

文章编号: 1006-2475(2018)09-0005-06

# 基于多粒度特征表示的知识图谱问答

申 存<sup>1 2 3</sup>, 黄廷磊<sup>2 3</sup>, 梁 霄<sup>2 3</sup>(1. 中国科学院大学电子电气与通信工程学院, 北京 100049; 2. 中国科学院电子学研究所, 北京 100190;  
3. 中国科学院空间信息处理与应用系统技术重点实验室, 北京 100190)

**摘要:** 近年来, 基于知识图谱的问答系统逐渐成为学术界和工业界的研究和应用热点方向, 而传统方法通常存在效率不高以及未充分利用数据信息的问题。针对以上问题, 本文将中文知识图谱问答分为实体抽取和属性选择 2 个子任务, 采用双向长短期记忆条件随机场(Bi-LSTM-CRF)模型来进行实体识别, 并提出一种多粒度特征表示的属性选择模型。该模型采用字符级别以及词级别分别对问句和属性进行嵌入表示并通过编码器进行编码, 对于属性同时还引入热度编码的信息。通过不同粒度文本表示的结合, 并对问句和属性进行相似度计算, 最终该系统在 NLPCC-ICCPOL 2016 KBQA 数据集上取得了 73.96% 的 F1 值, 能够较好地完成任务。

**关键词:** 知识图谱; 问答系统; 实体抽取; 属性选择

中图分类号: TP391

文献标识码: A

doi: 10.3969/j.issn.1006-2475.2018.09.002

## Knowledge Graph Question Answering Based on Multi-granularity Feature Representation

SHEN Cun<sup>1 2 3</sup>, HUANG Ting-lei<sup>2 3</sup>, LIANG Xiao<sup>2 3</sup>(1. School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing 100049, China;  
2. Institute of Electronics, Chinese Academy of Sciences, Beijing 100190, China;  
3. Key Laboratory of Technology in Geo-Spatial Information Processing and Application System, Chinese Academy of Sciences, Beijing 100190, China)

**Abstract:** Recently, knowledge graph question answering has gradually become the focus of academic and industrial circles. However, traditional methods often have problems of inefficiency and insufficient use of data information. In order to solve the problems above, this paper divides the Chinese knowledge graph question answering into two sub-tasks: entity extraction and property selection. The Bi-LSTM-CRF model is used to identify entities, and a multi-granularity feature representation model is proposed to carry out property selection. The model utilizes character-level and word-level to represent questions and properties and encode them through the encoder. At the same time, it also introduces the one-hot information for the properties. Through the combination of multi-granularity text representations and the similarity calculation of questions and properties, the system finally achieves a 73.96% F1 value on the NLPCC-ICCPOL 2016 KBQA data set, which finishes the knowledge graph question and answer task successfully.

**Key words:** knowledge graph; question answering system; entity extraction; property selection

## 0 引 言

开放域知识图谱问答是一项具有挑战性的任务, 其旨在结合知识图谱为自然语言问题提供相应的答案。近年来, 大规模的高质量知识图谱发展迅速, 并在许多领域得到了广泛的应用, 典型的包括如 Free-

base<sup>[1]</sup>、DBpedia<sup>[2]</sup> 等英文知识图谱以及 Zhishi.me<sup>[3]</sup>、XLore<sup>[4]</sup> 等中文知识图谱。由于知识的结构化形式, 知识图谱已经成为开放领域问答的重要资源, 越来越多的研究工作也集中在知识图谱问答上<sup>[5-6]</sup>。对于知识图谱问答, 其主要挑战是对问句的语义理解, 因为给定的问句是自然语言的形式, 而知识图谱

收稿日期: 2018-03-15

基金项目: 国家自然科学基金资助项目(61725105, 61331017)

