



# 计算机科学

COMPUTER SCIENCE

## 基于学术知识图谱的辅助创新技术研究

钟将, 尹红, 张剑

引用本文

钟将, 尹红, 张剑. 基于学术知识图谱的辅助创新技术研究[J]. 计算机科学, 2022, 49(5): 194-199.

ZHONG Jiang, YIN Hong, ZHANG Jian. Academic Knowledge Graph-based Research for Auxiliary Innovation Technology[J]. Computer Science, 2022, 49(5): 194-199.

---

## 相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[合注意力机制与几何信息的特征融合框架](#)

Feature Fusion Framework Combining Attention Mechanism and Geometric Information

计算机科学, 2022, 49(5): 129-134. <https://doi.org/10.11896/jsjcx.210300180>

[基于深度学习的自动调制识别研究](#)

Automatic Modulation Recognition Based on Deep Learning

计算机科学, 2022, 49(5): 266-278. <https://doi.org/10.11896/jsjcx.211000085>

[基于 QRNN 的网络协议模糊测试用例过滤方法](#)

Testcase Filtering Method Based on QRNN for Network Protocol Fuzzing

计算机科学, 2022, 49(5): 318-324. <https://doi.org/10.11896/jsjcx.210300281>

[基于时空自适应图卷积神经网络的脑电信号情绪识别](#)

EEG Emotion Recognition Based on Spatiotemporal Self-Adaptive Graph Convolutional Neural Network

计算机科学, 2022, 49(4): 30-36. <https://doi.org/10.11896/jsjcx.210900200>

[大数据驱动的社会经济地位分析研究综述](#)

Big Data-driven Based Socioeconomic Status Analysis:A Survey

计算机科学, 2022, 49(4): 80-87. <https://doi.org/10.11896/jsjcx.211100014>

# 基于学术知识图谱的辅助创新技术研究

钟 将<sup>1</sup> 尹 红<sup>1</sup> 张 剑<sup>2</sup>

1 重庆大学计算机学院 重庆 400044

2 重庆西信天元数据资讯有限公司 重庆 401121

**摘 要** 计算机领域知识快速更新且存在较多歧义,导致学生自主创新时难以找到合理的解决方案。作为辅助创新工具,智能问答系统可以协助学生更快地把握学科发展前沿,精准地找出解决问题的方法。在大规模科技文献库上构建科研知识图谱,实现了辅助学生创新的智能问答系统。为了减小查询问句中噪声实体带来的影响,提出一种基于辅助任务的意图信息增强神经网络(Auxiliary Task Enhanced Intent Information for Question Answering in Computer Domain, ATEI-QA)。相比传统方法,该方法能够更精确地提取问句意图信息,减小噪声实体给意图识别带来的影响。在计算机领域数据集和通用数据集上与3个主流模型开展了对比实验,结果表明所提模型在领域数据集上的MAP和MRR值平均提升了3.27%和1.72%,在通用数据集上的MAP和MRR值平均提升了4.37%和2.81%。

**关键词**: 智能问答; 知识图谱; 意图识别; 深度学习

**中图法分类号** TP391

## Academic Knowledge Graph-based Research for Auxiliary Innovation Technology

ZHONG Jiang<sup>1</sup>, YIN Hong<sup>1</sup> and ZHANG Jian<sup>2</sup>

1 College of Computer Science, Chongqing University, Chongqing 400044, China

2 Chongqing Xixintianyuan Data Information Co., Ltd., Chongqing 401121, China

**Abstract** Due to the rapid updating of computer knowledge with many ambiguities, it is difficult for students to seek reasonable solutions for independent innovation. As an auxiliary innovation tool, intelligent question-answering system can help students to grasp the frontier of subject development, find out solutions for problems faster and precisely. In this paper, a knowledge graph of scientific research is constructed based on a large-scale database of scientific and technological documents, which realizes an intelligent question answering system for assisting students in innovation. In order to reduce the influence of noisy entities on query questions, this paper proposes an auxiliary task enhanced intent information for question answering in computer domain (ATEI-QA). Compared with the traditional method, this method can extract the question intention information more accurately and further reduce the influence of noisy entity on intention recognition. Additionally, we conduct a series of experimental studies on computer and common datasets, and compare with three mainstream methods. Finally, experimental results demonstrate that our model achieves significant improvements against with three baselines, improving MAP and MRR scores by average of 3.27%, 1.72% in the computer dataset and 4.37%, 2.81% in the common dataset respectively.

