-0077.2020.07.005)]
  - [111] Tang X, Chen L, Cui J, Wei BG. Knowledge representation learning with entity descriptions, hierarchical types, and textual relations. Information Processing & Management, 2019, 55(3): 809–822. [doi: [10.1016/j.ipm.2019.01.005](https://doi.org/10.1016/j.ipm.2019.01.005)]
  - [112] Yih WT, Chang MW, He XD, Gao JF. Semantic parsing via staged query graph generation: Question answering with knowledge base. In: Proc. of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th Int'l Joint Conf. on Natural Language Processing. Beijing: Association for Computational Linguistics (ACL), 2015. 1321–1331. [doi: [10.3115/v1/P15-1128](https://doi.org/10.3115/v1/P15-1128)]
  - [113] Blanco R, Ottaviano G, Meij E. Fast and space-efficient entity linking for queries. In: Proc. of the 8th ACM Int'l Conf. on Web Search and Data Mining (WSDM). Shanghai: Association for Computing Machinery, 2015. 179–188. [doi: [10.1145/2684822.2685317](https://doi.org/10.1145/2684822.2685317)]
  - [114] Pappu A, Blanco R, Mehdad Y, Stent A, Thadani K. Lightweight multilingual entity extraction and linking. In: Proc. of the 10th ACM Int'l Conf. on Web Search and Data Mining (WSDM). Cambridge: Association for Computing Machinery, 2017. 365–374. [doi: [10.1145/3018661.3018724](https://doi.org/10.1145/3018661.3018724)]