作者简介: 申存(1992-), 男, 四川广安人, 中国科学院大学电子电气与通信工程学院、中国科学院电子学研究所硕士研究生, 研究方向: 自然语言处理, 知识图谱; 黄廷磊(1971-), 男, 研究员, 博士生导师, 博士, 研究方向: 数据挖掘, 大数据分析; 梁霄(1981-), 男, 助理研究员, 博士, 研究方向: 数据组织管理, 知识图谱。

是结构化的信息存储,两者的表述存在差异,需要对问句和知识图谱的文本进行深入的语义关联,以从知识图谱中选取与问句表述最为匹配的三元组作为候选答案。例如给定问句“你知道哈姆雷特是哪个国家的电影吗?”,首先需要从知识图谱中确定问句所包含的主题实体“哈姆雷特(1964年美国电影)”,然后从实体的属性中选出与表述“是哪个国家的电影”最为相关的属性“制片地区”以得到三元组“哈姆雷特(1964年美国电影)||制片地区||美国”。该过程主要包含2部分工作:实体抽取以及属性选择。

实体抽取主要是从问句中识别出实体提及并链接至知识图谱的过程。目前传统的研究主要通过搜索知识图谱中每个问句的  $n$  元语法( $n$ -gram)来实现实体抽取<sup>[8-9]</sup>,这种方法通常需要较大的搜索空间。Berant 等人<sup>[5]</sup>使用语言学工具来完成实体抽取工作,而通常语言学工具依赖于逻辑表达式以及预定义规则,不具有广泛的适应性。

属性选择则是在抽取问句的实体之后,从实体的所有属性中选取与问句描述最为匹配的属性,并将属性值作为候选答案。这一任务的难点在于自然语言问句的表述与属性的文本表述存在差异性,如何有效地将两者关联并选出置信度最高的属性直接影响到结果的准确率。在相关工作中,深度学习的方法正逐渐应用于知识图谱问答的属性选择中。Yih 等人<sup>[10]</sup>利用卷积神经网络(Convolutional Neural Network, CNN)对问句和属性进行字符级别三元语法的建模。Golub 等人<sup>[9]</sup>将关系和问句作为字符序列进行处理并提出基于注意力的长短期记忆(Long Short-Term Memory, LSTM)网络的方法。Yin 等人<sup>[11]</sup>采用注意力池化方法来学习属性的嵌入表示。这些属性选择方法都利用神经网络分别学习问句和属性的向量表示,然后计算向量之间的相似度作为其语义相似度。这些方法在实验中仅使用词级别嵌入,没有充分利用实验数据的语义信息。且与英文知识图谱问答不同的是,在中文里单个汉字通常也含有特定的语义,因此可以考虑结合字符级别的嵌入进行属性选择的实验。

针对上述问题,在实体抽取任务中,本文采用双向长短期记忆条件随机场(Bi-LSTM-CRF)模型来进行实体识别,取得了较高的链接至知识图谱的准确性。对于属性选择任务,本文采用结合字符级别、词级别以及属性级别的文本嵌入的多粒度方法,充分利用图谱的可用信息,并与其他模型方法进行对比,以验证方法的有效性。

## 1 相关工作

基于知识图谱的问答在自然语言处理领域拥有很长的研究历史。早在20世纪60年代,就有研究人员针对领域内小规模知识库进行问答系统的研究以回答领域内的一些专业问答。早期的研究主要采用语义解析(Semantic Parsing)的方法<sup>[5,10,12]</sup>,其主要思想是按照特定的文法,将自然语言问句转化为等价的逻辑表达式,以完成对知识库的查询。除此之外,信息检索方法<sup>[6,8,13]</sup>也常用于知识图谱问答,与语义解析不同的是,它将问句转化为了检索问题。该方法从知识图谱中搜索问句中涉及的所有相关信息,并构建排序算法从候选答案中选择最佳候选答案。相比于语义解析,其优点是不必手动设计词汇表且领域迁移能力较强。Bordes 等人<sup>[8]</sup>的结果表明,信息检索方法在问答结果中也有着较好的表现。