**Keywords** Question answering, Knowledge graph, Intent recognition, Deep learning

## 1 引言

当前高校大力实施“双一流”建设,其中创新人才培养是各大高校的工作重点。将科学研究和创新能力培养有机融合,让学生围绕特定主题开展自主研发的研究型学习模式,是目前最广泛采用的创新人才培养模式。近年来,大数据智能化领域的新技术、新模式、新概念不断涌现,若能够根据研究问题精准挖掘出最合适的解决方案,可以大幅提高学习效率,减轻教师的负担。

自动化问答(Question Answering, QA)是自然语言处理和信息检索领域的一个重要分支,旨在为用户输入的提问

提供自然流畅且正确的答案<sup>[1]</sup>。目前,自动化问答的答案来源多种多样,如网络文档、问答社区、知识库等。按照数据源的不同,可将问答分为几种不同类型,比如机器阅读理解、回答选择以及知识库回答等。

知识图谱是由多种关系和多种实体构成的知识库,具有结构化的知识。由于现有大多数问题可通过简单推理得到问题答案,而知识图谱三元组本身特有的实体-关系-实体特性正好为答案推理提供了便利,因此,基于知识图谱的问答(Question Answering over Knowledge Graphs, KGQA)方法<sup>[2-3]</sup>成为了问答领域的研究热点。

为了辅助学生开展自主创新活动,减轻教师在创新型

到稿日期:2021-04-19 返修日期:2021-09-08

基金项目:重庆市高等教育教学改革研究重大项目(191003);教育部新工科研究与实践项目(E-JSJRJ20201335)

This work was supported by the Major Project of Chongqing Higher Education Teaching Reform Research (191003) and New Engineering Research and Practice Project of the Ministry of Education (E-JSJRJ20201335).

通信作者:钟将(zhongjiang@cqu.edu.cn)

学习过程中的工作负担,本文设计了基于辅助任务意图信息增强的问答模型 ATEI-QA,实现了一个基于学术知识图谱的智能问答系统。KGQA 系统首先会给定自然语言问句和图谱,然后对问题进行智能分析并利用知识图谱推理得到正确答案,其具体过程如图 1 所示。

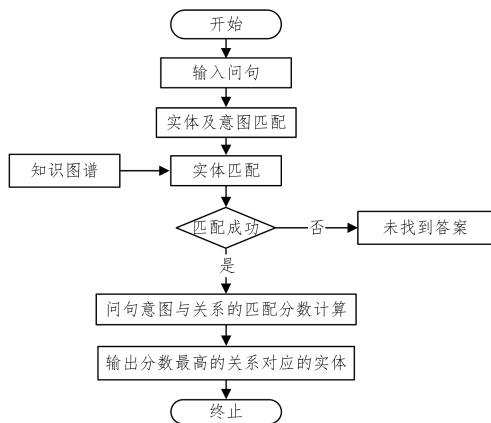


图 1 基于知识图谱的问答流程图

Fig. 1 Flowchart of KG QA

## 2 相关工作

问答任务最早采用人工定义的模板和规则对问句进行解析,如语义和模式相结合的语义模板<sup>[4]</sup>。由语言学专家专门制定的模板和规则,可移植性低,当将模板应用在特定领域时,需要重新制定,导致大量的人力、物力浪费。随着语言表示模型的不断进步,领域研究者提出了基于神经网络语义解析和基于信息检索两种研究方法。

