

基于框架语义和图结构的阅读理解答案抽取方法

杨陟卓, 许玲玲, 张虎, 李茹

引用本文

杨陟卓, 许玲玲, 张虎, 李茹. [基于框架语义和图结构的阅读理解答案抽取方法](#)[J]. 计算机科学, 2023, 50(8): 170-176.

YANG Zhizhuo, XU Lingling, Zhang Hu, LI Ru. [Answer Extraction Method for Reading Comprehension Based on Frame Semantics and GraphStructure](#) [J]. Computer Science, 2023, 50(8): 170-176.

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)

[基于多粒度实体异构图的篇章级事件抽取方法](#)

Document-level Event Extraction Based on Multi-granularity Entity Heterogeneous Graph
计算机科学, 2023, 50(5): 255-261. <https://doi.org/10.11896/jsjcx.220300154>

[基于会话式机器阅读理解模型的事件抽取方法](#)

Event Extraction Method Based on Conversational Machine Reading Comprehension Model
计算机科学, 2023, 50(2): 275-284. <https://doi.org/10.11896/jsjcx.220400271>

[基于菌群优化的近邻传播聚类算法研究](#)

Study on Affinity Propagation Clustering Algorithm Based on Bacterial Flora Optimization
计算机科学, 2022, 49(5): 165-169. <https://doi.org/10.11896/jsjcx.210800218>

[融入句子中远距离词语依赖的图卷积短文本分类方法](#)

Graph Convolutional Networks with Long-distance Words Dependency in Sentences for Short Text Classification
计算机科学, 2022, 49(2): 279-284. <https://doi.org/10.11896/jsjcx.201200062>

[神经问题生成前沿综述](#)

Frontiers in Neural Question Generation:A Literature Review
计算机科学, 2021, 48(6): 159-167. <https://doi.org/10.11896/jsjcx.201100013>

基于框架语义和图结构的阅读理解答案抽取方法

杨陟卓¹ 许玲玲¹ 张虎¹ 李茹^{1,2}

1 山西大学计算机与信息技术学院 太原 030006

2 山西大学计算智能与中文信息处理教育部重点实验室 太原 030006

摘要 机器阅读理解是自然语言处理领域最具挑战性的任务之一。随着深度学习技术的不断发展以及大规模 MRC 数据集的发布,机器阅读理解模型的性能不断刷新记录。但是以往的模型在逻辑推理、深层语义理解等方面仍存在不足。为解决上述问题,提出了一种基于框架语义和图结构的阅读理解答案抽取方法。该方法首先利用汉语框架网匹配与问句语义相关的候选句;其次提取问题和候选句中的实体,以实体在句子中的依存句法和语义关系构建实体关系图;最后将实体关系图引入图注意网络进行逻辑推理,实现阅读理解答案抽取。在 DuReader-robust 数据集上的实验结果表明,所提方法取得了比基线模型更好的效果。

关键词: 机器阅读理解;CFN;实体关系图;图注意网络;句法关系

中图法分类号 TP391

Answer Extraction Method for Reading Comprehension Based on Frame Semantics and Graph Structure

YANG Zhizhuo¹, XU Lingling¹, Zhang Hu¹ and LI Ru^{1,2}

1 School of Computer and Information Technology, Shanxi University, Taiyuan 030006, China

2 Key Laboratory of Computational Intelligence and Chinese Information Processing of Ministry of Education, Shanxi University, Taiyuan 030006, China

Abstract Machine reading comprehension is one of the most challenging tasks in the field of natural language processing. With the continuous development of deep learning technology and the release of large-scale MRC datasets, the performance of MRC models keep breaking records. However, the previous models still have shortcomings in logical reasoning and deep semantic understanding. In order to solve the above problems, this paper proposes a reading comprehension answer extraction method based on frame semantics and graph structure. This method first uses Chinese FrameNet to match candidate sentences related to question semantics. Secondly, the entities in the questions and candidate sentences are extracted, and the entity relationship graph is constructed based on the dependent syntax and semantic relationships of entities in the sentences. Finally, the entity relationship graph is introduced into the graph attention network for logical reasoning, so as to realize the extraction of reading comprehension answers. Experiment results on Dureader-robust dataset show that the proposed method achieves better results than the baseline model.