近年来,随着人工智能的发展,神经网络的方法也开始应用于知识图谱问答,并取得了相比于传统方法更好的实验结果。在实体抽取中,Bordes 等人<sup>[8]</sup>和 Golub 等人<sup>[9]</sup>搜索给定问题的所有  $n$  元语法单词,然后链接到知识图谱。Berant 等人<sup>[5]</sup>使用语言学工具,其很大程度上依赖于逻辑表达式和预定义规则。Dai 等人<sup>[14]</sup>将中心实体映射回问句文本作为标注数据,并构建双向门控循环单元条件随机场(Bi-GRU-CRF)序列标注模型以进行实体识别。Yin 等人<sup>[11]</sup>则采用 Bi-LSTM-CRF 序列标注模型来提高该方法的性能。对于属性选择任务,Bordes 等人<sup>[15]</sup>首先采用深度学习的方法并取得了较好的实验结果,之后各种基于深度学习的模型也逐渐涌现出来。这些方法大都是将给定问题和候选关系分别映射到向量,然后计算向量之间的相似度作为它们的语义相似度。在 Dai 等人<sup>[14]</sup>的研究中,属性被视为整体的符号,并采用 TransE<sup>[16]</sup>学习的预训练向量初始化。文献[9]采用了字符级别表示法,以减小参数的大小并提高处理未登录词的鲁棒性,而 Yin 等人<sup>[11]</sup>则提出了注意力机制的最大池化的 CNN 模型。

## 2 基本框架

### 2.1 问题定义与建模

给定目标问句,实体抽取的目的是找到实体提及并正确链接至知识图谱,从而得到主题实体和候选属性  $C_p = \{\text{prop}_1, \text{prop}_2, \dots, \text{prop}_n\}$ 。属性选择的目的是识别问句中表述的属性,即找到与问句中除主题实体以外的文本描述最相符的实体属性。通常将属性选择任务考虑为排序问题,对于问句  $q$  候选属性集合

$C_p$  中的每个属性  $p$ , 模型计算其与问题  $S(q, p)$  的语义相似度, 并且选择置信度最高的属性作为候选, 有:

$$p^+ = \arg \max S(q, p) \quad (1)$$

## 2.2 知识图谱问答流程

对于中文知识图谱问答, 其流程如图 1 所示, 主要包括以下几个步骤: 1) 对知识图谱进行规范化处理, 主要包括删除三元组中属性之间的空格、前缀以及后缀, 对于某些属性和属性值相同的无效三元组, 直接删除该条记录; 2) 对问句进行主题实体的识别, 并利用规范化的知识图谱进行实体链接; 3) 检索知识图谱得到链接实体的所有候选属性, 并利用设计的属性选择模型进行比较排序, 选出置信度最高的候选属性, 得到预测答案。

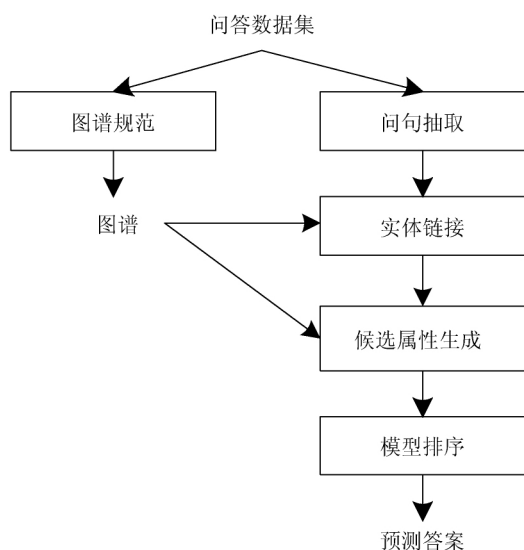


图1 知识图谱问答流程图

## 3 实体链接模型

对于实体抽取模型, 其主要难点是识别问句中的主题实体, 以便进一步链接至知识图谱, 检索相应候选属性, 以确定最终答案。笔者发现该任务与命名实体识别任务有着很大的共同点, 因此本文采用由 Bi-LSTM 和 CRF 模型组成的 Bi-LSTM-CRF 模型<sup>[17]</sup>。基本思想是使用 LSTM 层来考虑先前的输入特征并从 CRF 层获得句子级别标签信息。因此, 输出是一个最佳的标签序列, 而不是相互独立的标签。