其中,基于神经网络语义解析的方法致力于将非结构化自然语言问句转换成可执行的查询语言(如 SPARQL, Fun-QL 等),然后在知识图谱中执行查询语句,最终的获得答案。基于信息检索的方法从语义角度出发获取答案,首先识别出自然语言问句中的实体词,将其链接到知识图谱中,然后将以实体为中心的知识子图中的实体视为问句的候选答案;最后对候选答案所链接的关系与问句意图进行相关性建模,并根据相关性确定最终答案。由于目前已有数据集几乎为问句-答案对,较少为问题-查询语言对,因此,本文采用基于信息检索的方法进行问答任务的研究。

Yao 等<sup>[5]</sup>于 2014 年提出采用依存句法树识别问句意图,主要对问句中的问题词、问题类型、主题词和中心词进行建模,将其转换成问句图,并根据主题词从知识图谱中构造对应的候选答案集合,然后联合问句图中的问句意图特征与候选答案特征以捕捉问句-候选答案间的语义关联性。采用依存句法树提取特征在一定程度上依赖于语言学的知识和特定的规则,不能有效地捕捉问句的全局信息,因此,研究者提出利用表示学习方法将问句和答案映射至同一语义空间进行语义匹配,获得最终的正确答案。Bordes 等<sup>[6]</sup>首次提出学习问句语义表示,所有词语的语义表示之和代表问句意图,同时考虑利用候选答案的类型、实体-候选答案路径和对应实体子图与实体间的相关性来学习候选答案的语义表示。随着深度学习的发展,Dong 等<sup>[7]</sup>于 2015 年提出利用多列卷积神经网络从与标准答案相关的多个方面识别意图;为捕捉问句和答案间的交互信息,Hao 等<sup>[8]</sup>于 2017 年提出利用交互注意力模型进一步提取问句意图信息和答案信息;Xu 等<sup>[9]</sup>为消除知识图谱不完整对问答推理带来的影响,提出引入外部知识(如维基百科)来提高正确推理准确率;近年来,Sun 等<sup>[2]</sup>提出在由知识图谱与外部知识构成的异构图上进行意图识别,学习意图表示。由于已有方法大多通过外部语料知识来解决图谱不完整的问题,但外部知识的收集整理是一项艰巨且耗时的任务,因此,Saxena 等<sup>[10]</sup>提出将全连接图谱合并至当前知识图谱中,探索如何更有效地解决知识图谱不完整的问题,并结合知识图谱补全的思想以解决图谱不完整带来的问题。从上述代表性的研究方法可以发现,目前的方法缺乏对语义信息的挖掘。因此,Jiang 等<sup>[11]</sup>提出利用 BERT 来深度挖掘问句和图谱的语义信息;同年,Kacupaj 等<sup>[12]</sup>采用图注意力网络(Graph Attention),借助多任务学习的方法对语义信息进行更深层次的挖掘;为了更有效地利用知识图谱专有的特征,Xiong 等<sup>[13]</sup>提出了一种综合深度学习框架的方法来进行语义挖掘,并设计了 4 种相似度计算模型来处理较为复杂的问句。Wu 等<sup>[14]</sup>发现,现有研究方法大多集中在关系匹配上,较少关注问题的内部结构,因此提出采用基于图卷积的关系型语义解析模型<sup>[15]</sup>来提取问句的全局语义。