Keywords Machine reading comprehension, CFN, Entity relationship graph, Graph attention network, Syntactic relations

1 引言

机器阅读理解(Machine Reading Comprehension, MRC)的目标是使计算机理解文章语义并正确回答与文本相关的问题。近年来,随着深度学习技术的不断发展以及大规模 MRC 数据集的发布,大量优秀的模型不断被提出和改进,特别是在以 Transformer 和 BERT 为代表的大规模预训练模型发布之后, MRC 开始转向端到端的深度神经网络结构,这些模型

可以更好地发现自然语言中的潜在特征,模型的性能也不断刷新记录。

由于模型受到计算和存储性能的限制,在面对多篇章的长文本时无法对整体进行语义建模,导致其无法有效建立长距离语义关联。并且,通过对大量 MRC 模型的研究发现,当前的诸多阅读理解模型大多在文章、问题和候选答案之间使用注意力机制对信息进行交互和提取。虽然通过注意力机制能够交互文章与问题,并从中提取出一些重要的信息,但是它

到稿日期:2022-06-08 返修日期:2022-10-10

基金项目:国家自然科学基金(61936012, 62176145);山西省基础研究计划项目面上基金项目(20210302123469)

This work was supported by the National Natural Science Foundation of China (61936012, 62176145) and Shanxi Province Basic Research Program General Fund Project(20210302123469).

通信作者:杨陟卓(yangzhizhuo@sxu.edu.cn)

可能会过度关注某些特定的词语而遗漏某些重要词语,忽略了对句子整体语义的理解。此外,以往的模型在构图时,没有充分利用句子以及句中实体间丰富的语义关联进行建模,导致模型在逻辑推理时缺乏有效的语义线索。如图1机器阅读理解中文示例,在多段落的长文本中,并不是每个句子都对问题的解答有效,有的句子甚至对问题的解答存在一定的干扰,因此,问题相关句的筛选对于多篇章阅读理解任务是必要的;另外,通过分析句子中要素之间的句法依存关系可以获取句子的重要信息,通过问题关键词与文章中要素的关联可以追踪到句子之间的关系,进而获取与问题相关文本的整体语义,分析推理得到答案。从图1可以看出,句子要素的句法依存关系、句子整体结构以及句子之间的关系对问题的解答都尤为重要。

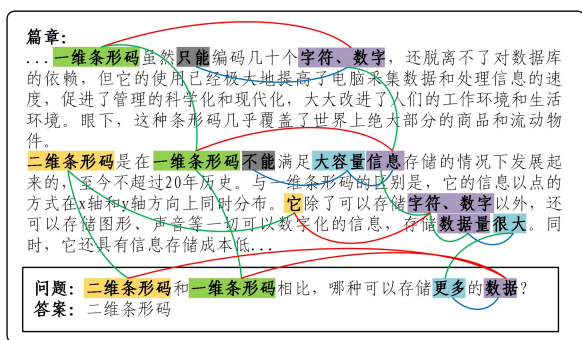


图1 机器阅读理解中文示例

Fig. 1 Chinese examples of machine reading comprehension

本文认为,句子整体语义理解是解决阅读理解问答的关键。针对上述问题,本文提出一种基于框架语义和图结构的阅读理解答案抽取方法。该方法首先利用汉语框架网筛选出与问句语义相关的候选句;其次提取问题和候选句中的实体,以实体在句子中的依存句法和多种语义结构关系构建实体关联图;最后将实体关联图融入图注意力网络进行逻辑推理,从而实现阅读理解答案抽取。该方法既解决了无法建模长距离语义依赖的问题,也可以通过关系丰富的实体关联图捕获文章的全局信息,综合分析推理篇章整体语义得到答案。

2 相关研究

2.1 机器阅读理解

机器阅读理解是让机器学会阅读和理解文章,根据给定的问题,从相关文章中找到答案。机器阅读理解任务涉及知识推理、语言理解以及摘要生成等多种复杂的技术,因此机器阅读理解是一项非常具有挑战性的任务。借助深度学习技术,机器阅读理解也取得了突破式的发展。在机器阅读理解任务上,Hermann等^[1]提出的Attentive Reader和Impatient Reader首次将深度学习技术应用于机器阅读理解。人们发现深度学习技术在上下文信息获取方面具有较强的优越性,这之后,深度学习技术被广泛地应用于机器阅读理解领域。Kadlec等^[2]提出的ASReader将注意力机制应用于机器阅读理解,使用注意力机制从上下文中选择概率最高的词作为最终答案。Dhingra等^[3]提出的GARReader采用新的门控注意力机制,使人们开始针对不同的任务设计新的注意力机制来