形式上, 输入一个问句序列  $X = \{X_1, X_2, \dots, X_n\}$ ,  $y = \{y_1, y_2, \dots, y_n\}$  表示  $X$  的标签序列,  $P_{n \times k}$  表示概率矩阵, 其中  $k$  是标签类型的数量。最佳标签序列可通过最大化如下目标函数来获得。

$$S(X, y) = \sum_{i=0}^n A_{y_i, y_{i+1}} + \sum_{i=0}^n P_{i, y_i} \quad (2)$$

其中  $P_{i, j}$  是第  $i$  个单词被标记为第  $j$  个标签的概率,  $A$  是状态转移矩阵, 其中元素  $A_{i, j}$  是从第  $i$  个标签转移

到第  $j$  个标签的概率。

整体实体抽取模型如图 2 所示。问句以字序列文本嵌入, 并与额外的特征连接作为循环层的输入。本文采用代表文字边界特征的一个热点向量来进行说明。循环层为双向 LSTM, 其将前向和后向隐层表示的输出连接并投影到每个标签以计算得分, CRF 层主要用于解决参数偏倚问题。

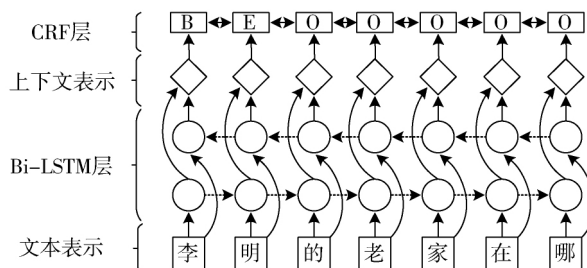


图2 实体抽取模型

## 4 多粒度特征表示属性选择模型

属性选择中, 其主要难点是问句中属性描述可能与知识图谱中的属性名称存在着较大的差异, 例如“请问红楼梦是什么时候写的?”就需要与知识图谱中的“创作年代”相关联。因此, 如何设计有效的模型以最高的置信度将问句与属性进行关联是本文研究重点。对此, 本文提出一种多粒度特征表示模型, 采用 GRU 编码器来获得问句和属性的隐层表示。在文本表示中, 模型考虑字符级别和词级别以获得更丰富的语义信息。最后使用余弦相似度来计算问句与属性之间的语义相似度度量。

与英文不同的是, 中文里单个的汉字通常具有语义, 而英文孤立的字母通常并没有具体的含义。因此在中文知识图谱问答中, 对于属性  $p$ , 本文考虑不同的粒度来表示特征: 字符级别、词级别以及属性级别。字级别的建模将属性拆为单个汉字进行嵌入式表达, 词级别则是通常意义下的进行分词后引入词嵌入, 而属性级别的表示是将属性整体视为唯一符号。3 种类型的属性表示包含不同层次的抽象意义, 各粒度层次都有其自身的优缺点。在实验中属性级别表示采用随机初始化, 它更多地关注全局信息, 但其存在数据稀疏性的缺点。单词级别更注重局部信息, 如单词和短语等。然而, 这 2 个级别都受到未登录词问题的影响, 字符级别没有这样的问题, 并且通常在预测正确的实体和属性方面也有着较高的准确率。

以下详细介绍多粒度特征表示模型。为了利用属性不同粒度的信息并将其结合, 本文采用嵌套的连接方法, 将预训练的字嵌入和词嵌入经过 GRU 的编码并进行组合, 其网络表示如图 3 所示。

