

# 华中科技大学

## 研究生（文献综述与选题）报告

题目：知识图谱的链接预测研究

学	号	<u>M201973167</u>		
姓	名	<u>高也</u>		
专	业	<u>计算机软件与理论</u>		
指	导	教	师	<u>何琨</u>
院（系、所）	<u>计算机科学与技术</u>			

华中科技大学研究生院制

## 填表注意事项

- 一、本表适用于攻读硕士学位研究生选题报告、学术报告，攻读专业硕士学位研究生实践环节报告，攻读博士学位研究生文献综述、选题报告、论文中期进展报告、学术报告等。
- 二、以上各报告内容及要求由相关院（系、所）做具体要求。
- 三、以上各报告均须存入研究生个人学籍档案。
- 四、本表填写要求文句通顺、内容明确、字迹工整。

一、 课题的来源、目的、意义、国内外研究现状

二、 预计需达到的要求、技术指标，预计的技术关键、技术方案和主要实验研究情况

三、 课题研究进展计划

# 一、 课题的来源、目的、意义、国内外研究现状

## 1.1 课题的来源

知识是对某个主题确信的认知，并且这些认知拥有潜在的能力为特定目的而使用。意指透过经验或联想，而能够熟悉进而了解某件事情；这种事实或状态就称为知识。另一方面，人的思维适合把有关系的事物联系在一起，以提高我们的认知能力。笛卡尔在《探寻真理的指导原则》的原则十一和原则十二都提到了这种关联的必要性：“在查看了若干单纯命题之后，要想从中得出其他推论的话，不妨以连续的毫不间断的思维运动把那些命题通观一遍，考虑它们相互之间的关系，也不防择出若干来尽可能清楚地全面加以构想——只有这样，我们的认识才可以更加确定得多，心灵的认识能力才可以大为提高”“最后，应该充分利用悟性、想象、感觉和记忆所提供的一切助力，或者用以清楚地单纯直观命题，或者用以恰当地比较所求事物与已认识事物，以便得知所求，或者用以发现那些应该彼此关联的事物，从而使人的奋勉努力之所及不致有所遗漏”。

1989 年，蒂姆·伯纳斯·李(Tim Berners-Lee)发明了万维网，把网页文本通过超链接的方式联系在一起。受 1960 年提出的语义网络(Semantic network)的启发,1998 年语义网(Semantic Web)的概念被提出，试图将万维网网页文本之间的链接改进为语义之间的链接。2006 年，蒂姆·伯纳斯·李提出语义网的本质是要建立开放数据之间的链接。2012 年，谷歌公司为了改进自己的搜索引擎，引入知识图谱的概念。可见，知识图谱的提出和网络技术的发展息息相关。

近年来知识图谱技术的发展和应用一直在迅猛发展。许多知识图谱项目，比如 Freebase[1]，DBpedia[2]，YAGO[3]和 NELL[4]已经在现实生活中有了很多应用，比如语义划分(semantic parsing)[5][6]，命名实体识别(named entity disambiguation)[7][8]，信息抽取(information extraction)[9][10]，智能问答(question answering)[11][12]。一个知识图谱是由许多实体(节点)和关系(各种类型的边)组成的多重关系图。每一条边用一个三元组的形式表示(头实体 h，关系 r，尾实体 t)，意义为两个实体被一种关系联系起来，比如(姚明，夫妻关系，叶莉)。

## 1.2 课题的目的与意义

知识图谱中，知识以三元组的形式表示，形式为(主语 subject，谓语 predicate，宾语 object)或者(头实体 head，关系 relation，尾实体 tail)。比如，“姚明的妻子是叶莉”表示为(姚明，夫妻关系，叶莉)，“钱玄同的儿子是钱三强”被表示为(钱玄同，父子关系，钱三强)。三元组意味着知识图谱可以用图(graph)的形式来表示，其中头实体和尾实体用图中的节点来表示，关系用图中连接头实体节点和尾实体节点的边来表示。