解决问题。Seo等^[4]提出的双向注意力流(Bi-Directional Attention Flow, BiDAF)网络在句子的不同粒度上对上下文和查询进行编码得到字符向量和词向量,通过双向的注意力机制提取问题和文章的信息。Wang等^[5]提出门控自匹配网络R-NET,将问题感知的段落表示与问题本身进行自匹配,进而从整篇文章中动态收集段落中的单词证据,并将其与当前段落单词的相关证据以及匹配的问题信息编码到段落表示中。之后Yu等^[6]提出的OANet使用卷积神经网络对上下文的局部结构进行捕捉,使用自注意力对单词进行交互,极大地提高了模型的训练速度和推理速度。

此外,机器阅读理解中大规模数据集的发布也对其发展有很大的推动作用。2016年,Rajpurkar等^[7]提出SQuAD1.1数据集,该数据集是第一个真正意义上的大规模抽取式机器阅读理解数据集,在抽取式阅读理解的发展中具有里程碑式的意义。Dunn等^[8]采用远程监督方法构造了SearchQA数据集,通过搜索引擎获取干扰和真实的上下文,更加贴近人类生活,有较高的实用性。之后,国内学者也开始陆续提出中文领域的MRC数据集。2021年,Tang等^[9]提出了一个全新的中文机器阅读理解数据集DuReader-robust。该数据集中的全部样例均为真实应用场景下的自然文本,旨在从过敏感、过稳定与泛化能力3个方面评估现有MRC模型在实际应用场景中的鲁棒性和泛化性。

2.2 图神经网络

近年来,一些学者将图结构数据应用到深度学习任务中,在这种大背景下,图注意力网络(Graph Attention Networks, GAT)^[10]应运而生。2018年,Peng等^[11]根据词共现关系构建图结构,并将其应用在文本分类的任务中,且取得了较好的结果,对后面的研究有很大启发。随后,Song等^[12]以句间实体共现关系、句内不同实体间的关系,以及指代关系进行构图,结合图循环网络(Graph Recurrent Network, GRN)应用在机器阅读理解中。在Song等工作的基础上,Cao等^[13]引入了更多边的类型,采用ELMo上下文嵌入向量特征。Fang等^[14]提出层次化图神经网络,根据实体、句子、段落3种不同力度创建了实体、句子、段落、问题4种类型的节点,并根据节点间的7种关系构建了异构图,随后使用图注意力网络实现节点间的信息传播和整合,在HotpotQA数据集上的实验表明了这种方法的有效性。Qiu等^[15]提出的DFGN模型设计了一个动态融合层,通过图注意力网络在实体图上进行节点之间的消息传递,但是在图结构构建上节点关系较为单一,没有考虑实体之间的依存句法以及同近义词的关系,并且没有体现出每种关系对问题解答的重要程度。受以上研究的启发,本文将图神经网络应用于机器阅读理解中。

已有方法存在以下缺陷:多段落的长文本构建的图结构包含的信息范围过于宽泛,构建的模型冗余信息较多,以往的方法对句子整体语义理解不够深入。而在机器阅读理解任务中,需要从多个段落中问题相关句的整体语义中找到与问题关系更密切的答案。本文利用汉语框架网^[16](Chinese FrameNet, CFN)对问题相关句进行筛选。CFN可以挖掘文本中隐含的语义信息,避免答案在这一阶段被遗漏。在构图过程中,丰富的实体关系图可以有效反映出文本中实体之间

的关联度,获得问题相关句的整体语义。最后,将图结构融入图注意力网络中进行逻辑推理,经过多次迭代后,得到各个句子和节点的语义表示,经由预测模块处理后得到最终的答案。

3 基于框架语义和图结构的阅读理解答案抽取方法

本文提出的模型模仿人类做机器阅读理解问答任务时的推理过程,模型框架如图2所示。整个模型包括5个模块:句子选择模块、实体关系图构建模块、文本编码模块、逻辑推理模块和答案预测模块。各个模块实现的功能为:

(1)句子选择模块:根据所提问题对多段落的篇章句子进行

框架语义匹配和句子相似性匹配,筛选出与问句相关的句子。

(2)实体关系图构建模块:通过实体之间的位置信息、关键实体关联关系、依存句法关系以及语义相似度综合获取实体之间的关系,并对其进行建模形成图表示。

(3)文本编码模块:对句子选择模块得到的句子集和输入的问题进行语义编码表示。

(4)逻辑推理模块:模仿人类的推理行为,结合图注意力网络和实体关系图,在关系图图中节点之间完成信息传播并进行动态融合。

(5)答案预测模块:输出所给问题的答案。

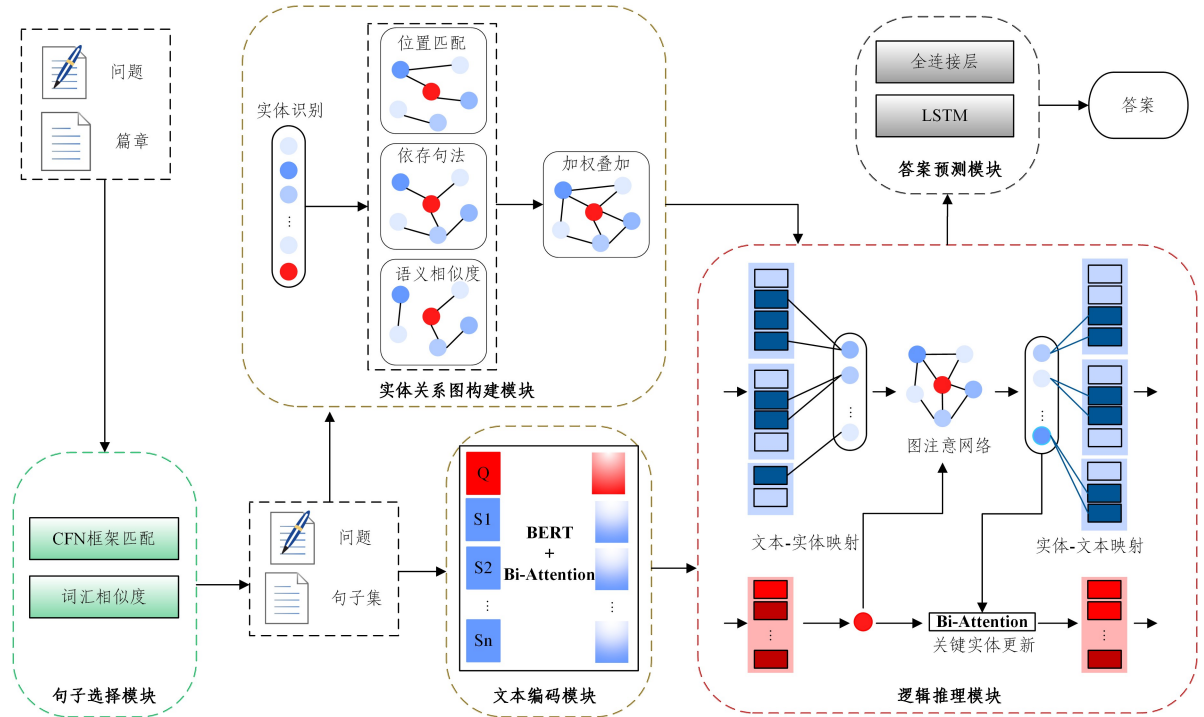


图2 模型框架示意图

Fig. 2 Schematic diagram of model frame

3.1 句子选择模块

在机器阅读理解中,文章中并非每个句子都与问题相关,因此在句子选择模块中,通过结合语义框架和词汇相似度匹配从语义上筛选出与问题相关的句子集。其中,基于词汇相似度的方法在浅层语义层面筛选与问题相关的句子,没有考虑句子语义深层特征;基于框架匹配的方法通过分析句子中的目标词所属语义场景及相关论元的关系,把与问题有深层语义关联的句子筛选出来,弥补了基于词汇相似度的方法的不足。该方法不仅有效避免了答案在这一阶段被遗漏,也可以解决问题不相关文本对下一步推理的干扰,降低了模型解答问题的难度。

3.1.1 基于词汇相似度的方法

对输入的问题 Q 和篇章中的句子 S 去停用词得到 $Q = \{k_1, k_2, \dots, k_m\}$, $S = \{\omega_1, \omega_2, \dots, \omega_n\}$ 。利用 Word2Vec^[17]对 Q 中的关键词 k_i 与 S 中的关键词 ω_j 进行词汇相似度计算,如式(1)所示,得到匹配值大于指定阈值的词汇匹配数 $sumword$ 。

$$\cos(\vec{k}_i, \vec{\omega}_j) = \frac{\vec{k}_i \cdot \vec{\omega}_j}{|\vec{k}_i| \cdot |\vec{\omega}_j|} \quad (1)$$

其中, \vec{k}_i 与 $\vec{\omega}_j$ 分别为 k_i 与 ω_j 的词向量。

随后对 $sumword_i$ 进行归一化处理,得到基于词汇相似度匹配的句子得分,如式(2)所示:

$$score_{sim} = \frac{sumword_i}{\{\max_{1 \leq i \leq n} sumword_i - \min_{1 \leq i \leq n} sumword_i\}} \quad (2)$$

