



MIRA Book for Representation Learning

[draft 2020/11/25]

作者：张占秋、王学良、蔡健宇、陈佳俊、贺华瑞、邱哲睿、刘云飞

组织：MIRA Lab

目录

第一部分 总更新记录	1
第二部分 知识图谱	3
1 知识图谱概述 (内容待补充)	4
2 知识图谱基础 (内容待补充)	5
2.1 图与网络	5
2.2 概率图模型	6
2.3 自然语言处理	7
3 知识图谱构建	8
3.1 知识图谱构建概述 (内容待补充)	8
3.2 知识融合发展历史 (内容待补充)	9
3.3 实体链接 (内容待补充)	10
3.4 关系抽取 (张占秋-2020/11/17)	11
3.5 知识图谱构建前沿 (内容待补充)	16
4 知识图谱嵌入	17
4.1 知识图谱嵌入概述 (陈佳俊-2020/11/11)	17
4.2 知识表示发展历史 (内容待补充)	25
4.3 基于平移距离的知识图谱嵌入 (蔡健宇-2020/11/10)	26
4.4 基于张量分解的知识图谱嵌入 (张占秋-2020/11/17)	34
4.5 基于神经网络的知识图谱嵌入 (内容待补充)	41
4.6 知识图谱嵌入前沿 (内容待补充)	42
5 知识图谱推理 (内容待补充)	43
5.1 知识图谱推理概述	43
5.2 知识推理发展历史	44
5.3 基于规则的知识图谱推理	45
5.4 基于图结构的知识图谱推理	46
5.5 基于表示学习的知识图谱推理	47
5.6 知识图谱推理前沿	48

6 知识图谱应用（内容待补充）	49
6.1 基于知识图谱的问答	49
6.2 基于知识图谱的聊天机器人	50
6.3 结合知识图谱的推荐系统	51
6.4 结合知识图谱的信息检索	52

第一部分

总更新记录

2020/11/25 更新: 【NeurIPS 2020】Duality-Induced Regularizer for Tensor Factorization Based Knowledge Graph Completion 错误订正 (张占秋)

2020/11/17 更新: 将公众号文章《远程监督在关系抽取中的应用》以及【NeurIPS 2020】Duality-Induced Regularizer for Tensor Factorization Based Knowledge Graph Completion 更新至此书 (张占秋)

2020/11/11 更新: 将公众号文章《知识图谱嵌入模型简介》从本地 markdown 文件迁移至此处 (陈佳俊)

2020/11/11 更新: 梳理目录 (贺华瑞)

2020/11/10 更新: 将公众号文章《可建模语义分层的知识图谱补全方法》从本地 markdown 文件迁移至此处 (蔡健宇)

第二部分

知识图谱

第一章 知识图谱概述（内容待补充）

内容待补充。

第二章 知识图谱基础（内容待补充）

2.1 图与网络

内容待补充。

2.2 概率图模型

内容待补充。

2.3 自然语言处理

内容待补充。

第三章 知识图谱构建

3.1 知识图谱构建概述（内容待补充）

内容待补充。

3.2 知识融合发展历史（内容待补充）

内容待补充。

3.3 实体链接（内容待补充）

内容待补充。

3.4 关系抽取（张占秋-2020/11/17）

【公众号文章】远程监督在关系抽取中的应用

3.4.1 什么是关系抽取

知识就是力量。使人类及机器能够更好地利用知识是提升工作效率及实现人工智能的关键所在，也催生了包括知识表示、知识推理在内的一系列研究。作为一切对知识的利用的基础，我们首先要获取知识，即知识抽取。

与我们一般通过阅读书籍来获取知识一样，知识抽取的主要数据来源是文本（如新闻、小说、维基百科等）。通常来讲，我们可以直接从文本中获取的知识包含两类：实体 (entity) 和关系 (relation) [1]，这也对应了两类知识抽取任务：实体识别 (entity recognition) 和关系抽取 (relation extraction)。下表给出了一个从句子中进行知识抽取的示例。在本文中，我们将主要关注关系抽取。

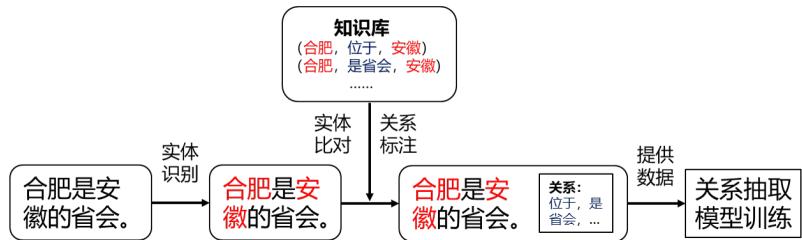
ID	文本	实体	关系
1	北京是中国的首都，具有悠久的历史。	北京，中国	首都
2	比尔盖茨是微软的创始人。	比尔盖茨，微软	创始人
3	合肥位于安徽中部，是安徽省会。	合肥，安徽	位于，是省会

顾名思义，关系抽取的目的就是从文本数据中抽取出关系。直观来讲，关系必然存在于两个（或多个实体）之间，比如在“合肥位于安徽中部”这一文本中，“位于”是“合肥”和“安徽”之间的关系，而撇开实体或者单看某单一实体时不会有关系这一概念的出现。从上表第三个例子又可以看出，“合肥”和“安徽”之间同时具有“位于”和“省会”两个关系，即一个文本中的实体对之间事实上可以存在多个不同的关系。因此，在不考虑发现未知的新关系时，我们可以将关系抽取定义成一个给定实体对情况下的多标签分类 (multi-label classification) 任务，其中的标签即为实体对之间的关系。

3.4.2 关系抽取的难点

从上一节对关系抽取任务的定义可以看出，要训练一个关系抽取模型，数据中应同时包含对实体对和关系的标注。实体的标注对应了实体识别任务，目前可以通过较为成熟的命名实体识别 (NER) 等技术来进行高质量自动标注。但关系的标注通常较为困难，一些句子中甚至不会显式的出现定义好的关系的相近描述。如“合肥是安徽的省会”一句中事实上是蕴含了“位于”这一关系，但句子中并没有显式地出现“位于”的相关描述，我们需要通过一些人类的常识进行推断。因此，在构建精确标注的关系抽取数据集时，大量人工标注通常是必不可少的，这就导致数据集的构建成本非常高昂。

此外，如 Mintz 等人指出，在特定领域的语料上进行关系标注而训练得到的关系抽取模型通常具有偏置 (bias) [2]。比如，发表于 ACL2020 的一篇文章 [3] 中对关系抽取中存在的性别偏见 (Gender Bias) 进行了讨论，感兴趣的同学可以去读一下这篇文章。



在上文中我们提到了仅利用给定语料上精确标注数据集进行关系抽取模型训练存在的两个问题，即标注成本高和存在偏置。那么该怎样解决这些问题呢？

3.4.2.1 标注成本高问题

如果我们继续采用监督学习 (supervised learning) 来对模型进行训练，那这个问题就很难回避。尽管我们可以通过少样本学习 (few shot learning) 等技术来充分利用已有的少量精确标注数据，但这些标注本身就更容易产生偏置，这就会导致第二个问题变得更加突出。

因此，要想从根本上缓解标注成本高问题，我们应该考虑放弃监督学习，转而采用弱监督、无监督，或者探索一种新的学习方式。

3.4.2.2 偏置问题

偏置问题产生的主要原因是在若干个特定领域的标注语料库上进行训练，因此解决偏置问题最简单粗暴的做法自然是选择覆盖面较广的数据。幸运地是，我们并不缺少这种数据：现在网络、书刊等媒体上充斥着大量文本，这些文本获取成本低且覆盖领域广，如果能够有效利用这些数据，那偏置问题将在很大程度上得到缓解。然而，这些数据又转而面临标注成本高问题。

3.4.3 远程监督

为同时解决以上两个问题，Mintz 等人于 2009 年提出了一种能够在未经精确标注的文本上进行有效学习的新的学习范式---远程监督。

3.4.3.1 远程监督的基本假设

如果一对实体之间具有某种关系，那么所有包含这对实体的句子都将具有这个关系的含义。

3.4.3.2 远程监督的基本流程

从上文介绍中可以看出，利用远程监督进行关系抽取模型的学习，我们只需要收集文本，在文本中识别实体对，然后与知识库中进行比对标注关系即可。其基本流程可用下例来表示。

很明显，远程监督方法极大地扩充了关系抽取能够利用的数据量，其标注成本极低，覆盖面可以极广。如果标注的每个句子都是正确的，那么问题就归结于一个相对较为简单的文本分类任务。

但是，细心的读者可能已经发现了，我们假定存在给定实体对的句子就能表示这对实体之间的所有关系，这势必会引入错误的标注。如“合肥”和“安徽”之间同时存在“位于”、“是省会”两个关系，那对于“合肥位于安徽”这一句子，我们也会同时给他标注“位于”和“是省会”两个关系。但很明显，这句话并没有表达“是省会”这一关系的含义。如果模型在这错误的标注上进行训练并将这个模式“记住”，即“ A 位于 B ”蕴含了 A 是 B 的省会，那在遇到“黄山位于安徽”这一句子时，它就会认为“黄山”和“安徽”之间也具有“是省会”这一关系，这明显是不合理的。因此，利用远程监督进行关系抽取的关键在于如何消除错误标注样本对模型训练的影响。

3.4.4 远程监督关系抽取模型

为了消除错误标注样本的影响，远程监督关系抽取模型主要采用了两种方法：错误标注样本筛除法和错误标注样本转正确标注法。

在本节中，我们将从上述两种方法出发简单介绍几个经典的模型。受篇幅限制，我们在本文中仅介绍这些模型的核心思路，技术细节请参照原文。因此类模型大都考虑句子级文本，在下文中涉及到数据的描述时我们不再对“句子”和“文本”进行区分。

3.4.4.1 错误标注样本筛除

PCNN

可以设想，在远程监督框架下，如果我们每次都将一个带标注的句子作为模型训练的样本，那甚至可以说训练过程中所使用的大部分样本都是被错误标注的，这将会极大地影响模型在实际应用中的效果。

因此，Zeng 等在 PCNN [4] 一文中提出将多实例学习 (multi-instance learning) 应用于远程监督中。其采用 expressed-at-least-once [5] 假设：

当一对实体之间存在某个关系，那么包含这对实体的句子中至少有一个能够表达这个关系。

在上述假设下，Zeng 等不再将带标注的句子逐个送入模型进行训练，而是将具有相同实体对和关系标注的所有（也可以是一部分，但直观上来讲越多越好）句子看成一个整体，称为包 (bag)，然后将标注的关系作为整个包的标签进行训练。如下图

这样一来，即便存在句子是被错误标注为了某关系的，但包中有这么多句子，我们总能找到一个句子是具有这个关系的吧！再退一步，就算包中所有句子都不表达所标注的关系，那结果也不会变得更坏了。因此从整体上来看，错误标注的几率被降低。

显然，接下来的工作就是如何从包中选择出那个被正确标注的样本来进行接下来的训练了。PCNN 一文中采取的方式是挑选使得条件概率 $p(\text{包的标签} | \text{句子})$ 最大的那个句子作为正确标注样本，具体实现细节在此不再赘述。

其他模型



PCNN 每次只选取包中一个句子作为正确标注样本的做法存在一个问题：如果包中有多个句子是被正确标注的，那么它们中的大部分都将被舍弃，从而造成了数据的浪费。为此，一系列基于注意力机制 (attention mechanism) 的模型被提出 [6,7,8]，它们通过对包中的句子进行注意力权重的分配来同时选择多个句子。若包中存在多个被正确标注的句子，那么它们的注意力权重都会较高，从而都会在最终关系预测中发挥作用；此外，被正确标注的置信度（即注意力权重）越高，发挥的作用将越大。这种软选择 (soft selection) 机制有效地缓解了 PCNN 中硬选择 (hard selection) 带来的数据浪费问题，从而可能在相同样本量的情况下达到更高的性能。