由于知识图谱中经常存在链接缺失，所以我们需要经常进行知识图谱中链接的预测(link prediction)和补全任务。为了解决这个问题，一种被称为知识图谱嵌入(knowledge

graph embedding)的技术被提了出来并获得了广泛的关注[13][14][15][16][17][18][19]。它的核心思想是把知识图谱中的实体(节点)和关系(边)映射到连续的向量空间中,以便简化运算同时保留知识图谱原有的特征。这些实体和关系的嵌入可以进一步用在其他任务中,比如知识图谱的补全(KG completion)[14][15],关系抽取(relation extraction)[20][21],实体分类(entity classification)[13][22],实体解析(entity resolution)[13][18]。

### 1.3 国内外研究现状

目前绝大多数知识图谱嵌入技术基于观测到的事实(fact)。给定一个知识图谱,这些方法首先把实体和关系在一个连续向量空间表示出来,然后定义一个评分函数(scoring function)来计算每一个事实的可信度。接下来通过使所有事实的评分函数之和最大化,就可以得到实体和关系在连续向量空间中的嵌入。

假设我们给定一个知识图谱包含  $n$  个实体和  $m$  个关系。知识图谱中观测到的事实被储存在一个三元组集合  $D^+ = \{(h, r, t)\}$ 。每个三元组由一个头实体  $h$ , 一个尾实体  $t$ , 一个关系  $r$  组成。知识图谱嵌入目的是把实体和关系嵌入到低维连续向量空间中,以便简化知识图谱的运算。

一个典型的知识图谱嵌入一般分为三步:(i)表示实体和关系;(ii)定义评分函数;(iii)训练实体和关系表示。第一步用来定义实体和关系在连续向量空间中的表示形式。实体一般表示为向量,比如,向量空间中确定的点[13][14][15][16][19]。近期工作[34]进一步考虑了实体的不确定性,使用多变量高斯分布来模拟它们。关系一般被视为向量空间的运算,它可以表示成向量[14][15]、矩阵[16][18]、张量[19]、多变量高斯分布[35]。在第二步,对每个事实  $(h, r, t)$  定义一个评分函数  $f_r(h, t)$ , 以衡量这个事实的可靠性。知识图谱中观测到的事实应该比没有观测到的拥有更高的评分。最后,通过训练得到实体和关系的嵌入坐标,使得所有事实的评分函数之和尽可能大。

主流的嵌入模型大致可以分为两类:翻译距离模型和语义匹配模型。前者使用基于距离的评分函数,后者使用基于相似度的评分函数。下面我们将介绍几种常见的知识图谱嵌入模型。

#### 1.3.1 翻译距离模型

TransE 模型[14]是最有代表性的翻译模型。它把实体和关系都表示成相同向量空间的向量,比如  $R^d$ 。给定一个事实  $(h, r, t)$ , 关系被解释成一个翻译向量  $\mathbf{r}$ , 使  $\mathbf{h} + \mathbf{r} \approx \mathbf{t}$ , 如图 1 所示。评分函数被定义为  $\mathbf{h} + \mathbf{r}$  和  $\mathbf{t}$  的距离。比如,

$$f_r(h, t) = -\|\mathbf{h} + \mathbf{r} - \mathbf{t}\|_{l_2}$$

这个评分函数的值需要被优化得尽可能大。

虽然 TransE 函数非常简单有效，在处理一对多、多对一、多对多关系[15][16]时非常乏力。以一对多关系为例。给定一个关系  $r$  和头实体  $h$ ，如果  $h + r \approx t_i$ ，则  $t_1 = t_2 = \dots = t_n$ 。这意味着，在一对多关系中，所有尾实体的嵌入坐标将非常接近，尽管它们的含义各不相同——这显然是我们不希望看到的。类似的问题同样存在于多对一关系和多对多关系中。

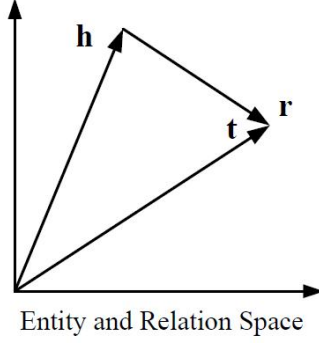


图 1. TransE 图示

为了解决 TransE 模型在处理一对多、多对一、多对多关系上的乏力，一种有效的策略是 让一个实体在不同的关系中有不同的表示。TransH 模型[15]就是在这种思想下提出来的，它引入了一个和关系有关的超平面，如图 2 所示。TransH 同样将实体建模成向量，但是把每一个关系  $r$  建模到一个超平面上，记  $w_r$  为这个超平面的法向量。给定一个事实  $(h, r, t)$ ，头实体  $h$  和尾实体  $t$  首先被投影到一个超平面上，

