



Towards Foundation Models for Knowledge Graph Reasoning

ULTRA: 一种学习 通用和可转移的图表示的方法



2024/9/29

I Introduction

起因：KG 上设计基础模型的主要挑战是学习这种可转移的表示，希望这些表示可以在具有任意实体和关系词汇的任何图上进行推理

主要研究目标：是找到具有任意实体和关系词汇的图之间可转移的不变性。利用和学习这种不变性将使 KG 推理的基础模型能够进行预训练和微调范式

- (1) 即使数据集之间的关系不同，关系之间的交互也可能相似且可转移;
- (2) 初始关系表示可以以这种交互为条件，绕过任何输入特征。

Ultra?

- 一种学习通用和可转移的图表示的方法。ULTRA 将关系表示构建为以交互为条件的函数。
- 这种调节策略允许预先训练的 ULTRA 模型以任何关系词汇归纳泛化到任何看不见的 KG，并在任何图形上进行微调。

I Introduction

相关知识:

基础模型 (FM)：以自我监督的方式在大型数据集上训练的骨干模型

零样本学习 (Zero-Shot Learning)：在没有看到某些类别的样本的情况下，使模型能够识别或预测这些类别的性能。

链接预测 (Link Prediction)：预测知识图谱中可能存在的新链接或关系的任务。

归纳学习 (Inductive Learning)：在训练时使用固定的实体和关系集合，但在推理时能够推广到新的实体和关系。

条件关系表示 (Conditional Relation Representations)：基于关系图和查询条件动态生成的关系表示。

标签技巧 (Labeling Trick)：一种用于增强GNNs性能的技术，通过为图中的节点分配唯一的标签来打破对称性。

预训练 (Pre-training)：在大量数据上训练模型以学习通用特征的过程。

微调 (Fine-tuning)：在特定任务上调整预训练模型的参数以提高性能的过程。

多模态学习 (Multimodal Learning)：结合并处理来自多种模态（如文本、图像、声音）的数据的学习方式。

• **Mean Reciprocal Rank (MRR)**：三元组的排名得分的倒数的平均值。

• **Hits@n**：在前 n 个预测中正确预测的百分比。



I Introduction

相关知识（现有的归纳方法）

NBFNet，全称为Neural [Bellman-Ford](#) Networks，是基于论文《Neural Bellman-Ford Networks: A General Graph Neural Network Framework for Link Prediction》实现的一个强大框架

技术分析

NBFNet巧妙地将经典的Bellman-Ford最短路径算法思想融入到图神经网络中，形成了一种既具有传统路径搜索的直观性和解释性，又拥有图神经网络的强大表达力和学习能力的综合框架。通过这种设计，NBFNet能够在多个GPU或分布式环境中高效运行，支持PyTorch和TorchDrug生态，为开发者提供了灵活且高性能的解决方案。

应用场景

NBFNet的应用领域广泛，尤其适合于两个关键场景：

1. **知识图谱补全**：在如FB15k-237、WN18RR和ogbl-biokg这样的标准基准上，NBFNet展现出卓越的性能，对于增强知识图谱的完整性至关重要。
2. **同构图链接预测**：在诸如Cora、CiteSeer、PubMed等图数据集上的出色表现，使得NBFNet成为社交网络分析、推荐系统等领域的重要工具。

项目特点

- **兼容性**：无缝集成PyTorch环境，提供对多GPU和分布式计算的支持。
- **通用性**：无论是知识图谱还是普通的同构图，NBFNet都能应对自如。
- **性能强劲**：在多个任务上的实验结果显示，NBFNet能够达到或超越当前先进的链接预测模型。
- **易于理解**：结合传统图理论与深度学习，NBFNet提供了较好的模型解释性，利于调试和应用。
- **社区与文档**：详尽的安装指南与配置文件确保快速上手，而丰富的实验案例则帮助用户深入理解模型机制。




I Introduction

相关知识:

GraIL

论文题目: Inductive Relation prediction by subgraph Reasoning (子图推理的归纳关系预测)

问题: 知识图谱中关系预测的主导范式涉及对实体和关系的潜在表征（即嵌入）的学习和操作。这些基于嵌入的方法并没有明确地捕捉知识图谱背后的组成逻辑规则，并且仅限于转导设置，在该设置中，在训练过程中必须知道完整的实体集。

方法: 提出了一个基于 图神经网络  的关系预测框架GraIL，该框架推理了局部子图结构，并且在学习与实体无关的关系语义时具有很强的归纳偏见。相较于基于嵌入的模型，GraIL是自然归纳的，并且可以在训练后推广到看不见的实体和图。

I Introduction

相关知识:

RED-GNN

Knowledge Graph Reasoning with Relational Directed Graph

Yongqi Zhang¹ Quanming Yao^{1,2}

¹4Paradigm Inc.