3.4.4.2 错误标注样本转正确标注

无论怎样进行样本的筛选，部分确定被错误标注的样本总是可以被看成在数据集中剔除掉了。那么，这些错误标注的样本真的对远程监督关系抽取不起作用吗？（此句借用了 Shang 等论文 [9] 的题目）至少在一些论文的作者看来，答案是否定的。

比如，Shang 等 [9] 利用无监督聚类的方法来为错误标注的样本重新分配新的标签；而 Wu 等 [10] 认为每个句子的现有标注和其应有标注之间存在映射关系，于是通过学习转移矩阵来模拟该映射，进而将每个句子的标注都转换成其应有标注。

通过将错误标注样本转换为正确标注，此类模型在一定程度上扩充了数据量，也为提升关系抽取的质量提供了新的可能性。但要注意的是，转换后的样本标注可能还是错的，有时仍需要进行进一步的筛选。

3.4.5 总结

在不考虑发现未知的新关系时，关系抽取本质上是一个多标签分类任务。但因任务的特殊性，关系抽取常面临标注数据不足等问题。远程监督的提出在一定程度上解决了

这些问题，但它同时也引入了错误标注样本。因此，如何在存在错误标注样本的数据上学习一个优秀的分类器成为远程监督关系抽取模型的关键。

- [1] Ji, S., Pan, S., Cambria, E., Marttinen, P., & Yu, P. S. (2020). A Survey on Knowledge Graphs: Representation, Acquisition and Applications. 1–26.
- [2] Mintz, M., Bills, S., Snow, R., & Jurafsky, D. (2009). Distant supervision for relation extraction without labeled data. Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2 - ACL-IJCNLP '09, 2(2005), 1003.
- [3] Gaut, A., Sun, T., Tang, S., Huang, Y., Qian, J., ElSherief, M., Zhao, J., Mirza, D., Belding, E., Chang, K.-W., & Wang, W. Y. (2020). Towards Understanding Gender Bias in Relation Extraction. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 1(1), 2943–2953.
- [4] Zeng, D., Liu, K., Chen, Y., & Zhao, J. (2015). Distant supervision for relation extraction via Piecewise Convolutional Neural Networks. *Conference Proceedings - EMNLP 2015: Conference on Empirical Methods in Natural Language Processing*, September, 1753–1762.
- [5] Riedel, S., Yao, L., & McCallum, A. (2010). Modeling relations and their mentions without labeled text. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 6323 LNAI(PART 3), 148–163.
- [6] Lin, Y., Shen, S., Liu, Z., Luan, H., & Sun, M. (2016). Neural relation extraction with selective attention over instances. *4th Annual Meeting of the Association for Computational Linguistics, ACL 2016 - Long Papers, 4, 2124–2133.
- [7] Han, X., Yu, P., Liu, Z., Sun, M., & Li, P. (2018). Hierarchical relation extraction with coarse-to-fine grained attention. Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, EMNLP 2018, 2236–2245.
- [8] Ye, Z.-X., & Ling, Z.-H. (2019). Distant Supervision Relation Extraction with Intra-Bag and Inter-Bag Attentions. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 1, 2810–2819.
- [9] Shang, Y., Huang, H.-Y., Mao, X.-L., Sun, X., & Wei, W. (2020). Are Noisy Sentences Useless for Distant Supervised Relation Extraction? Proceedings of the AAAI Conference on Artificial Intelligence, 34(05), 8799–8806.
- [10] Wu, S., Fan, K., & Zhang, Q. (2019). Improving Distantly Supervised Relation Extraction with Neural Noise Converter and Conditional Optimal Selector. Proceedings of the AAAI Conference on Artificial Intelligence, 33, 7273–7280.

3.5 知识图谱构建前沿（内容待补充）

内容待补充。

第四章 知识图谱嵌入

4.1 知识图谱嵌入概述 (陈佳俊-2020/11/11)

知识图谱 (Knowledge Graphs) 是大规模语义网络知识库，采取符号化的知识表示方式，利用三元组来描述具体的知识，并以有向图的形式对其进行表示和存储，具有语义丰富、结构友好、易于理解等优点。由于在表达人类先验知识上具有优良的特性，知识图谱近年来在自然语言处理、问答系统、推荐系统等诸多领域取得了广泛且成功的应用。

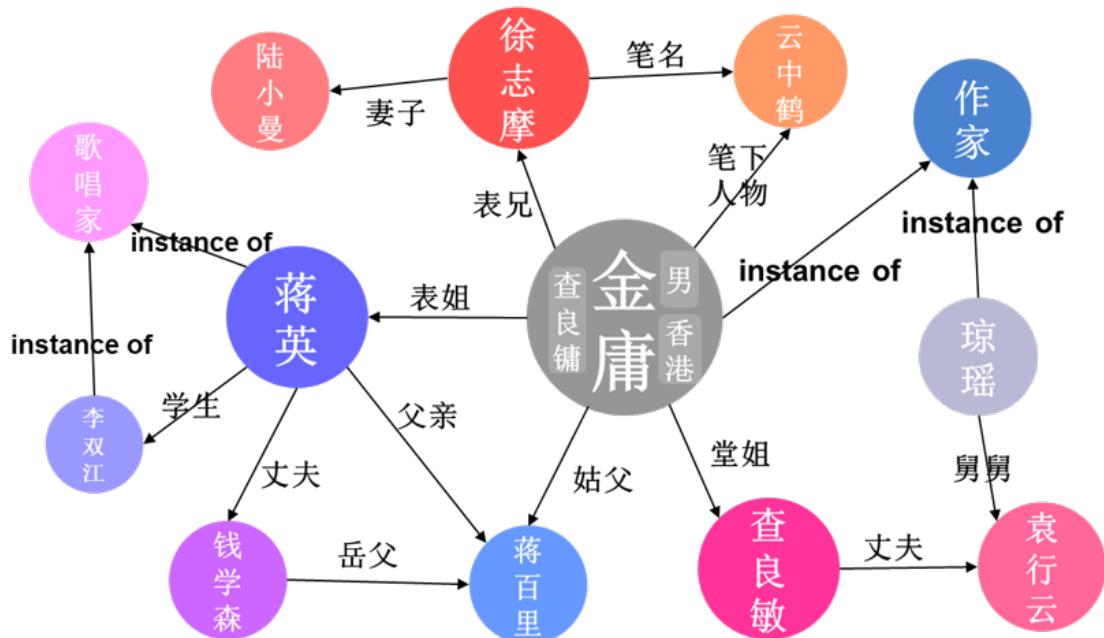


图 4.1: 知识图谱示意图

然而，知识图谱通常存在链接缺失问题，这限制了知识图谱在相关下游任务中的应用。为解决该问题，知识图谱补全任务应运而生。知识图谱补全旨在根据知识图谱中已有事实推断出新的事实，从而使得知识图谱更完整。

知识图谱嵌入 (Knowledge Graph Embedding) 是解决知识图谱补全问题的重要方法之一，它通过将知识图谱中的实体 (Entity) 和关系 (Relation) 嵌入到连续向量空间，从而在方便计算的同时保留知识图谱中的结构信息。知识图谱嵌入模型大致可以分为三类：

1. 基于距离的模型 (Distance-based Models)
2. 双线性模型 (Bilinear Models)
3. 神经网络模型 (Neural Network Models)

本文将首先简要介绍知识图谱补全任务，然后回顾总结各类知识图谱嵌入模型，最后就模型的统一评测进行补充说明。

4.1.1 知识图谱补全任务

4.1.1.1 符号定义

知识图谱通常以三元组 (Triple) 来表示知识，三元组的一般格式为 (头实体, 关系, 尾实体)，如“金庸的表兄是徐志摩”表示成三元组为 (金庸, 表兄, 徐志摩)，其中金庸为头实体，徐志摩为尾实体，表兄是这两个实体之间的关系。我们使用小写的字母 h, r, t 来分别表示头实体、关系和尾实体， (h, r, t) 表示知识图谱中的一个三元组。相应地，小写加粗字母 $\mathbf{h}, \mathbf{r}, \mathbf{t}$ 分别表示头实体、关系和尾实体对应的嵌入向量。向量 \mathbf{h} 的第 i 个元素记作 $[\mathbf{h}]_i$ 。 k 表示嵌入向量的维度。 $\circ : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}^n$ 记作两个向量间的哈达玛积 (Hadamard product)，其定义为：

$$[\mathbf{a} \circ \mathbf{b}]_i = [\mathbf{a}]_i \cdot [\mathbf{b}]_i$$

另外，我们使用 $\|\cdot\|_1, \|\cdot\|_2$ 来分别表示 ℓ_1 和 ℓ_2 范数。

4.1.1.2 补全任务与性能指标

知识图谱补全 (Knowledge Graph Completion)，是根据知识图谱中已有的三元组去对未知三元组进行预测。我们使用 $(h, r, ?)$ 表示已知头实体和关系来预测尾实体， $(?, r, t)$ 表示已知关系和尾实体来预测头实体，其中 $(h, r, ?)$ 和 $(?, r, t)$ 被称为查询 (Query)。实际上，知识图谱补全任务，本质上给定查询下的排序问题。比如，对于图 2 所示的查询 (陆小曼, 丈夫, ?)，返回的是预测答案的排序列表，正确答案排名越高，预测的效果就越好。

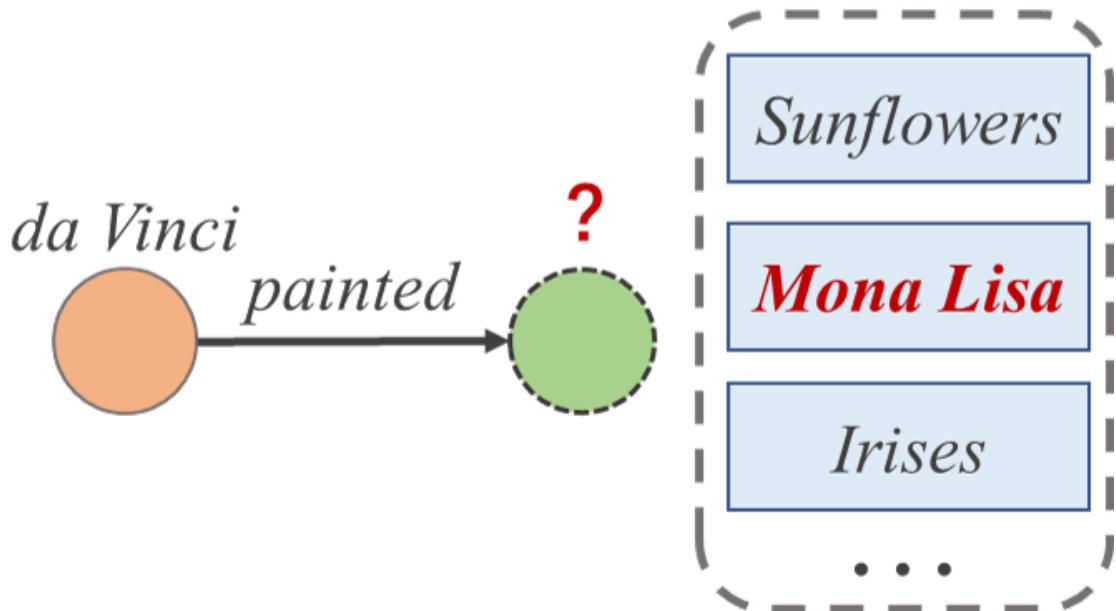


图 4.2: 查询示意图

对于知识图谱补全任务，常用的性能评测指标有：MR (Mean Rank)，MRR (Mean Reciprocal Rank)，和 Hits@N。

1. MR 是针对一系列查询 Q 返回排名的平均值:

$$\text{MR} = \frac{1}{|Q|} \sum_{i=1}^{|Q|} \text{rank}_i$$

2. MRR 是针对一系列查询 Q 返回排名取倒数之后的平均值:

$$\text{MRR} = \frac{1}{|Q|} \sum_{i=1}^{|Q|} \frac{1}{\text{rank}_i}$$

3. Hits@N 是排名在前 N 名所占的比例:

$$\text{Hits@N} = \frac{1}{|Q|} \sum_{i=1}^{|Q|} \mathbb{I}_{x \leq N}(\text{rank}_i)$$