## 3 模型设计

基于辅助任务的意图信息增强模型的整体结构如图 2 所示,该模型由 3 个核心部分组成。

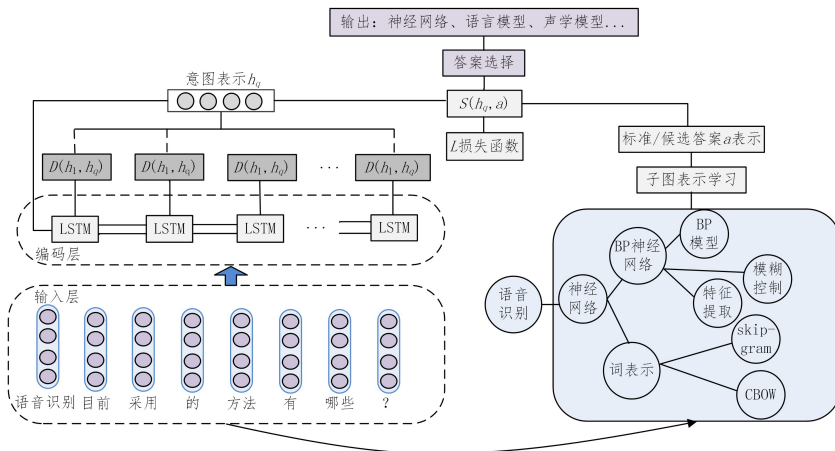


图 2 ATEI-QA 模型整体架构

Fig. 2 Overall model architecture of ATEI-QA

(1)问句意图表示:利用双向长短期记忆网络(Bi-directional Long Short-Term Memory, Bi-LSTM)和辅助任务学习问句意图信息的表示;

(2)子图表示学习模块:学习候选答案构成的子图的表示;

(3)答案选择模块:根据意图-候选答案匹配分数选择问句的最终答案。

### 3.1 问句意图表示

该模块主要负责提取问句意图信息,学习意图表示。问句通常由两部分组成,一部分为实体词,即问句的主题词;另一部分为非实体词,即用于描述与主题词关系的词语,这里将其称作关系词,同时将其语义作为问句意图。为精准提取问句意图信息,该模块在意图表示学习过程中考虑了关系词的句子级别语义信息。因此,在整个意图表示学习过程中,将问句作为模型输入,采用 Bi-LSTM 学习每个词语的语义表示,同时引入辅助任务来增强意图信息的学习。

(1)词语表示学习:该过程将问句中的词语建模为分布式表示。每个词语通过嵌入 $E^w \in \mathbb{R}^{v \times d}$ 映射为表达 $e_w$ ,其中 $v$ 为词汇量大小, $d$ 为词表示的维度。将 $e_w$ 送入 Bi-LSTM,得到词语表示 $h_w$ :

$$h_w = [h_w^b, h_w^e] \quad (1)$$

(2)意图表示学习:该过程利用主题词与非主题词之间的依赖关系学习意图分布式表示。为减小主题词对意图表示学习的影响,引入以互信息为损失函数的辅助任务。Hjelm等<sup>[16]</sup>认为数据在粗粒度层次携带的信息应该与细粒度层次携带的信息尽可能接近,当应用到问句时,则认为问句中每个词语的语义表示都尽可能接近问句整体语义表示。但在问答任务中,意图识别模块更关注关系词而非主题词,以此降低词对问答意图的干扰。因此,让问句意图表示 $h_q$ 更大程度地接近关系词,更小程度地接近主题词。采用鉴别函数 $D$ 来刻画问句意图与词语之间的接近程度:

$$D(h_j, h_q) = \sigma(h_j^T W h_q) \quad (2)$$

其中, $h_j$ 为第 $j$ 个词语与问句意图 $h_q$ 之间的接近程度。同时,定义了损失函数 $L$ ,它将结合答案选择模块的损失函数对模型进行训练。

$$L = \frac{1}{N+M} \left( \sum_{i=1}^N \log D(h_i, h_q) + \sum_{j=1}^M \log D(h_j, h_q) \right) \quad (3)$$

其中, $N, M$ 分别表示主题词与关系词数量。

### 3.2 子图表示学习

对于问句在知识图谱中对应的每个候选答案,采用子图表示学习方法对其进行编码。通常,问句对应的答案为图谱中的单个实体,若单独将实体与意图进行比较,则不能把知识图谱的知识引入候选答案中。为此考虑了候选答案在二阶范围内的实体,并将它们形成的子图抽取出来作为候选答案表示学习的输入。

在该模块中,采用图卷积神经网络(Graph Convolutional Network, GCN)<sup>[17]</sup>学习子图中每个实体的向量表示,其具体学习方式如下:

$$h_i^{(l+1)} = \sigma \left( \sum_{j \in S_i} f(h_i^{(l)}, h_j^{(l)}) \right) \quad (4)$$