²Department of Electronic Engineering, Tsinghua University
zhangyongqi@4paradigm.com, qyaoaa@mail.tsinghua.edu.cn

CSDN @露葵025

摘要

知识图推理旨在从已有的知识中推断出新的事实。基于关系路径的方法在文献中显示出较强的可解释性和归纳推理能力。然而，在KG中捕获复杂拓扑(Capturing complex topology)时，路径是有限的。本文引入了一种新的关系结构，即由重叠关系路径组成的关系有向图(r-digraph)来捕获KG的结构信息。由于有向图表现出比路径更复杂的结构，因此在r-digraph上构建和学习是具有挑战性的。在这里，我们提出了一种图神经网络，即RED-GNN，通过使用GNN的变体学习关系有向图来解决上述挑战。明确地RED-GNN递归编码多个具有共享边的r-digraph，并通过依赖于查询的注意力权重来选择强相关边。我们展示了在用看不见的实体推理KG和通过r-digraph完成KG基准、RED-GNN的效率以及在r-digraph可解释的依赖性方面的显著收获。

可以仅推广到具有新节点的图形，但不能推广到新的关系类型，因为节点表示是作为固定关系词汇表的函数构造的。

I Introduction

FB15k-237 数据集

Cross3domain 数据集

相关知识:

WN18RR 数据集

- 数据规模:
 - 关系 (Relations) : 11 种
 - 实体 (Entities) : 40,943 个
 - 训练集 (Training set) : 86,835 个三元组
 - 验证集 (Validation set) : 3,034 个三元组
 - 测试集 (Test set) : 3,134 个三元组
- 数据样本示例:
 - ("Arthur", "is_related_to", "Camelot")
 - ("Merlin", "has_ability", "Magic")

ML任务形式化

- 任务: 链接预测 (Link Prediction), 也称为实体预测 (Entity Prediction) 或实体排名 (Entity Ranking)。
- 形式化表述:
 - 给定一个三元组 $(h, r, ?)$ 或 $(?, r, t)$, 预测未知的实体。具体而言:
 - 头实体预测: 已知尾实体 t 和关系 r , 预测头实体 h 。
 - 尾实体预测: 已知头实体 h 和关系 r , 预测尾实体 t 。
- 评价指标:
 - Mean Reciprocal Rank (MRR): 三元组的排名得分的倒数的平均值。
 - Hits@n: 在前 n 个预测中正确预测的百分比。

数据样本特征描述

- 样本形式: 与 WN18RR 类似, 每个数据样本也由三元组 (h, r, t) 表示。
- 数据规模:
 - 关系 (Relations) : 237 种
 - 实体 (Entities) : 14,541 个
 - 训练集 (Training set) : 271,115 个三元组
 - 验证集 (Validation set) : 17,535 个三元组
 - 测试集 (Test set) : 20,466 个三元组
- 数据样本示例:
 - ("Paris", "is_capital_of", "France")
 - ("Einstein", "discovered", "Theory of Relativity")

ML任务形式化

- 任务: 链接预测 (Link Prediction)。
- 形式化表述:
 - 与 WN18RR 类似, 给定一个三元组 $(h, r, ?)$ 或 $(?, r, t)$, 预测未知的实体。

数据样本特征描述

- 样本形式: 同样使用三元组 (h, r, t) 表示数据样本。
- 数据规模:
 - 关系 (Relations) : 445 种
 - 实体 (Entities) : 9,282 个
 - 训练集 (Training set) : 45,739 个三元组
 - 验证集 (Validation set) : 1,044 个三元组
 - 测试集 (Test set) : 1,249 个三元组
- 数据样本示例:
 - ("Inception", "directed_by", "Christopher Nolan")
 - ("The Beatles", "performed", "Hey Jude")

ML任务形式化

- 任务: 跨域链接预测 (Cross-domain Link Prediction)。
- 形式化表述:
 - 给定一个跨多个领域的三元组 $(h, r, ?)$ 或 $(?, r, t)$, 预测未知的实体。
- 评价指标:
 - Mean Reciprocal Rank (MRR)
 - Hits@n

这三个数据集各自代表了知识图谱嵌入领域中的不同挑战和应用场景:

- **WN18RR 和 FB15k-237:** 主要针对可逆关系问题, 提供更具挑战性的任务以提高模型的泛化能力。
- **Cross3domain:** 强调跨域的知识图谱嵌入, 测试模型的跨域学习能力和知识迁移能力。



I Introduction

知识图谱和归纳学习

实体 \mathcal{V} (nodes)

关系 \mathcal{R} (边类型)

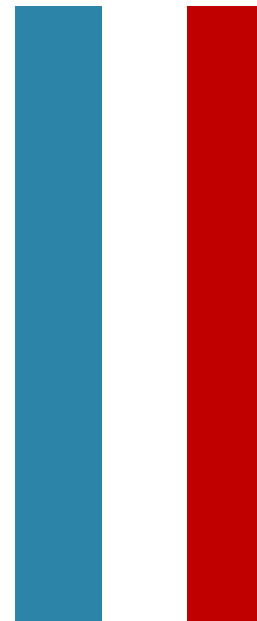
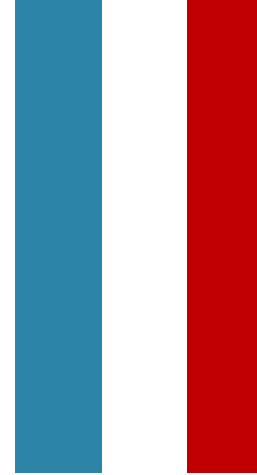
三元组 (边) $\mathcal{E}=(\mathcal{V}\times\mathcal{R}\times\mathcal{V})$

知识图谱 \mathcal{G} 是一个元组 $\mathcal{G}=(\mathcal{V},\mathcal{R},\mathcal{E})$.