其中, 如果 $\text{rank}_i \leq N$, $\mathbb{I}_{x \leq N}(\text{rank}_i) = 1$, 否则 $\mathbb{I}_{x \leq N}(\text{rank}_i) = 0$ 。

在一些实际应用中, MRR 能够比 MR 更好地反映排名的综合情况。例如, 在一次测试中共有 4 次查询, 模型 A 返回的排名结果序列为 [1,1,1,40001], 模型 B 返回的排名结果序列为 [10000,10000,10000,10000], 从 MR 的角度, 模型 A 效果略差于模型 B, 但从 MRR 的角度, 模型 A 的效果远好于模型 B。在实际应用中, 可能的场景是模型给出若干个最可能的结果供人选择, 我们可以发现排名 10000 的结果和排名 40001 的结果由于排名非常靠后, 很难出现在供人选择的若干结果中, 这两个结果虽然排名值相差 30001, 但是他们对实际应用的贡献都接近 0。从这个角度讲, 使用 MRR 来评价这两个模型更为合适。

4.1.1.3 知识图谱补全数据集

知识图谱补全任务上, 目前常用的数据集有 WN18RR, FB15k-237 和 YAGO3-10 等, 它们分别是 WN18, FB15k 和 YAGO3 的子集。WN18 和 FB15k 数据集早期被用于知识图谱补全任务, 但是后来有研究者发现这两个数据集存在一定的测试集泄漏问题, 在这两个数据集上, 简单的基于规则的方法就能够达到先进模型的效果, 所以 WN18 和 FB15k 渐渐不被研究者使用。表 1 显示了知识图谱补全数据集的具体统计信息, 其中数据集大小指的是三元组的数量。

数据集	实体数量	关系数量	训练集大小	验证集大小	测试集大小
FB15k	14,951	1,345	483,142	50,000	59,071
WN18	40,943	18	141,442	5,000	5,000
FB15k-237	14,541	237	272,115	17,535	20,466
WN18RR	40,943	11	86,835	3,034	3,134
YAGO3-10	123,182	37	1,079,040	5,000	5,000

图 4.3: 知识图谱补全数据集

4.1.2 知识图谱嵌入模型

知识图谱嵌入模型的设计通常需要三步：1) 定义实体和关系的表示形式；2) 定义衡量三元组合理性的打分函数；3) 训练学习实体和关系的嵌入表示 [13]。打分函数值越高，代表三元组的合理性越高，即正确的可能性越大。在训练学习实体和关系的嵌入表示时，优化目标是使得知识图谱中已有三元组得分尽可能比未出现的三元组得分要高。根据打分函数的定义形式，可以将知识图谱嵌入模型大致分为基于距离的模型、双线性模型和神经网络模型。部分知识图谱嵌入模型的打分函数对比见表 2。

Model	Score Function $f_r(\mathbf{h}, \mathbf{t})$	Parameters
TransE [2]	$-\ \mathbf{h} + \mathbf{r} - \mathbf{t}\ _{1/2}$	$\mathbf{h}, \mathbf{r}, \mathbf{t} \in \mathbb{R}^k$
TransR [3]	$-\ \mathbf{M}_r \mathbf{h} + \mathbf{r} - \mathbf{M}_t \mathbf{t}\ _2$	$\mathbf{h}, \mathbf{t} \in \mathbb{R}^d, \mathbf{r} \in \mathbb{R}^k, \mathbf{M}_r \in \mathbb{R}^{k \times d}$
RotatE [4]	$-\ \mathbf{h} \circ \mathbf{r} - \mathbf{t}\ _2$	$\mathbf{h}, \mathbf{r}, \mathbf{t} \in \mathbb{C}^k, \ [\mathbf{r}]_z\ = 1$
HAKE [1]	$-\ \mathbf{h}_m \circ \mathbf{r}_m - \mathbf{t}_m\ _2 - \lambda \ \sin((\mathbf{h}_p + \mathbf{r}_p - \mathbf{t}_p)/2)\ _1$	$\mathbf{h}_m, \mathbf{t}_m \in \mathbb{R}^k, \mathbf{r}_m \in \mathbb{R}_+^k, \mathbf{h}_p, \mathbf{r}_p, \mathbf{t}_p \in [0, 2\pi]^k, \lambda \in \mathbb{R}$
RESCAL [5]	$\mathbf{h}^\top \mathbf{M}_r \mathbf{t}$	$\mathbf{h}, \mathbf{t} \in \mathbb{R}^k, \mathbf{M}_r \in \mathbb{R}^{k \times k}$
DistMult [6]	$\mathbf{h}^\top \text{diag}(\mathbf{r}) \mathbf{t}$	$\mathbf{h}, \mathbf{r}, \mathbf{t} \in \mathbb{R}^k$
Hole [7]	$\mathbf{r}^\top (\mathbf{h} \star \mathbf{t})$	$\mathbf{h}, \mathbf{r}, \mathbf{t} \in \mathbb{R}^k$
ComplEx [8]	$\text{Re}(\mathbf{h}^\top \text{diag}(\mathbf{r}) \bar{\mathbf{t}})$	$\mathbf{h}, \mathbf{r}, \mathbf{t} \in \mathbb{C}^k$
MLP [9]	$\mathbf{w}^\top \tanh(\mathbf{M}^1 \mathbf{h} + \mathbf{M}^2 \mathbf{r} + \mathbf{M}^3 \mathbf{t})$	$\mathbf{w}, \mathbf{h}, \mathbf{r}, \mathbf{t} \in \mathbb{R}^k, \mathbf{M}^1, \mathbf{M}^2, \mathbf{M}^3 \in \mathbb{R}^{k \times k}$
ConvE [10]	$f(\text{vec}(f([\bar{\mathbf{r}}, \bar{\mathbf{h}}] * \omega)) \mathbf{W}) \mathbf{t}$	$\mathbf{h}, \mathbf{r}, \mathbf{t} \in \mathbb{R}^k$

图 4.4: 知识图谱嵌入模型打分函数对比。其中 \circ 表示哈达玛积 (Hadamard product), f 表示激活函数, \star 表示循环相关运算 (Circular Correlation Operation), $*$ 表示 2D 卷积操作, w 表示卷积层的滤波器 (Filter), $\bar{\cdot}$ 表示 ComplEx 模型中的复向量取共轭和 ConvE 模型中的实向量的 2D reshaping。

4.1.2.1 基于距离的模型

基于距离的模型将关系建模成从头实体到尾实体的距离变换，并通过变换后的距离差来定义打分函数。

TransE [2] 假设实体和关系满足 $\mathbf{h} + \mathbf{r} \approx \mathbf{t}$, 其中 $\mathbf{h}, \mathbf{r}, \mathbf{t} \in \mathbb{R}^k$, k 表示向量维度, 定义打分函数为：

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

其中, 我们使用 $\|\cdot\|_{1/2}$ 来代表在该打分函数中使用 ℓ_1 范数或者 ℓ_2 范数皆可。

然而, TransE 模型难以处理多对一、一对多、多对多问题。以一对多问题为例, 如存在两个三元组 $(h_1, r_1, t_1), (h_1, r_1, t_2)$, 在 TransE 模型下, 我们有 $\mathbf{h}_1 + \mathbf{r}_1 \approx \mathbf{t}_1, \mathbf{h}_1 + \mathbf{r}_1 \approx \mathbf{t}_2$, 由此可得 $\mathbf{t}_1 \approx \mathbf{t}_2$ 。这可能是不合理的, 例如, 我们有两个三元组为 (金庸, 作品, 倚天屠龙记) 和 (金庸, 作品, 天龙八部), 从 TransE 模型的角度, 要求倚天屠龙记和天龙八部这两个实体的嵌入向量非常相近, 然而倚天屠龙记和天龙八部显然是不同的作品。

通过让实体在不同关系下拥有不同的嵌入向量, 可以一定程度解决这个问题。TransR [3] 提出针对每个关系 r 设计一个投影矩阵 \mathbf{M}_r , 实体在不同关系下的表示为：

$$\mathbf{h}_\perp = \mathbf{M}_r \mathbf{h}, \quad \mathbf{t}_\perp = \mathbf{M}_t \mathbf{t}$$

通过这种方式，实体在不同关系下可以有不同的表示，最后的打分函数定义为：

$$f_r(\mathbf{h}, \mathbf{t}) = -\|\mathbf{h}_\perp + \mathbf{r} - \mathbf{t}_\perp\|_2 = -\|\mathbf{M}_r \mathbf{h} + \mathbf{r} - \mathbf{M}_t \mathbf{t}\|_2$$

其中， $\mathbf{h}, \mathbf{t} \in \mathbb{R}^d$, $\mathbf{r} \in \mathbb{R}^k$, $\mathbf{M}_r \in \mathbb{R}^{k \times d}$ 。

另一方面，为了能够更好地建模知识图谱中的关系模式，RotatE [4] 把关系定义成复空间中头实体到尾实体的旋转变换 (Rotation)，并假设实体和关系满足 $\mathbf{h} \circ \mathbf{r} \approx \mathbf{t}$ ，其中 $\mathbf{h}, \mathbf{r}, \mathbf{t} \in \mathbb{C}^k$, k 表示向量维度，RotatE 的打分函数为：

$$f_r(\mathbf{h}, \mathbf{t}) = -\|\mathbf{h} \circ \mathbf{r} - \mathbf{t}\|_2$$

最近，HAKE [1] 模型使用极坐标系建模知识图谱中普遍存在的语义分层现象，利用模长部分建模分属不同层级的实体，利用角度部分建模属于同一层级的实体，在主流数据集上的性能超越了现有基于距离的模型。

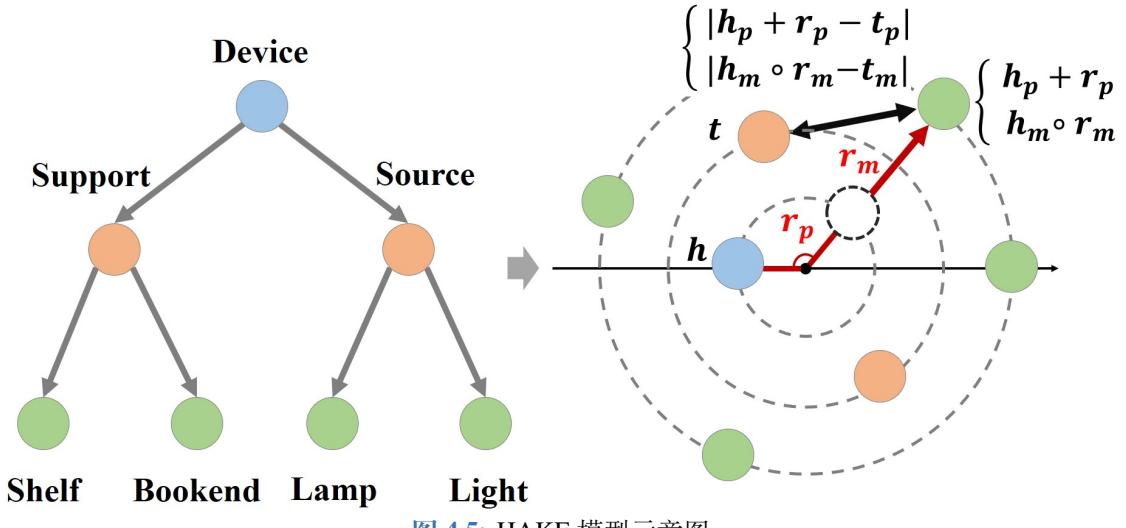


图 4.5: HAKE 模型示意图

4.1.2.2 双线性模型

双线性模型采用乘积形式的打分函数来衡量实体和关系的语义相关性。

RESCAL [5] 定义打分函数为：

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

可以看作是 $\mathbf{h}, \mathbf{t} \in \mathbb{R}^k$ 的双线性函数，其中矩阵 $\mathbf{M}_r \in \mathbb{R}^{k \times k}$ 。为了降低计算量，DistMult [6] 将矩阵 \mathbf{M}_r 设计成对角矩阵，HolE [7] 则使用循环相关运算 (Circular Correlation Operation) 来代替中间的关系矩阵乘积。这些方法计算方式的示意图见图 5。

