

# 基于结构增强的知识图谱实体对齐

# CCKS2022

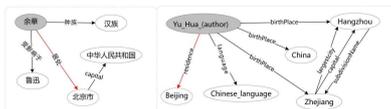
张富晓, 董姜媛, 李丽双

大连理工大学

计算机科学与技术学院

## 引言

不同知识图谱中的等价实体通常具有不同的表面形式或邻域, 如下图所示, 这导致对齐模型使用图卷积网络 (Graph Convolutional Networks, GCN) 学习实体表示时会引入大量噪声。另外, 关系三元组表述的事实具有很强的方向性, 而GCN不能编码实体的方向信息。为了应对这些问题, 该文提出了一种基于结构增强的知识图谱实体对齐模型 (Structure-Enhanced Entity Alignment, SEEA), 模型使用长短期记忆网络 (Long Short-Term Memory Network, LSTM) 抽取实体名称中的语义信息来减轻结构差异带来的负面影响, 并编码关系三元组中实体的方向信息以学习更好的实体表示, 同时通过获取实体间的结构相似信息提高模型的对齐性能。模型在 DBP15K 的三个子数据集上进行实验, 在 Hits@1 上分别取得了 92.0%、94.1% 和 97.5% 的结果。



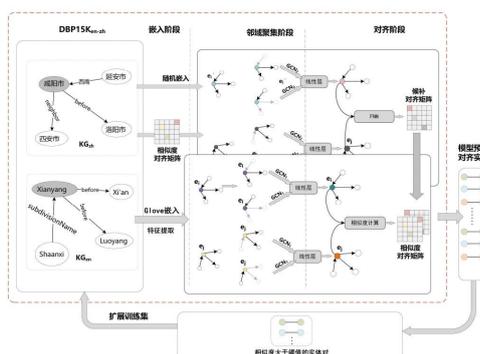
## 方法

### • 整体框架

本文提出了一种基于结构增强的知识图谱实体对齐模型 SEEA, 模型架构如图所示。

### • 嵌入阶段

对于知识图谱中的实体名称序列, 本文采用单词级嵌入, 将实体名称的每个单词映射到低维向量空间中。使用预训练的 GloVe 词向量对实体名称进行嵌入, 并利用 LSTM 对嵌入序列进行编码作为实体初始表示。



### • 邻域聚集阶段

为了编码关系三元组中实体的方向信息, 并将实体的方向信息添加到实体表示中, 将实体在知识图谱一阶邻域中的头实体集合和尾实体集合分别使用 GCN 进行聚集, 得到实体的一阶头嵌入和一阶尾嵌入, 然后通过全连接网络直接相加。

### • 对齐阶段

实体对齐是将源知识图谱中的实体匹配目标知识图谱中等价实体的过程。模型将向量内积作为两个实体的相似度得分, 并将该相似度得分作为判断实体是否对齐的依据。

### • 数据集扩展阶段

半监督学习模型在训练过程中, 既可以使用有标注的数据进行训练, 还可以利用大量未标注的数据提升模型效果。为了克服对齐实体数据不足的问题, 本文使用半监督 Bootstrapping 方法来扩展数据集规模。

## 实验与结果

### • 数据集介绍

该模型在实体对齐标准数据集 DBP15K 上展开实验, 包括三个子数据集: DBP15Kzh-en、DBP15Kja-en 和 DBP15Kfr-en。

## • 总体结果

该模型在标准数据集 DBP15k 上的结果如下表所示, 充分说明了 SEEA 模型的优越性。

模型	DBP15K <sub>zh-en</sub>		DBP15K <sub>ja-en</sub>		DBP15K <sub>fr-en</sub>	
	Hits@1	Hits@10	Hits@1	Hits@10	Hits@1	Hits@10
TransE <sup>(1)</sup>	73.5	91.9	71.9	93.2	71.0	94.1
BootEA <sup>(2)</sup>	62.9	84.8	62.2	85.4	65.3	87.4
GN <sup>(3)</sup>	67.9	78.5	74.0	87.2	89.4	95.2
HGN <sup>(4)</sup>	72.0	85.7	76.6	89.7	89.2	96.1
DGMC <sup>(5)</sup>	80.1	87.5	84.8	89.7	93.3	96.0
RAGA <sup>(6)</sup>	87.3	-	90.9	-	96.6	-
ERMC <sup>(7)</sup>	90.3	94.6	94.2	96.4	96.2	98.2
SEEA	92.0	95.3	94.1	96.4	97.5	98.7

## • 消融实验

为了评估的模型各部分的有效性, 并对模型结构进行纵向功能验证, 实验结果充分验证了各个模块的有效性。

模型	DBP15K <sub>zh-en</sub>		DBP15K <sub>ja-en</sub>		DBP15K <sub>fr-en</sub>	
	Hits@1	Hits@10	Hits@1	Hits@10	Hits@1	Hits@10
SEEA (w/o LS)	90.5	94.8	91.6	95.3	97.0	98.4
SEEA (w/o DR)	89.2	94.0	91.4	94.8	95.8	98.1
SEEA (w/o M)	80.6	95.3	84.9	96.6	90.4	98.7
SEEA	92.0	95.3	94.1	96.4	97.5	98.7

## 结论

本文提出了一种基于结构增强的知识图谱实体对齐模型 (SEEA), 该模型使用 LSTM 抽取语义信息作为初始实体表示, 并且能够将关系三元组中实体的方向信息添加到实体表示中, 学习更好地实体表示。在对齐过程中, 模型引入一种细化策略来学习实体之间的结构相似信息, 提高模型对齐效果。实验表明, 本文提出的 SEEA 模型在 DBP15K 的三个子数据集上的对齐性能优于主流基线模型。

## 致谢

感谢大连市科技创新基金(2020JJ26GX035)对本研究的资助。