问题表述

(h,r,t) head 实体 h 连接到 Tail 实体 t 按关系 r .

II Methods

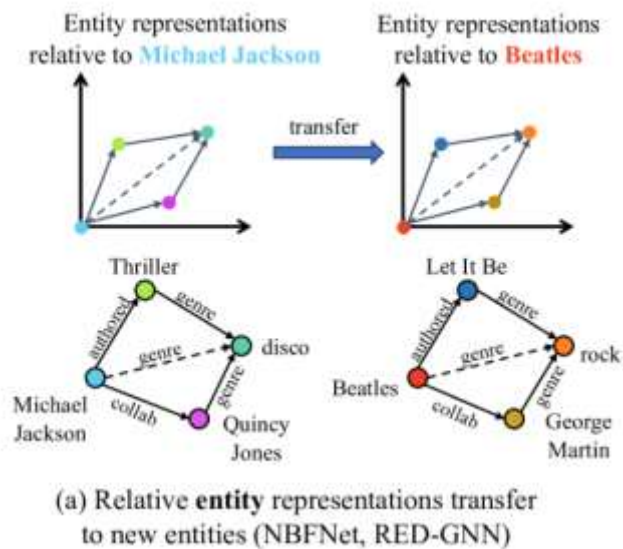


使用不同实体和关系词汇表进行归纳推理的主要挑战是找到 *可转移的不变性*

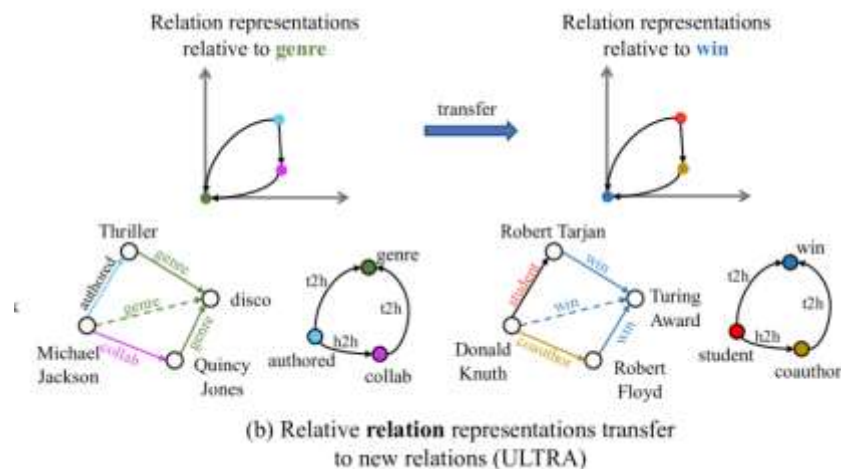
关系交互：尾对头（T2H）、头对头（H2H）、头对尾（H2T）和尾对尾（T2T）

给定一个查询 $(h, q, ?)$ （为典型的 KG 推理任务 $(h, q, ?)$ ）在图表上 G ,

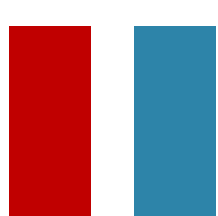
我们可以利用关系结构和四个基本关系的**不变性**来获得看不见的推理图的关系表示。作为典型的 KG 推理任务 $(h, q, ?)$ 以查询关系为条件 q , 则可以构建与查询相关的所有关系的表示 q 通过在关系图上使用标记技巧。这种**相对关系表示**不需要任何输入特征, 自然可以推广到任何多关系图。



归纳模型中使用的**相对实体**表示推广到新实体



基于关系图的**相对关系**表示可以推广到新的关系和实体



Ultra 采用三步算法,

(1) 构建关系图 (提升原始图表 G 到关系图 G_r)

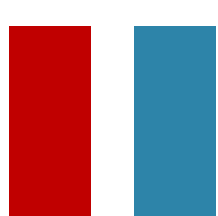
将原始图谱 G 中的每个关系类型作为一个节点, 构建一个新图谱 G_r , 其中的边表示原始图中关系之间的交互 (如尾到头、头到头等