ComplEx [8] 模型将 DistMult 模型扩展到复数空间，从而更好地建模反对称和可逆关系，它的打分函数为：

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

其中， $\mathbf{h}, \mathbf{r}, \mathbf{t} \in \mathbb{C}^k$, $\bar{\mathbf{t}}$ 代表对向量 \mathbf{t} 取共轭， $\operatorname{Re}(\cdot)$ 代表取实部。

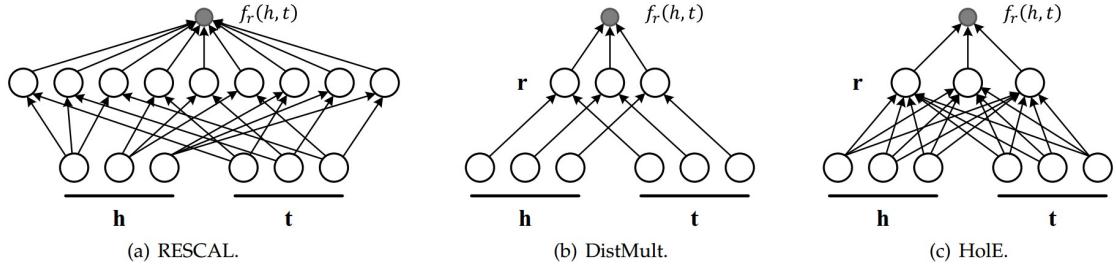


图 4.6: 双线性模型计算方式示意图

4.1.2.3 神经网络模型

神经网络模型将 $\mathbf{h}, \mathbf{r}, \mathbf{t}$ 同时输入神经网络, 来判断三元组的打分。早期的工作如 MLP [9] 使用全连接神经网络来对给定的三元组进行打分。近年来, ConvE [10] 等工作使用卷积神经网络 (Convolutional Neural Network) 来定义打分函数。由于知识图谱是一种图结构的数据, 近期也有一些工作将图神经网络 (Graph Neural Network) 应用于知识图谱嵌入。

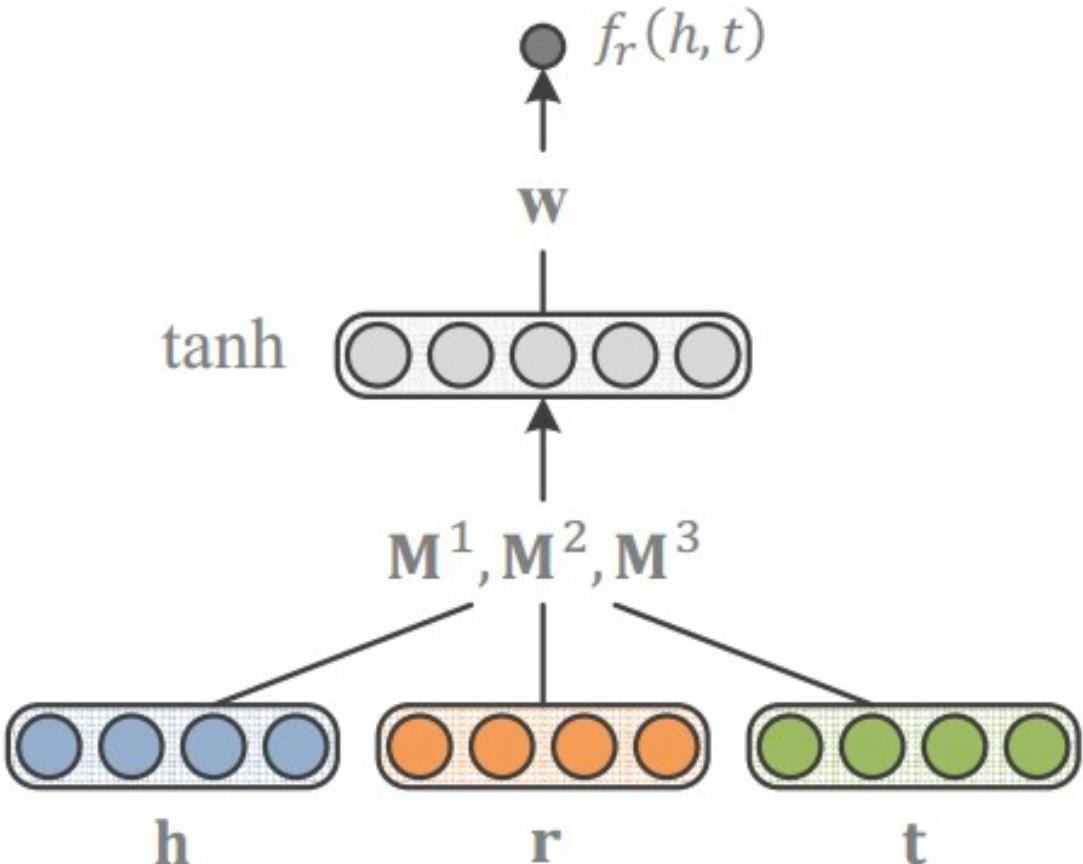


图 4.7: MLP 模型计算示意图

4.1.3 模型统一评测

近年来, 大量的知识图谱嵌入模型被提出, 各种方法层出不穷, 模型效果的实际对比有待研究。近期, 有两篇工作对现有的知识图谱嵌入模型进行了一些总结和重新评测, 揭示了各个模型性能在统一评测下的效果, 也为我们提供了对这些模型更为清晰的认识。

Zhiqing Sun 等人 [11] 指出，使用不恰当的评测方式，可能会造成实验结果虚高。在进行效果评测时，可能会有一些三元组得分和正确答案的得分相同，这时候针对这些相同得分的结果集合 \mathcal{T} 可以有不同的排序策略：

1. TOP：将正确答案排在 \mathcal{T} 的最前面
2. BOTTOM：将正确答案排在 \mathcal{T} 的最后面
3. RANDOM：给正确答案在 \mathcal{T} 中随机分配一个位置

值得注意的是，TOP 的排序策略在某些情况下是不恰当的，会导致结果虚高。一个直观的例子是，让一个模型对所有的候选三元组打分都输出 0，这时正确答案和所有候选三元组得分相同，在 TOP 排序策略下，该模型所有正确答案都排名第一，评测性能达到顶峰。但实际上，由于候选三元组得分相同，该模型在预测时无法判别哪个为正确答案。由此可见，TOP 的排序策略会导致结果虚高。文章指出，RANDOM 的排序策略是最为合理的 [11]，这和我们实际的认知也是相符合的，即给定多个相同得分的候选者，我们通常采用随机的方式来挑选。采用统一的 RANDOM 策略进行评测之后，可以发现有些模型尤其是部分神经网络模型的性能显著下降，这表明有些模型由于不恰当的评测方式，导致了结果的虚高，这是我们需要注意甄别的。

另外，Daniel Ruffinelli 等人 [12] 对各种知识图谱嵌入模型中使用到的训练方式进行了系统的实验和研究，发现一些早期的模型如 RESCAL 在调整训练方式和进行更大的超参搜索之后，也能够达到或者超过现有先进模型的效果。这也显示了模型性能的提升未必是方法具有很大的优势，训练技巧也可能在其中起到不小的作用，我们需要通过统一评测去认知各个方法中真正有价值的思路和想法。

4.1.4 总结

本文首先简要介绍了知识图谱补全任务，包括评测方式、评测指标和评测数据集。知识图谱嵌入是处理知识图谱补全任务的重要方法之一。本文按照基于距离的模型、双线性模型和神经网络模型的分类方式，对知识图谱嵌入模型进行简要介绍。最后，本文结合两篇最新工作，指出模型统一评测的重要性。限于篇幅原因，本文只进行了简要的总结，更多细节请参考相应的原论文，以下给出文中参考文献的详细列表。

4.1.5 未来展望

笔者认为知识图谱嵌入模型未来值得研究的方向可能包括：

1. 知识图谱嵌入模型的可解释性。知识图谱嵌入模型获取的嵌入向量中蕴含的模式难以被人类理解，具有黑盒特性。目前也有不少基于规则的知识图谱推理方法应用于补全任务，能够获取可解释的规则，但这些模型的实验性能大多不如先进的知识图谱嵌入模型。如何在保证性能的同时，提升知识图谱嵌入模型的可解释性，是一个值得研究的问题。
2. 知识图谱动态更新。真实场景下的知识是实时动态更新的，比如电商、金融等领域的知识图谱都需要进行实时更新来保证知识的实时性。在知识图谱需要动态更新

的场景下，如何快速地迁移知识图谱嵌入模型来适应新数据，是一个值得探讨的问题。

4.1.6 参考文献

- [1] Zhanqiu Zhang, Jianyu Cai, Yongdong Zhang, and Jie Wang. 2020. Learning Hierarchy-Aware Knowledge Graph Embeddings for Link Prediction. In AAAI.
- [2] Antoine Bordes, Nicolas Usunier, Alberto Garcia-Durán, Jason Weston, and Oksana Yakhnenko. 2013. Translating Embeddings for Modeling Multi-relational Data. In NIPS.
- [3] Yankai Lin, Zhiyuan Liu, Maosong Sun, Yang Liu, and Xuan Zhu. 2015. Learning Entity and Relation Embeddings for Knowledge Graph Completion. In AAAI.
- [4] Zhiqing Sun, Zhi-Hong Deng, Jian-Yun Nie, and Jian Tang. 2019. RotatE: Knowledge Graph Embedding by Relational Rotation in Complex Space. In ICLR.
- [5] Maximilian Nickel, Volker Tresp, and Hans-Peter Kriegel. 2011. A three-way model for collective learning on multi-relational data. In ICML.
- [6] Bishan Yang, Scott Wen-tau Yih, Xiaodong He, Jianfeng Gao, and Li Deng. 2015. Embedding Entities and Relations for Learning and Inference in Knowledge Bases. In ICLR.
- [7] Maximilian Nickel, Lorenzo Rosasco, and Tomaso Poggio. 2016. Holographic Embeddings of Knowledge Graphs. In AAAI.
- [8] Théo Trouillon, Johannes Welbl, Sebastian Riedel, Éric Gaussier, and Guillaume Bouchard. 2016. Complex Embeddings for Simple Link Prediction. In ICML.
- [9] Xin Luna Dong, Evgeniy Gabrilovich, Jeremy Heitz, Wilko Horn, Ni Lao, Kevin Murphy, Thomas Strohmann, Shaohua Sun, and Wei Zhang. 2014. Knowledge vault: A webscale approach to probabilistic knowledge fusion. In SIGKDD.
- [10] Tim Dettmers, Pasquale Minervini, Pontus Stenetorp, and Sebastian Riedel. 2018. Convolutional 2d knowledge graph embeddings. In AAAI.
- [11] Daniel Ruffinelli, Samuel Broscheit, and Rainer Gemulla. 2020. You CAN Teach an Old Dog New Tricks! On Training Knowledge Graph Embeddings. In ICLR.
- [12] Zhiqing Sun, Shikhar Vashishth, Soumya Sanyal, Partha Pratim Talukdar, and Yiming Yang. 2020. A Re-evaluation of Knowledge Graph Completion Methods. In ACL.
- [13] Quan Wang, Zhendong Mao, Bin Wang, and Li Guo. 2017. Knowledge Graph Embedding: A Survey of Approaches and Applications. In TKDE.

