

基于知识协同微调的低资源知识图谱补全方法*

张宁豫^{1,2}, 谢辛^{1,2}, 陈想^{1,2}, 邓淑敏^{1,2}, 叶宏彬^{1,2}, 陈华钧^{1,2}



¹浙江大学 AZFT 知识引擎实验室, 浙江 杭州 310028)

²浙江大学 杭州国际科创中心, 浙江 杭州 310028)

通信作者: 陈华钧, E-mail: huajunsir@zju.edu.cn

摘要: 知识图谱补全能让知识图谱变得更加完整. 现有的知识图谱补全工作大多会假设知识图谱中的实体或关系有充足的三元组实例. 然而, 在通用领域, 存在大量长尾三元组; 在垂直领域, 较难获得大量高质量的标注数据. 针对这一问题, 提出了一种基于知识协同微调的低资源知识图谱补全方法. 通过已有的结构化知识来构造初始的知识图谱补全提示, 并提出一种协同微调算法来学习最优的模板、标签和模型的参数. 所提方法同时利用了知识图谱中的显式结构化知识和语言模型中的隐式事实知识, 且可以同时应用于链接预测和关系抽取两种任务. 实验结果表明, 该方法在 3 个知识图谱推理数据集和 5 个关系抽取数据集上都取得了目前最优的性能.

关键词: 低资源; 知识图谱补全; 链接预测; 关系抽取; 预训练语言模型

中图法分类号: TP18

中文引用格式: 张宁豫, 谢辛, 陈想, 邓淑敏, 叶宏彬, 陈华钧. 基于知识协同微调的低资源知识图谱补全方法. 软件学报, 2022, 33(10): 3531–3545. <http://www.jos.org.cn/1000-9825/6628.htm>

英文引用格式: Zhang NY, Xie X, Chen X, Deng SM, Ye HB, Chen HJ. Knowledge Collaborative Fine-tuning for Low-resource Knowledge Graph Completion. Ruan Jian Xue Bao/Journal of Software, 2022, 33(10): 3531–3545 (in Chinese). <http://www.jos.org.cn/1000-9825/6628.htm>

Knowledge Collaborative Fine-tuning for Low-resource Knowledge Graph Completion

ZHANG Ning-Yu^{1,2}, XIE Xin^{1,2}, CHEN Xiang^{1,2}, DENG Shu-Min^{1,2}, YE Hong-Bin^{1,2}, CHEN Hua-Jun^{1,2}

¹(AZFT Joint Laboratory for Knowledge Engine, Zhejiang University, Hangzhou 310028, China)

²(ZJU-Hangzhou Global Scientific and Technological Innovation Center, Hangzhou 310028, China)

Abstract: Knowledge graph completion can make the knowledge graph more complete. Unfortunately, most of existing methods on knowledge graph completion assume that the entities or relations in the knowledge graph have sufficient triple instances. Nevertheless, there are great deals of long-tail triple sin general domains. Furthermore, it is challenging to obtain a large amount of high-quality annotation data in vertical domains. To address these issues, a knowledge collaborative fine-tuning approach is proposed for low-resource knowledge graph completion. The structured knowledge is leveraged to construct the initial prompt template and the optimal templates, labels, and model parameters are learnt through a collaborative fine-tuning algorithm. The proposed method leverages the explicit structured knowledge in the knowledge graph and the implicit triple knowledge from the language model, which can be applied to the tasks of link prediction and relation extraction. Experimental results show that the proposed approach can obtain state-of-the-art performance on three knowledge graph reasoning datasets and five relation extraction datasets.

Key words: low-resource; knowledge graph completion; link prediction; relation extraction; pre-trained language model

知识图谱是一种用图结构来描述知识和建模世界万物之间关联关系的大规模语义网络, 是大数据时代知识表示的重要方式之一. 知识图谱已被广泛应用于各种智慧信息系统中, 辅助深度理解人类语言和支持推理,

* 基金项目: 国家自然科学基金(91846204, U19B2027)

本文由“智慧信息系统新技术”专题特约编辑邢春晓研究员、王鑫教授、张勇副研究员、于戈教授推荐.

收稿时间: 2021-07-20; 修改时间: 2021-08-30; 采用时间: 2021-12-24; jos 在线出版时间: 2022-02-22

提升人机问答的用户体验, 促进智能化软件开发^[1-3]等。同时, 知识图谱在金融、农业、电商、医疗健康、环境保护等大量垂直领域也都得到广泛的应用^[4-6]。然而, 现有的知识图谱并不完整, 存在语义信息缺失问题。知识图谱补全工作能让知识图谱变得更加完整, 一般可通过对已有知识图谱的链接预测或基于文本的关系抽取等方法来实现^[7]。

现有的知识图谱补全工作大多假设知识图谱中的实体或关系有充足的三元组实例来训练以得到向量表示, 因此需要大量的人工标注样本。但在实际场景的知识图谱中, 大量的实体或关系仅具有非常少的三元组, 即存在低资源问题。对于大量垂直领域, 如新基建、工业制造、医疗健康等, 由于其领域的专业性, 大量的关系只有非常少的训练语料。如果进一步考虑常识性知识图谱和多模态跨媒体知识图谱的构建, 则会加剧低资源问题。知识图谱构建的低资源问题已经严重制约了图谱构建的效率和性能。如图 1 所示, 在关系预测的推理任务中, 图谱中频率较高的关系明显优于频率较低的关系; 在关系抽取的补全任务中, 预测精度随着关系样本数目的减少而大幅下降。因此, 研究低资源知识图谱补全问题, 有助于促进自动化知识图谱构建的发展, 具有广泛的学术和应用价值。

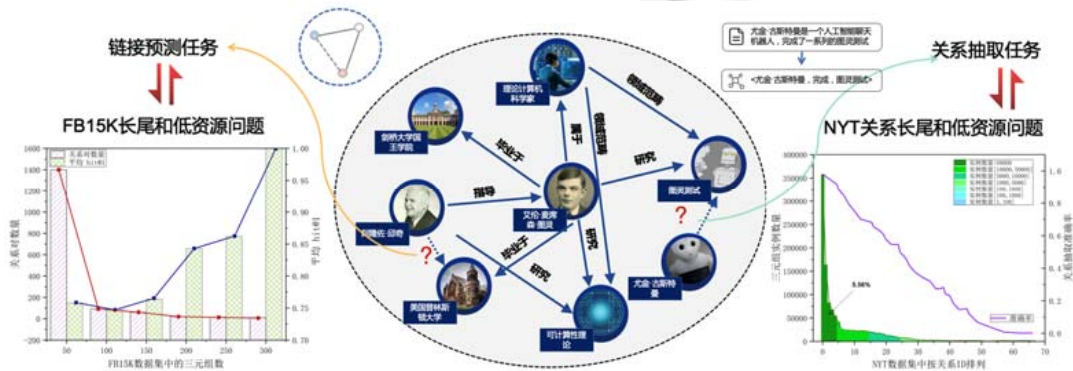


图 1 知识图谱补全的低资源问题: 以链接预测和关系抽取为例

低资源知识图谱补全吸引了诸多研究学者, 其方法一般可以划分为基于知识迁移的方法^[8-10]和基于元学习的方法^[11,12]。基于知识迁移方法的核心思路在于通过知识图谱的不同关系之间存在的显式或隐式的关联, 将头部高频样本的知识迁移到尾部的低频样本中去, 进而改进低资源知识图谱补全的效果。然而, 对于那些关联较弱的关系甚至于没有关联的关系, 这类方法较难迁移知识, 且迁移学习可能存在负迁移效应^[9]。基于元学习的方法主要通过构造若干少样本来学习元任务, 利用元任务学到的知识经验来指导新任务的学习, 使其具有学会学习的能力。然而, 元学习需要构建大量的元任务, 且元任务这一设定使其较难适应复杂多变的真实场景^[12]。

近年来, OpenAI、北京智源研究院、华为公司等分别提出了超大规模预训练语言模型 GPT-3^[13]、悟道^[14]、盘古^[15], 其仅通过给定提示词(prompt)和少量样本实例即可实现较好的任务性能, 给低资源学习带来新思路。Gao 等学者提出的提示词微调方法 LM-BFF^[16]进一步展现出预训练语言模型的少样本学习能力。一方面, 基于提示词微调的方法可以减小预训练和微调之间的任务差异, 并极大地提升了低资源场景下的模型性能; 另一方面, 已有工作^[17,18]表明, 预训练语言模型中已经学到了部分事实知识, 而提示词微调可以激活并利用语言模型中的隐式知识。因此, 显然可以借助预训练语言模型提示词微调来提升低资源知识图谱补全性能。然而, 不同的提示词对模型的性能有较大影响, 且与自然语言处理任务不同, 知识图谱本身就存在着大量的结构化知识约束。