目的: 为了捕捉知识图谱中不同关系之间的交互模式。

过程: 将每个关系类型作为节点, 构建一个新图, 称为关系图。

关系图中的边表示原始图中关系之间的交互, 包括四种基本类型: 尾到头 (t2h)、头到头 (h2h)、头到尾 (h2t)、尾到尾 (t2t)。

原理: 通过关系图, 可以形式化地表示关系之间的相互作用, 为学习关系表示提供结构化信息。



(2) 条件关系表示 (获取相对关系表示 $\mathbf{R}_{q|}(q, \mathcal{G}_r)$ 以 Query 关系为条件 q 在 Relation Graph 中 \mathcal{G}_r)

使用图神经网络 (GNN) 和标签技巧 (Labeling Trick) 在关系图 \mathcal{G}_r 上获得每个关系的唯一表示, 这些表示是相对于查询关系的。

给定一个查询 $(h, q, ?)$ 和一个关系图 \mathcal{G}_r , 然后我们获得 d 以查询关系 q 为条件的 \mathcal{G}_r -维节点表示 $\mathbf{R}_{q|} \in \mathbb{R}^{|\mathcal{R}| \times d}$ (对应于原始图 \mathcal{G} 中的所有边缘类型 \mathcal{R})。实际上, 我们通过应用标签技巧来实现条件反射, 通过指示器 r 函数初始化节点 $\mathcal{G}_{r|q}$, 并使用消息传递 GNN \mathcal{G}_r

目的: 学习能够泛化到任意知识图谱的关系表示。

过程: 使用图神经网络 (GNN) 在关系图上学习每个关系的独特表示。

通过一个标签技巧 (Labeling Trick), 初始化查询关系节点并应用消息传递机制, 获得关系表示。

原理: 这种方法允许模型动态地生成关系表示, 这些表示是相对于查询关系的, 从而提高了表示的适应性和泛化能力。

$$\begin{aligned} \mathbf{h}_{v|q}^0 &= \text{INDICATOR}_r(v, q) = \mathbb{1}_{v=q} * \mathbf{1}^d, \quad v \in \mathcal{G}_r \\ \mathbf{h}_{v|q}^{t+1} &= \text{UPDATE} \left(\mathbf{h}_{v|q}^t, \text{AGGREGATE} \left(\text{MESSAGE}(\mathbf{h}_{w|q}^t, \mathbf{r}) \mid w \in \mathcal{N}_r(v), r \in \mathcal{R}_{fund} \right) \right) \end{aligned}$$

(3) 实体级链接预测（使用关系表示 \mathbf{R}_q 作为起始关系特征，在原始图形上运行归纳链接预测 \mathcal{G} ）

利用上一步得到的每个关系的条件表示，运行归纳链接预测器，如NBFNet，来预测知识图谱中的链接

目的：利用获得的关系表示来预测知识图谱中的链接。

过程：使用修改后的NBFNet（Neighborhood-Based Forgetting Network）作为链接预测器。

在预测过程中，头节点被初始化为查询关系向量，而其他节点则初始化为零。

原理：通过利用学习到的关系表示，模型能够更好地理解实体间的潜在联系，从而提高链接预测的准确性。

$$\begin{aligned}\mathbf{h}_{v|u}^0 &= \text{INDICATOR}_e(u, v, q) = \mathbb{1}_{u=v} * \mathbf{R}_q[q], \quad v \in \mathcal{G} \\ \mathbf{h}_{v|u}^{t+1} &= \text{UPDATE}\left(\mathbf{h}_{v|u}^t, \text{AGGREGATE}\left(\text{MESSAGE}(\mathbf{h}_{w|u}^t, g^{t+1}(\mathbf{r})) \mid w \in \mathcal{N}_r(v), r \in \mathcal{R}\right)\right)\end{aligned}$$

$$\mathcal{L} = -\log p(u, q, v) - \sum_{i=1}^n \frac{1}{n} \log(1 - p(u'_i, q, v'_i))$$

III Experiments

III-A Datasets & Evaluation

为了评估 ULTRA 作为 KG 推理基础模型的质量，我们探讨了以下问题：

- (1) 预先训练的 ULTRA 是否能够以零镜头的方式归纳泛化到看不见的 KG？
- (2) 在特定数据集上微调 ULTRA 是否有任何好处？
- (3) 单个预训练的 ULTRA 模型与在每个目标数据集上从头开始训练的模型相比如何？
- (4) 预训练组合中的图表越多，性能是否越好？