4.2 知识表示发展历史（内容待补充）

内容待补充。

4.3 基于平移距离的知识图谱嵌入（蔡健宇-2020/11/10）

【公众号文章】可建模语义分层的知识图谱补全方法

4.3.1 简介

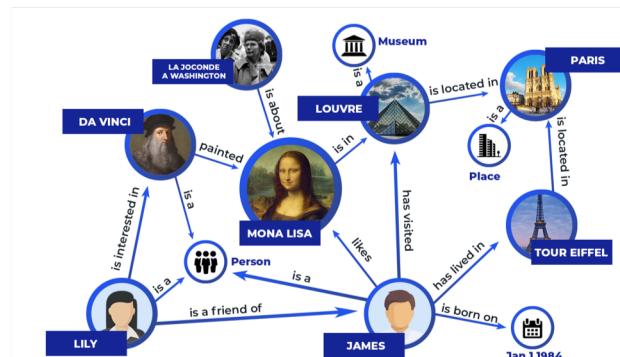
近些年，知识图谱 (Knowledge Graph) 在自然语言处理、问答系统、推荐系统等诸多领域取得了广泛且成功的应用。然而，现有知识图谱普遍存在链接缺失问题。为解决该问题，知识图谱补全任务应运而生。目前的知识图谱补全模型可分为多个流派，而基于距离的模型是其中重要一派。这类模型可以建模对称、互逆与复合等不同的抽象关系模式 (Relation Pattern)，却难以对知识图谱中普遍存在的语义分层 (Semantic Hierarchies) 现象进行有效建模。

本文接下来所介绍的工作 HAKE 使用极坐标系对语义分层现象进行建模，并在主流数据集上超越了这一流派中现有性能最好的方法。论文的标题为 Learning Hierarchy-Aware Knowledge Graph Embeddings for Link Prediction，发表于 AAAI 2020。论文代码已开源至 Github:<https://github.com/MIRALab-USTC/KGE-HAKE>。

4.3.2 知识图谱与补全任务

知识图谱是一种以图 (Graph) 的形式存储人类知识的大规模语义网络。知识图谱中的节点 (Node) 表示实体 (Entity)，边 (Edge) 表示关系 (Relation)。在知识图谱中，事实 (Fact) 一般以三元组 (Triple) 的形式进行表示。三元组一般记为 <头实体，关系，尾实体> (< head entity, relation, tail entity >)。

例如，下图中的 <达芬奇 (Da Vinci), 绘制 (painted), 蒙娜丽莎 (Mona Lisa) > 就是一个三元组。



然而，知识图谱却普遍存在着三元组缺失（即链接缺失）问题。因缺失三元组的规模通常较大，人工寻找这些三元组代价高昂。因此，研究者们设计了知识图谱补全任务 (Knowledge Graph Completion)，旨在通过机器学习模型自动补全缺失三元组。由于该任务并不会新增实体与关系，只是补全已有实体间的链接 (Link)，因此也被称为链接预测 (Link Prediction) 任务。

4.3.3 相关工作

现有知识图谱补全模型首先将实体与关系映射为低维嵌入（knowledge graph embeddings，可为向量/矩阵/张量），然后通过一个以低维嵌入为输入的打分模型为给定三元组进行打分，从而确定任意三元组真实存在的概率。

现有的模型主要分为以下三类：

- 基于距离的模型 (Distance-based Models)
- 双线性模型 (Bilinear Models)
- 神经网络模型 (Neural Network Models)

本文所介绍的 HAKE 模型属于“基于距离的模型”这一流派。

基于距离的模型将关系视为从头实体到尾实体的映射，并通过计算映射后的头实体与尾实体之间的距离来得到三元组的得分，代表性模型包括 TransE 与 RotatE 等。该流派之前的模型能够对一些抽象的关系模式进行建模，包括对称关系 (Symmetry)，互逆关系 (Inversion) 和复合 (Composition) 关系等。

TransE 将一个真实三元组 (h, r, t) 建模为 $h + r \approx t$ ，其中 $h, r, t \in \mathbb{R}^k$ ， k 表示向量维度，它可以建模互逆关系与复合关系，却难以建模对称关系。RotatE 将一个真实三元组 (h, r, t) 建模为 $h \circ r \approx t$ ，其中 $h, r, t \in \mathbb{C}^k$ ， $\circ : \mathbb{R}^k \times \mathbb{R}^k \rightarrow \mathbb{R}^k$ 表示向量间的 Hadamard 积，即 $[a \circ b]_i = [a]_i \cdot [b]_i$ ，它能够建模上述三种关系。

然而，这些模型难以建模在知识图谱中普遍存在的语义分层现象。

4.3.4 模型介绍

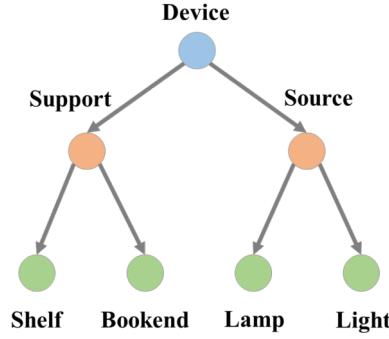
4.3.4.1 语义分层现象

语义分层 (Semantic Hierarchies) 现象在知识图谱中普遍存在。以下面两个三元组为例，

1. <棕榈树 (palm), 上位词 (_hypernym), 树 (tree)>
2. <兰开斯特 (Lancaster), 位于 (located_in), 英格兰 (England)>

实体 <树> 与 <英格兰> 更加抽象，属于更高的范畴，因而在语义分层中属于高层级的实体；实体 <棕榈树> 与 <兰开斯特> 更加具体，因而在语义分层中属于低层级的实体。进一步，我们发现知识图谱中的语义分层现象可以抽象为树型 (Tree) 结构。其中

1. 处于最高语义层级的实体对应着树的根节点
2. 拥有更高的语义层级的实体更加接近根节点
3. 语义层级更低的实体更加接近叶子节点
4. 处于相同语义层级的实体到根节点的距离相同



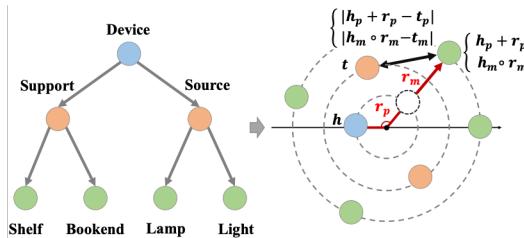
也就是说，树中一个节点的深度 (Depth) 反映了其对应实体的语义层级。基于以上观察，我们将知识图谱中的不同实体分为两类：

1. 分属不同语义层级的实体，例如 < 哺乳动物 (mammal)> 与 < 狗 (dog)>；< 奔跑 (run)> 与 < 移动 (move)>。
2. 属于相同语义层级的实体，例如 < 玫瑰 (rose)> 与 < 牡丹 (peony)>；< 卡车 (truck)> 与 < 货车 (lorry)>。

若要建模知识图谱的语义层级，一个模型必须能够建模上述两类实体。

4.3.4.2 HAKE 模型

HAKE (Hierarchy-Aware Knowledge Graph Embedding) 模型使用极坐标系建模上述表示语义层级的树型结构。



极坐标系 (Polar Coordinate System) 是一个二维坐标系统，包括半径坐标和角度坐标两部分。该坐标系统中任意位置可由一个半径 r 和一个角度 ρ 来表示。该模型将不同实体的嵌入向量的相同维都映射到同一个极坐标系中，包括模长和角度两部分。其中，

1. 模长部分用于建模分属不同层级的实体；
2. 角度部分用于建模属于同一层级的实体。

下面对这两部分进行详细介绍。

模长部分 受到前文所述树型结构的启发，我们使用极坐标系中的模长来表示树中节点的深度。即在语义层级中，我们使用模长部分来区分那些分属不同层级的实体。

对于一个三元组 (h, r, t) ， $\mathbf{h}_m, \mathbf{t}_m \in \mathbb{R}^k$ 与 $\mathbf{r}_m \in \mathbb{R}_+^k$ 分别为头实体、尾实体与关系向量，我们的建模如下：

$$\mathbf{h}_m \circ \mathbf{r}_m \approx \mathbf{t}_m,$$

对应的距离函数 (Distance Function) 为：

$$d_{r,m}(\mathbf{h}_m, \mathbf{t}_m) = \|\mathbf{h}_m \circ \mathbf{r}_m - \mathbf{t}_m\|_2$$

角度部分 位于同一个圆上的点（模长相同）可以有不同的夹角。受其启发，我们使用极坐标系中的角度来区分树中处于同一深度的不同节点。即，在语义层级中，我们使用角度部分来区分那些属于同一层级的不同实体。

对于一个三元组 (h, r, t) , $\mathbf{h}_p, \mathbf{r}_p, \mathbf{t}_p \in [0, 2\pi]^k$ 分别为头实体、关系与尾实体向量，我们的建模如下：

$$\mathbf{h}_p + \mathbf{r}_p \approx \mathbf{t}_p,$$

对应的距离函数 (Distance Function) 为：

$$d_{r,p}(\mathbf{h}_p, \mathbf{t}_p) = \|\sin((\mathbf{h}_p + \mathbf{r}_p - \mathbf{t}_p) / 2)\|_1$$

模长 + 角度 综合模长部分与角度部分，我们得到

$$\begin{cases} \mathbf{h}_m \circ \mathbf{r}_m \approx \mathbf{t}_m, \\ \mathbf{h}_p + \mathbf{r}_p \approx \mathbf{t}_p. \end{cases}$$

而 $\mathbf{h} = [\mathbf{h}_m, \mathbf{h}_p]$ 、 $\mathbf{r} = [\mathbf{r}_m, \mathbf{r}_p]$ 与 $\mathbf{t} = [\mathbf{t}_m, \mathbf{t}_p]$ 分别表示综合后的头实体、关系与尾实体向量。

综合后的距离函数为

$$d_r(\mathbf{h}, \mathbf{t}) = d_{r,m}(\mathbf{h}_m, \mathbf{t}_m) + \lambda d_{r,p}(\mathbf{h}_p, \mathbf{t}_p).$$

得分函数 (Score Function) 为

$$f_r(\mathbf{h}, \mathbf{t}) = -d_r(\mathbf{h}, \mathbf{t}).$$

基准模型 ModE 为了更加公平地进行对比实验，我们仅利用模长信息设计一个新的模型—ModE，它将作为基准模型之一与 HAKE 进行对比。

4.3.5 实验与分析

4.3.5.1 数据集

在本次实验中，我们选择在 WN18RR、FB15k-237 与 YAGO3-10 这三个数据集上进行实验。其中，

1. WN18RR，主要包含两种类型的关系：(a) 对称关系，如 `_similar_to`，该关系类型连接的头尾实体属于同一语义层级；(b) 非对称关系，如 `_hypernym`，该关系连接的头尾实体属于不同语义层级。

2. FB15k-237, 包含的关系数量最多, 关系类型最复杂, 仅有部分关系表示语义层级。
3. YAGO3-10, 该数据集和FB15k-237类似, 包含许多具有高入度(Indegree)的关系, 即在同一个关系下, 一个头/尾实体可能对应着大量的尾/头实体。例如, (, hasGender, male) 拥有超过 1,000 个正确的头实体, 但这些头实体的含义却可能相差悬殊, 因此建模难度更高。</li

4.3.5.2 实验结果

HAKE 与基准模型 ModE 在 WN18RR、FB15k-237 与 YAGO3-10 上的实验对比结果如下图所示。

	WN18RR				FB15k-237				YAGO3-10			
	MRR	H@1	H@3	H@10	MRR	H@1	H@3	H@10	MRR	H@1	H@3	H@10
TransE	.226	-	-	.501	.294	-	-	.465	-	-	-	-
DistMult	.43	.39	.44	.49	.241	.155	.263	.419	.34	.24	.38	.54
ConvE	.43	.40	.44	.52	.325	.237	.356	.501	.44	.35	.49	.62
ComplEx	.44	.41	.46	.51	.247	.158	.275	.428	.36	.26	.40	.55
RotatE	.476	.428	.492	.571	.338	.241	.375	.533	.495	.402	.550	.670
ModE	.472	.427	.486	.564	.341	.244	.380	.534	.510	.421	.562	.660
HAKE	.497	.452	.516	.582	.346	.250	.381	.542	.545	.462	.596	.694