$$h_{\perp} = h - w_r^T h w_r, \quad t_{\perp} = t - w_r^T t w_r$$

在这个超平面上，头实体的投影  $h_{\perp}$  与关系  $r$  的和希望与尾实体的投影  $t_{\perp}$  尽可能接近，即  $h_{\perp} + r \approx t_{\perp}$ 。评分函数因此被定义为

$$f_r(h, t) = -\|h_{\perp} + r - t_{\perp}\|_2^2$$

可以看到，这个评分函数和 TransE 的评分函数非常相似。通过引入和关系相关的超平面机制，TransH 使得一个实体在不同的关系中有不同的投影位置。

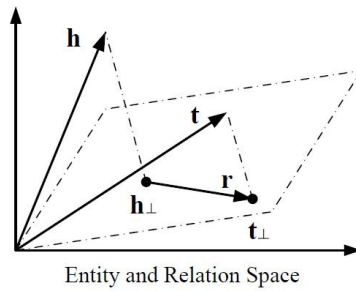


图 2. TransH 图示

TransR 模型[16]和 TransH 模型的思想相似，不过 TransR 的策略是引入和关系有关的空间，而不是 TransH 中的超平面。在 TransR 中，实体被表示成实体空间  $R^d$  中的向量，关系被表示成关系空间  $R^k$  中的翻译向量。给定一个事实  $(h, r, t)$ ，TransR 首先把头实体  $h$  和尾实体  $t$  投影到和关系  $r$  相关的关系空间中，

$$h_{\perp} = M_r h, \quad t_{\perp} = M_r t$$

在这里  $M_r \in R^{k \times d}$  是一个从实体空间到和  $r$  有关的关系空间的一个投影矩阵。类似的，我们定义评分函数

$$f_r(h, t) = -\|h_{\perp} + r - t_{\perp}\|_2^2$$

TransR 的机制如图 3 所示。虽然 TransR 在模拟复杂关系上十分有效，但是由于 TransR 为每一个关系引入了一个投影矩阵，所以对于每一个参数需要  $O(dk)$  个参数。这导致 TransR 失去了 TransE 和 TransH 的简洁性和高效性。TransR 还有更加复杂的版本[23][24]。在这些版本中，每个关系和两个矩阵相关联，一个投影头实体，另一个投影尾实体。

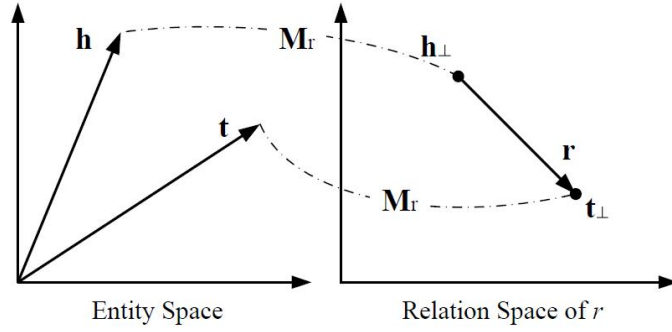


图 3. TransR 图示

TransD 模型[25]简化了 TransR 模型，通过将投影矩阵分解成两个向量。特别的，对于每个事实  $(h, r, t)$ ，TransD 模型引入投影向量  $w_h, w_t \in R^d$  且  $w_r \in R^k$ 。两个投影矩阵因此被定义为

$$M_r^1 = w_r w_h^T + I, M_r^2 = w_r w_t^T + I$$

这两个投影矩阵被相应地作用于头实体  $h$  和尾实体  $t$  以得到它们的投影，

$$h_{\perp} = M_r^1 h, t_{\perp} = M_r^2 t$$

评分函数的定义类似于之前的 TransR。TransD 需要  $O(nd+mk)$  个参数，且比 TransR 更有效率。

### 1.3.2 语义匹配模型

RESCAL 模型[13] (又被称为双线性模型[17]) 将每个实体和一个向量关联起来，以捕捉

它隐含的语义。每一个关系被表示成一个模拟成对关系的矩阵。一个事实  $(h, r, t)$  的分数被定义成一个双线性函数

$$f_r(h, t) = h^T M_r t$$

其中  $h, t \in R^d$  是实体的向量表示， $M_r \in R^{d \times d}$  是关联头实体  $h$  和尾实体  $t$  的向量。这个分数捕捉了所有头实体  $h$  和尾实体  $t$  的成对的关系，其中每个关系需要  $O(d^2)$  个参数。RESCAL 模型的图示如图 4 所示。