其中, $h_i^l \in \mathbb{R}^d$ 为实体 $i$ 在 GCN 第 $l$ 层的语义表示, $S_i$ 表示

指向实体 $i$ 的直接邻居节点集合。在学习过程中,函数 $f$ 用于计算两个实体之间的相关性,并通过求和的方式累积每个邻居实体对中心实体提供的语义信息,最后通过一个激活函数 $\sigma(\cdot)$ 对表示的每一维取值进行变换,如 $\text{ReLU} = \max(0, \cdot)$ 。

通过以上过程,模型成功学习到每个实体的语义表示。为获得候选答案所形成的子图表示,文中将子图中所有的实体表示进行平均,从而学习到语义信息更丰富的候选答案表征 $a$ 。

### 3.3 答案选择模块

答案选择模块主要有模型训练与答案推理两个子模块。在模型训练模块,首先通过得分函数 $S$ 计算意图-候选答案的匹配得分:

$$S(h_q, a) = h_q^T \cdot a \quad (5)$$

然后设计损失函数 $L'$ ,确保问句与对应答案的匹配分数尽可能高,与非答案的匹配分数尽可能低,同时结合问句意图表示模块的损失函数 $L$ 对模型进行训练。因此,模型的整体损失函数为:

$$\ell = \lambda L + (1-\lambda) L' \quad (6)$$

$$L' = \sum_{i=1}^K \sum_{\bar{a} \in \bar{A}(a_i)} \max(0, m - S(h_q^i, a_i) + S(h_q^i, \bar{a})) \quad (7)$$

其中, $\lambda$ 表示平衡因子, $K$ 表示训练样本数, $\bar{a}$ 表示非正确答案集合 $\bar{A}$ 的一个非正确答案, $m$ 为边缘值。最后,以最小化损失函数为目标进行模型训练。

完成模型训练后,即可通过模型完成问答任务,此时答案推理模块就计算意图与候选答案匹配得分,随后根据得分进行排序,确定最终答案 $a$ :

$$a = \arg \max_{a' \in ans} S(q, a') \quad (8)$$

其中, $ans$ 表示候选答案集合。

## 4 实验结果及分析

### 4.1 实验设置

#### 4.1.1 数据集

本文主要对人工智能、大数据、智能制造等领域的最新的学术成果构建了学术知识图谱。构建知识图谱主要有两种方法:一种是自顶向下的方法,首先构造好本体概念层次,即模式层,然后构建数据层,即从文本中抽取实体和关系对模式层进行填充;另一种方法是从底向上,首先对数据进行分析,然后通过信息抽取技术得到实体和关系,最后进行汇总,逐渐抽取上层模式层。本文知识图谱构建结合领域特性,采用自顶向下的方法进行。

针对创新过程中涉及的文献,我们首先采用了自顶向下的方法定义模式层,其包括处理任务、处理方法、处理对象以及性能指标4类实体,并包括包含、应用、对比和同指4种关系。在实体和关系的确定过程中,利用了本团队研发的实体识别模型和关系抽取模型在近十年计算机领域的科技文件上进行了实体识别和关系抽取两个任务,并基于提取出的实体和关系自动构建了计算机领域的创新知识图谱,图3给出了图谱示例。



图3 计算机知识图谱示例

Fig. 3 Example of computer knowledge graph

本文根据已构建的创新知识图谱,通过人工的方式构建了标准的问答语料。该语料包含 2 000 个计算机领域的问答题 CF-QA(Computer Field-Question Answer),并采用 6:2:2 的方式将其划分为训练集、验证集和测试集。最终的数据集详细信息如表 1 所列。

表 1 数据集基本统计信息

Table 1 Basic statistics of dataset

	Question avg-len	Answer avg-len
Train	12.23	6.07
Validation	11.93	6.36
Test	12.01	6.12

数据集集中的问答题实例如下:

问句:查询语句为:当前序列建模方法有哪些?