我们可以看到

1. ModE 模型的参数量和 TransE 相同, 且模型同样简单, 但在各个数据集上的性能都远高于它;
2. HAKE 模型在各个数据集上的表现都显著优于现有的最佳模型。

4.3.5.3 可视化分析

关系的可视化分析

关系的模长部分 基于上文中关于树型结构的论述, 我们对于实体的嵌入向量的表现有如下的期望:

1. 位于更高语义层级的实体更加靠近树的根节点, 故模长更小;
2. 位于更低语义层级的实体更加靠近树的叶节点, 故模长更大。

由于关系连接着头尾实体, 我们对于关系的嵌入向量的表现有如下期望:

1. 如果头实体的语义层级更高, 而尾实体的语义层级更低, 那么我们期望

$$\mathbf{r}_m = \mathbf{t}_m / \mathbf{h}_m > 1$$

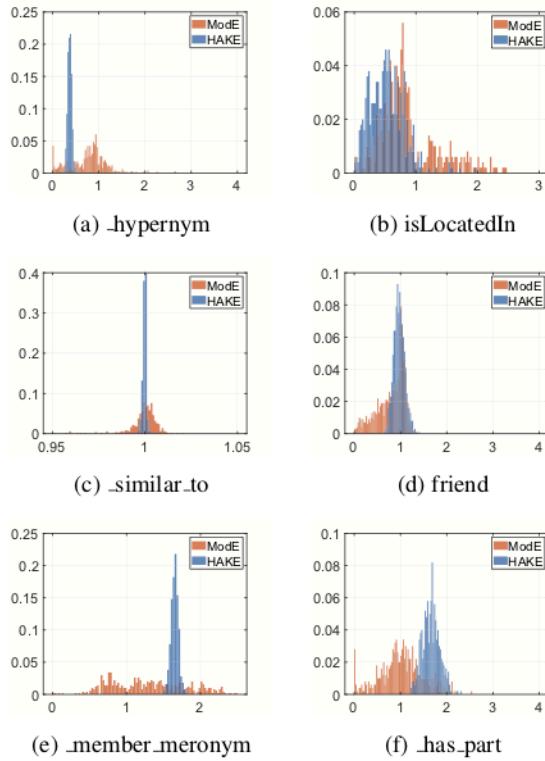
2. 如果头实体的语义层级更低, 而尾实体的语义层级更高, 那么我们期望

$$\mathbf{r}_m = \mathbf{t}_m / \mathbf{h}_m < 1$$

3. 如果头尾实体位于相同的语义层级, 那么我们期望

$$\mathbf{r}_m = \mathbf{t}_m / \mathbf{h}_m \approx 1$$

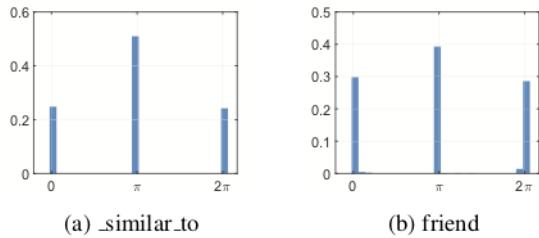
为了验证上述猜想，我们进行了如下实验分析。首先，我们从三个数据集中选取了一些代表性的关系。接着，对于这些关系在 ModE 与 HAKE 模型中的模长部分向量，我们绘制了每一维度的取值的分布直方图，如下图所示：



对于图中的六个关系，我们将其分为三组。图 (a) (b) 中的关系所连接的头实体的语义层级低于尾实体；图 (c) (d) 中的关系所连接的头尾实体的语义层级相同；图 (e) (f) 中的关系所连接的头实体的语义层级高于尾实体。

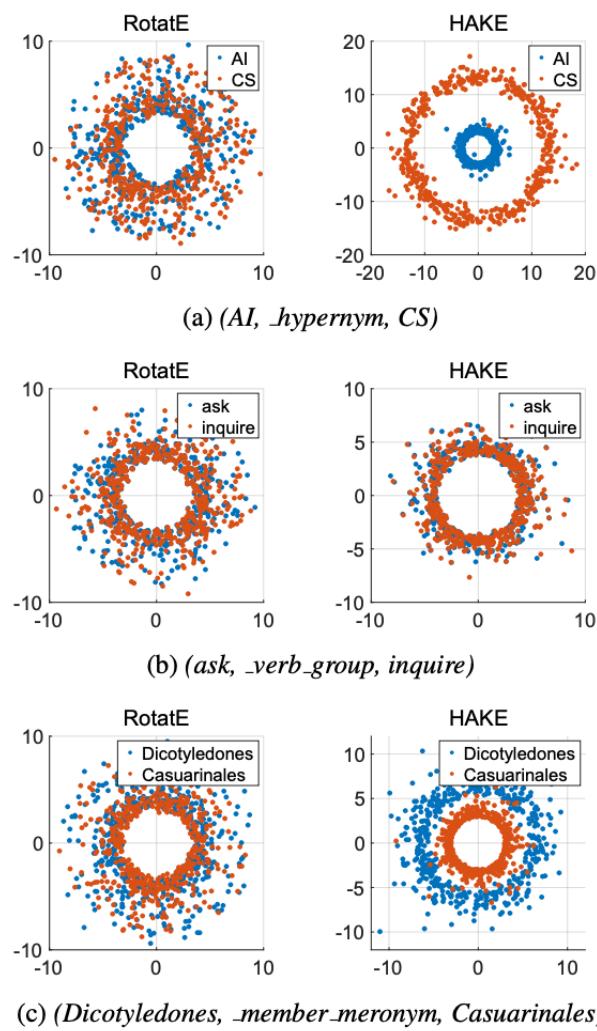
从图中可以分析得出，当头实体语义层级更高时，关系的模长总体大于 1；当尾实体语义层级更高时，关系的模长总体小于 1；当头尾实体的语义层级相同时，关系的模长总体接近 1。这样的实验结果与上述猜想完全一致。我们还发现，与 ModE 相比，HAKE 的模长嵌入向量的取值分布更为集中，方差更小，这表明 HAKE 能够更加清晰准确地对语义层级进行建模。

关系的模角度部分 在上述图 (c) (d) 中，关系的模长都接近于 1，因此利用模长部分难以对该关系连接的实体进行区分。因此，我们求助于 HAKE 的角度部分。我们绘制了两种关系向量每一维度的取值的分布直方图，如下图所示。属于相同语义层级的不同实体，我们从 WN18RR 和 FB15k-237 中各选取了一种关系，并绘制了它们每一维度的取值的分布直方图，如下图所示。



从图中可以看到，它们每一维度的取值主要集中于三个值附近：0、 π 与 2π ，并且取值接近 π 的维度占比为40%-50%。也就是说，这些头尾实体中接近一半的维度取值都相差 π 。因此位于同一语义层级的实体可以被角度部分区分。

实体可视化分析 我们接下来对实体进行可视化分析。在这部分的实验中，为了更加清晰的展示出通过 HAKE 建模得到的分层效果，我们将 HAKE 与 RotatE 进行对比，将两种模型得到的实体向量的不同维度都绘制成同一个二维平面上的点。我们从 WN18RR 中选取了三个不同类型的三元组，得到如下的散点图。



图(a)中的三元组中,头实体位于更低的语义层级;图(b)中,头尾实体的语义层级相同;图(c)中,头实体位于更高的语义层级。需要注意的是,为了更加直观地显示出分

层效果，我们绘制散点图时对于每一个点的模长使用了对数缩放操作。因此，图中更大的半径实际上表示更小的模长。

从图中可以看到，在头尾实体分属不同语义层级的场景下（图(a)与(c)），HAKE 的散点图表现出了更加明显的分层效果，而在 RotatE 的散点图中，头尾实体则难以依靠半径进行区分。对于头尾实体属于相同语义层级的场景下，表示头尾实体的点应该具有大致相同的半径。在此场景下，HAKE 依然表现的更好，因为 HAKE 中的实体模长分布的方差更小。

4.3.6 总结

本文介绍了一个可建模语义分层的知识图谱补全模型：HAKE。该模型使用极坐标系对语义分层进行建模。其中，模长部分用于建模分属不同语义层级的实体；角度部分用于建模属于相同语义层级的实体。实验结果表明 HAKE 的性能优于现有的性能最好的方法。进一步分析结果显示，训练得到的模型中模长与角度的表现与预期相符，能够很好地对语义层级进行建模。

4.4 基于张量分解的知识图谱嵌入 (张占秋-2020/11/17)

【NeurIPS 2020】 Duality-Induced Regularizer for Tensor Factorization Based Knowledge Graph Completion

4.4.1 简介

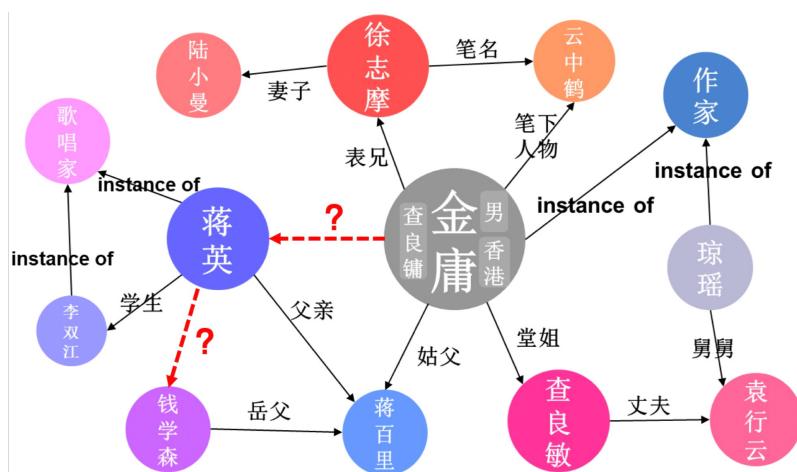
近年来，张量分解模型凭借模型简洁、计算速度快等优点在知识图谱补全任务上取得了令人瞩目的成就。但是，这些模型较易受到过拟合的影响，在性能上通常落后于其他类型的模型。为解决过拟合问题，包括 L2 正则，N3 正则 [1] 在内的多种正则项被提出，但这些正则项又在性能或者适用范围上存在明显缺陷。为此，我们基于知识图谱补全模型之间的对偶性，为张量分解模型提出了一种新的正则项—DURA。该正则项可以广泛地应用于多种不同的张量分解知识图谱补全模型，且能够显著提升模型性能。

该工作题为 Duality-Induced Regularizer for Tensor Factorization Based Knowledge Graph Completion，已发表于 NeurIPS 2020。论文已上传至 arXiv，代码已开源至 Github。

4.4.1.1 知识图谱补全

知识图谱由大量的事实三元组构成，其存储了结构化的人类知识。例如，下图中的（金庸，笔下人物，云中鹤）和（金庸，姑父，蒋百里）都是事实三元组。

尽管常用的知识图谱具有十亿乃至以上级别的三元组，但其中仍会存在大量的三元组缺失。例如，下图中“蒋英”和“钱学森”之间缺失了“丈夫”关系，“金庸”和“蒋英”之间缺失了“表姐”关系。因知识图谱的规模通常极为庞大，人工对图谱进行补全代价高昂。因此，基于知识图谱中的已知三元组来自动预测缺失三元组的知识图谱补全 (KGC) 技术近年来备受关注。



4.4.1.2 形式化描述

给定一个由实体构成的集合 \mathcal{E} 和一个由关系构成的集合 \mathcal{R} , 一个知识图谱 $\mathcal{K} = \{(e_i, r_j, r_k)\} \subset \mathcal{E} \times \mathcal{R} \times \mathcal{E}$ 是一个三元组的集合, 其中 e_i 是第 i 个实体, r_j 是第 j 个关系。通常 e_i 和 e_j 也分别称为头实体和尾实体, 有时也会被写成 h_i (head) 和 t_j (tail)。知识图谱补全的目标是根据 \mathcal{K} 中已知的三元组来预测未知的正确三元组。

