基于查询路径排序的知识库问答系统

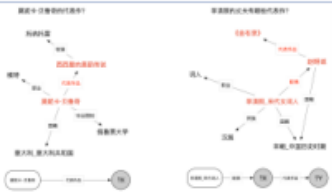
宋鹏程, 单丽莉, 孙承杰, 林磊, 王明江
哈尔滨工业大学

I. 摘要

目前，对中文知识库问答系统的研究大多针对简单问题展开，对于复杂问题缺少有效的解决方案。为了对简单问题与复杂的多约束问题进行统一处理，并提高系统的准确性， 本文提出了一种基于查询路径排序的知识库问答系统。系统采用基于LambdaRank算法构建的排序模型，对查询路径按照与问题的相关度大小进行排序，选择与问题相关度最高的路径用于抽取答案。同时，本文还应用了一种融合方法以提高实体识别的准确性。本文所构建的系统在CCKS2019 KBQA任务与CCKS2020 KBQA任务上都取得了较好的效果。

II. 介绍

- 查询路径：相互关联的三元组在图上表现为一条路径，因此将与问题相关的三元组组合称为查询路径。
- 简单问题：即问题结构简单通过知识库中的一个三元组即可获取答案。
- 复杂问题：也称为多约束问题，即需要知识库中多个三元组进行推理，并根据约束条件不断的进行筛选才能够获取答案。



III. 基于LambdaRank算法的排序模型

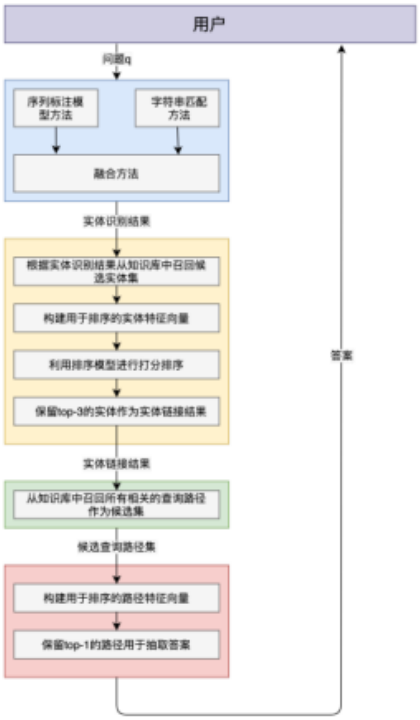
LambdaRank算法是一种基于pair-wise策略的排序方法，关注的是两个查询路径间相对顺序的准确性。算法在进行反向传播时，在梯度上增加了归一化折损累计增益变量，用以影响每次参数更新的方向和强度，从而保证top-k结果的准确性。

- 模型结构：由两个前馈神经网络模型构成的孪生网络模型，其中，前馈神经网络模型为输入的查询路径计算相关度打分。
- 模型输入：对应于同一问题的两条查询路径$query_i$和$query_j$的特征向量x_i和y_j。
- 模型输出：$query_i$比$query_j$更相关的概率
- 损失函数：交叉熵损失函数。

IV. 基于融合方法的实体识别过程

利用基于字符串匹配方法的实体识别结果对基于序列标注模型的实体识别结果进行修正，提高结果的召回率。

V.系统框架



用于实体排序的特征：

- 实体与问题间的相似度特征：字面相似度；语义相似度
- 实体自身的特征：实体属性与问题的相似度；实体在知识图谱中的出度与入度

用于路径排序的特征：

- 路径中实体与问题相关度的平均值
- 路径中实体属性与问题相关度的平均值
- 路径与问题的语义相似度特征
- 路径与问题的字面相似度特征

VI. 实验

实体识别实验

Table 1. 使用不同指称识别方法时在测试集上的效果

Method	Acc	Recall	F1
模型标注方法	0.8787	0.8841	0.8779
字符串匹配方法	0.2581	**0.9620**	0.3848
融合方法	**0.9151**	0.9404	**0.9195**

Table 2. 使用不同序列标注模型时在测试集上的融合效果

Model	Acc	Recall	F1
Bi-LSTM+CRF	0.7481	0.8028	0.7745
Bert+CRF	0.8175	0.8745	0.8435
SpanBert+CRF	**0.9151**	**0.9404**	**0.9195**

实体链接实验

Table 3. 测试集上实体链接结果的召回率指标

Model	Recall@3
our model	0.9114
our model w/o 实体上下文特征	0.8745
our model w/o 实体出度入度特征	0.8175

查询路径排序实验

Table 4. 不同问答系统在CCKS2019 KBQA任务上的效果

Model	F1
(Luo et al.,2019)	0.7354
(Zhou et al.,2019)	0.7305
(Yang et al.,2019)	0.7045
(Cao et al.,2019)	0.6768
our model	0.7313

VII. 结论

本文构建了一种基于查询路径排序的知识库问答系统，相比于目前已有的方法，本文所构建的系统实现了对简单问题与复杂问题的统一处理，并且极大的减少了用于排序的特征数量，降低了系统复杂性，取得了较为优异的效果。