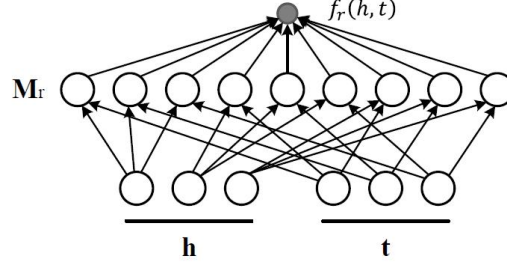


图 4. RESCAL 模型图示

DistMult 模型[26]简化了 RESCAL，通过将  $M_r$  限定为对角矩阵。对于每一个关系  $r$ ，该模型引入一个嵌入向量  $r \in R^d$  且要求  $M_r = \text{diag}(r)$ 。评分函数因此被定义为

$$f_r(h, t) = h^T \text{diag}(r) t$$

这个评分函数仅仅捕捉头实体  $h$  和尾实体  $t$  在相同维度上的关系，因此将参数数量减少到每个关系  $O(d)$  个参数 (DistMult 模型的图示如图 5 所示)。然而，由于  $h^T \text{diag}(r) t = t^T \text{diag}(r) h$ ，这个过度简化的模型仅仅能处理对称关系，这无法适用于一般的知识图谱。

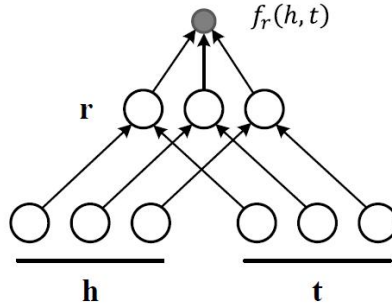


图 5. DistMult 模型图示

全息嵌入 (Holographic Embeddings, HoIE) 模型。HoIE 模型[27]结合了 RESCAL 模型表达力的优势和 DistMult 模型的简洁性。它把实体和关系都用向量空间  $R^d$  中的向量表示。给定一个事实  $(h, r, t)$ ，实体表示首先被组合成  $h * t \in R^d$  通过使用以下运算[27]，即

$$[h * t]_i = \sum_{k=0}^{d-1} [h]_k \cdot [t]_{(k+i) \bmod d}$$

这个组合而成的向量接下来和关系匹配，以此来计算这个事实的评分，

$$f_r(h, t) = r^T (h * t)$$

$h * t$  这个运算将成对的实体做了一个压缩 (如图 6 所示)。所以 HoIE 模型每个关系仅仅需要  $O(d)$  个参数，这比 RESCAL 模型速度快很多。同时，由于  $h * t \neq t * h$ ，不具有传递性，所以 HoIE 模型能够和 RESCAL 模型一样模拟非对称关系。



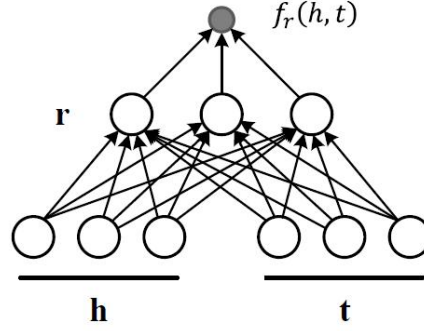


图 6. Ho1E 图示

复数嵌入模型(Complex Embeddings, ComplEx)。ComplEx 模型[29]扩展了 DistMult 模型, 通过引入复数嵌入以便更好地模拟非对称关系。在 ComplEx 模型中, 实体和关系嵌入不在实数空间而在复数空间里, 比如  $C^d$ 。一个事实的分数被定义为

$$f_r(h, t) = \text{Re}(h^T \text{diag}(r) \bar{t})$$

其中  $\bar{t}$  是  $t$  的共轭,  $\text{Re}(\cdot)$  是取一个复数值的实部。这个评分函数是非对称的, 所以非对称关系的事实交换头实体和尾实体可以得到不同的分数。[27]中提到每一个 ComplEx 模型都有一个等价的 Ho1E 模型, 反过来说, Ho1E 模型可以被视为 ComplEx 模型的一种特殊情况。