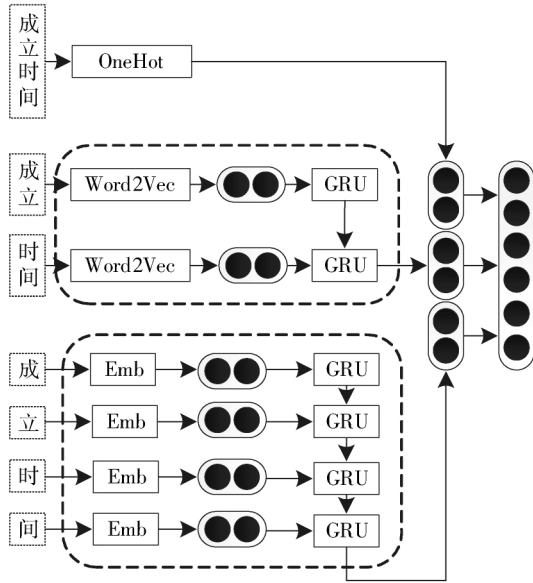


图3 多粒度特征表示模型

实验中文本的词嵌入和字嵌入采用的是基于维基百科语料进行 Word2Vec 预训练字词向量<sup>[18]</sup>。对于属性字序列  $\{c_1^p, c_2^p, \dots, c_n^p\}$ , 其字级别嵌入的定义为:

$$\vec{v}_i^c = W_c^T v_i^c \quad (3)$$

其中  $W_c^T \in \mathbb{R}^{|V_c| \times d_c}$  是词汇量为  $|V_c|$  的字符嵌入矩阵,  $v_i^c$  为字符  $c_i$  的独热表示 (One-hot Representation),  $d_c$  为字向量的维度。最终通过 GRU 对字符进行编码得到:

$$r_p^c = \text{ENC}_c([\vec{v}_1^c, \vec{v}_2^c, \dots, \vec{v}_n^c]) \quad (4)$$

同理对于属性词序列  $\{w_1^p, w_2^p, \dots, w_n^p\}$ , 其词级别的嵌入定义为:

$$\vec{v}_i^w = W_w^T v_i^w \quad (5)$$

其中  $W_w^T \in \mathbb{R}^{|V_w| \times d_w}$  是词汇量为  $|V_w|$  的词嵌入矩阵,  $v_i^w$  为词  $w_i$  的独热表示,  $d_w$  为词向量的维度。编码后同样有:

$$r_p^w = \text{ENC}_w([\vec{v}_1^w, \vec{v}_2^w, \dots, \vec{v}_n^w]) \quad (6)$$

最终得到 3 种不同粒度的属性表示为:

$$r_p = \begin{bmatrix} r_p^o \\ r_p^w \\ r_p^c \end{bmatrix} \quad (7)$$

其中  $r_p^o$  为属性的独热表示。同样对于问句, 本文对其类似地采用字级别和词级别的编码组合, 得到:

$$r_q = \begin{bmatrix} r_q^w \\ r_q^c \end{bmatrix} \quad (8)$$

其中  $r_q^w$  为问句词序列的 GRU 编码,  $r_q^c$  为问句字符序列的 GRU 编码。最终采用余弦相似度计算属性表示  $r_p$  和问句表示  $r_q$  的关联得分:

$$S(q, p) = \cos(r_q, r_p) \quad (9)$$

其中余弦相似度函数  $\cos$  定义为:

$$\cos(a, b) = \frac{(a \bullet b)}{|a| |b|} \quad (10)$$

## 5 实验分析

### 5.1 实验数据

本文实验使用的数据集为 NLPCC-ICCPOL 2016 KBQA 数据集。该数据集是目前最大的公开中文知识图谱问答数据集, 其包含大约 4300 万个三元组和 600 万个实体。该知识图谱的三元组大部分来自百度百科的属性表格。在数据集中, 14609 条问答对作为训练数据以及 9870 条问句作为测试数据。