数据集

我们对 57 个不同的 KG 进行了广泛的评估，并报告了 KG 完成任务的非饱和结果。数据集可分为三组

在训练和推理时具有 **固定实体和关系集** 的转导数据集

归纳实体数据集在推理时具有 **新实体** 但是有 **固定关系** 的数据集

归纳实体关系和数据集其中 **推理的实体和关系都是新的** 数据集

III-A Datasets & Evaluation

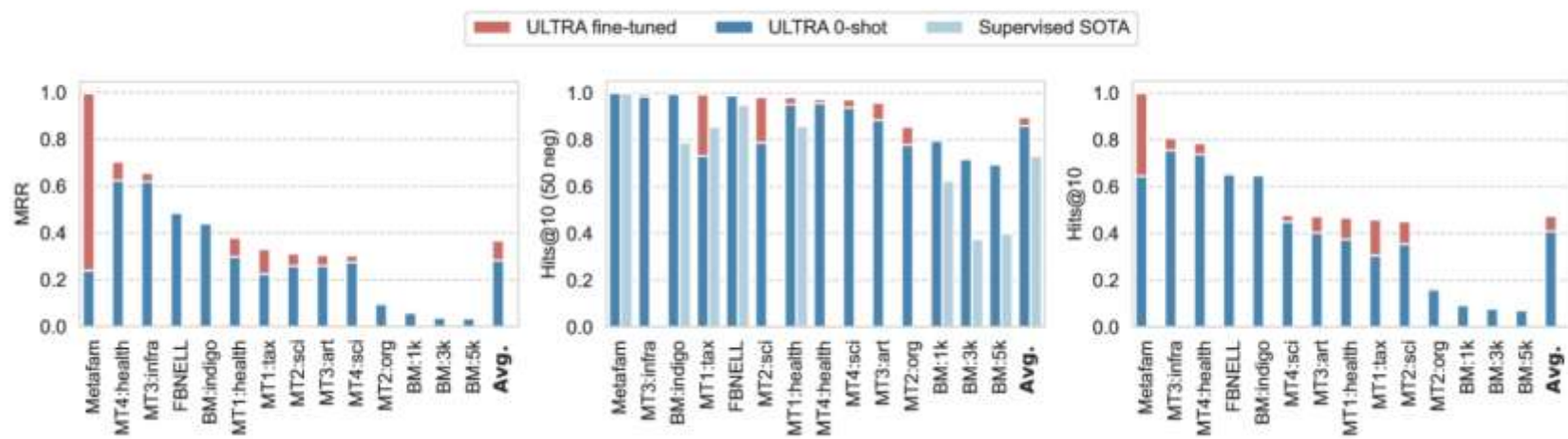
评估指标：主要使用平均倒数排名（Mean Reciprocal Rank, MRR）和Hits@10作为性能指标。

实验设置：对ULTRA进行了零样本（zero-shot）和微调（fine-tuned）两种情况的评估。

零样本情况下，使用预训练的模型直接在未见过的图谱上进行测试。

微调情况下，使用特定数据集的训练集对预训练模型进行额外的训练。

基线比较：与多个**强基线模型**进行比较，包括特定数据集上训练的模型。



III-A Datasets & Evaluation

主要结果:

在零样本推理中, ULTRA在多个数据集上的表现超过了或与特定数据集上训练的强基线相当。
微调进一步提高了ULTRA在大多数数据集上的性能。

Model	Inductive (e) + (e, r) (27 graphs)	Transductive e (40 graphs)	Total Avg (3 graphs)	Pretraining (8 graphs)	Inductive (e) + (e, r)				
	MRR	H@10	MRR	H@10	MRR	H@10	MRR	H@10	Hits@10 (50 negs)
Supervised SOTA	0.342		0.482		0.348		0.494		0.344
ULTRA 0- shot	0.435		0.603		0.312		0.458		0.395
ULTRA fine- tuned	0.443		0.615		0.379		0.543		0.422