针对上述问题, 本文提出了一种新颖的知识协同微调的低资源知识图谱补全方法(KnowCo-Tuning), 其通过已有的结构化知识来构造初始的知识图谱补全提示, 并通过一种协同微调算法来学习最优的模板与标签以及模型的参数。本文方法同时利用了知识图谱中的显式结构化知识和语言模型中的隐式知识, 且可以同时应用于链接预测和关系抽取任务。我们在 3 个知识图谱推理和 5 个关系抽取数据集中验证了本文的模型效果。实

验结果表明, 模型取得了目前最优的低资源知识图谱补全性能。

本文第 1 节对低资源知识图谱补全的方法进行概述, 并系统地介绍链接预测和关系抽取相关的方法。第 2 节介绍本文的知识协同微调方法的模型结构和优化算法。第 3 节介绍实验, 包含链接预测和关系抽取两部分。最后, 在第 4 节总结全文, 并对未来值得关注的研究方向进行初步的探讨。

1 相关工作

1.1 低资源知识图谱补全

知识图谱补全可以通过知识图谱的链接预测^[19]或者从语料文本中抽取新的关系^[20]来实现。对于链接预测, 学术界提出了各种不同的方法来将实体和关系编码到一个低维的连续空间当中。TransE^[21]将三元组 (h, r, t) 中的关系 r 看作从 h 到 t 的转换过程。TransE 取得了很好的结果并具有很多拓展的工作, 包括 TransR^[22]、TransD^[23]、TransH^[24]等。RESCAL^[25]提出了基于矩阵分解的知识图谱补全方法, 其采用双线性形式作为知识图谱嵌入的评分函数。DistMult^[26]分析了矩阵编码关系并简化了 RESCAL 方法, 而 ComplEx^[27]将 DistMult 扩展到了复数域中。ConvE^[28]使用了卷积神经网络来作为评分函数。Analogy^[29]训练了嵌入实体的类别属性和关系的隐变量表示。KGGAN^[30]提出了一种基于对抗生成网络的知识图谱嵌入方法。RotatE^[31]将每个关系定义为从复数向量空间中的源实体到目标实体的变换旋转。KG-BERT^[32]将三元组中实体和关系的描述作为输入然后利用 BERT 来计算三元组的得分。在低资源场景中, Xiong 等人^[33]提出了一个 1-shot(每个关系给一个训练样本)关系学习框架, 其通过学习到的嵌入表示以及单跳图结构来学习一个匹配矩阵。Chen 等人^[11]提出 MetaR 框架来进行少样本情景下的知识图谱中的链接预测。Zhang 等人^[17]提出一种 IterE 方法以连续地学习知识图谱嵌入与规则, 其有效地提高了低资源实体嵌入的质量及其链接预测的性能。Zhang 等人^[34]提出了一种建模异构图表示和融合不同的支持三元组的低资源链接预测方法。Sheng 等人^[35]提出了一种自适应注意力机制的低资源知识图谱推理算法。Zhang 等人^[9]提出关系对抗网络 RAN, 以将头部关系的知识迁移到尾部关系, 从而提升低资源知识图谱补全的效果。Zhang 等人^[12]提出一种基于建模关系间语义关联的低资源知识图谱补全模型 GRL。

关系抽取旨在基于给定的上下文文本来判别实体间存在的关系。为了降低标注成本, 早期学者提出了基于知识图谱自动生成标注实例的远程监督方法^[36]。近年来, 低资源情境下的关系抽取越来越受到学者的关注。Zhang 等人^[8]提出基于图神经网络和知识图谱嵌入的方法, 将头部高频关系的知识迁移到尾部低频关系以改善低资源关系抽取效果。Gao 等人^[37]提出利用基于混合注意力的原型网络进行少样本关系抽取。Ye 等人^[38]提出基于原型网络的多层次匹配和聚合进行少样本关系抽取。Gao 等人^[39]提出基于关系孪生网络的少样本关系抽取框架。Yu 等人^[40]提出基于多原型嵌入的方法进行少样本关系三元组抽取。Soares 等人^[41]提出利用 BERT 进行从关系语句到关系表示的映射学习, 不需要任何知识图谱或注释的关系表示的方法来实现少样本关系抽取。Geng 等人^[42]提出了一种基于元学习的少样本关系抽取框架, Qu 等人^[43]提出了一种基于关系后验分布的贝叶斯元学习模型来进行少样本关系抽取。Dong 等人^[44]提出了一种基于关系语义指导的元学习少样本关系抽取方法。然而, 上述方法大多受到关系约束或样本个数的限制。本文的方法放宽了那些可能与目标关系完全不同来源的关系标签约束, 且可用于任意 Shot(每个关系所包含的训练样本)的情况, 更加适用于真实场景。

1.2 预训练语言模型微调

预训练微调已经成为了自然语言处理的新范式, 并在大量的任务中取得了优异的效果。超大规模预训练模型 GPT-3 甚至具备一定程度的少样本学习能力, 其模型能在多任务学习以及少样本学习中表现出杰出的性能。然而, GPT-3 严重依赖于上下文学习^[45,46], 以此来迁移知识适应到下游的任务当中。此外, GPT-3 模型有着 1.75 MB 的参数, 由于算力的限制使得在实际应用场景中无法使用该模型。

为了更好地将大型语言模型应用于自然语言理解任务当中, 最近的工作主要探讨了利用提示对下游自然语言处理任务进行微调, 以更好地利用来自预训练语言模型的知识。具体来说, 基于提示预测可将下游任务视为掩码语言建模(MLM)问题, 语言模型可以像掩码词预训练一样, 输出预测的词汇即可。提示包含两部分,

第一是模板, 由自然语言组成用来提示模型的输出; 第二是标签词, 表达了如何将模型预测出的词汇转化成每一个类别分数. 比如, PET^[47]首次通过离散搜索的方式构造模板和标签词. LM-BFF^[14]使用 T5^[48]模型来生成模板以及在词表中搜索标签词. 然而, 这些方法仍需额外地生成模型, 且其标签词搜索过程的计算复杂度很高, 尤其是在涉及多输出类的任务上. Auto-Prompt^[49]提出一个基于梯度来选择标签词和模板的搜索方法, 而这需要大量的样本来进行训练. P-tuning^[45]、Prefix-tuning^[50]以及 OPTIPROMPT^[51]学习连续的模板, 其十分有效并且优于离散提示搜索. 以上方法主要适用于自然语言处理任务并在多个基准数据集上取得最优效果. 不同于这些工作, 本文主要借助外部知识来构建提示, 并在知识图谱补全任务中同时考虑知识图谱的显式知识和语言模型隐式知识.

低资源知识补全任务与上述自然语言处理任务存在一些显著差异. 首先, 该任务通常涉及大量关系标签而不是像其他自然语言处理任务中的几个类. 第二, 知识图谱补全任务中的关系类型(例如, “别名”“出生地”等)一般都有特定的语义信息, 很难指定一个模型词表中单一的词来表示它. 第三, 知识图谱包含大量的结构化先验知识, 知识图谱补全任务中的实体是自然的提示词, 正好适合作为基于模板微调方法中的模板. 结合以上分析, 如何获取知识图谱补全的标签词具有一定的挑战性, 并且之前的工作也没有考虑如何基于已有的结构化知识和数据获取最优的标签词汇. 据我们所知, 本文所提方法是第一个基于提示微调的低资源知识图谱补全方法. 此外, 对比之前的方法, 本文还放开了低资源的设定, 每个关系(类)均可以包含任意标注数量(比如 8 或 16 等任意数量)的样本, 而非指定的 K -shot (每个关系都包含 K 个训练样本), 因此更加适用于真实场景.

2 知识协同微调方法: KnowCo-Tuning

我们首先在第 2.1 节介绍低资源知识图谱补全的任务定义, 并在第 2.2 节和第 2.3 节介绍基于知识指导的提示模板构建和标签组合. 然后在第 2.4 节引入了模板和标签微调, 它为知识图谱补全任务构建自适应提示, 以便更好地利用预训练语言模型中所存储的事实知识. 最后在第 2.5 节介绍协同微调算法. 此外, KnowCo-Tuning 还引入了一个新的训练目标: 实体判别任务, 以进一步促进模型的训练. 整体模型如图 2 所示.