3.1.2 基于框架匹配的方法

自2004年起,山西大学针对汉语自身的特点,以 Fillmore的框架语义学^[18]为理论基础,参照加州大学伯克利分校的 FrameNet^[19]创建了汉语框架语义网。CFN依据汉语真实语料,构建了供计算机使用的汉语词汇语义知识库。“框架”(Frame)可以被理解为人们在理解某一语言时,大脑中已有的认知结构,是想象、实践、信仰等概念的图解表征。一个框架包含一类词语,这一类词语表示一个认知结构,框架元素是组成框架的语义单元,目标词是在一个具体句子中能够激起某一框架的词。CFN中还建立了框架之间的复杂联系。在机器阅读理解任务中,应用CFN可以更有效地满足其深层语义推理的需求。

首先根据问题 Q 中的词语扩展目标词;然后通过CFN将目标词激起的框架与句子 S 中的目标词激起的框架进行

语义匹配,如果相同或者两个框架之间的距离小于等于 2,则匹配数 $sumframe_i$ 加 1。同样地,对 $sumframe_i$ 进行归一化处理,得到基于框架匹配的句子得分,如式(3)所示:

$$score_{frame} = \frac{sumframe_i}{\left\{ \max_{1 \leq i \leq n} sumframe_i, \min_{1 \leq i \leq n} sumframe_i \right\}} \quad (3)$$

基于框架匹配候选句抽取示例如图 3 所示,问题 Q 中“远征军”激起的框架与候选句 S 目标词“军队”激起的框架一致,且与“军事成员”激起的框架之间存在继承关系;目标词“死”激起的框架与“伤亡”激起的框架一致。问题 Q 与候选句 S 涉及的语义场景非常相似,因此将该句筛选出来作为候选句。

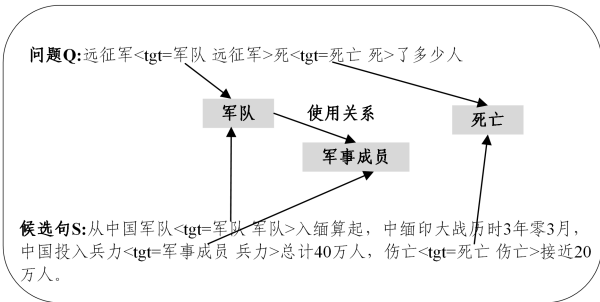


图 3 基于框架匹配候选句抽取示例

Fig. 3 Example of candidate sentence extraction based on frame matching

之后,根据词汇相似度和框架匹配的贡献度分配权重 ω_1 和 ω_2 ,综合两种方法,得到句子相关性得分,如式(4)所示:

$$S = \omega_1 * score_{sim} + \omega_2 * score_{frame} \quad (4)$$

其中, $0 \leq \omega_1, \omega_2 \leq 1$, 且 $\omega_1 + \omega_2 = 1$ 。将句子相关性得分 S_i 大于指定阈值 μ 的句子作为候选句子集,输入实体关系图构建模块和文本编码模块。

3.2 实体关系图构建模块

本文采用斯坦福大学的 Stanford CoreNLP 工具包^[20]对句子中名词、名词短语以及它们之间的句法依存关系进行识别,以下将名词、名词短语统称为实体。本文采用无向加权图,以实体作为节点,节点之间存在关系则连接一条边。本文构造实体关系图时采用了以下规则。

(1)句子位置关系边:根据中文分句的习惯,位于同一个句子中的实体之间会存在一定的关系,因而连接边;另外,不同句子中相同的实体之间连接边。

(2)关键实体关系边:问题中的实体对问题的解答至关重要,因此本文通过查询的方式,找出问题中出现的实体,并统计每个实体在句子集中出现的次数,将出现次数最多的实体作为关键实体,将关键实体与其余实体之间连接边。

(3)依存句法边:若两个实体间存在依存句法关系,则连接边。

(4)语义关系边:通过 HowNet^[21]计算两个实体间的语义相似度,若两个实体之间语义相似度大于指定阈值 δ ,则连接边。

另外,在构图中将上述 5 种规则连接的边根据重要度分配权值,并可相互叠加。例如,两个实体之间共同存在多种关系,则叠加权值,表示两个实体间关联程度较高。

3.3 文本编码模块

本文通过预训练模型 BERT^[22]和双向注意力机制^[4]来实现文本编码,使用双向注意力机制进一步加强问题与句子集之间的交互作用。采用 BERT 的输入格式,具体操作如式(5)~式(7)所示:

$$input = [CLS] + Q + [SEP] + S + [SEP] \quad (5)$$

$$(\bar{Q}, \bar{S}) = Bert(input) \quad (6)$$

$$(Q_0, S_0) = BiAttention(\bar{Q}, \bar{S}) \quad (7)$$

$\bar{Q} \in \mathbb{R}^{L \times d_1}$, $\bar{S} \in \mathbb{R}^{M \times d_1}$ 为通过预训练模型 BERT 后得到的问题与句子集的表示,其中 L 和 M 分别为问题与句子集的长度, d_1 为 BERT 隐藏层大小。 $Q_0 \in \mathbb{R}^{L \times d_2}$, $S_0 \in \mathbb{R}^{M \times d_2}$ 为通过双向注意力机制后得到的问题与句子集的表示,其中 d_2 为输出嵌入大小。