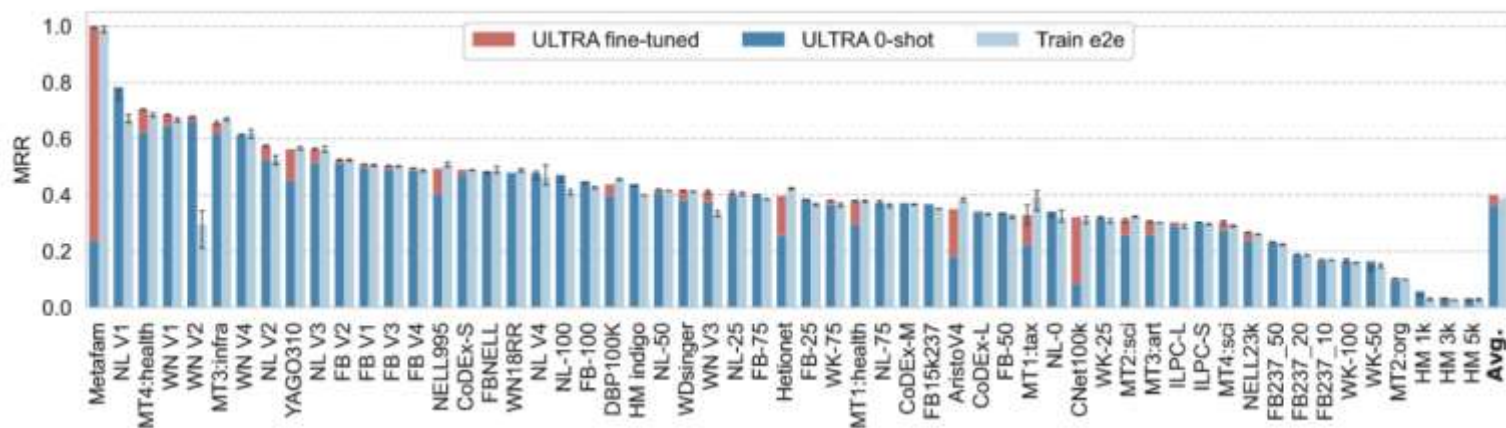
III-A Datasets & Evaluation

在按数据集类别分组的 57 个图表的完整集上进行零拍摄和微调的 ULTRA 结果。微调特别有助于较大的转导数据集，并将总平均 MRR 提高 10%。此外，我们将从头开始训练的特定于数据集的 ULTRA 模型在每个图形上的平均性能报告为 (训练 e2e)。

Model	Inductive e, r (23 graphs)	Inductive e (18 graphs)	Transductive (13 graphs)	Total Avg (54 graphs)	Pretraining					
	MRR	H@10	MRR	H@10	MRR	H@10	MRR	H@10	MRR	H@10
ULTRA (train e2e)	0.392	0.552	0.402	0.559	0.384	0.545	0.393	0.552	0.403	0.562
ULTRA o-shot	0.345	0.513	0.431	0.566	0.312	0.458	0.366	0.518	-	-
ULTRA fine-tuned	0.397	0.556	0.442	0.582	0.379	0.543	0.408	0.562	0.407	0.568

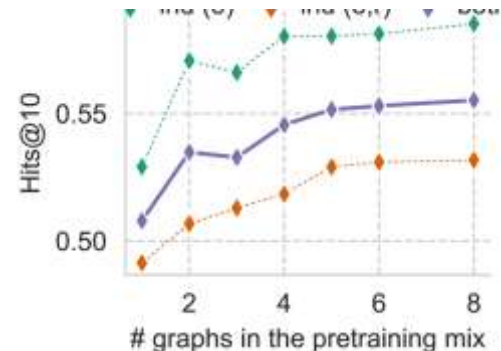
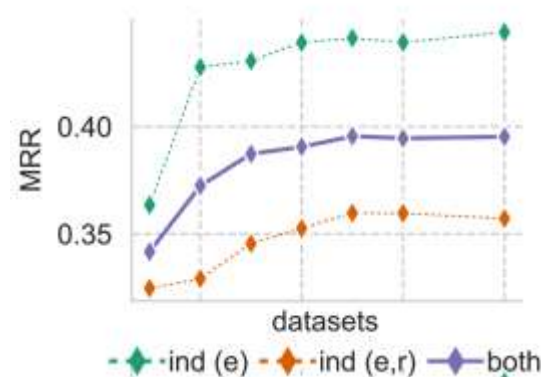
III-A Datasets & Evaluation

消融实验(验证一个系统、模型或理论的关键组成部分对整体系统性能的影响。)



零样本和微调的 ULTRA 每个数据集性能与在每个数据集上从头开始训练模型 (训练 e2e) 的比较。单个预训练模型的零样本性能与从头开始训练相当，而微调可产生总体最佳结果。

在零样本情况下，单个预训练的 ULTRA 模型的性能几乎与经过训练的单独模型相当，落后于较大的转导图上的模型，并且在归纳数据集上表现出更好的性能。微调预先训练的 ULTRA 总体上显示出最佳性能，并且与在每个目标图上从头开始训练模型相比，所需的计算资源要少得多。



III-A Datasets & Evaluation

消融实验(验证一个系统、模型或理论的关键组成部分对整体系统性能的影响。)

研究归纳推理性能如何取决于训练混合物。虽然主要的 ULTRA 模型是在三个图的混合上训练的，但在这里我们训练了更多的模型，将训练集中的 KGs 数量从单个 FB15k237 变为 8 个转导 KGs 的组合（更多细节见附录 C）。为了公平比较，我们仅在归纳数据集（总共 41 张图）上评估零镜头模式下的预训练模型。结果如图 1 所示。6 中，我们观察到混合中具有三个以上图形的性能饱和。我们假设获得更高的推理性能与模型容量、规模和优化有关。我们将这项研究与更有原则的方法一起留给为未来的工作选择训练前组合。

Model	Inductive <i>e, r</i> (23 graphs)	Inductive <i>e</i> (18 graphs) (13 graphs)	Transductive (54 graphs)	Total Avg (3 graphs)	Pretraining					
	MRR	H@10	MRR	H@10	MRR	H@10	MRR	H@10	MRR	H@10
ULTRA	0.345	0.513	0.431	0.566	0.312	0.458	0.366	0.518	0.407	0.568
- no etypes in rel. graph	0.292	0.466	0.389	0.539	0.258	0.409	0.316	0.477	0.357	0.517
- no etypes, - uncond. GNN (ones)	0.187	0.328	0.262	0.430	0.135	0.257	0.199	0.345	0.263	0.424
- no etypes, - uncond. GNN (random)	0.177	0.309	0.250	0.417	0.138	0.255	0.192	0.332	0.266	0.433