我们通常将知识图谱补全问题建模成一个给定头 (尾) 实体和关系的尾 (头) 实体预测问题。如在问题 $(e_i, r_j, ?)$ 中给定 (e_i, r_j) 来预测尾实体。

知识图谱补全模型通常先将每个实体 e_i 与关系 r_j 映射为低维嵌入 \mathbf{e}_i 和 \mathbf{r}_j (knowledge graph embeddings, 可为向量/矩阵/张量), 然后通过一个以低维嵌入为输入的打分模型 $s(e_i, r_j, e_k) : \mathcal{E} \times \mathcal{R} \times \mathcal{E} \rightarrow \mathbb{R}$ 为给定三元组 (e_i, r_j, r_k) 进行打分, 从而确定任意三元组真实存在的概率 (即置信度)。如果一个补全模型足够好, 那在 $(e_i, r_j, ?)$ 这一问题中, 能够使得三元组成立的尾实体 e_k 得分应该高于其他实体。

4.4.1.3 知识图谱补全模型

两类重要模型

基于距离的模型 (distance-based models) 和基于张量分解的模型 (tensor factorization-based models) 是两类重要的知识图谱补全模型。

基于距离的模型使用闵可夫斯基距离来度量三元组的合理性, 其一般具有如下形式。

$$s(e_i, r_j, e_k) = -\|\Gamma(e_i, r_j, e_k)\|_p, p = 1, 2,$$

其中 Γ 是与模型相关的函数。

虽然这类模型能够达到较好的性能, 但在建模复杂的关系模式 (如一对多和多对一关系) 时仍然存在困难。

基于张量分解的模型将知识图谱视为部分可观测的三阶张量 $\mathcal{X} \in \{0, 1\}^{|\mathcal{E}| \times |\mathcal{R}| \times |\mathcal{E}|}$, 当 (e_i, r_j, e_k) 是正确三元组时, $\mathcal{X}_{ijk} = 1$, 否则 $\mathcal{X}_{ijk} = 0$ 。因此, 知识图谱补全被建模为一个张量补全问题。

假设 \mathcal{X}_j 表示 \mathcal{X} 的第 j 个前向切片, 即第 j 个关系的邻接矩阵。一般情况下, 基于张量分解的知识图谱补全模型将 \mathcal{X}_j 近似分解为 $\mathcal{X}_j \approx \text{Re}(\bar{\mathbf{H}}\mathbf{R}_j\mathbf{T}^\top)$, 其中 $\mathbf{H}(\mathbf{T})$ 的第 $i(k)$ 行是 $\mathbf{e}_i(\mathbf{e}_k)$, \mathbf{R}_j 是第 j 个关系对应的嵌入矩阵, $\text{Re}(\cdot)$ 和 \cdot^\top 分别是复矩阵的实部和共轭 (实矩阵的实部和共轭是它本身)。也就是说, 打分函数被定义为 $s(e_i, r_j, e_k) = \text{Re}((\bar{\mathbf{e}}_i\mathbf{R}_j\mathbf{e}_k^\top))$ 。基于张量分解的模型的目的是寻找矩阵 $\mathbf{H}, \mathbf{R}_1, \dots, \mathbf{R}_{|\mathcal{R}|}, \mathbf{T}$, 使得 $\text{Re}(\bar{\mathbf{H}}\mathbf{R}_j\mathbf{T}^\top)$ 可以近似 \mathcal{X}_j 。

从理论上讲, 张量分解模型表现力强, 能够很好地处理复杂的关系。然而, 它们往往面临严重的过拟合问题, 无法达到最佳性能。因此, 我们通常会在训练模型时加入正则。

令 $\hat{\mathcal{X}}_j = \text{Re}(\bar{\mathbf{H}}\mathbf{R}_j\mathbf{T}^\top)$, 而 $\hat{\mathcal{X}}$ 是一个第 j 个前向切片为 $\hat{\mathcal{X}}_j$ 的张量, 则正则化后的基于张量分解的知识图谱补全模型可以写成

$$\min_{\hat{\mathcal{X}}_1, \dots, \hat{\mathcal{X}}_{|\mathcal{R}|}} \sum_{j=1}^{|\mathcal{R}|} L(\mathcal{X}_j, \hat{\mathcal{X}}_j) + \lambda g(\hat{\mathcal{X}}), \quad (4.1)$$

其中 $\lambda > 0$ 是固定参数, $L(\mathcal{X}_j, \hat{\mathcal{X}}_j)$ 衡量 \mathcal{X}_j 与 $\hat{\mathcal{X}}_j$ 之间的差异, 而 g 是正则函数。

面向张量分解知识图谱补全的正则

针对张量分解模型, 目前流行的正则项主要包括 L_2 正则和 $N3$ 正则。

L_2 正则是最常用的正则项, 具有如下形式

$$g(\hat{\mathcal{X}}) = \|\mathbf{H}\|_F^2 + \|\mathbf{T}\|_F^2 + \sum_{j=1}^{|\mathcal{R}|} \|\mathbf{R}_j\|_F^2,$$

其又被称为平方 Frobenius 正则。从上式这可以看出, L_2 正则能够应用于绝大部分张量分解模型, 但实验表明, 其并不一定能够带来稳定的性能提升 [2]。

$N3$ 正则 [1] 即张量 3 阶核范数 (tensor nuclear-3 norm) [3] 正则, 其适用于 \mathbf{R}_j 为对角矩阵的模型, 具有如下形式

$$g(\hat{\mathcal{X}}) = \sum_{d=1}^D (\|\mathbf{h}_{:d}\|_3^3 + \|\mathbf{r}_{:d}\|_3^3 + \|\mathbf{t}_{:d}\|_3^3),$$

其中 $\mathbf{h}_{:d}$ 和 $\mathbf{t}_{:d}$ 分别是矩阵 \mathbf{H} 和 \mathbf{T} 的第 d 列。假设 $\tilde{\mathbf{R}}$ 是一个由 \mathbf{R}_j 的对角元组成第 j 行的矩阵, 则 $\mathbf{r}_{:d}$ 是矩阵 $\tilde{\mathbf{R}}$ 的第 d 列。实验表明, N_3 正则在 \mathbf{R}_j 为对角矩阵时 (如 CP 和 ComplEx) 能带来稳定的性能提升, 但其不适用于 \mathbf{R}_j 为非对角阵的更一般情况 (如 RESCAL)。

因此, 我们还需要一个适用范围广且能带来稳定性能提升的正则项。

4.4.2 对偶诱导正则 DURA

基本 DURA

考虑一个知识图谱补全问题 $(h_i, r_j, ?)$, 即已知头实体和关系来预测尾实体。假设用 $f_j(i, k) = s(h_i, r_j, t_k)$ 来衡量给定三元组 (h_i, r_j, t_k) 的置信度, 则上一节中提到的基于张量分解的模型的打分函数可以写成

$$f_j(i, k) = \text{Re}(\bar{\mathbf{h}}_i \mathbf{R}_j \mathbf{t}_k^\top) = \text{Re}(\langle \mathbf{h}_i \bar{\mathbf{R}}_j, \mathbf{t}_k \rangle),$$

它首先通过一个线性变换 $\bar{\mathbf{R}}_j$ 对实体嵌入 \mathbf{h}_i 做一个映射, 然后使用内积的实部来度量 $\mathbf{h}_i \bar{\mathbf{R}}_j$ 和 \mathbf{t}_k 之间的相似性。我们可以用另一个常用的相似性度量欧几里德距离可以代替上述等式中的内积相似性。从而得到一个与之对应的基于距离的模型:

$$f_j^E(i, k) = -\|\mathbf{h}_i \bar{\mathbf{R}}_j - \mathbf{t}_k\|_2^2.$$

因平方打数函数和无平方的形式是等价的, 我们事实上得到了以下对偶性: 对于现有的基于张量分解的知识图谱补全模型 (原模型), 通常还有另一个与之相对应的基于距

离的知识图谱补全模型（对偶模型）。具体地，原始模型和对偶模型之间的对偶关系又可以表示为

$$f_j^E(i, k) = -\|\mathbf{h}_i \bar{\mathbf{R}}_j - \mathbf{t}_k\|_2^2 \quad (4.2)$$

$$= -\|\mathbf{h}_i \bar{\mathbf{R}}_j\|_2^2 - \|\mathbf{t}_k\|_2^2 + 2\operatorname{Re}(\langle \mathbf{h}_i \bar{\mathbf{R}}_j, \mathbf{t}_k \rangle) \quad (4.3)$$

$$= 2f_j(i, k) - \|\mathbf{h}_i \bar{\mathbf{R}}_j\|_2^2 - \|\mathbf{t}_k\|_2^2. \quad (4.4)$$

假设 \mathcal{S} 是所有已知正确三元组的集合。当我们训练一个基于距离的模型时，我们通常会在 \mathcal{S} 上最大化 $f_j^E(i, k)$ ，而从上式我们知道

$$\begin{aligned} \max f_j^E(i, k) &= \min -f_j^E(i, k) \\ &= \min -2f_j(i, k) + \|\mathbf{h}_i \bar{\mathbf{R}}_j\|_2^2 + \|\mathbf{t}_k\|_2^2. \end{aligned} \quad (4.5)$$

很显然， $\min -2f_j(i, k)$ 恰好是未加正则的张量分解模型的目标。因此，对偶关系为我们提供了如下正则项

$$\sum_{(h_i, r_j, t_k) \in \mathcal{S}} \|\mathbf{h}_i \bar{\mathbf{R}}_j\|_2^2 + \|\mathbf{t}_k\|_2^2, \quad (4.6)$$

我们称上述正则项为基本 DURA (Basic DURA)。

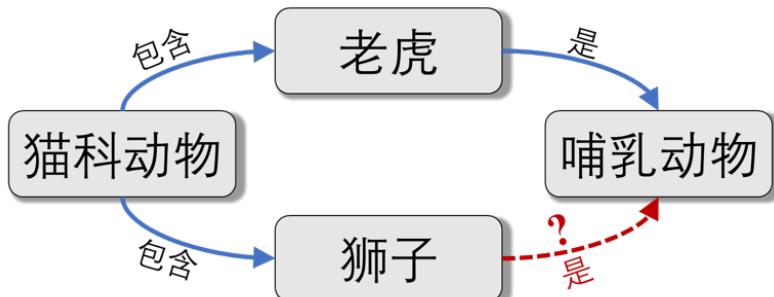
很明显，基本 DURA 适用与 \mathbf{R}_j 为非对角阵的倾向，适用范围相较 N3 正则更广。

基本 DURA 为何有效

我们将从直觉上给出基本 DURA 为何能够有效提升张量分解知识图谱补全模型的性能。如果多个尾实体能够使得三元组 $(h_i, r_j, ?)$ 成立，那么 DURA 将促使它们具有相似的嵌入 (embeddings)。

首先，我们可以断言，如果多个尾实体能够使得三元组 $(h_i, r_j, ?)$ 成立，他们应该具有相近的语义。如在下图中，我们已知（猫科动物，包含，老虎）和（猫科动物，包含，狮子）两个三元组，则实体“老虎”和“狮子”在语义上是相近的。

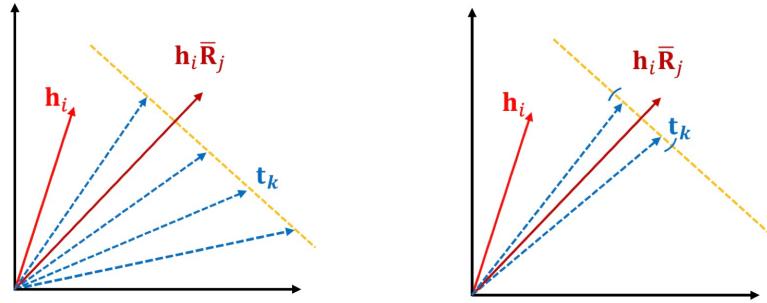
进一步，我们可以断言，如果两个实体具有相近的语义，那么它们应该具有相似的嵌入。如果“老虎”和“狮子”具有相似的嵌入，我们又已知（老虎，是，哺乳动物）这一三元组，那么我们就可以对（狮子，是，哺乳动物）这一未知三元组进行补全。