答案:RNN,LSTM,GRU,Transformer

#### 4.1.2 评价指标

由于问答最终可归结为一个排序问题,因此,本文采用排序模型常用的评价指标 MAP 和 MRR。MAP 主要衡量平均准确率,它反映了模型在全部问答题上的性能,模型得到的相关答案越靠前,MAP 值可能就越高;MRR 衡量了问题对应的正确答案的排序位置的分数。MAP 和 MRR 可分别表示为:

$$MAP = \frac{1}{|Q|} \sum_{q \in Q} AvgP(q) \quad (9)$$

$$MRR = \frac{1}{|Q|} \sum_{q=1}^{|Q|} \frac{1}{rank(q)} \quad (10)$$

其中,|Q|为问题总数,AvgP(q)是对于一个问题的准确率,1/rank(q)为问题的第一个正确答案出现位置的倒数。

#### 4.1.3 参数设置

在模型训练过程中,词嵌入层的初始化是通过在大语料库上训练 word2vec 实现的。本文采用 Adam 优化器对参数进行优化,并利用了深度学习网络常用的学习率和权重衰减取值。除优化器所涉及的参数外,BiLSTM 和 GCN 网络的参数设置如表 2 所列。

表 2 实验参数设置

Table 2 Experimental parameter settings

实验参数设置	取值
word embedding size	300
LSTM/GCN hidden size	200
LSTM layer	1
GCN layer	3
learning rate	0.0001
lr decay	0.05

## 4.2 模型性能对比分析

为验证模型的性能,分别在通用数据集 TrecQA<sup>[18]</sup> 和计算机领域数据集 CF-QA 上进行了实验。TrecQA 数据集是 TREC 比赛提供的公开问答数据,其中包含两种形式的数据集,分别是 TRAIN 和 TRAIN-ALL。TRAIN 数据集的有效性由人工审核保证,而 TRAIN-ALL 数据集的有效性是通过正则匹配问题和答案自动验证保证。本文同时在这两个数据集上进行实验<sup>[19]</sup>。

将所提模型与已有模型在相同数据集上进行对比。参与对比的模型包括 3 个经典模型及最新的基于交互注意力机制的模型。1) Yang 等<sup>[20]</sup> 提出使用卷积神经网络对问句进行编码,同时引入了注意力机制学习词级别的语义信息;2) Yu 等<sup>[21]</sup> 提出利用基于层次残差 LSTM 的方法对问句进行编码;3) Sachan 等<sup>[22]</sup> 提出利用生成模型在已有问句的基础上生成更多的问句,再利用答案选择模块选择答案;4) Huang<sup>[19]</sup> 提出采用交互注意力机制捕捉问答语句之间的交互信息,进一步学习候选答案的语义表示。

对比模型与所提模型在两个数据集上的整体性能如表 3、表 4 所列。其中,ATEI-QA 表示模型引入了辅助任务进行模型训练,Un-ATEI-QA 表示模型未引入辅助任务进行模型训练。上述两个模型在数据集上的实验结果用于验证辅助任务的有效性。从表 3、表 4 可以看出,在通用领域和特定领域,所提模型的性能相比基线模型有了一定的提升。相比 4 个对比模型,ATEI-QA 在 TRAIN-ALL 和 TRAIN 两个数据集上的 MAP 值与 MRR 值平均提升 3.84% 和 1.56%,2.48% 和 2.64%;在 CF-QA 数据集上的 MAP 值与 MRR 值平均提升 2.48% 和 1.56%。同时,相比最经典的模型,无论是否引入辅助任务,所提模型的性能均有一定的提升,其 MAP 值和 MRR 值最高提升了 7.2% 和 4.79%。