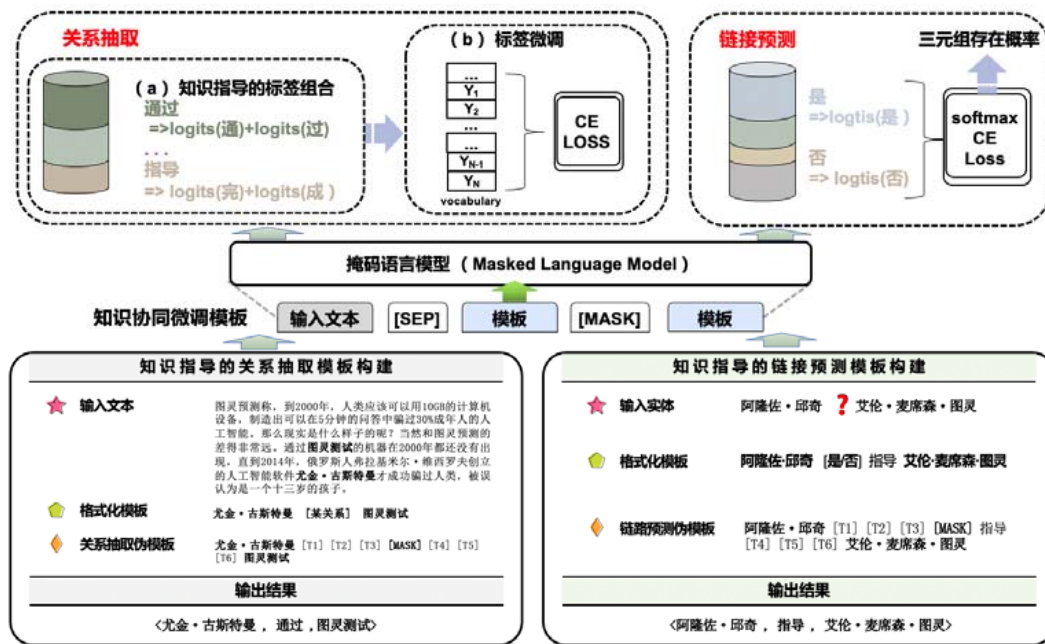


图 2 知识协同微调

图 2 中, T_i 和 Y_i 均为词典中已有的未使用或特殊字符, 且均为可学习的字符. 本文将输入文本和模板一同送入到掩码语言模型中. 针对链接预测任务, 基于模型在 [MASK] 的输出来判断三元组成立的概率并与真实标签计算交叉熵损失函数. 针对关系抽取任务, 本文将模型在 [MASK] 上预测的类别基于标签映射函数转化为类别标签, 并与真实属性标签计算交叉熵损失函数.

2.1 任务定义和背景

令 $X_m = \{x_1, x_2, \dots, X_s, \dots, X_o, x_N\}$ 是一个输入序列, 其中, x_n 是序列中第 n 个输入. 在低资源知识图谱补全任务中: 链接预测(输入只有头尾实体和关系的文本描述, 没有描述关系上下文的句子)的目标是判断头实体 X_s 和尾实体 X_o 之间的关系 r 是否成立. 而对于关系抽取任务(输入是包含头尾实体的描述关系上下文的句子), 任务目标是预测头实体 X_s 和尾实体 X_o 之间的关系 r , 其中, $r \in R$ 并且 R 是数据集预先定义的关系集合. 本文的方法统一了模型输入, 可以同时应用于链接预测和关系抽取. 针对低资源知识图谱补全场景, 基于提示词微调的方法将分类任务转换成掩码语言模型 (MLM) 的字符预测任务, 它需要解决以下两个核心问题.

- (1) 如何构建模板句子 T 以引导模型预测对应的字符.
- (2) 如何构造特定的标签词 $\mathcal{M}(Y)$.

本文将模板和标签词汇统一称为提示(prompt) \mathcal{P} . \mathcal{V} 指的是语言模型的词汇表, 而 $\mathcal{M}: Y \rightarrow \mathcal{V}$ 是标签关系到标签词汇的映射函数.

预训练微调(比如使用 [CLS] 向量作为分类器的输入)的输入通常为 $\tilde{X}_m = [\text{CLS}]X_m[\text{SEP}]$. X_m 会被转换成固定的 token 序列 \tilde{X}_m , 之后, 基于语言模型 \mathcal{L} 将 \tilde{X}_m 转化为一个向量序列 $\{\mathbf{h}_k \in \mathbb{R}^d\}$. 但是对于本文方法 KnowCo-Tuning 来说, 输入 X_{prompt} 中包含之前预训练任务特有的特殊字符 [MASK], 即本文将输入转为以下形式:

$$X_{\text{prompt}} = [\text{CLS}] X_m [\text{SEP}] T [\text{SEP}].$$

将 X_{prompt} 输入到预训练模型当中, 本文可以得到在特殊字符 [MASK] 位置上预测的离散分布 $p([\text{MASK}] | X_{\text{prompt}})$. 然后, 本文使用映射函数 \mathcal{M} 就可以得到每一个关系类别 y 对应的概率如下:

$$p(y | X_{\text{prompt}}) = \sum_{w \in \mathcal{M}(y)} p([\text{MASK}] = w | X_{\text{prompt}}),$$

其中, $\mathcal{M}(y)$ 表示的是每一类 y 对应的标签词汇. 特别地, 对于关系抽取, 本文采用特殊的字符 $(\langle e1 \rangle, \langle /e1 \rangle)$ 来标记头尾实体而不使用额外的信息, 比如实体类别.

2.2 知识指导的模板构建

本文首先基于知识图谱中的结构化知识为知识图谱补全任务生成模板. 事实上, 预训练语言模型可以通过手工构建模板句子而无需任何额外训练来预测正确的关系^[15,16]. 遵循此范式, 本文基于实体知识构造模板, 以利用语言模型经过大规模语料预训练存储的隐式知识. 比如, 本文定义知识图谱补全任务特定的模板 T 为

$$T = X_s \text{ is the } [\text{MASK}] \text{ of } X_o,$$

其中, X_s 和 X_o 分别是头尾实体. 显然, 模板构造非常简单, 且具有一定程度的可解释性和直观性.

2.3 知识指导的标签组合

知识图谱补全任务中关系标签词汇本身具有大量的语义信息. 受此启发, 本文考虑使用一对多的映射函数 \mathcal{M} 来更好地表达标签的语义信息特点(比如“父”“母”都可以代表“祖辈”这一关系). 所以映射函数 \mathcal{M} 可以表达为

$$\mathcal{M}(Y_j) = \{v_1, \dots, v_k\}, v \in \mathcal{V},$$

其中, v 表示在语言模型字典中的字. 通过该映射函数, 本文可以利用关系词本身来建立特定的标签词. 比如对于“出生地”这个关系, 本文将其表达为“出”“生”和“地”3 个字来表示. 那么, 本文可以将标签概率组合为如下形式:

$$p(y|X_{\text{prompt}}) = \frac{\exp(W_{M(y)} \cdot h_{[\text{MASK}]})}{\sum_{y' \in Y} \exp(W_{M(y')} \cdot h_{[\text{MASK}]})}$$

其中, $h_{[\text{MASK}]}$ 是语言模型 \mathcal{L} 输出 [MASK] 位置对应的特征向量, 其代表了模型 \mathcal{L} 在该位置预测的字含有的语义. $W_{M(y)}$ 代表了标签词汇对应的特征矩阵, 两者相乘得到了语言模型预测词汇与特征词汇的相似程度的度量指标. 上文图 2 中已展示出本文 KnowCo-Tuning 离散标签词汇的映射过程, 其训练过程简单而有效并且在原理上避免了预训练和下游任务之间训练目标不一致的问题.

2.4 模板和标签微调

由于知识驱动的模板仍然是人工构建的离散句子, 且标签组合在标签词较短时无法较好地表达标签含义. 因此, 本文进一步提出了模板和标签微调技术, 以进一步提升模板的质量, 进一步可以让 KnowCo-Tuning 应用于更多的下游任务中.