表明消融模型难以达到相同的预训练性能，并且在所有图组中都表现出较差的零镜头泛化性能，例如，在具有同质关系图和随机初始化节点状态的模型上，高达 48% 的相对 MRR 下降（0.192 对 0.366）使用无条件 R-GATv2 编码器。因此，我们假设条件表示（在关系和实体级别上）对于链接预测任务的可转移表示至关重要，这些任务通常需要成对表示来打破邻域对称性。

III-A Datasets & Evaluation

实验过程

1. 预训练： ULTRA在多个知识图谱上进行预训练，以学习通用的关系表示。
2. 零样本推理： 在多个数据集上进行链接预测，评估模型在未见过的图谱上的性能。
3. 微调： 对预训练的ULTRA模型进行微调，以适应特定数据集的特点。
4. 性能比较： 将ULTRA的性能与现有技术进行比较，展示其在不同数据集上的优势。

实验结果

1. 零样本性能： ULTRA在多个数据集上展示了出色的零样本推理能力，其性能往往与或优于特定数据集上训练的基线模型。
2. 微调性能： 微调后的ULTRA模型在大多数数据集上进一步提升了性能。
3. 泛化能力： 实验结果表明，ULTRA具有良好的泛化能力，能够在多种不同类型的知识图谱上进行有效的推理。（存在饱和点）

III-A Datasets & Evaluation

(1) 预先训练的 ULTRA 是否能够以零样本的方式归纳泛化到看不见的 KG?

通过在57个不同的知识图谱上进行链接预测实验，发现单个预训练的ULTRA模型在各种规模的未见图谱上的零样本（zero-shot）推理性能通常与在特定图谱上训练的强大基线相当或更好。这表明ULTRA确实能够在没有看到特定知识图谱的情况下进行有效的归纳和泛化。

(2) 在特定数据集上微调 ULTRA 是否有任何好处?

对预训练的ULTRA模型进行微调可以进一步提高其在特定数据集上的性能。微调后的ULTRA模型在大多数数据集上的性能得到了提升，这证实了微调对于适应特定任务和数据集是有益的。

(3) 单个预训练的 ULTRA 模型与在每个目标数据集上从头开始训练的模型相比如何?

通过广泛的实验验证了预训练的ULTRA模型与从头开始训练的模型相比，在多个数据集上的性能表现。实验结果表明，预训练的ULTRA模型在零样本推理性能上通常与或优于特定数据集上训练的基线模型。这显示了预训练模型在泛化能力和效率上的优势。

(4) 预训练组合中的图表越多，性能是否越好?

进行了消融研究，探讨了预训练数据集的多样性对模型性能的影响。研究发现，包含更多图表（即不同类型和领域的知识图谱）的预训练组合可以提高模型的泛化能力。然而，也指出了性能提升可能存在饱和点，即超过一定数量的图表后，性能提升可能不再显著。这暗示了在设计预训练策略时需要平衡数据集的多样性和训练的效率。

IV Conclusion



■ 干了什么

- 提出了 ULTRA，这是一种学习通用和可转移的图表示的方法，可以作为构建 KG 推理基础模型的方法之一。ULTRA 利用关系结构和条件关系表示的不变性，支持对任何多关系图进行训练和推理，而无需任何输入特征。

■ 实验表现

- 从实验上讲，即使在零样本制度下，单个预训练的 ULTRA 模型在 1k-120k 节点的 50+ 图表上也优于最先进的定制监督基线15%。微调 ULTRA 的采样效率很高，并进一步提高了平均性能10%。我们希望 ULTRA 有助于寻找归纳和可转移表示，其中单个预训练模型可以归纳泛化到任何图并执行各种下游任务。



```
[6] %pip install torch==2.1.0 --index-url https://download.pytorch.org/whl/cu118
%pip install torch-scatter==2.1.2 torch-sparse==0.6.18 torch-geometric==2.4.0
%pip install ninja easydict pyyaml
# 因为他这个是容器 每次（指每次开）使用都要重新装一下环境
```

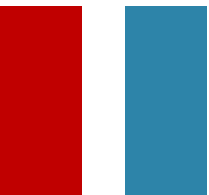
```
[7] %cd /content/drive/MyDrive/ULTRA-main
    %pwd
    /content/drive/MyDrive/ULTRA-main
    '/content/drive/MyDrive/ULTRA-main'
```

```
[8] !python script/run.py\
    -c config/transductive/inference.yaml\
    -dataset CoDexSmall --epochs 0 --bpe null --gpus [0]\
    --ckpt /content/drive/MyDrive/ULTRA-main/ckpts/ultra_4g.pth
```

```
Done!
08:44:49 CoDExSmall dataset
08:44:49 #train: 32888. #valid: 1827. #test: 1828
08:44:55 >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
08:44:55 Evaluate on valid
Load rpsmm extension. This may take a while...
08:46:51 mr: 38.5285
08:46:51 mrr: 0.477777
08:46:51 hits@1: 0.37329
08:46:51 hits@3: 0.529557
08:46:51 hits@10: 0.676793
08:46:51 >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
08:46:51 Evaluate on test
08:47:00 mr: 42.3884
08:47:00 mrr: 0.463778
08:47:00 hits@1: 0.360503
08:47:00 hits@3: 0.514497
08:47:00 hits@10: 0.665208
```