表 3 TrecQA 实验结果

Table 3 Experimental results on TrecQA

Models	TRAIN-ALL		TRAIN	
	MAP	MRR	MAP	MRR
Yang <sup>[20]</sup>	0.7495	0.8109	0.7417	0.8102
Yu <sup>[21]</sup>	0.7613	0.8312	0.7701	0.8218
Sachan <sup>[22]</sup>	0.7980	0.8540	—	—
IKAAS <sup>[19]</sup>	<b>0.8237</b>	<b>0.8681</b>	0.7843	0.8544
Un-ATEI-QA	0.8013	0.8519	<b>0.7985</b>	0.8482
ATEI-QA	0.8215	0.8523	0.7914	<b>0.8579</b>

表 4 CF-QA 实验结果

Table 4 Experimental results on CF-QA

Model	CF-QA	
	MAP	MRR
Yang <sup>[20]</sup>	0.7301	0.7918
Yu <sup>[21]</sup>	0.7577	0.8053
Sachan <sup>[22]</sup>	0.7611	0.8172
IKAAS <sup>[19]</sup>	0.7705	0.8013
Un-ATEI-QA	0.7762	0.7814
ATEI-QA	<b>0.7824</b>	<b>0.8221</b>

从提升效果来看,所提模型在计算机领域的提升效果

逊于通用领域的提升效果,这是由于本文构建的知识图谱相比 DBpedia 是一个较小的知识图谱,在进行图表示学习时训练数据相对较少,导致表示所包含的语义信息不够充足,当与意图进行匹配时,匹配结果并非完全准确。同时,观察 Un-ATEI-QA 模型与 ATEI-QA 模型在两个数据集上的实验效果发现,Un-ATEI-QA 模型的性能均低于几个基准模型和 ATEI-QA 模型,原因在于其他基准模型都从多个维度进行了意图识别;而 ATEI-QA 的实验效果提升较为明显,表明了该模型可充分地识别问句意图。

上述在计算机领域和通用领域数据集上的测试结果表明,采用辅助任务增强机制可以提升问句意图的识别率。

### 4.3 智能问答系统应用

为了支撑学生开展自主创新活动,针对研究型学习过程中的常见需求实现了辅助创新系统。表 5 列出了常见的问答任务和典型的问句样例。

表 5 典型的智能问答任务

Table 5 Typical task of intelligent QA

常见任务	典型问句
查找解决问题的方法	最新的人脸识别方法?
查找方法的用途	遗传算法有哪些新应用?
查找包含方法文献	深度学习的最新文献?
查找研究人员	异常检查领域的知名专家?
查找研究机构	哪些机构擅长 NLP 研究?
发现热点问题	AI 领域的热点研究有哪些?

图 4 给出了查找并解决某个特定问题的方法时的结果视图。当使用者利用自然语言描述待求解的问题“卷积神经网络有哪些最新应用”时,在模型初始阶段,“卷积神经网络”作为实体被识别出来作为主题词,且“有哪些最新应用”被识别出来作为关系词,即意图。根据问句主题词在知识图谱上定位相应实体,计算意图与所包含的边即关系之间的匹配分数,筛选出最有可能的答案。从页面可看出,智能问答系统不仅列出了常见的方法,同时给出了这些方法的学术来源文献。



图 4 智能问答效果展示

Fig. 4 Display of intelligent QA

**结束语** 本文从人工智能、大数据、智能制造等领域的最新学术成果内容中提取知识三元组,构建了计算机领域知识图谱,并利用该图谱对常见的问题进行智能解答,是传统科技文献资源服务的有益补充。在构建问答模型时,引入了辅助任务学习问句意图表示,并在特定领域和通用领域数据集上进行了实验。实验结果表明,所提模型较为准确地提取出了问句意图,并且能同时应用于特定领域和通用领域。目前,模型仅是简单地从问句出发进行意图识别,并未考虑外部信息对意图识别带来的影响。因此,在下一步工作中,我们将结合用户行为、使用环境等信息进行研究,进一步提高意图识别的准确率。

### 参考文献

[1] PRAGER J M. Open-Domain Question-Answering [J]. *Foundation and Trends in Information Retrieval*, 2006, 1(2): 91-231.