2.4.1 自适应模板微调

对于不同任务, 不同标签需要不同的模板, 这不仅费时费力, 还容易陷入局部最优, 即可能模板符合训练集但却与测试集文本不一致而导致模型性能不佳. 因此, 本文在模板中增加一些随机初始化的字符, 然后固定其他已有的字符参数来学习这些特殊字符, 以得到最优的模板. 给定模板 $\mathcal{T} = \{X_s, [T_{i:i}], [\text{MASK}], [T_{i+1:m}], X_o\}$, 本文将模板转化为

$$\{e(X_s), e([T_{i:i}]), e([\text{MASK}]), e([T_{i+1:m}]), e(X_o)\}.$$

本文通过预训练模型将输入的字符串转化成为向量, 对于新加入的字符, 本文采用随机初始化:

$$\{e(X_s), h_1, \dots, h_i, e([\text{MASK}]), h_{i+1}, \dots, h_m, e(X_o)\},$$

其中, $h_i (1 \leq i \leq j)$ 是可学习的词向量参数. 通过训练这些任务特有的参数, 可以在连续空间找到更加适合下游任务(如链接预测、关系抽取)的模板句子. 最终, 本文通过计算损失函数 \mathcal{J} 以优化这些任务特有的词向量:

$$\hat{h}_{i:m} = \arg \min_h \mathcal{J}(\mathcal{L}(X_{\text{prompt}}, y)).$$

不同于之前的 P-tuning 方法^[41], 其采用双向 LSTM 模型以及两层 MLP 来优化模板向量, 本文没有增加任何额外的参数就可以有效地获得高可用的模板句子.

2.4.2 自适应标签微调

对于每一个类别 $y \in Y$, 以前的方法 LM-BFF^[12] 基于预训练语言模型 \mathcal{L} 的条件似然去寻找 k 个能够最大化 $p(y)$ 的标签词. 然而仅仅依靠语言模型去搜索标签词汇会产生以下问题: (1) 当标签词汇变多时, 标签词汇搜索空间呈指数增长, 而知识图谱补全任务的类别(关系)非常多, 这会导致产生 $O(k^Y)$ 的搜索复杂度, 从而极其低效. (2) 该方法中搜索标签词汇需要花费大量的时间和计算资源. 其他的优化方法, 比如 AutoPrompt^[45] 通过字梯度搜索算法得到标签词汇, 这也需要大量的资源且在少样本场景下效果不佳.

考虑到关系标签包含一定程度的语义信息, 本文提出自适应标签微调以在连续的词汇空间中搜索最佳标签词. 具体来说, 给定关系标签 $Y = \{Y_1, Y_2, \dots, Y_m\}$, 与离散空间的基本版本可变数量的标签词 $\{v_1, v_2, \dots, v_k\}$ 相比, KnowCo-Tuning 将标签词汇定义为可学习的特殊词, 其表示为

$$\mathcal{M}(Y_j) = \{v_j\}, v_j \in \mathcal{V}_{\text{label}},$$

其中, $\mathcal{V}_{\text{label}}$ 为标签词汇表, 且 $\mathcal{V}_{\text{label}} \cap \mathcal{V} = \emptyset$. 具体实现中, 本文将 BERT 模型原先词表中的 “[unused]” 字替换成为本文的特殊标签字(对于不存在 unused 字的词表, 如 GPT-2 词表, 本文在词表中加入额外的字来表示特殊字).

2.5 协同学习算法

下面介绍协同学习算法的过程. 算法 1 显示了 KnowCo-Tuning 的具体优化过程. 首先基于知识驱动的模板构建和标签组合得到初始的提示(prompt), 并随机初始化其余部分待优化模板词和标签. 然后固定原始模

型的所有参数, 并只优化模板特殊字以及标签词特殊字 $\{h_1, \dots, h_m, \dots, h_{m+n}\}$ (3-7 行). 最后, 优化模型全部的参数(8-12 行). KnowCo-Tuning 在构造模板时引入了知识图谱显式的结构先验, 在后续优化过程中又充分利用了预训练语言模型中的隐式事实知识, 且无需搜索标签词. 特别地, 本文方法没有额外增加任何参数, 并充分利用了已有知识图谱和语言模型的能力.

算法 1. KnowCo-Tuning 算法.

设: $\mathcal{J}(\theta)$: 目标函数

t : 训练步数

θ_t : 在步数 t 时模型参数

α : 学习率

β : 学习率

$h \in \theta$: 模板句子和标签词汇对应词嵌入参数

1: 随机初始化 h_0

2: 初始化训练步数 $t \leftarrow 0$

3: 当 h_t 未收敛:

4: $t \leftarrow t+1$

5: $g_t \leftarrow \nabla_{\theta} \mathcal{J}_t(\theta_{t-1})$

6: $h_t \leftarrow \text{AdamW}(g_t, \alpha)$

7: 结束第 1 阶段训练

8: 当 θ_t 未收敛:

9: $t \leftarrow t+1$

10: $g_t \leftarrow \nabla_{\theta} \mathcal{J}_t(\theta_{t-1})$

11: $\theta_t \leftarrow \text{AdamW}(g_t, \beta)$

12: 结束第 2 阶段训练

由于本文的 KnowCo-Tuning 方法中的部分提示是随机初始化的, 提示句子的词嵌入应该相互依赖而非互相独立. 因此, 本文引入一个辅助用于训练的实体判别器优化目标. 总的来说, 本文有两个目标函数: 关系判别目标函数 \mathcal{J}_R 和实体判别目标函数 \mathcal{J}_E .

关系判别目标函数是知识图谱补全任务的主要目标函数, 其目的在于让模型能够拟合在本文知识图谱补全任务的数据集上. 给定输入 (X_{in}, \mathcal{T}) , 可以得到模型真实输入 X_{prompt} 从而计算损失函数:

$$\mathcal{J}_R = CE(p(y | X_{\text{prompt}})),$$

其中, CE 代表着交叉熵损失函数. 为了保留预训练语言模型本身预测掩码字的能力从而缓解灾难性遗忘, 且让部分随机初始化的提示非互相独立, KnowCo-Tuning 增加了一个实体判别目标来训练预训练语言模型通过句子中的其他信息来理解实体的意思. 本文在模型输入中随机掩盖一个实体, 然后利于预训练任务目标让语言模型去预测该实体. 设 x' 为随机掩盖实体 X_e 中的字, x^m 为除实体外的一系列输入, 本文通过计算以下公式得到 \mathcal{J}_E :

$$q(x^m | x', y) = \text{Softmax}(\mathcal{L}(x', y)),$$

$$\mathcal{J}_E = \sum_{m \in M} BCE(q(x^m | x', y)),$$

其中, $\mathcal{L}(x', y)$ 为语言模型在掩盖字上输出的每一个字的分数, 通过 Softmax 变换得到模型预测每一个字出现的概率. BCE 表示二值交叉熵损失函数, 这里使用了 Softmax 加上 BCE 的组合, 为的是参数能够更新词表上所有的字. 通过优化 \mathcal{J}_E 模型可以由上下文学习到丰富的实体语义信息. 最终通过以下公式得到最终的目标函数:

$$\mathcal{J} = \mathcal{J}_R + \lambda \mathcal{J}_E,$$

其中, λ 是一个超参数.

3 实验

本节详细介绍了本文方法在两个知识图谱上的补全任务: 链接预测和关系抽取上的表现. 本文测试了各个领域不同规模的低资源链接预测和关系抽取任务. 实验结果分析表明, 本文提出的方法 KnowCo-Tuning 在全量以及低资源情境下均表现较好.

3.1 数据集介绍

对于链接预测, 本文选取了几个标准的知识图谱补全数据集, 其中包含了多种类别的 Umls^[52]、基于 WordNet^[53] 的 WN18RR 以及基于 Freebase^[54] 的 FB15K-237. FB15K 是大规模知识图谱 Freebase 的子集. 而 FB15K-237 中删去了可以反向的三元组, 反向三元组对于模型来说更加难以判别. WN18RR 相比于 FB15K, 实体数量更多, 而关系类型更少, 同时形成的三元组数量也更少, 每一个节点连接较少, 更加复杂. 而 Umls 数据集由于是领域数据集, 所以实体相较于前两者较少. 具体详情见表 1.

对于关系抽取, 本文选择了各种不同的数据集, 包括了句子级别的抽取数据集, 如 TACRED-Revisit^[55]、SemEval-2010 Task 8^[56]、Wiki 80^[57] 和 ChemProt^[58]. 其中, ChemProt 是蛋白质领域的关系抽取任务. 此外, 也在对话级别的抽取任务中进行了测试, 比如 DialogRE^[59]. 各个关系抽取数据集的详细统计情况见表 2.