然而，由于内积相似度度量的特性，在一个基于张量分解的知识补全模型中，即使

尾实体具有完全不同的嵌入，三元组 (h_i, r_j, t_k) 也可以具有相同的打分（如下左图所示）。这在一定程度上会影响对未知三元组的预测。

在加入 DURA 之后，我们事实上通过欧氏距离引入了 $\mathbf{h}_i \bar{\mathbf{R}}_j$ 和 \mathbf{t}_k 应该较为相近的先验，因此尾实体的嵌入会被限制在一个较小的范围内（如下右图所示），从而有利于知识图谱的补全。



DURA

由内积的性质我们可知 $\text{Re}(\bar{\mathbf{h}}_i \mathbf{R}_j \mathbf{t}_k^\top) = \text{Re}(\bar{\mathbf{t}}_k \bar{\mathbf{R}}_j^\top \mathbf{h}_i^\top)$ ，利用与基本 DURA 相似的推导过程，我们可以从对偶性中得到另一个正则项

$$\sum_{(h_i, r_j, t_k) \in \mathcal{S}} \|\mathbf{t}_k \mathbf{R}_j^\top\|^2 + \|\mathbf{h}_i\|^2. \quad (4.7)$$

类似地，当多个头实体能够使得三元组 $(?, r_j, t_k)$ 成立时，该正则项能够使得这些头实体的嵌入较为接近。

将两个不同方向的正则项相结合，我们得到对偶诱导正则 (DUality-induced RegulArizer, DURA) 的最终形式

$$\sum_{(h_i, r_j, t_k) \in \mathcal{S}} \left[\|\mathbf{h}_i \bar{\mathbf{R}}_j\|_2^2 + \|\mathbf{t}_k\|_2^2 + \|\mathbf{t}_k \mathbf{R}_j^\top\|_2^2 + \|\mathbf{h}_i\|_2^2 \right]. \quad (4.8)$$

DURA 的理论性质

当我们放宽正则项的求和条件 $(h_i, r_j, t_k) \in \mathcal{S}$ 至所有可能的实体和关系，DURA 具有如下形式

$$|\mathcal{E}| \sum_{j=1}^{|\mathcal{R}|} (\|\mathbf{H} \bar{\mathbf{R}}_j\|_F^2 + \|\mathbf{T}\|_F^2 + \|\mathbf{T} \mathbf{R}_j^\top\|_F^2 + \|\mathbf{H}\|_F^2), \quad (4.9)$$

其中 $|\mathcal{E}|$ and $|\mathcal{R}|$ 分别是实体和关系的数目。

如果我们进一步假设关系嵌入矩阵 \mathbf{R}_j 为对角阵，我们有如下定理

在上述定理中， $\|\hat{\mathcal{X}}\|_*$ 是张量 $\hat{\mathcal{X}}$ 的 2 阶核范数 (nuclear 2-norm) [引]用 nuclear]，即矩阵迹范数 (trace norm) 在张量情形下的推广。

上述定理说明，在给定一些假设的情况下，DURA 事实上给出了张量 $\hat{\mathcal{X}}$ 的 2 阶核范数的一个上界。

Theorem 1. Suppose that $\hat{\mathcal{X}}_j = \mathbf{H}\mathbf{R}_j\mathbf{T}^\top$ for $j = 1, 2, \dots, |\mathcal{R}|$, where $\mathbf{H}, \mathbf{T}, \mathbf{R}_j$ are real matrices and \mathbf{R}_j is diagonal. Then, the following equation holds

$$\min_{\hat{\mathcal{X}}_j = \mathbf{H}\mathbf{R}_j\mathbf{T}^\top} \frac{1}{\sqrt{|\mathcal{R}|}} \sum_{j=1}^{|\mathcal{R}|} (\|\mathbf{H}\mathbf{R}_j\|_F^2 + \|\mathbf{T}\|_F^2 + \|\mathbf{T}\mathbf{R}_j^\top\|_F^2 + \|\mathbf{H}\|_F^2) = \|\hat{\mathcal{X}}\|_*$$

The minimization attains if and only if $\|\mathbf{h}_{:,d}\|_2 \|\mathbf{r}_{:,d}\|_2 = \sqrt{|\mathcal{R}|} \|\mathbf{t}_{:,d}\|_2$ and $\|\mathbf{t}_{:,d}\|_2 \|\mathbf{r}_{:,d}\|_2 = \sqrt{|\mathcal{R}|} \|\mathbf{h}_{:,d}\|_2$, $\forall d \in \{1, 2, \dots, D\}$, where $\mathbf{h}_{:,d}$, $\mathbf{r}_{:,d}$, and $\mathbf{t}_{:,d}$ are the d -th columns of \mathbf{H} , \mathbf{R} , and \mathbf{T} , respectively.

4.4.3 实验结果

知识图谱补全任务

我们在三个常用知识图谱补全数据集 WN18RR、FB15k-237、YAGO3-10 上验证了 DURA 的有效性。所考虑的基于张量分解的知识图谱补全模型包括目前流行的 CP、ComplEx [4] 和 RESCAL [5]。所采用的评价指标包括 MRR (Mean Reciprocal Rank)，H@N (Hits at N, N=1,10)，越高的 MRR 和 H@N 代表越好的性能。实验结果如下表所示。

	WN18RR			FB15k-237			YAGO3-10		
	MRR	H@1	H@10	MRR	H@1	H@10	MRR	H@1	H@10
RotatE	.476	.428	.571	.338	.241	.533	.495	.402	.670
MuRP	.481	.440	.566	.335	.243	.518	-	-	-
HAKE	.497	.452	.582	.346	.250	.542	.546	.462	.694
TuckER	.470	.443	.526	.358	.266	.544	-	-	-
CP	.438	.414	.485	.333	.247	.508	.567	.494	.698
RESCAL	.455	.419	.493	.353	.264	.528	.566	.490	.701
ComplEx	.460	.428	.522	.346	.256	.525	.573	.500	.703
CP-DURA	.478	.441	.552	.367	.272	.555	.579	.506	.709
RESCAL-DURA	.498	.455	.577	.368	.276	.550	.579	.505	.712
ComplEx-DURA	.491	.449	.571	.371	.276	.560	.584	.511	.713

可以看出，DURA 为三个模型带来了稳定的性能提升。尤其值得一提的是，RESCAL 作为最早被提出的知识图谱补全模型之一（2011 年），因其参数量较大而较易过拟合，在过去的几年中表现一直较差。但在加入 DURA 后，RESCAL 在 WN18RR 上取得了 MRR 为 0.498 的好成绩，超越了之前的所有模型。

与其他正则的对比

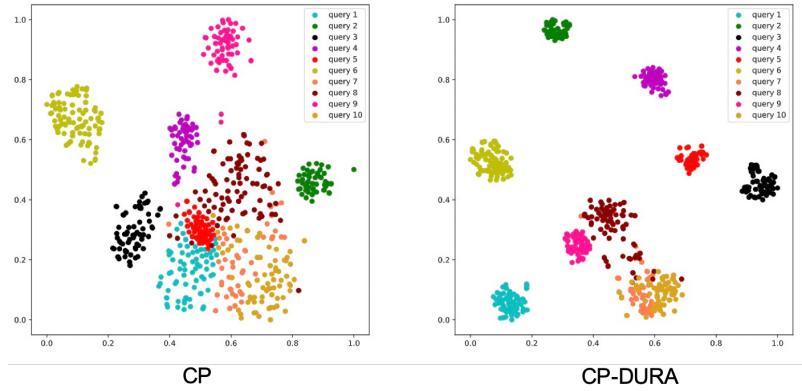
我们在 CP、ComplEx 和 RESCAL 三个模型上对比了 DURA、 L_2 （即 FRO）、 N_3 三种正则的性能，结果如下表所示。注意 N_3 无法应用于 RESCAL 模型。

	WN18RR			FB15k-237			YAGO3-10		
	MRR	H@1	H@10	MRR	H@1	H@10	MRR	H@1	H@10
CP-FRO*	.460	-	.480	.340	-	.510	.540	-	.680
CP-N3	.470	.430	.544	.354	.261	.544	.577	.505	.705
CP-DURA	.478	.441	.552	.367	.272	.555	.579	.506	.709
ComplEx-FRO*	.470	-	.540	.350	-	.530	.570	-	.710
ComplEx-N3	.489	.443	.580	.366	.271	.558	.577	.502	.711
ComplEx-DURA	.491	.449	.571	.371	.276	.560	.584	.511	.713
RESCAL-FRO	.397	.363	.452	.323	.235	.501	.474	.392	.628
RESCAL-DURA	.498	.455	.577	.368	.276	.550	.579	.505	.712

从上表中可以看出，相比 L_2 和 N_3 正则，DURA 在适用范围和有效性上都具有明显优势。

可视化

为验证“如果多个尾实体能够使得三元组 $(h_i, r_j, ?)$ 成立，那么 DURA 将促使它们具有相似的嵌入 (embeddings)”这一描述，我们利用 T-SNE 对加入 DURA 前后的实体嵌入进行了可视化。结果如下图所示，其中每一个数据点代表一个尾实体，而相同颜色的数据点代表这些尾实体能够使得同一 $(h_i, r_j, ?)$ 成立。



从上图可以看出，在加入 DURA 后，能够使得同一 $(h_i, r_j, ?)$ 成立的尾实体将倾向于拥有更为相似的嵌入。

4.4.4 总结

在知识图谱补全任务上，基于张量分解模型和基于距离的模型之间具有对偶性质，利用该对偶性，我们提出了一种面向张量分解模型的对偶诱导正则 DURA。DURA 能够广泛应用于多种不同的张量分解模型，且能够带来稳定而又显著的性能提升。

4.4.5 参考文献

- [1] Lacroix, T., Usunier, N. & Obozinski, G.. (2018). Canonical Tensor Decomposition for Knowledge Base Completion. Proceedings of the 35th International Conference on Machine Learning, in PMLR 80:2863-2872
- [2] Ruffinelli, D., Broscheit, S., & Gemulla, R. (2019). You CAN teach an old dog new tricks! on training knowledge graph embeddings. In International Conference on Learning Representations.
- [3] Friedland, S., & Lim, L. H. (2018). Nuclear norm of higher-order tensors. Mathematics of Computation, 87 (311), 1255-1281.
- [4] Trouillon, T., Welbl, J., Riedel, S., Gaussier, E. & Bouchard, G.. (2016). Complex Embeddings for Simple Link Prediction. Proceedings of The 33rd International Conference on Machine Learning, in PMLR 48:2071-2080.
- [5] Nickel, M., Tresp, V., & Kriegel, H. P. (2011). A three-way model for collective learning on multi-relational data. Proceedings of the 28th International Conference on Machine Learning, in PMLR 11:809-816

4.5 基于神经网络的知识图谱嵌入（内容待补充）

内容待补充。

4.6 知识图谱嵌入前沿（内容待补充）

内容待补充。

第五章 知识图谱推理（内容待补充）

5.1 知识图谱推理概述

内容待补充。

5.2 知识推理发展历史

内容待补充。

5.3 基于规则的知识图谱推理

内容待补充。

5.4 基于图结构的知识图谱推理

内容待补充。

5.5 基于表示学习的知识图谱推理

内容待补充。

5.6 知识图谱推理前沿

内容待补充。

5.6.1 基于元学习的少样本知识图谱推理

5.6.2 基于强化学习的知识图谱推理

5.6.3 时序知识图谱推理

5.6.4 图神经网络与知识图谱推理

第六章 知识图谱应用（内容待补充）

6.1 基于知识图谱的问答

内容待补充。

6.2 基于知识图谱的聊天机器人

内容待补充。

6.3 结合知识图谱的推荐系统

内容待补充。

6.4 结合知识图谱的信息检索

内容待补充。