3.4 逻辑推理模块

在本模块中,将已编码的问题 Q_0 、句子集 S_0 和已构建的实体关系图引入逻辑推理模块中,通过几次更新迭代,找出潜在答案的文本范围。

由于实体关系图中的实体是从文本中的某一个句子中提取得到的,任一实体都对应着与自身相关的文本范围,因此将文本与实体相对应是本模块的第一步。我们构造一个 $u \times v$ 的二进制矩阵 M ,用于选择与实体相关的文本范围。 u 表示文本中的 token 数量, v 表示文本的实体数量,当文本中的第 i 个 token 在第 j 个实体包含的文本范围内,则 $M_{i,j}$ 为 1,否则为 0。接着对每一个实体所包含的 token 向量进行最大池化和平均池化,将两者得到的特征向量进行拼接,得到每个实体的嵌入表示如式(8)所示:

$$E^{t-1} = [e_1^{t-1}, e_2^{t-1}, \dots, e_v^{t-1}] \quad (8)$$

本文从问题出发,根据问题中的关键词进行推理,使用问题嵌入向量与实体嵌入向量之间的注意力网络得到每个实体与问题之间的相关度,如式(10)所示。然后将相关度与对应的实体嵌入相乘,进而鼓励与问题相关度高的实体,并惩罚其余实体,从而更新实体表示,如式(12)所示:

$$\tilde{Q}^{t-1} = MeanPooling(Q^{t-1}) \quad (9)$$

$$Att_i^t = \frac{\tilde{Q}^{t-1} V e_i^{t-1}}{\sqrt{d_2}} \quad (10)$$

$$W^t = Sigmoid(Att_i^t) \quad (11)$$

$$\tilde{E}^{t-1} = [w_1^t e_1^{t-1}, w_2^t e_2^{t-1}, \dots, w_v^t e_v^{t-1}] \quad (12)$$

其中, V 是线性变换矩阵, Q^{t-1} 是通过文本编码模块得到的问题表示,通过平均池化得到新的问题向量 \tilde{Q}^{t-1} 。

接着,为了将上一阶段得到的 \tilde{E}^{t-1} 在实体关系图上传播信息,使节点获取其邻居节点信息,根据相关度来分配不同的权重。本文采用图注意力网络计算出两个实体之间的注意力分数,如式(14)所示。最后,通过对每个节点的列求和,使其包含从邻居节点接收到的总信息,进而更新每个节点的表示,如式(15)、式(16)所示:

$$h_i^t = A_i e_i^{t-1} + b_i \quad (13)$$

$$Att_{i,j}^t = \frac{\exp(LeakyReLU(W_i^t [h_i^t, h_j^t]))}{\sum_k \exp(LeakyReLU(W_i^t [h_i^t, h_k^t]))} \quad (14)$$

$$e_i = \text{ReLU}(\sum_j \text{Attn}_{ij} h_j') \quad (15)$$

$$E' = [e_1', e_2', \dots, e_v'] \quad (16)$$

其中,假设 i 为任意一实体, j 为实体 i 的相邻实体, $A_i \in \mathbb{R}^{d_2} \times 2d_2$, $W_i^T \in \mathbb{R}^{2d_2}$ 为可训练的线性变换参数。

通过上述过程完成了第一次推理,此后需要更新问题中的关键实体嵌入表示作为开始实体表示进行下一步的推理。本文利用双向注意力机制来实现,计算式如式(17)所示。同时,返回每个节点中的实体对应的文本范围也尤为重要,本文同样使用上文提到的二进制矩阵 \mathbf{M} 将文本 token 返回,再通过长短期记忆网络(Long Short-Term Memory, LSTM)^[23] 进行融合后得到下一轮推理的句子集表示 S' ,如式(18)所示:

$$Q' = \text{Bi-Attention}(Q'^{-1}, E') \quad (17)$$

$$S' = \text{LSTM}(S'^{-1}, \mathbf{M}E'^T) \quad (18)$$

3.5 答案预测模块

经过几轮的动态推理后,问题与文章已基本交互融合,然后通过长短期记忆网络 and 全连接层预测答案的开始和结束位置 $P_{\text{start}}, P_{\text{end}}$,最后使用交叉熵损失函数对模型进行优化, $T_{\text{start}}, T_{\text{end}}$ 为参考答案的开始和结束位置,计算式如式(19)一式(21)所示:

$$L_{\text{start}} = -\log P_{\text{start}} T_{\text{start}} \quad (19)$$

$$L_{\text{end}} = -\log P_{\text{end}} T_{\text{end}} \quad (20)$$

$$\text{Loss} = L_{\text{start}} + L_{\text{end}} \quad (21)$$

4 实验结果及分析

4.1 数据集与实验参数

本文的实验数据集为 DuReader-robust,训练集包括约 13000 个样本,验证集包括 1300 个样本。DuReader-robust 数据集数据来源均为真实场景下的用户问题,领域涉及医疗、交通、文化等,该数据集问题更加复杂,对模型的推理能力要求更高,是首个关注阅读理解鲁棒性的中文数据集。

在实验训练过程中使用自适应矩估计(Adam)优化方法动态调整每个参数的学习率,初始学习率为 2×10^{-5} ,批处理大小为 16,共训练 35 轮。预训练模型 BERT 采用 bert-base-chinese 模型,在句子选择阶段, ω_1, ω_2 设置为 0.5, μ 设置为 0.1。

4.2 实验结果

4.2.1 对比实验

为了验证本文方法的有效性,本文使用预训练模型以及机器阅读理解任务中的代表性模型进行对比实验。在预训练模型对比实验中,将文本和问题进行数据预处理,接着通过预训练模型得到文本表示,最后通过预测层进行答案 span 预测。对比的基线模型如下。

Albert 模型^[24]:该模型引入词嵌入的因式分解和交叉层的参数共享这两种参数精简技术,解决了内存受限和训练速度受限的问题。本文采用 Albert-base 模型。

Ernie 模型^[25]:该模型通过更改 BERT 模型结构,采用短语和实体级别的 mask,将知识和语言语义信息融合,增强了语义的表示。本文采用 Ernie 2.0 base 模型。

BERT 模型^[22]:该模型是一种 Transformer 的双向编码器,模型的输入序列为句子对所对应的 token 嵌入、位置嵌入

以及分割嵌入,采用 Mask Language Model 和 Next Sentence Prediction 两种训练子任务,针对阅读理解任务进行微调。本文采用 bert-base-chinese 模型。

BERT_wwm 模型^[26]:该模型是由哈工大讯飞联合实验室提出的针对中文的全码预测(Whole Word Masking, WWM)。它更改了 BERT 模型中预训练阶段的训练样本生成策略,用 mask 标签替换一个完整的词而不是字。

RoBERTa 模型^[27]:该模型在 BERT 模型的基础上采用动态掩码的方式训练模型,取消了下一句预测部分,改进了优化函数,使用更大批次及更多数据来提升模型的性能。本文对比实验用到其 base 版本。

XLNet 模型^[28]:该模型结合自回归和自编码语言模型,在 Transformer-XL 基础上提出广义自回归方法,解决了 BERT 预训练和微调之间不匹配的问题,并且能够更好地处理长文本。本文采用 XLNet-base 模型。

DFGN 模型^[15]:该模型将篇章和问题输入文本编码层,结合构建的实体关系图在动态融合块中进行推理更新,经由预测层返回答案 span。

本文实验采用精准匹配(Exact Match, EM)和 F1 值作为定量评价指标。EM 值计算模型预测的答案和真实答案完全一致的百分比,是一个相对严格的评估指标。F1 值指模型的输出答案与真实答案的覆盖率,通过准确率和召回率计算得到的调和平均值。模型对比的结果如表 1 所列,可以看出本文方法的 EM 值和 F1 值均超越了所有对比的基线模型。本文利用汉语框架网对问题相关句进行筛选,一些与问题无关的候选句在这一阶段被剔除,降低了模型的推理难度。此外,在构图过程中,本文将更加丰富的实体以及语义关系引入图结构,这些语义关系可以有效反映问题与候选答案之间的关联,因此本文的方法相比预训练模型和代表性的 MRC 方法取得了更好的答题效果。

表 1 模型对比实验结果

Table 1 Results of model comparison experiment

(单位: %)

Model	EM	F1
Albert	68.45	80.82
Ernie	69.02	80.84
BERT	68.87	80.87
BERT_wwm	69.01	81.07
RoBERTa	70.64	81.12
XLNet	70.68	81.21
DFGN	70.59	82.92
Our model	71.07	83.80