表 1 知识图谱推理数据集统计

数据集名称	实体数	关系数	训练集(随机采样)	验证集	测试集
Umls	135	46	329	652	661
WN18RR	40 943	18	88	3 034	3 134
FB15K-237(mini)	14 541	237	1 896	17 535	2 000

表 2 关系抽取数据集统计

数据集名称	关系数	样本数	无关系样本
TACRED-Revisit	42	106 264	79.5%
Wiki 80	80	56 000	-
SemEval-2010 Task 8	19	10 717	17.4%
DialogRE	36	1 788	-
ChemProt	13	10 065	-

3.2 实验设定

在完全监督的设定中, 采用了完整的 $\mathcal{D}_{\text{train}}$ 微调; 然而, 在低资源场景下, 传统的微调方法效果较差. 此外, 基于小数据集进行微调可能会导致模型训练的不稳定, 且在不同数据集下, 结果方差很大. 因此本文借鉴 LM-BFF^[12] 的设定, 通过使用 5 个不同随机种子得到的随机采样的小样本 $\mathcal{D}_{\text{train}}$, 并计算 5 次实验的均值来评估模型的性能. 在每次实验中, 使用固定的种子集合获得不同的样本取平均得到结果. 对于 $s \in \mathcal{S}_{\text{seed}}$, 得到对应的训练验证集合 $\{\mathcal{D}_{\text{train}}^s, \mathcal{D}_{\text{dev}}\}$. 在此基础上以一些超参数设定进行网格搜索并且在验证集 \mathcal{D}_{dev} 上选取最好的参数. 实验所使用的基础模型是无大小写区分的 BERT-Large^[58] 模型, 特别地, 本文使用 SCIBERT^[60] 在 ChemProt 数据集上进行实验. 需要注意的是, 本文的设定不同于以往少样本设定 (N -way K -shot), 对于每个关系可以使用任意数量的训练样本而无需构造元任务, 关系之间也无需有任何关联, 因此更加适合真实场景.

3.3 主要实验结果

如表 3 所示, 本文比较了多个知识图谱链接预测的模型在低资源场景(每个关系 8 个样本, 8-shot)下的性能. 本文发现, 所提出的模型的效果优于传统的知识图谱补全方法, 且优于两个低资源补全基线方法 RAN^[9] 和 GRL^[12] (采用 DisMult 作为底层模型), 这证明了所提出模型的优越性. 此外, KnowCo-Tuning 在 Umls 上的提升较小, 而在 FB15K-237 等数据集上相较于 KG-BERT 在 hits 10 指标中提升了 5%, 这在少样本情况下有着巨大的提升. 在 FB15K-237 上的提升主要归结于 KnowCo-Tuning 方法能够充分利用 BERT 在原预训练语料中

获得的隐式事实知识。

如表4所示, 对于关系抽取, 本文比较了常规微调和一些数据集上的最先进基线模型. 发现在低资源和全量样本的设置下, KnowCo-Tuning 方法训练出来的模型在所有关系抽取任务上都优于传统的微调方法. 在全量样本中, KnowCo-Tuning 与传统微调的方法相比, 平均获得了约 2.9% 的提升, 尤其是在 TACRED-Revisit 和 DialogRE 数据集中, 本文方法分别领先于最先进模型 GDPNet^[61] 0.6% 和 1.1%. 注意到, KnowCo-Tuning 的实现相对简单, 该模型能够促进其他更复杂场景的关系抽取的任务性能并提升已有的基线模型性能.

在低资源少样本场景下, KnowCo-Tuning 表现出了惊人的效果. 首先, 相较于常规的微调操作, KnowCo-Tuning 由于利用了预训练模型的大量先验知识, 在每个关系仅有 8 个样本的时候获得了高达 28.9% 的提升(平均 20.6%). 其次, 随着 K 从 8 增加到 32, 本文的 KnowCo-Tuning 比传统的微调方法的提升逐渐减少. 并且, 需要注意的是, 该方法可以通过训练自动找到更合适的模板和标签, 这使其适用于其他领域.

表 3 低资源知识图谱链接预测结果(8-shot)

模型/数据集	Umls		WN18RR		FB15K-237 (mini)	
	MR (raw)	Hits@10 (raw)	MR (raw)	Hits@10 (raw)	MR (raw)	Hits@10 (raw)
TransE ^[19]	38.7	0.32	19 988.3	0.000 4	5 834.2	0.093 7
TransR ^[20]	43.1	0.30	20 241.8	0.002 2	5 920.2	0.091 0
TransD ^[21]	43.8	0.30	20 371.2	0.000 6	5 915.2	0.09
TransH ^[22]	41.7	0.31	19 979.5	0.001 5	6 024.8	0.080 5
DistMult ^[24]	59.6	0.18	20 785.5	0.000 3	5 705.0	0.006 0
Complex ^[25]	59.2	0.21	20 701.1	0.000 0	6 372.7	0.005 0
RotatE ^[29]	67.3	0.05	20 120.2	0.000 3	7 273.8	0.006 7
KG-BERT ^[30]	35.2	0.33	2 720.8	0.11	2 020.7	0.047
RAN ^[7]	34.2	0.32	3 150.2	0.07	3 120.5	0.007
GRL ^[10]	34.5	0.33	2 913.3	0.09	2 913.5	0.030
KnowCo-Tuning	27.4 _(-6.8)	0.34 _(+0.01)	1557.1 _(-1163.7)	0.127 _(+0.017)	733.4 _(-1287.3)	0.101 _(+0.0073)

表 4 低资源关系抽取, Wiki 80 为准确率, 其他数据集为 $F1$, 每个关系给定 8、16、32 个样本的实验结果

数据集	模型	$K=8$	$K=16$	$K=32$	Full
TACRED-Revisit	Fine-Tuning	7.6	16.4	25.8	75.0
	GDPNet ^[62]	7.9	17.8	26.1	77.2
	KnowCo-Tuning	26.9 _(+19.3)	31.2 _(+14.8)	32.0 _(+6.2)	77.8 _(+2.8)
SemEval	Fine-Tuning	28.8	45.5	65.4	88.1
	KnowCo-Tuning	53.6 _(+24.8)	68.9 _(+23.4)	77.2 _(+11.8)	89.4 _(+1.3)
Wiki 80	Fine-Tuning	47.6	59.4	69.9	87.5
	KnowCo-Tuning	69.5 _(+21.9)	75.7 _(+16.3)	80.2 _(+10.3)	88.0 _(+0.5)
ChemProt	Fine-Tuning	30.0	42.0	51.3	80.5
	KnowCo-Tuning	58.9 _(+28.9)	60.6 _(+18.6)	64.7 _(+13.4)	81.5 _(+1.0)
DialogRE	Fine-Tuning	26.1	40.8	47.7	57.3
	GDPNet ^[62]	23.6	38.5	47.1	64.9
	KnowCo-Tuning	39.5 _(+13.4)	46.5 _(+5.7)	51.2 _(+3.5)	66.0 _(+8.7)

3.4 消融实验

为了验证方法中不同组件的有效性, 本文进行了消融研究实验, 如图 3 所示. 观察到 KnowCo-Tuning 在没有自适应提示微调的情况下性能有所下降, 这表明了该方法模块的有效性. 此外, 观察到自适应标签微调对 KnowCo-Tuning 非常有益, 尤其是在低资源场景下. 同时, 随着 K 的增加, 模块提升逐渐减小.

表 5 KnowCo-Tuning 消融实验

方法名称	$K=8$	$K=16$	$K=32$	Full
Fine-Tuning	22.8	45.7	65.4	88.1
KnowCo-Tuning	53.6	68.9	77.2	89.4
-ED	52.3 _(-1.3)	67.1 _(-1.8)	76.0 _(-1.2)	88.2 _(-0.8)
-自适应模板微调	52.8 _(-0.8)	68.3 _(-0.6)	76.8 _(-0.4)	88.8 _(-0.6)
-自适应标签微调	47.1 _(-6.5)	63.3 _(-5.6)	74.3 _(-2.9)	87.8 _(-1.6)

4 分析和讨论

4.1 不同预训练语言模型

由于在主要实验过程中只关注于以预测遮蔽词汇为预训练任务的自编码语言模型(比如 BERT), 为了探讨方法的泛化性, 进一步地在自回归语言模型, 比如 GPT-2 中进行了实验. 如图 3 所示, 使用 GPT-2 生成模型作为预训练模型可以和 BERT-Large 的效果相媲美, 这显示出本文的方法能够更好地发挥 GPT-2 模型的潜力, 使其在语言理解任务上与 BERT 模型表现一致. 同时, GPT-2 在低资源情景下常规微调方式优于 BERT, 这表明, 传统的微调方法在 BERT 中具有更大的劣势.