语义匹配能量模型(Semantic Matching Energy, SME)。SME 模型[18]使用神经网络结构来计算语义匹配。给定一个事实  $(h, r, t)$ , 它首先将实体和关系投影到输入层的嵌入向量上。在隐藏层, 关系  $r$  和头实体  $h$  结合得到  $g_u(h, r)$ , 和尾实体  $t$  结合得到  $g_v(h, r)$ 。一个事实的评分被定义为  $g_u$  和  $g_v$  的内积,

$$f_r(h, t) = g_u(h, r)^T g_v(t, r)$$

SME 模型有两个版本, 一个线性的版本和一个双线性的版本。线性 SME 定义为

$$\begin{aligned} g_u(h, r) &= M_u^1 h + M_u^2 r + b_u \\ g_v(t, r) &= M_v^1 t + M_v^2 r + b_v \end{aligned}$$

双线性 SME 定义为

$$\begin{aligned} g_u(h, r) &= (M_u^1 h) \circ (M_u^2 r) + b_u \\ g_v(t, r) &= (M_v^1 t) \circ (M_v^2 r) + b_v \end{aligned}$$

其中,  $M_u^1, M_u^2, M_v^1, M_v^2 \in R^{d \times d}$  是权重矩阵,  $b_u, b_v \in R^d$  是所有关系共享的偏差向量。图 7 是 SME 模型的图示。

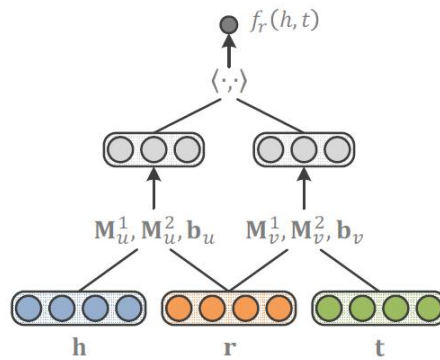


图 7. SME 模型的图示

神经张量网络(Neural Tensor Network, NTN)模型。NTN 模型[19]是另一个神经网络结

构。给定一个事实，它首先将实体投影到输入层的嵌入向量。然后，头实体和尾实体  $h, t \in R^d$  与和关系相关的张量  $\underline{M}_r \in R^{d \times d \times k}$  结合，接下来映射到一个非线性隐藏层中。最后，和关系相关的线性输出层给出以下评分函数

$$f_r(h, t) = r^T \tanh(h^T \underline{M}_r t + M_r^1 h + M_r^2 t + b_r)$$

其中  $M_r^1, M_r^2 \in R^{k \times d}$  是和关系相关的权重矩阵， $b_r \in R^k$  是和关系相关的偏差向量。双线性张量积  $h^T \underline{M}_r t$  可以得出一个向量。NTN 模式的图示如图 8 所示。如果设定  $\underline{M}_r = 0$  和  $b_r = 0$ ，NTN 模型可以退化成一个单层模型。NTN 模型或许是当前最具有表达力的模型，但是它每个关系需要  $O(d^2 k)$  个参数，所以当处理大型知识图谱时不够简单和快速。

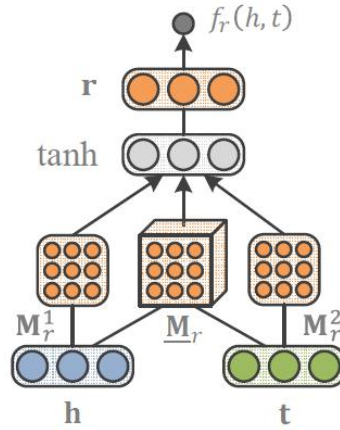


图 8. NTN 模型的图示

多层感知机模型(Multi-Layer Perceptron, MLP)。MLP 模型[30]是一个简化的方案，其中每一个关系和实体都和一个向量相关联。MLP 模型的图示如图 9 所示。给定一个事实  $(h, r, t)$ ， $h, r, t$  的嵌入向量在输入层被连接起来，然后映射到非线性的隐藏层中。评分由线性输出层产生，即

$$f_r(h, t) = w^T \tanh(M^1 h + M^2 r + M^3 t)$$

其中， $M^1, M^2, M^3 \in R^{d \times d}$  是第一层的权重， $w \in R^d$  是第二层的权重，这些都是被所有关系共享的。