- [115] Bordes A, Usunier N, Chopra S, Weston J. Large-scale simple question answering with memory networks. arXiv: 1506.02075, 2015.
- [116] Yang MC, Duan N, Zhou M, Rim HC. Joint relational embeddings for knowledge-based question answering. In: Proc. of the 2014 Conf. on Empirical Methods in Natural Language Processing (EMNLP). Doha: Association for Computational Linguistics (ACL), 2014. 645–650. [doi: [10.3115/v1/D14-1071](https://doi.org/10.3115/v1/D14-1071)]
- [117] Yang MC, Lee DG, Park SY, Rim HC. Knowledge-based question answering using the semantic embedding space. Expert Systems with Applications, 2015, 42(23): 9086–9104. [doi: [10.1016/j.eswa.2015.07.009](https://doi.org/10.1016/j.eswa.2015.07.009)]
- [118] Dai ZH, Li L, Xu W. CFO: Conditional focused neural question answering with large-scale knowledge bases. In: Proc. of the 54th Annual Meeting of the Association for Computational Linguistics (ACL). Berlin: Association for Computational Linguistics (ACL), 2016. 800–810. [doi: [10.18653/v1/P16-1076](https://doi.org/10.18653/v1/P16-1076)]
- [119] Dong L, Wei FR, Zhou M, Xu K. Question answering over freebase with multi-column convolutional neural networks. In: Proc. of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th Int'l Joint Conf. on Natural Language Processing. Beijing: Association for Computational Linguistics (ACL), 2015. 260–269. [doi: [10.3115/v1/P15-1026](https://doi.org/10.3115/v1/P15-1026)]
- [120] Hao YC, Zhang YZ, Liu K, He SZ, Liu ZY, Wu H, Zhao J. An end-to-end model for question answering over knowledge base with cross-attention combining global knowledge. In: Proc. of the 55th Annual Meeting of the Association for Computational Linguistics (ACL). Vancouver: Association for Computational Linguistics (ACL), 2017. 221–231. [doi: [10.18653/v1/P17-1021](https://doi.org/10.18653/v1/P17-1021)]
- [121] Lukovnikov D, Fischer A, Lehmann J, Auer S. Neural network-based question answering over knowledge graphs on word and character level. In: Proc. of the 26th Int'l Conf. on World Wide Web (WWW). Perth: ACM, 2017. 1211–1220. [doi: [10.1145/3038912.3052675](https://doi.org/10.1145/3038912.3052675)]
- [122] Yin WP, Yu M, Xiang B, Zhou BW, Schütze H. Simple question answering by attentive convolutional neural network. In: Proc. of the 26th Int'l Conf. on Computational Linguistics: Technical Papers. Osaka: The COLING 2016 Organizing Committee, 2016. 1746–1756.
- [123] Huang X, Zhang JY, Li DC, Li P. Knowledge graph embedding based question answering. In: Proc. of the 12th ACM Int'l Conf. on Web Search and Data Mining (WSDM). Melbourne: Association for Computing Machinery, 2019. 105–113. [doi: [10.1145/3289600.3290956](https://doi.org/10.1145/3289600.3290956)]
- [124] Di Noia T, Ostuni VC, Tomeo P, Di Sciascio E. SPrank: Semantic path-based ranking for top-N recommendations using linked open data. ACM Trans. on Intelligent Systems and Technology, 2017, 8(1): 9. [doi: [10.1145/2899005](https://doi.org/10.1145/2899005)]
- [125] Yu X, Ren X, Sun YZ, Gu QQ, Sturt B, Khandelwal U, Norick B, Han JW. Personalized entity recommendation: A heterogeneous information network approach. In: Proc. of the 7th ACM Int'l Conf. on Web Search and Data Mining (WSDM). New York: Association for Computing Machinery, 2014. 283–292. [doi: [10.1145/2556195.2556259](https://doi.org/10.1145/2556195.2556259)]
- [126] Catherine R, Cohen W. Personalized recommendations using knowledge graphs: A probabilistic logic programming approach. In: Proc. of the 10th ACM Conf. on Recommender Systems (RecSys). Boston: Association for Computing Machinery, 2016. 325–332. [doi: [10.1145/2959100.2959131](https://doi.org/10.1145/2959100.2959131)]
- [127] Ostuni VX, Di Noia T, Mirizzi R, Di Sciascio E. Top-N recommendations from implicit feedback leveraging linked open data. In: Proc. of the 5th Italian Information Retrieval Workshop (IIR). Roma: CEUR-WS, 2014. 20–27.
- [128] Palumbo E, Rizzo G, Troncy R. entity2rec: Learning user-item relatedness from knowledge graphs for Top-N item recommendation. In: Proc. of the 11th ACM Conf. on Recommender Systems (RecSys). Como: Association for Computing Machinery, 2017. 32–36. [doi: [10.1145/3109859.3109889](https://doi.org/10.1145/3109859.3109889)]
- [129] Zhang FZ, Yuan NJ, Lian DF, Xie X, Ma WY. Collaborative knowledge base embedding for recommender systems. In: Proc. of the 22nd ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining (KDD). San Francisco: Association for Computing Machinery, 2016. 353–362. [doi: [10.1145/2939672.2939673](https://doi.org/10.1145/2939672.2939673)]
- [130] Wang HW, Zhang FZ, Xie X, Guo MY. DKN: Deep knowledge-aware network for news recommendation. In: Proc. of the 2018 Conf. on World Wide Web. Lyon: Association for Computing Machinery, 2018. 1835–1844. [doi: [10.1145/3178876.3186175](https://doi.org/10.1145/3178876.3186175)]
- [131] Wang HW, Zhang FZ, Zhao M, Li WJ, Xie X, Guo MY. Multi-task feature learning for knowledge graph enhanced recommendation. In: Proc. of the 2019 Conf. on World Wide Web (WWW). San Francisco: Association for Computing Machinery, 2019. 2000–2010. [doi: [10.1145/3308558.3313411](https://doi.org/10.1145/3308558.3313411)]
- [132] Ai QY, Azizi V, Chen X, Zhang YF. Learning heterogeneous knowledge base embeddings for explainable recommendation. Algorithms, 2018, 11(9): 137. [doi: [10.3390/a11090137](https://doi.org/10.3390/a11090137)]
- [133] Chowdhury G, Srilakshmi M, Chain M, Sarkar S. Neural factorization for offer recommendation using knowledge graph embeddings. In: Proc. of the SIGIR 2019 Workshop on eCommerce, co-located with the 42nd Int'l ACM SIGIR Conf. on Research and Development in Information Retrieval. Paris: CEUR-WS, 2019.
- [134] Sha X, Sun Z, Zhang J. Hierarchical attentive knowledge graph embedding for personalized recommendation. arXiv: 1910.08288, 2019.
- [135] Ni CC, Liu KS, Torzec N. Layered graph embedding for entity recommendation using wikipedia in the Yahoo! Knowledge graph. In:



- Proc. of the 2020 Companion of the Web Conf. Taipei: ACM, 2020. 811–818. [doi: [10.1145/3366424.3383570](https://doi.org/10.1145/3366424.3383570)]
- [136] Zelenko D, Aone C, Richardella A. Kernel methods for relation extraction. *The Journal of Machine Learning Research*, 2003, 3: 1083–1106.
  - [137] Zeng DJ, Liu K, Lai SW, Zhou GY, Zhao J. Relation classification via convolutional deep neural network. In: *Proc. of the 25th Int'l Conf. on Computational Linguistics (COLING)*. Dublin: Association for Computational Linguistics, 2014. 2335–2344.
  - [138] Riedel S, Yao LM, McCallum A. Modeling relations and their mentions without labeled text. In: *Proc. of the 2010 European Conf. on Machine Learning and Knowledge Discovery in Databases*. Barcelona: Springer, 2010. 148–163. [doi: [10.1007/978-3-642-15939-8\\_10](https://doi.org/10.1007/978-3-642-15939-8_10)]
  - [139] Surdeanu M, Tibshirani J, Nallapati R, Manning CD. Multi-instance multi-label learning for relation extraction. In: *Proc. of the 2012 Joint Conf. on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*. Jeju Island: Association for Computational Linguistics (ACL), 2012. 455–465.
  - [140] Weston J, Bordes A, Yakhnenko O, Usunier N. Connecting language and knowledge bases with embedding models for relation extraction. In: *Proc. of the 2013 Conf. on Empirical Methods in Natural Language Processing (EMNLP)*. Seattle: Association for Computational Linguistics (ACL), 2013. 1366–1371.
  - [141] Han X, Liu ZY, Sun MS. Joint representation learning of text and knowledge for knowledge graph completion. *arXiv*: 1611.04125, 2016.
  - [142] Han X, Liu ZY, Sun MS. Neural knowledge acquisition via mutual attention between knowledge graph and text. In: *Proc. of the 32nd AAAI Conf. on Artificial Intelligence (AAAI)*. New Orleans: AAAI Press, 2018. 4832–4839.
  - [143] Lei K, Chen DY, Li YL, Du N, Yang M, Fan W, Shen Y. Cooperative denoising for distantly supervised relation extraction. In: *Proc. of the 27th Int'l Conf. on Computational Linguistics (COLING)*. Santa Fe: Association for Computational Linguistics, 2018. 426–436.
  - [144] Zhang Z, Zhuang FZ, Qu M, Lin F, He Q. Knowledge graph embedding with hierarchical relation structure. In: *Proc. of the 2018 Conf. on Empirical Methods in Natural Language Processing (EMNLP)*. Brussels: Association for Computational Linguistics (ACL), 2018. 3198–3207. [doi: [10.18653/v1/D18-1358](https://doi.org/10.18653/v1/D18-1358)]
  - [145] Toutanova K, Chen DQ. Observed versus latent features for knowledge base and text inference. In: *Proc. of the 3rd Workshop on Continuous Vector Space Models and Their Compositionality*. Beijing: Association for Computational Linguistics (ACL), 2015. 57–66. [doi: [10.18653/v1/W15-4007](https://doi.org/10.18653/v1/W15-4007)]
  - [146] Le P, Dymetman M, Renders JM. LSTM-based mixture-of-experts for knowledge-aware dialogues. In: *Proc. of the 1st Workshop on Representation Learning for NLP (Rep4NLP)*. Berlin: Association for Computational Linguistics (ACL), 2016. 94–99. [doi: [10.18653/v1/W16-1611](https://doi.org/10.18653/v1/W16-1611)]
  - [147] Zhu WY, Mo KX, Zhang Y, Zhu ZB, Peng XZ, Yang Q. Flexible end-to-end dialogue system for knowledge grounded conversation. *arXiv*: 1709.04264, 2017.
  - [148] Huang HZ, Heck LP, Ji H. Leveraging deep neural networks and knowledge graphs for entity disambiguation. *arXiv*: 1504.07678, 2015.
  - [149] Fang W, Zhang JW, Wang DL, Chen Z, Li M. Entity disambiguation by knowledge and text jointly embedding. In: *Proc. of the 20th SIGNLL Conf. on Computational Natural Language Learning (CoNLL)*. Berlin: Association for Computational Linguistics (ACL), 2016. 260–269. [doi: [10.18653/v1/K16-1026](https://doi.org/10.18653/v1/K16-1026)]
  - [150] Krompaß D, Baier S, Tresp V. Type-constrained representation learning in knowledge graphs. In: *Proc. of the 14th Int'l Semantic Web Conf. (ISWC)*. Bethlehem: Springer, 2015. 640–655. [doi: [10.1007/978-3-319-25007-6\\_37](https://doi.org/10.1007/978-3-319-25007-6_37)]
  - [151] Cochez M, Ristoski P, Ponzetto SP, Paulheim H. Global RDF vector space embeddings. In: *Proc. of the 16th Int'l Semantic Web Conf. (ISWC)*. Vienna: Springer International Publishing, 2017. 190–207. [doi: [10.1007/978-3-319-68288-4\\_12](https://doi.org/10.1007/978-3-319-68288-4_12)]
  - [152] Ristoski P, Paulheim H. RDF2Vec: RDF graph embeddings for data mining. In: *Proc. of the 15th Int'l Semantic Web Conf. Kobe*: Springer, 2016. 498–514. [doi: [10.1007/978-3-319-46523-4\\_30](https://doi.org/10.1007/978-3-319-46523-4_30)]
  - [153] Chen MH, Tian YT, Chang KW, Skiena S, Zaniolo C. Co-training embeddings of knowledge graphs and entity descriptions for cross-lingual entity alignment. In: *Proc. of the 27th Int'l Joint Conf. on Artificial Intelligence (IJCAI)*. Stockholm: IJCAI.org, 2018. 3998–4004. [doi: [10.24963/ijcai.2018/556](https://doi.org/10.24963/ijcai.2018/556)]
  - [154] Chen MH, Tian YT, Yang MH, Zaniolo C. Multilingual knowledge graph embeddings for cross-lingual knowledge alignment. In: *Proc. of the 26th Int'l Joint Conf. on Artificial Intelligence (IJCAI)*. Melbourne: AAAI Press, 2017. 1511–1517.
  - [155] Gentile AL, Ristoski P, Eckel S, Ritze D, Paulheim H. Entity matching on Web tables: A table embeddings approach for blocking. In: *Proc. of the 20th Int'l Conf. on Extending Database Technology (EDBT)*. Venice: OpenProceedings.org, 2017. 510–513.
  - [156] Sun ZQ, Hu W, Li CK. Cross-lingual entity alignment via joint attribute-preserving embedding. In: *Proc. of the 16th Int'l Semantic Web Conf. (ISWC)*. Vienna: Springer International Publishing, 2017. 628–644. [doi: [10.1007/978-3-319-68288-4\\_37](https://doi.org/10.1007/978-3-319-68288-4_37)]
  - [157] Tay Y, Luu AT, Hui SC, Brauer F. Random semantic tensor ensemble for scalable knowledge graph link prediction. In: *Proc. of the 10th*