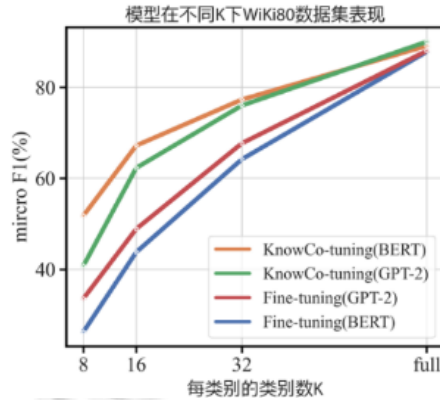


图 3 基于 GPT-2 和 BERT 的 KnowCo-Tuning 实验结果

4.2 知识协同微调得到的模板和提示词

在本文提出的 KnowCo-Tuning 中模板词和标签词是作为随机初始化的词向量 $\{h_1, \dots, h_m, \dots, h_{m+n}\}$ 存在的, 且其在训练过程中根据梯度自动微调优化. 为了理解优化好的标签词到底是什么, 通过计算 L_2 距离, 采样了在词表 \mathcal{V} 中距离标签词汇最近的 3 个词汇. 针对数据集 Wiki 80 的标签的向量表示进行 t -SNE 可视化以展示这些标签词的具体含义. 比如“交通枢纽(place_served_by_transport_hub)”该关系在图 4 中以红星的形式表达, 并且最近的 3 个词汇分别是“机场(airport)”“飞过(flown)”和“小飞机场(aerodrome)”. 这表明, 本文训练得到的标签词符合直觉且有效.

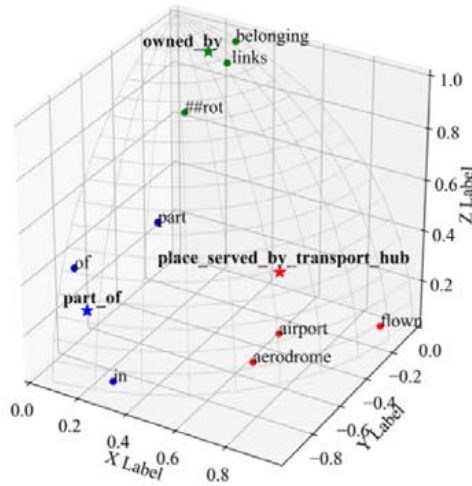


图 4 在 KnowCo-Tuning 中对 Wiki 80 数据集学习的几种关系表示的 t-SNE 可视化

4.3 KnowCo-Tuning和Fine-Tuning比较

上述结果和分析证明,使用 KnowCo-Tuning 微调的语言模型相较于 Fine-Tuning 可以更好地利用模型中存储的外部知识.这主要归功于以下几个不同点.

(1) KnowCo-Tuning 没有引入新的网络架构与参数,而 Fine-Tuning 需要加入新的分类层并且以[CLS]的特征向量作为文本表示输入到分类层中得到类别预测概率,这需要额外的参数且低效.

(2) KnowCo-Tuning 与预训练任务高度相似,将微调和预训练统一结合.而 Fine-Tuning 的方式仅仅利用了[CLS]的特征而抛弃了其序列输入的特征.

(3) KnowCo-Tuning 利用了知识图谱中结构知识构造了初始的提示模板,同时利用了知识图谱中的显式结构化知识和语言模型的隐式知识,最大程度地发挥了知识对于下游任务的作用.

总而言之,结果显示了本文提出的 KnowCo-Tuning 作为新颖的微调范式不仅性能优越,而且同时还具有拓展到其他任务的能力.

5 结 论

本文提出了一种基于知识协同微调的低资源知识图谱补全方法(KnowCo-Tuning),该方法可基于不同的预训练语言模型进行学习,并应用于链接预测和关系抽取任务,在8个数据集上均达到了最优效果.本文还分析了协同微调得到的提示词和模板以及本文方法与传统微调的区别.总的来说,本文方法不需要任何显式或隐式关联先验,且能够适用于不同数量标注样本的场景,具有一定的应用价值.在未来,本文计划将所提方法与半监督学习算法相结合,以更好地利用大量无标注数据.此外,本文也将尝试将知识图谱中的规则注入本文所提模型中,以指导构造出更好的提示模板.

References:

- [1] Wang F, Liu J, Liu B, Qian T, Xiao Y, Peng Z. Survey on construction of code knowledge graph and intelligent software development. *Ruan Jian Xue Bao/Journal of Software*, 2020, 31(1): 47–66 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/5893.htm> [doi: 10.13328/j.cnki.jos.005893]
- [2] Xing S, Liu M, Peng X. Automatic code semantic tag generation based on software knowledge graph. *Ruan Jian Xue Bao/Journal of Software* (in Chinese with English abstract). Published online first. <http://www.jos.org.cn/1000-9825/6369.htm> [doi: 10.13328/j.cnki.jos.006369]
- [3] Liu B, Wang X, Liu P, Li S, Zhang X, Yang Y. KGDB: Knowledge graph database system with unified model and query language. *Ruan Jian Xue Bao/Journal of Software*, 2021, 32(3): 781–804 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/6181.htm> [doi: 10.13328/j.cnki.jos.006181]
- [4] Wang X, Zou L, Wang C, Peng P, Feng Z. Research on knowledge graph data management: A survey. *Ruan Jian Xue Bao/Journal of Software*, 2019, 30(7): 2139–2174 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/5841.htm> [doi: 10.13328/j.cnki.jos.005841]
- [5] Yang Y, Xu B, Hu J, Tong M, Zhang P, Zheng L. Accurate and efficient method for constructing domain knowledge graph. *Ruan Jian Xue Bao/Journal of Software*, 2018, 29(10): 2931–2947 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/5552.htm> [doi: 10.13328/j.cnki.jos.005552]
- [6] Wu X, Sheng S, Jiang T, Bu C, Wu M. Huapu-CP: From knowledge graphs to a data central-platform. *Acta Automatica Sinica*, 2020, 46(10): 2045–2059 (in Chinese with English abstract). <http://www.aas.net.cn/cn/article/doi/10.16383/j.aas.c200502>
- [7] Zhang N, Chen X, Chen J, Deng X, Ruan W, Wu C, Chen H. Location based link prediction for knowledge graph. *Journal of Chinese Information Processing*, 2018, 32(4): 80–86, 129 (in Chinese with English abstract). <http://jcip.cipsc.org.cn/CN/abstract/abstract2552.shtml>
- [8] Zhang N, Deng S, Sun Z, Wang G, Chen X, Zhang W, Chen H. Long-tail relation extraction via knowledge graph embeddings and graph convolution networks. In: Burstein J, Doran C, Solorio T, eds. *Proc. of the 2019 Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Volume 1 (Long and Short Papers)*. Association for Computational Linguistics, 2019. 3016–3025.