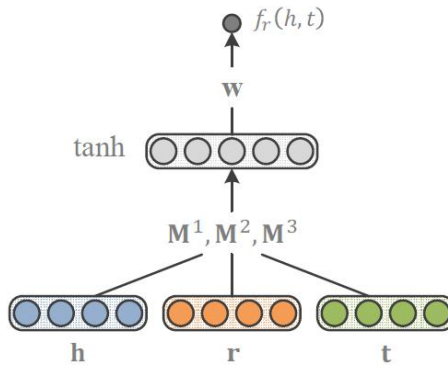


图 9. MLP 模型的图示

## 二、预计需达到的要求、技术指标，预计的技术关键、技术方案和主要实验研究情况

### 2.1 预计需达到的要求和技术指标

知识图谱的链接预测是指预测某个实体是否和另外一个实体具有某种关系。比如，给定关系和尾实体  $(r, t)$ ，预测某个头实体  $h$  是否属于这个三元组；或者给定关系和头实体  $(h, r)$ ，预测某个尾实体  $t$  是否属于这个三元组。链接预测在知识图谱补全任务中非常重要。链接预测有时也被称为实体预测[31]或者实体排序[18]。类似的思想也可用于预测两个实体是否具有某种关系，即  $(h, ?, t)$ ，这通常被称为知识图谱的关系预测[32][33]。

由于我们之前已经训练出了  $(h, r, t)$  的数值并且设计好了评分函数  $f_r(h, t)$ ，链接预测仅仅通过一个排序过程就可以实现。比如我们打算预测一个三元组中欠缺的头实体  $(?, r, t)$ 。为了预测头实体，我们把知识图谱中的所有头实体  $h'$  作为候选答案，然后为每一个  $(h', r, t)$  计算评分  $f_r(h', t)$ 。把这些分数按照降序评分可以得到包含所有候选答案的排序表。欠缺的尾实体和欠缺的关系也可以用类似的方法预测。

为了检验链接预测的有效性，一种常见的方法是记录这些排序表中的正确答案，检测这些正确答案是否排在错误答案的前面。正确答案的排序越高，说明模型的效果越好。常用的检测指标有正确答案的平均排序(mean rank)，正确答案的平均倒数排序(mean reciprocal rank)，正确答案在前  $n$  名的命中率(Hits@n)等等。我们希望本课题在相同数据集下，至少能达到 TransE 的技术指标。