谢谢！

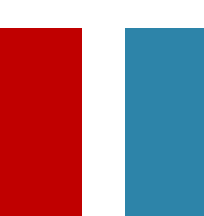




Towards Foundation Models for Knowledge Graph Reasoning

知识图谱推理的基础模型

ULTRA

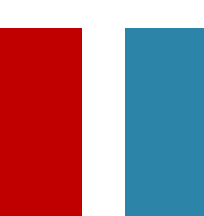


论文的主要研究目标

是找到能够跨越具有任意实体和关系词汇表的图谱的可转移的不变性，并利用这些不变性来实现知识图谱推理的基础模型。

论文核心

找到具有任意实体和关系词汇的图之间可转移的不变性。利用和学习这种不变性将使 KG 推理的基础模型能够进行预训练和微调范式。其中在一个或多个图上用特定的关系集合进行训练的单个模型将能够零样本学习转移到任何具有完全不同的关系和关系模式的新的看不见的图。



可转移的不变性

ULTRA 将**关系表示**构建成为以交互为条件的函数。

找到具有任意实体和关系词汇的图之间可转移的不变性

1. 关系图的构建

首先，ULTRA模型会构建一个关系图，其中每个节点代表知识图谱中的一个关系类型，而边则表示这些关系类型之间的交互。这些交互是基于知识图谱的结构，而不是基于具体的实体实例。

2. 交互类型

在关系图中，交互类型包括尾到头 (t2h)、头到头 (h2h)、头到尾 (h2t) 和尾到尾 (t2t)。这些交互类型捕捉了知识图谱中不同关系如何相互连接的方式。

3. 条件关系表示

利用图神经网络 (GNN)，ULTRA模型会在关系图上学习每个关系类型的表示。这些表示是“条件”的，因为它们是基于关系之间的交互来学习的。换句话说，每个关系的表示不仅取决于它自身，还取决于它与其他关系的交互。

4. 函数化表示

这种表示可以被视为一种函数，输入是关系的交互模式，输出是该关系的条件表示。这个函数是通过图神经网络学习得到的，它能够捕捉到关系之间的复杂和动态的交互模式。



预训练和微调范式

利用和学习这种不变性将使 KG 推理的基础模型能够进行预训练和微调范式，

预训练是在一个大型的、多样化的数据集上进行的，学习通用的特征和模式。这些特征和模式可以转移到各种不同的任务上。

微调是在预训练模型的基础上，针对特定的任务或数据集进行额外的训练，以提高模型在该任务上的性能。在预训练阶段，ULTRA模型学习了丰富的、通用的关系表示，这些表示在结构上是通用的。

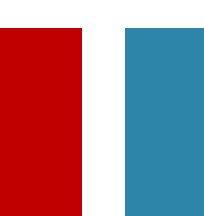


零样本学习

在一个或多个图上用特定的关系集合进行训练的单个模型，将能够零样本学习转移到任何具有完全不同的关系和关系模式的新的看不见的图

新的知识图谱：新的数据集，它包含了实体（节点）和这些实体之间的关系，在模型的训练过程中没有被使用过，不同的领域或上下文，不同的结构特征

链接预测怎么推理的：在预测过程中，头节点被初始化为查询关系向量，而其他节点则初始化为零。这种方法使得模型能够利用学习到的关系表示来预测知识图谱中的链接。



方法步骤

构建关系图

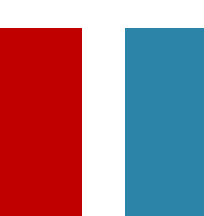
将原始图谱G中的每个关系类型作为一个节点，构建一个新图谱Gr，其中的边表示原始图中关系之间的交互（如尾到头、头到头等）

条件关系表示

应用一个标签技巧（例如，将查询关系节点初始化为一个特殊的向量），并使用GNN进行消息传递。为每个关系类型生成一个相对于查询关系的条件表示。

实体级链接预测

利用条件关系表示作为输入特征，结合一个链接预测模型（如NBFNet）。在链接预测过程中，头节点被初始化为查询关系向量，其他节点初始化为零。



方法步骤

预训练

在一个或多个知识图谱上训练ULTRA模型（在包含多种关系和实体的数据集上训练模型），以学习通用的关系表示。

微调

在特定数据集上进一步训练ULTRA模型，以提高其在该数据集上的性能(调整模型参数以适应特定任务或数据集的特点)。

零样本推理

直接将预训练的ULTRA模型应用于新的、未见过的知识图谱。

谢谢！