- [9] Zhang N, Deng S, Sun Z, Chen J, Zhang W, Chen H. Relation adversarial network for low resource knowledge graph completion. In: Huang Y, King I, Liu T-Y, van Steen M, eds. Proc. of the WWW 2020. ACM/IW3C2, 2020. 1–12.
- [10] Deng S, Zhang N, Kang J, Zhang Y, Zhang W, Chen H. Meta-learning with dynamic-memory-based prototypical network for few-shot event detection. In: Caverlee J, Hu X, Lalmas M, Wang W, eds. Proc. of the WSDM 2020: The 13th ACM Int'l Conf. on Web Search and Data Mining. ACM, 2020. 151–159.
- [11] Chen M, Zhang W, Zhang W, Chen Q, Chen H. Meta relational learning for few-shot link prediction in knowledge graphs. In: Inui K, Jiang J, Ng V, Wan X, eds. Proc. of the 2019 Conf. on Empirical Methods in Natural Language Processing and the 9th Int'l Joint Conf. on Natural Language Processing, EMNLP-IJCNLP 2019. Association for Computational Linguistics, 2019. 4216–4225.
- [12] Zhang Y, Zhang X, Wang J, Liang H, Lei W, Sun Z, Jatowt A, Yang Z. Generalized relation learning with semantic correlation awareness for link prediction. In: Proc. of the 35th AAAI Conf. on Artificial Intelligence, AAAI 2021, the 33rd Conf. on Innovative Applications of Artificial Intelligence, IAAI 2021, the 11th Symp. on Educational Advances in Artificial Intelligence, EAAI 2021. AAAI Press, 2021. 4679–4687.
- [13] Brown TB, Mann B, Ryder N, Subbiah M, Kaplan J, Dhariwal P, Neelakantan A, Shyam P, Sastry G, Askell A, Agarwal S, Herbert-Voss A, Krueger G, Henighan T, Child R, Ramesh A, Ziegler DM, Wu J, Winter C, Hesse C, Chen M, Sigler E, Litwin M, Gray S, Chess B, Clark J, Berner C, McCandlish S, Radford A, Sutskever I, Amodei D. Language models are few-shot learners. In: Larochelle H, Ranzato M, Hadsell B, Balcan M-F, Lin H-T, eds. Advances in Neural Information Processing Systems 33: Annual Conf. on Neural Information Processing Systems 2020, NeurIPS 2020. 2020.
- [14] Zhang Z, Gu Y, Han X, Chen S, Xiao C, Sun Z, Yao Y, Qi F, Guan J, Ke P, Cai Y, Zeng G, Tan Z, Liu Z, Huang M, Han W, Liu Y, Zhu X, Sun M. CPM-2: Large-scale cost-effective pre-trained language models. CoRR, abs/2106.10715, 2021.
- [15] Zeng W, Ren X, Su T, Wang H, Liao Y, Wang Z, Jiang X, Yang Z, Wang K, Zhang X, Li C, Gong Z, Yao Y, Huang X, Wang J, Yu J, Guo Q, Yu Y, Zhang Y, Wang J, Tao H, Yan D, Yi Z, Peng F, Jiang F, Zhang H, Deng L, Zhang Y, Lin Z, Zhang C, Zhang S, Guo M, Gu S, Fan G, Wang Y, Jin X, Liu Q, Tian Y. PanGu-alpha: LARGE-scale autoregressive pretrained chinese language models with auto-parallel computation. CoRR, abs/2104.12369, 2021.
- [16] Gao T, Fisch A, Chen D. Making pre-trained language models better few-shot learners. arXiv: 2012.15723 [cs], 2020.
- [17] Petroni F, Rocktäschel T, Riedel S, Lewis PSH, Bakhtin A, Wu Y, Miller AH. Language models as knowledge bases? In: Inui K, Jiang J, Ng V, Wan X, eds. Proc. of the 2019 Conf. on Empirical Methods in Natural Language Processing and the 9th Int'l Joint Conf. on Natural Language Processing, EMNLP-IJCNLP 2019. Association for Computational Linguistics, 2019. 2463–2473.
- [18] Roberts A, Raffel C, Shazeer N. How much knowledge can you pack into the parameters of a language model? In: Webber B, Cohn T, He Y, Liu Y, eds. Proc. of the 2020 Conf. on Empirical Methods in Natural Language Processing, EMNLP 2020. Association for Computational Linguistics, 2020. 5418–5426.
- [19] Zhang W, Paudel B, Wang L, Chen J, Zhu H, Zhang W, Bernstein A, Chen H. Iteratively learning embeddings and rules for knowledge graph reasoning. arXiv: 1903.08948, 2019.
- [20] Lin Y, Shen S, Liu Z, Luan H, Sun M. Neural relation extraction with selective attention over instances. In: Proc. of the ACL. 2016, 1: 2124–2133.
- [21] Bordes A, Usunier N, García-Durán A, Weston J, Yakhnenko O. Translating embeddings for modeling multi-relational data. In: Proc. of the NIPS. 2013. 2787–2795.
- [22] Lin Y, Liu Z, Sun M, Liu Y, Zhu X. Modeling relation paths for representation learning of knowledge bases. arXiv: 1506.00379, 2015.
- [23] Ji G, He S, Xu L, Liu K, Zhao J. Knowledge graph embedding via dynamic mapping matrix. In: Proc. of the ACL. 2015. 687–696.
- [24] Wang Z, Zhang J, Feng J, Chen Z. Knowledge graph embedding by translating on hyperplanes. In: Proc. of the AAAI. 2014. 1112–1119.
- [25] Nickel M, Tresp V, Kriegel H-P. A three-way model for collective learning on multi-relational data. In: Proc. of the ICML. 2011. 809–816.
- [26] Yang B, Yih W, He X, Gao J, Deng L. Embedding entities and relations for learning and inference in knowledge bases. In: Bengio Y, LeCun Y, eds. Proc. of the 3rd Int'l Conf. on Learning Representations, ICLR 2015. 2015.

- [27] Trouillon T, Welbl J, Riedel S, Gaussier É, Bouchard G. Complex embeddings for simple link prediction. In: Proc. of the ICML. 2016. 2071–2080.
- [28] Dettmers T, Minervini P, Stenetorp P, Riedel S. Convolutional 2D knowledge graph embeddings. arXiv: 1707.01476, 2018.
- [29] Liu H, Wu Y, Yang Y. Analogical inference for multi-relational embeddings. In: Proc. of the ICML. 2017. 2168–2178.
- [30] Cai L, Wang WY. KBGAN: Adversarial learning for knowledge graph embeddings. In: Proc. of the 2018 Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers). New Orleans: Association for Computational Linguistics, 2018. 1470–1480.
- [31] Sun Z, Deng Z-H, Nie J-Y, Tang J. RotatE: Knowledge graph embedding by relational rotation in complex space. In: Proc. of the 7th Int'l Conf. on Learning Representations, ICLR 2019. New Orleans: OpenReview.net, 2019.
- [32] Yao L, Mao C, Luo Y. KG-bert: Bert for knowledge graph completion. arXiv: 1909.03193 [cs], 2019.
- [33] Xiong W, Yu M, Chang S, Guo X, Wang WY. One-shot relational learning for knowledge graphs. In: Riloff E, Chiang D, Hockenmaier J, Tsujii J, eds. Proc. of the 2018 Conf. on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2018. 1980–1990.
- [34] Zhang C, Yao H, Huang C, Jiang M, Li Z, Chawla NV. Few-shot knowledge graph completion. In: Proc. of the 34th AAAI Conf. on Artificial Intelligence, AAAI 2020, the 32nd Innovative Applications of Artificial Intelligence Conf., IAAI 2020, the 10th AAAI Symp. on Educational Advances in Artificial Intelligence, EAAI 2020. New York: AAAI Press, 2020. 3041–3048.
- [35] Sheng J, Guo S, Chen Z, Yue J, Wang L, Liu T, Xu H. Adaptive attentional network for few-shot knowledge graph completion. In: Webber B, Cohn T, He Y, Liu Y, eds. Proc. of the 2020 Conf. on Empirical Methods in Natural Language Processing, EMNLP 2020. Association for Computational Linguistics, 2020. 1681–1691.
- [36] Ji G, Liu K, He S, Zhao J. Distant supervision for relation extraction with sentence-level attention and entity descriptions. In: Singh SP, Markovitch S, eds. Proc. of the 31st AAAI Conf. on Artificial Intelligence. San Francisco: AAAI Press, 2017. 3060–3066.
- [37] Gao T, Han X, Liu Z, Sun M. Hybrid attention-based prototypical networks for noisy few-shot relation classification. In: Proc. of the 33rd AAAI Conf. on Artificial Intelligence, AAAI 2019, the 31st Innovative Applications of Artificial Intelligence Conf., IAAI 2019, the 9th AAAI Symp. on Educational Advances in Artificial Intelligence, EAAI 2019. AAAI Press, 2019. 6407–6414.
- [38] Ye Z-X, Ling Z-H. Multi-level matching and aggregation network for few-shot relation classification. In: Korhonen A, Traum DR, Màrquez L, eds. Proc. of the 57th Conf. of the Association for Computational Linguistics, ACL 2019, Volume 1: Long Papers. Association for Computational Linguistics, 2019. 2872–2881.
- [39] Gao T, Han X, Xie R, Liu Z, Lin F, Lin L, Sun M. Neural snowball for few-shot relation learning. In: Proc. of the 34th AAAI Conf. on Artificial Intelligence, AAAI 2020, the 32nd Innovative Applications of Artificial Intelligence Conf., IAAI 2020, the 10th AAAI Symp. on Educational Advances in Artificial Intelligence, EAAI 2020. New York: AAAI Press, 2020. 7772–7779.
- [40] Yu H, Zhang N, Deng S, Ye H, Zhang W, Chen H. Bridging text and knowledge with multi-prototype embedding for few-shot relational triple extraction. In: Scott D, Bel N, Zong C, eds. Proc. of the 28th Int'l Conf. on Computational Linguistics, COLING 2020. Int'l Committee on Computational Linguistics, 2020. 6399–6410.
- [41] Soares LB, FitzGerald N, Ling J, Kwiatkowski T. Matching the blanks: Distributional similarity for relation learning. In: Korhonen A, Traum D R, Màrquez L, eds. Proc. of the 57th Conf. of the Association for Computational Linguistics, ACL 2019, Volume 1: Long Papers. Association for Computational Linguistics, 2019. 2895–2905.
- [42] Geng X, Chen X, Zhu KQ, Shen L, Zhao Y. MICK: A meta-learning framework for few-shot relation classification with small training data. In: D'Aquin M, Dietze S, Hauff C, Curry E, Cudré-Mauroux P, eds. Proc. of the CIKM 2020: The 29th ACM Int'l Conf. on Information and Knowledge Management, Virtual Event. ACM, 2020. 415–424.
- [43] Qu M, Gao T, Xhonneux L-PAC, Tang J. Few-shot relation extraction via bayesian meta-learning on relation graphs. In: Proc. of the 37th Int'l Conf. on Machine Learning, ICML 2020, Virtual Event. PMLR, 2020, 119: 7867–7876.
- [44] Dong B, Yao Y, Xie R, Gao T, Han X, Liu Z, Lin F, Lin L, Sun M. Meta-information guided meta-learning for few-shot relation classification. In: Scott D, Bel N, Zong C, eds. Proc. of the 28th Int'l Conf. on Computational Linguistics, COLING 2020. Int'l Committee on Computational Linguistics, 2020. 1594–1605.
- [45] Liu X, Zheng Y, Du Z, Ding M, Qian Y, Yang Z, Tang J. GPT understands, too. arXiv: 2103.10385 [cs], 2021.