### 5.2 实验设置

对于上述表示模型, 本文采用排序模型进行训练, 该方法驱动模型输出包含在训练集中的问题实体和问题谓词对的高分, 同时为不合理配对产生较低分数。在训练期间最小化的损失函数由下式给出:

$$-\sum_{(q, p) \in C} \max(0, S(q, p^-) - S(q, p^+) + \gamma) \quad (11)$$

因此在训练中, 模型主要关注负例和正例得分之差小于边界  $\gamma$  的数据对, 以使得正例和负例得分相差越大越好。

### 5.3 实验结果

对于实体识别模型, 本文采用的是 100 维的字符级向量。LSTM 隐层维度为 100, dropout 为 0.5, 学习率为 0.001, 本文采用反向传播算法来更新训练中的参数。实验中, 随机选取 10% 训练数据作为验证集<sup>[19]</sup>, 结果如表 1 所示, 可以看出在测试集上, 实体识别 F1 值为 97.36%, 取得了较好的识别率, 证明该模型的有效性, 也为属性选择实验提供了有效的实验结果。

表1 实体抽取实验结果

	准确率/%	召回率/%	F1/%
验证集	97.56	97.48	97.51
测试集	97.41	97.32	97.36

属性选择实验中, 本文对比了选择不同词向量和字向量时的实验结果, 如表 2 所示。可以看出词嵌入和字嵌入采用 50 维时并不能较好地表示汉字, 其实验结果甚至低于对比实验中单独采用 100 维的词向量的结果。对于本实验, 词向量和字向量分别取 100 维时得到最优实验结果, 而随着维度的增加, 属性选择模型的 F1 并没有明显的提升。因此, 最终本文选择词向量和字向量维度都为 100。其他参数如 GRU 编码器隐层维度为 200, dropout 设置为 0.3。

表 2 不同维度字向量与词向量实验结果

Word-Level	Char-Level/%		
	50 dims	100 dims	200 dims
50 dims	71.45	72.49	72.28
100 dims	72.34	73.96	73.90
200 dims	72.11	73.78	73.57

本文同时与 NLPCC 官方提供的基线模型以及只采用词级别嵌入表示并通过 GRU 进行编码的模型进行对比。实验结果如表 3 所示,可以看出本文模型在最终结果上比基线模型有了很大的提高,且与只采用词级别信息表示模型相比,结合字符级别、词级别以及独热编码信息的组合模型,更能充分对数据进行表示。相比于词级别模型,多粒度模型包含更加丰富的表示信息,其包含的字符级别模型可以更好地处理单个汉字的语义信息,例如对问句“列克星敦号航空母舰能载多少人?”,字级别信息的引入能够使得属性“人员编制”的置信度更加准确,同时对于未登录词,字符级别模型也能较好地进行处理;而独热编码更关注全局信息,对于问句“陈浩民的家庭成员有哪些?”,独热信息可以直接与属性“家庭成员”进行匹配。因此相比于词级别模型,多粒度表示模型可以达到比采用单一表示更好的实验结果。

表 3 问答实验结果对比 单位: %

模型	F1	Pre@1	Pre@2	Pre@5
NLPCC	52.48	52.48	60.46	67.33
Word-Level	71.60	71.60	76.38	79.55
本文	73.96	73.96	79.45	82.51

同时,本文也将实验结果与其他在该中文知识图谱问答数据集上进行实验的论文结果进行对比,如表 4 所示。前 3 名的结果分别为 82.47%、81.59%、79.57%,且作者在实验中基本都采用了一些预定义的规则以及集成方法对模型进行优化。本文在仅使用单一神经网络模型、结构尽量简单的情况下,也取得了较好的实验结果,验证了模型的有效性。

表 4 不同实验结果比较

模型	F1/%
PKU <sup>[20]</sup>	82.47
NUDT <sup>[21]</sup>	81.59
CCNU <sup>[22]</sup>	79.57
NEU	72.72
本文	73.96