- ACM Int'l Conf. on Web Search and Data Mining (WSDM). Cambridge: Association for Computing Machinery, 2017. 751–760. [doi: [10.1145/3018661.3018695](https://doi.org/10.1145/3018661.3018695)]
- [158] Wang YJ, Gemulla R, Li H. On multi-relational link prediction with bilinear models. In: Proc. of the 32nd AAAI Conf. on Artificial Intelligence (AAAI). New Orleans: AAAI Press, 2018. 4227–4324.
- [159] Chandrhas, Sharma A, Talukdar P. Towards understanding the geometry of knowledge graph embeddings. In: Proc. of the 56th Annual Meeting of the Association for Computational Linguistics (ACL). Melbourne: Association for Computational Linguistics (ACL), 2018. 122–131. [doi: [10.18653/v1/P18-1012](https://doi.org/10.18653/v1/P18-1012)]
- [160] Battaglia PW, Hamrick JB, Bapst V, *et al.* Relational inductive biases, deep learning, and graph networks. arXiv: 1806.01261, 2018.
- [161] Qu M, Tang J. Probabilistic logic neural networks for reasoning. In: Proc. of the 33rd Conf. on Neural Information Processing Systems (NIPS). Vancouver: Neural Information Processing Systems Foundation, 2019. 7710–7720.
- [162] Zhang YY, Chen XS, Yang Y, Ramamurthy A, Li B, Qi Y, Song L. Efficient probabilistic logic reasoning with graph neural networks. In: Proc. of the 8th Int'l Conf. on Learning Representations (ICLR). Addis Ababa: OpenReview.net, 2020.
- [163] Yang F, Yang ZL, Cohen WW. Differentiable learning of logical rules for knowledge base reasoning. In: Proc. of the 31st Int'l Conf. on Neural Information Processing Systems (NIPS). Long Beach: Curran Associates Inc., 2017. 2316–2325.

#### 附中文参考文献:

- [98] 金婧, 万怀宇, 林友芳. 融合实体类别信息知识图谱表示学习方法. 计算机工程, 2021, 47(4): 77–83. [doi: [10.19678/j.issn.1000-3428.0057353](https://doi.org/10.19678/j.issn.1000-3428.0057353)]
- [110] 杜文倩, 李弼程, 王瑞. 融合实体描述及类型的知识图谱表示学习方法. 中文信息学报, 2020, 34(7): 50–59. [doi: [10.3969/j.issn.1003-0077.2020.07.005](https://doi.org/10.3969/j.issn.1003-0077.2020.07.005)]



张天成(1969—), 男, 博士, 副教授, CCF 高级会员, 主要研究领域为教育大数据, 时空数据管理.



于明鹤(1989—), 女, 博士, 讲师, CCF 专业会员, 主要研究领域为数据库, 信息检索.



田雪(1998—), 女, 硕士生, CCF 学生会会员, 主要研究领域为数据挖掘, 知识发现, 知识图谱.



孙艳红(1997—), 女, 硕士生, CCF 学生会会员, 主要研究领域为教育大数据, 知识图谱.



孙相会(1997—), 男, 硕士生, CCF 学生会会员, 主要研究领域为自然语言理解, 表示学习.



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库理论与技术, 区块链.