- [46] Zhao Z, Wallace E, Feng S, Klein D, Singh S. Calibrate before use: Improving few-shot performance of language models. In: Meila M, Zhang T, eds. Proc. of the 38th Int'l Conf. on Machine Learning, ICML 2021, Virtual Event. PMLR, 2021, 139: 12697–12706.
- [47] Schick T, Schütze H. Exploiting cloze-questions for few-shot text classification and natural language inference. In: Merlo P, Tiedemann J, Tsarfay R, eds. Proc. of the 16th Conf. of the European Chapter of the Association for Computational Linguistics: Main Volume, EACL 2021. Association for Computational Linguistics, 2021. 255–269.
- [48] Raffel C, Shazeer N, Roberts A, Lee K, Narang S, Matena M, Zhou Y, Li W, Liu PJ. Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research*, 2020, 21: 140:1–140:67.
- [49] Shin T, Razeghi Y, IV RLL, Wallace E, Singh S. AutoPrompt: Eliciting knowledge from language models with automatically generated prompts. In: Webber B, Cohn T, He Y, Liu Y, eds. Proc. of the 2020 Conf. on Empirical Methods in Natural Language Processing, EMNLP 2020. Association for Computational Linguistics, 2020. 4222–4235.
- [50] Li XL, Liang P. Prefix-tuning: Optimizing continuous prompts for generation. CoRR, abs/2101.00190, 2021.
- [51] Zhong Z, Friedman D, Chen D. Factual probing is [mask]: Learning vs. learning to recall. In: Toutanova K, Rumshisky A, Zettlemoyer L, Hakkani-Tür D, Beltagy I, Bethard S, Cotterell R, Chakraborty T, Zhou Y, eds. Proc. of the 2021 Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2021. Association for Computational Linguistics, 2021. 5017–5033.
- [52] Bodenreider O. The unified medical language system (Umls): Integrating biomedical terminology. *Nucleic Acids Research*, 2004, 32 (Database issue): D267–D270.
- [53] Miller GA. WordNet: A lexical database for English. *Communications of the ACM*, 1995, 38: 39–41.
- [54] Bollacker K, 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. ACM, 2008. 1247–1250.
- [55] Alt C, Gabryszak A, Hennig L. TACRED revisited: A thorough evaluation of the tacred relation extraction task. In: Jurafsky D, Chai J, Schluter N, Tetreault JR, eds. Proc. of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020. Association for Computational Linguistics, 2020. 1558–1569.
- [56] Han X, Zhu H, Yu P, Wang Z, Yao Y, Liu Z, Sun M. FewRel: A large-scale supervised few-shot relation classification dataset with state-of-the-art evaluation. In: Riloff E, Chiang D, Hockenmaier J, Tsujii J, eds. Proc. of the 2018 Conf. on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2018. 4803–4809.
- [57] Hendrickx I, Kim SN, Kozareva Z, Nakov P, Séaghdha DÓ, Padó S, Pennacchiotti M, Romano L, Szipakowicz S. SemEval-2010 task 8: Multi-way classification of semantic relations between pairs of nominals. In: Erk K, Strapparava C, eds. Proc. of the 5th Int'l Workshop on Semantic Evaluation, SemEval@ACL 2010. The Association for Computer Linguistics, 2010. 33–38.
- [58] Kringelum J, Kjærulff SK, Brunak S, Lund O, Oprea TI, Taboureaux O. ChemProt-3.0: A global chemical biology diseases mapping. Database-The Journal of Biological Databases and Curation, 2016.
- [59] Yu D, Sun K, Cardie C, Yu D. Dialogue-based relation extraction. In: Jurafsky D, Chai J, Schluter N, Tetreault JR, eds. Proc. of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020. Association for Computational Linguistics, 2020. 4927–4940.
- [60] Beltagy I, Lo K, Cohan A. SciBERT: A pretrained language model for scientific text. In: Inui K, Jiang J, Ng V, Wan X, eds. Proc. of the 2019 Conf. on Empirical Methods in Natural Language Processing and the 9th Int'l Joint Conf. on Natural Language Processing, EMNLP-IJCNLP 2019. Association for Computational Linguistics, 2019. 3613–3618.
- [61] Xue F, Sun A, Zhang H, Chng ES. GDPNet: Refining latent multi-view graph for relation extraction. In: Proc. of the 35th AAAI Conf. on Artificial Intelligence, AAAI 2021, the 33rd Conf. on Innovative Applications of Artificial Intelligence, IAAI 2021, the 11th Symp. on Educational Advances in Artificial Intelligence, EAAI 2021, Virtual Event. AAAI Press, 2021. 14194–14202.
- [62] Devlin J, Chang M-W, Lee K, Toutanova K. BERT: Pre-training of deep bidirectional transformers for language understanding. In: Burstein J, Doran C, Solorio T, eds. Proc. of the 2019 Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Volume 1 (Long and Short Papers). Association for Computational Linguistics, 2019. 4171–4186.

附中文参考文献:

- [1] 王飞, 刘井平, 刘斌, 钱铁云, 肖仰华, 彭智勇. 代码知识图谱构建及智能化软件开发方法研究. 软件学报, 2020, 31(1): 47–66. <http://www.jos.org.cn/1000-9825/5893.htm> [doi: 10.13328/j.cnki.jos.005893]
- [2] 邢双双, 刘名威, 彭鑫. 基于软件知识图谱的自动化代码语义标签生成方法. 软件学报. 在线出版. <http://www.jos.org.cn/1000-9825/6369.htm> [doi: 10.13328/j.cnki.jos.006369]
- [3] 刘宝珠, 王鑫, 柳鹏凯, 李思卓, 张小旺, 杨雅君. KGDB: 统一模型和语言的知识图谱数据库管理系统. 软件学报, 2021, 32(3): 781–804. <http://www.jos.org.cn/1000-9825/6181.htm> [doi: 10.13328/j.cnki.jos.006181]
- [4] 王鑫, 邹磊, 王朝坤, 彭鹏, 冯志勇. 知识图谱数据管理研究综述. 软件学报, 2019, 30(7): 2139–2174. <http://www.jos.org.cn/1000-9825/5841.htm> [doi: 10.13328/j.cnki.jos.005841]
- [5] 杨玉基, 许斌, 胡家威, 仝美涵, 张鹏, 郑莉. 一种准确而高效的领域知识图谱构建方法. 软件学报, 2018, 29(10): 2931–2947. <http://www.jos.org.cn/1000-9825/5552.htm> [doi: 10.13328/j.cnki.jos.005552]
- [6] 吴信东, 盛绍静, 蒋婷婷, 卜晨阳, 吴明辉. 从知识图谱到数据中台: 华谱系统. 自动化学报, 2020, 46(10): 2045–2059. <http://www.aas.net.cn/cn/article/doi/10.16383/j.aas.c200502>
- [7] 张宁豫, 陈曦, 陈矫彦, 邓淑敏, 阮伟, 吴春明, 陈华钧. 基于位置的知识图谱链接预测. 中文信息学报, 2018, 32(4): 80–86. <http://jicp.cipsc.org.cn/CN/abstract/abstract2552.shtml>



张宁豫(1989—), 男, 博士, 副教授, CCF 专业会员, 主要研究领域为知识图谱, 自然语言处理.



邓淑敏(1995—), 女, 博士生, 主要研究领域为知识图谱, 事件抽取, 自然语言处理.



谢辛(1997—), 男, 硕士生, 主要研究领域为知识图谱, 关系抽取.



叶宏彬(1996—), 男, 博士生, 主要研究领域为关系抽取, 生成式语言模型, 知识图谱.



陈想(1994—), 男, 博士生, CCF 学生会会员, 主要研究领域为关系抽取, 多模态, 知识图谱.



陈华钧(1978—), 男, 博士, 教授, CCF 专业会员, 主要研究领域为知识图谱, 大数据系统, 自然语言处理.