## 6 结束语

本文提出了一种中文知识图谱问答的方法,有效地解决了其中实体抽取和属性选择 2 个主要任务。在实体抽取中,本文采用命名实体识别的方法,训练

Bi-LSTM-CRF 模型以获得问句主题实体。在属性选择任务中,本文提出多粒度特征表示模型,将字符级别、词级别的文本表示进行编码,并考虑属性的独热编码信息,将不同粒度的信息进行结合,充分利用数据特征。实验结果表明本文的模型在中文知识图谱问答中取得了较好的效果。

### 参考文献:

- [1] Bollacker K, Evans C, Paritosh P, et al. Freebase: A collaboratively created graph database for structuring human knowledge [C]// Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data. 2008: 1247-1249.
- [2] Auer S, Bizer C, Kobilarov G, et al. DBpedia: A nucleus for a Web of open data [C]// Proceedings of the 6th International Semantic Web Conference and the 2nd Asian Semantic Web Conference. 2007: 722-735.
- [3] Niu Xing, Sun Xinruo, Wang Haofen, et al. Zhishi. me: Weaving Chinese linking open data [C]// Proceedings of the 10th International Conference on the Semantic Web. 2011 2: 205-220.
- [4] Wang Zhi-chun, Wang Zhi-gang, Li Juan-zi, et al. Knowledge extraction from Chinese wiki encyclopedias [J]. Journal of Zhejiang University-Science C (Computers & Electronics), 2012, 13(4): 268-280.
- [5] Berant J, Chou A, Frostig R, et al. Semantic parsing on freebase from question-answer pairs [C]// Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing. 2013: 1533-1544.
- [6] Dong Li, Wei Furu, Zhou Ming, et al. Question answering over freebase with multi-column 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 1: 260-269.
- [7] Yu Lei, Hermann K M, Blunsom P, et al. Deep Learning for Answer Sentence Selection [DB/OL]. <https://arxiv.org/pdf/1412.1632v1.pdf>, 2014-12-04.
- [8] Bordes A, Usunier N, Chopra S, et al. Large-scale Simple Question Answering with Memory Networks [DB/OL]. <https://arxiv.org/pdf/1506.02075v1.pdf>, 2015-06-05.
- [9] Golub D, He Xiaodong. Character-level question answering with attention [C]// Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. 2016: 1598-1607.
- [10] Yih W, Chang M W, He Xiaodong, et al. Semantic parsing via staged query graph generation: Question answering with knowledge base [C]// Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing. 2015 1: 1321-1331.

- [11] Yin Wenpeng , Yu Mo , Xiang Bing , et al. Simple question answering by attentive convolutional neural network [C]// Proceedings of the 26th International Conference on Computational Linguistics. 2016: 1746-1756.
- [12] Cai Qingqing , Yates A. Large-scale semantic parsing via schema matching and lexicon extension [C]// Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics. 2013 ,1: 423-433.
- [13] Yao Xuchen , Van Durme B. Information extraction over structured data: Question answering with freebase [C]// Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics. 2014 ,1: 956-966.
- [14] Dai Zihang , Li Lei , Xu Wei. CFO: Conditional focused neural question answering with large-scale knowledge bases [C]// Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. 2016 ,1: 800-810.
- [15] Bordes A , Weston J , Usunier N. Open question answering with weakly supervised embedding models [C]// Proceedings of the 2014 European Conference on Machine Learning and Knowledge Discovery in Databases. 2014: 165-180.
- [16] Bordes A , Usunier N , Garcia-Duran A , et al. Translating embeddings for modeling multi-relational data [C]// Proceedings of the 2013 International Conference on Neural Information Processing Systems. 2013: 2787-2795.
- [17] Lample G , Ballesteros M , Subramanian S , et al. Neural architectures for named entity recognition [C]// Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2016: 260-270.
- [18] Dong Chuanhai , Zhang Jiajun , Zong Chengqing , et al. Character-based LSTM-CRF with radical-level features for Chinese named entity recognition [C]// Proceedings of the 24th International Conference on Computer Processing of Oriental Languages and the 5th CCF Conference on Natural Language Processing and Chinese Computing. 2016: 239-250.
- [19] 周志华. 机器学习 [M]. 北京: 清华大学出版社, 2016.
- [20] Lai Yuxian , Lin Yang , Chen Jiahao , et al. Open domain question answering system based on knowledge base [C]// Proceedings of the 24th International Conference on Computer Processing of Oriental Languages and the 5th CCF Conference on Natural Language Processing and Chinese Computing. 2016: 722-733.
- [21] Yang Fengyu , Gan Liang , Li Aiping , et al. Combining deep learning with information retrieval for question answering [C]// Proceedings of the 24th International Conference on Computer Processing of Oriental Languages and the 5th CCF Conference on Natural Language Processing and Chinese Computing. 2016: 917-925.
- [22] Xie Zhiwen , Zeng Zhao , Zhou Guangyou , et al. Knowledge base question answering based on deep learning models [C]// Proceedings of the 24th International Conference on Computer Processing of Oriental Languages and the 5th CCF Conference on Natural Language Processing and Chinese Computing. 2016: 300-311.