[2] SUN H, DHINGRA B, ZAHEER M, et al. Open Domain Question Answering Using Early Fusion of Knowledge Bases and Text [C] // *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, 2018: 4231-4242.

[3] TONG P, ZHANG Q, YAO J. Leveraging domain context for

question answering over knowledge graph [J]. *Data Science and Engineering*, 2019, 4(4): 323-335.

[4] LIANG Z P, JI Z, LIU X L. Research on Question and Answer System of Paper Template [J]. *Journal of Shenzhen University: Science and Technology Edition*, 2007, 24(3): 281-285.

[5] YAO X, VAN D B. Information extraction over structured data: Question answering with freebase [C] // *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, 2014: 956-966.

[6] BORDES A, CHOPRA S, WESTON J. Question Answering with Subgraph Embeddings [C] // *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2014: 615-620.

[7] DONG L, WEI F, ZHOU M, et al. Question answering over freebase with multicolumn convolutional neural networks [C] // *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*, 2015: 260-269.

[8] HAO Y, ZHANG Y, LIU K, et al. An end-to-end model for question answering over knowledge base with cross-attention combining global knowledge [C] // *Proceedings of the 55th An-*

- nual Meeting of the Association for Computational Linguistics, 2017:221-231.
- [9] XU K, REDDY S, FENG Y, et al. Question Answering on Freebase via Relation Extraction and Textual Evidence[C]// Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics(Long Papers), 2016:2326-2336.
- [10] SAXENA A, TRIPATHI A, TALUKDAR P. Improving Multi-hop Question Answering over Knowledge Graphs using Knowledge Base Embeddings[C]// Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 2020:4498-4507.
- [11] JIANG H, YANG B, JIN L, et al. A BERT-Bi-LSTM-Based Knowledge Graph Question Answering Method[C]// 2021 International Conference on Communications, Information System and Computer Engineering (CISCE). IEEE, 2021:308-312.
- [12] KACUPAJ E, PLEPI J, SINGH K, et al. Conversational question answering over knowledge graphs with transformer and graph attention networks[J]. arXiv:2104.01569, 2021.
- [13] XIONG H, WANG S, TANG M, et al. Knowledge Graph Question Answering with semantic oriented fusion model[J]. Knowledge-Based Systems, 2021, 221:106954.
- [14] WU P, WU Y, WU L, et al. Modeling Global Semantics for Question Answering over Knowledge Bases[J]. arXiv: 2101.01510, 2021.
- [15] SCHLICHTKRULL M, KIPF T N, BLOEM P, et al. Modeling relational data with graph convolutional networks[C]// European Semantic Web Conference. Cham:Springer, 2018:593-607.
- [16] HJELM R D, FEDOROV A, LAVOIE-MARCHILDON S, et al. Learning deep representations by mutual information estimation and maximization[J]. arXiv:1808.06670, 2018.
- [17] KIPF T N, WELING M. Semi-supervised classification with graph convolutional networks[J]. arXiv:1609.02907, 2016.
- [18] WANG M, SMITH N A, MITAMURA T. What is the Jeopardy model? A quasi-synchronous grammar for QA[C]// Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning(EMNLP-CoNLL). 2007:22-32.
- [19] HUANG W Y. Deep Neural Networks for Legal Question Answering Based on Knowledge Graph[D]. Beijing:University of Chinese Academy of Sciences, 2020.
- [20] YANG L, AI Q, GUO J, et al. aNMM:Ranking short answer texts with attention-based neural matching model[C]// Proceedings of the 25th ACM International Conference on Information and Knowledge Management, 2016:287-296.
- [21] YU M, YIN W, HASAN K S, et al. Improved neural relation detection for knowledge base question answering[J]. arXiv:1704.06194, 2017.
- [22] SACHAN M, XING E. Self-training for jointly learning to ask and answer questions[C]// Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics; Human Language Technologies, 2018:629-640.



**ZHONG Jiang**, born in 1974, Ph.D, professor. His main research interests include natural language processing, big data analysis and mining, cloud and network integration technology.

(责任编辑:李亚辉)