### 2.2 预计的技术关键、技术方案

对于静态知识图谱嵌入的建模目前已经较为完善和全面了。预计的技术关键可能在于知识推理方面。我们接下来将考虑能否将知识推理与现有知识图谱嵌入模型结合起来。

### 2.3 主要实验研究情况

目前尚且处于论文调研阶段。实验预计在 2021 年 1 月 1 日开始做。

### 三、 课题研究进展计划

(1) 2020.10.16 – 2020.12.31 广泛阅读国内外相关论文，对其中的算法进一步深入研究，领悟它们的思想实质，并发现它们之间的关联。

(2) 2021.1.1 – 2021.5.31 设计一系列技术方案，在理论上论证其可行性，筛选出可实施的有效技术方案。

(3) 2021.6.1 – 2021.9.30 写代码实现之前设计的技术方案，做实验检验其结果，同时尝试调整参数优化结果。

(4) 2021.10.1 – 2021.12.31 与国内外最新研究的研究结果进行对比，发现不足并进一步尝试改进。

(5) 2022.1.1 – 2022.6.30 撰写硕士学位论文，并完成硕士学位论文的答辩。

#### 四、参考文献

- [1] K. Bollacker, C. Evans, P. Paritosh, T. Sturge, and J. Taylor, “Freebase: A collaboratively created graph database for structuring human knowledge,” in Proc. of ACM SIGMOD Int. Conf. On Manage. Data, 2008, pp. 1247 – 1250.
- [2] J. Lehmann, R. Isele, M. Jakob, A. Jentzsch, D. Kontokostas, P. N. Mendes, S. Hellmann, M. Morsey, P. van Kleef, S. Auer et al., “DBpedia: A large-scale, multilingual knowledge base extracted from Wikipedia,” Semantic Web J., vol. 6, no. 2, pp. 167 – 195, 2015.
- [3] F. M. Suchanek, G. Kasneci, and G. Weikum, “YAGO: A core of semantic knowledge,” in Proc. 16th Int. Conf. on World Wide Web, 2007, pp. 697 – 706.
- [4] A. Carlson, J. Betteridge, B. Kisiel, B. Settles, E. R. Hruschka Jr, and T. M. Mitchell, “Toward an architecture for never-ending language learning,” in Proc. 24th AAAI Conf. Artif. Intell., 2010, pp. 1306 – 1313.
- [5] J. Berant, A. Chou, R. Frostig, and P. Liang, “Semantic parsing on Freebase from question-answer pairs,” in Proc. Conf. Empirical Methods Natural Language Process., 2013, pp. 1533 – 1544.
- [6] L. P. Heck, D. Hakkani-Tur, and G. Tur, “Leveraging knowledge graphs for web-scale unsupervised semantic parsing,” in INTERSPEECH, 2013, pp. 1594 – 1598.
- [7] D. Damjanovic and K. Bontcheva, “Named entity disambiguation using linked data,” in Proc. 9th Extended Semantic Web Conf., 2012.
- [8] Z. Zheng, X. Si, F. Li, E. Y. Chang, and X. Zhu, “Entity disambiguation with Freebase,” in Proc. IEEE/WIC/ACM Int. Joint Conf. Web Intell. Intell. Agent Technol., 2012, pp. 82 – 89.
- [9] R. Hoffmann, C. Zhang, X. Ling, L. Zettlemoyer, and D. S. Weld, “Knowledge-based weak supervision for information extraction of overlapping relations,” in Proc. 49th Annu. Meeting Assoc. Comput. Linguistics, 2011, pp. 541 – 550.
- [10] J. Daiber, M. Jakob, C. Hokamp, and P. N. Mendes, “Improving efficiency and accuracy in multilingual entity extraction,” in Proc. 9th Int. Conf. Semantic Syst., 2013, pp. 121 – 124.
- [11] A. Bordes, J. Weston, and N. Usunier, “Open question answering with weakly supervised embedding models,” in Proc. Joint Eur. Conf. Mach. Learn. Knowl. Discovery Databases, 2014, pp. 165 – 180.

- [12] A. Bordes, S. Chopra, and J. Weston, “Question answering with subgraph embeddings,” in *Proc. Conf. Empirical Methods Natural Language Process.*, 2014, pp. 615 – 620.
- [13] M. Nickel, V. Tresp, and H.-P. Kriegel, “A three-way model for collective learning on multi-relational data,” in *Proc. 28th Int. Conf. Mach. Learn.*, 2011, pp. 809 – 816.
- [14] A. Bordes, N. Usunier, A. Garc’ia-Dur’an, J. Weston, and O. Yakhnenko, “Translating embeddings for modeling multi-relational data,” in *Adv. Neural Inf. Process. Syst.*, 2013, pp. 2787 – 2795.
- [15] Z. Wang, J. Zhang, J. Feng, and Z. Chen, “Knowledge graph embedding by translating on hyperplanes,” in *Proc. 28th AAAI Conf. Artif. Intell.*, 2014, pp. 1112 – 1119.
- [16] Y. Lin, Z. Liu, M. Sun, Y. Liu, and X. Zhu, “Learning entity and relation embeddings for knowledge graph completion,” in *Proc. 29th AAAI Conf. Artif. Intell.*, 2015, pp. 2181 – 2187.
- [17] R. Jenatton, N. L. Roux, A. Bordes, and G. R. Obozinski, “A latent factor model for highly multi-relational data,” in *Adv. Neural Inf. Process. Syst.*, 2012, pp. 3167 – 3175.
- [18] A. Bordes, X. Glorot, J. Weston, and Y. Bengio, “A semantic matching energy function for learning with multi-relational data,” *Mach. Learn.*, vol. 94, no. 2, pp. 233 – 259, 2014.
- [19] R. Socher, D. Chen, C. D. Manning, and A. Y. Ng, “Reasoning with neural tensor networks for knowledge base completion,” in *Adv. Neural Inf. Process. Syst.*, 2013, pp. 926 – 934.
- [20] J. Weston, A. Bordes, O. Yakhnenko, and N. Usunier, “Connecting language and knowledge bases with embedding models for relation extraction,” in *Proc. Conf. Empirical Methods Natural Language Process.*, 2013, pp. 1366 – 1371.
- [21] S. Riedel, L. Yao, A. McCallum, and B. M. Marlin, “Relation extraction with matrix factorization and universal schemas,” in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics: Human Language Technol.*, 2013, pp. 74 – 84.
- [22] M. Nickel, V. Tresp, and H.-P. Kriegel, “Factorizing YAGO: Scalable machine learning for linked data,” in *Proc. 21st Int. Conf. World Wide Web*, 2012, pp. 271 – 280.
- [23] H.-G. Yoon, H.-J. Song, S.-B. Park, and S.-Y. Park, “A translation-based