( 上接第 4 页)

- [6] Lecun Y , Bottou L , Bengio Y , et al. Gradient-based learning applied to document recognition [J]. Proceedings of the IEEE , 1998 , 86( 11) : 2278-2324.
- [7] Ma Mingbo , Huang Liang , Xiang Bing , et al. Group sparse CNNs for question classification with answer sets [C]// Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. 2017 ,2: 335-340.
- [8] Gers F A , Schmidhuber J , Cummins F. Learning to forget: Continual prediction with LSTM [J]. Neural Computation , 2000 ,12( 10) : 2451-2471.
- [9] Bahdanau D , Cho K , Bengio Y. Neural Machine Translation by Jointly Learning to Align and Translate [DB/OL]. <https://arxiv.org/pdf/1409.0473v7.pdf>, 2016-05-49.
- [10] Ding Zixiang , Xia Rui , Yu Jianfei , et al. Densely Connected Bidirectional LSTM with Applications to Sentence Classification [DB/OL]. <https://arxiv.org/pdf/1802.00889v1.pdf>, 2018-02-03.
- [11] Graves A , Schmidhuber J. Framewise phoneme classification with bidirectional LSTM and other neural network architectures [J]. Neural Networks , 2005 ,18( 5-6) : 602-610.
- [12] Yang Zichao , Yang Diyi , Dyer C , et al. Hierarchical attention networks for document classification [C]// Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2016: 1480-1489.
- [13] Jain L C , Medsker L R. Recurrent Neural Networks: Design and Applications [M]. CRC Press , 1999.
- [14] Hochreiter S , Schmidhuber J. Long short-term memory [J]. Neural Computation , 1997 ,9( 8) : 1735-1780.
- [15] Graves A. Generating Sequences with Recurrent Neural Networks [DB/OL]. <https://arxiv.org/pdf/1308.0850v5.pdf> , 2014-06-05.
- [16] 张栋 , 李寿山 , 王晶晶. 基于问题与答案联合表示学习的半监督问题分类方法 [J]. 中文信息学报 , 2017 ,31( 1) : 1-7.
- [17] Mikolov T , Sutskever I , Chen Kai , et al. Distributed representations of words and phrases and their compositionality [C]// Proceedings of the 2013 Advances in Neural Information Processing Systems. 2013: 3111-3119.
- [18] Prechelt L. Early stopping-but when? [M]// Neural Networks: Tricks of the Trade. Springer , 1998: 55-69.
- [19] Srivastava N , Hinton G , Krizhevsky A , et al. Dropout: A simple way to prevent neural networks from overfitting [J]. Journal of Machine Learning Research , 2014 ,15( 1) : 1929-1958.