knowledge graph embedding preserving logical property of relations,” in Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics: Human Language Technol., 2016, pp. 907 – 916.

[24] D. Q. Nguyen, K. Sirts, L. Qu, and M. Johnson, “STransE: A novel embedding model of entities and relationships in knowledge bases,” in Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics: Human Language Technol., 2016, pp. 460 – 466.

[25] G. Ji, S. He, L. Xu, K. Liu, and J. Zhao, “Knowledge graph embedding via dynamic mapping matrix,” in Proc. 53rd Annu. Meeting Assoc. Comput. Linguistics 7th Int. Joint Conf. Natural Language Process., 2015, pp. 687 – 696.

[26] B. Yang, W.-t. Yih, X. He, J. Gao, and L. Deng, “Embedding entities and relations for learning and inference in knowledge bases,” in Proc. Int. Conf. Learn. Represent., 2015.

[27] K. Hayashi and M. Shimbo, “On the equivalence of holographic and complex embeddings for link prediction,” in arXiv: 1702.05563, 2017.

[28] T. A. Plate, “Holographic reduced representations,” IEEE Trans. Neural Netw., vol. 6, no. 3, pp. 623 – 641, 1995.

[29] T. Trouillon, J. Welbl, S. Riedel, E. Gaussier, and G. Bouchard, “Complex embeddings for simple link prediction,” in Proc. 33rd Int. Conf. Mach. Learn., 2016, pp. 2071 – 2080.

[30] X. Dong, E. Gabrilovich, G. Heitz, W. Horn, N. Lao, K. Murphy, T. Strohmann, S. Sun, and W. Zhang, “Knowledge vault: A webscale approach to probabilistic knowledge fusion,” in Proc. 20<sup>th</sup> ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2014, pp. 601 – 610.

[31] Y. Lin, Z. Liu, H. Luan, M. Sun, S. Rao, and S. Liu, “Modeling relation paths for representation learning of knowledge bases,” in Proc. Conf. Empirical Methods Natural Language Process., 2015, pp. 705 – 714.

[32] R. Xie, Z. Liu, and M. Sun, “Representation learning of knowledge graphs with hierarchical types,” in Proc. 25th Int. Joint Conf. Artif. Intell., 2016, pp. 2965 – 2971.

[33] Y. Lin, Z. Liu, H. Luan, M. Sun, S. Rao, and S. Liu, “Modeling relation paths for representation learning of knowledge bases,” in Proc. Conf. Empirical Methods Natural Language Process., 2015, pp. 705 – 714.

[34] K. Guu, J. Miller, and P. Liang, “Traversing knowledge graphs in vector space,” in Proc. Conf. Empirical Methods Natural Language Process., 2015, pp.

318 – 327.

[35] S. He, K. Liu, G. Ji, and J. Zhao, “Learning to represent knowledge graphs with Gaussian embedding,” in Proc. 24th ACM Int. Conf. Inf. Knowl. Manage., 2015, pp. 623 – 632.