4.2.2 消融实验

为验证模型所采用的各种策略对模型性能的影响,本文进行了消融实验,实验结果如表 2 所列。其中, -CFN&Word2Vec_{selected sentences} 表示模型在句子选择模块中没有采用基于 CFN 和词汇相似度的方法进行问题相关句筛选; -Syntactic relation_{Edge} 表示模型在图中没有加入依存句法关系边; -Semantic similarity_{Edge} 表示模型在图中没有加入语义关系边; -Keyword query_{Edge} 表示模型在图中没有加入关键实体关系边; -Edge weighted 表示模型在构建图结构时每种边都没有赋予权值且不可叠加。

表2 模型消融实验结果

Table 2 Results of model ablation experiment

Model	EM	F1
Our model	71.07	83.80
-CFN&Word2Vec _{selected sentences}	70.59	82.91
-Syntactic relation _{Edge}	69.87	82.56
-Semantic similarity _{Edge}	70.27	83.32
-Keyword query _{Edge}	70.19	82.80
-Edge weighted	69.79	82.57

从表2中可以看出,基于框架匹配以及句子相似性匹配的方法能够有效捕获与问题有关的句子,减少与问题不相关的文本对下一步推理的干扰,同时也缩小了下一步推理的答案范围,降低了模型解答问题的难度。当模型中加入实体之间的依存句法关系和语义相似关系后,模型的语义理解能力获得了显著的提升。另外,通过关键实体关系可以有效建立问句与候选句之间的语义关联,问题中的关键实体对问题中的解答非常重要。最后,相较于无权重的实体图,构建实体加权图更能反映出实体之间的关联程度,在推理过程中更能够关注到与相关候选句之间的逻辑关系。综上所述,本文提出的实体关系图中各种类型的边均有效提升了模型的阅读理解能力,对问题的解答有很大的帮助。

4.2.3 迭代层数对模型性能的影响

为了得到模型在 DuReader-rubust 数据集上的最优性能,本文设置了不同的推理层数进行训练。通过实验可以得到如图4所示的结果。当 Layer 等于1时,模型性能没有达到最优,说明只进行一次迭代不足以将需要深层逻辑推理得到的答案成功召回;当 Layer 大于2时,模型的性能逐步下降,表明存在过度推理的现象。因此,本文模型通过进行两次迭代(Layer=2)后,模型的性能达到最佳,EM值与F1值最高。

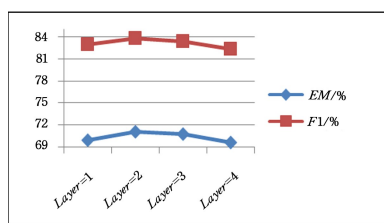


图4 推理层数对模型性能的影响

Fig. 4 Influence of reasoning layers on model performance

结束语 为解决机器阅读理解任务中出现的模型推理能力、深层语义理解能力不足的问题,本文提出一种基于框架语义和图结构的阅读理解答案抽取方法。首先利用汉语框架匹配与问句语义相关的候选句;然后提取问题和候选句中的实体,以实体在句子中的依存句法和语义关系构建实体关联图;最后将其融入图注意网络进行逻辑推理,从而实现阅读理解答案抽取。在引入图结构进行推理时,增加了系统的计算量,未来的工作将针对不同的阅读理解问题引入不同的图结构,保留在逻辑推理时起主导作用的语义关系,去除一些非关键语义关联,提升系统的答案抽取效率。此外,本文还要探索更深语义层次的阅读理解任务(如高考阅读理解),增强模型的泛化能力。

参考文献

- [1] HERMANN K M, TOMÁVS K, GREFFENSTETTE E, et al. Teaching Machines to Read and Comprehend[C]//Proceedings of the 29th International Conference on Neural Information Processing Systems. New York: Curran Associate Inc, 2015: 1693-1701.
- [2] KADLEC R, SCHMID M, BA JGAR O, et al. Text Understanding with the Attention Sum Reader Network[C]//Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, USA: Association for Computational Linguistics, 2016: 908-918.
- [3] DHINGRA B, LIU H, YANG Z, et al. Gated-attention readers for text comprehension[J]. arXiv:1606.01549, 2016.
- [4] SEO M, KEMBHAVI A, FARHADI A, et al. Bidirectional Attention Flow for Machine Comprehension [J]. arXiv: 1611.01603, 2016.
- [5] WANG W, YANG N, WEI F, et al. Gated self-matching networks for reading comprehension and question answering[C]//Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2017: 189-198.
- [6] YU A W, DOHAND, LUONG M T, et al. QANet: Combining Local Convolution with Global Self-Attention for Reading Comprehension[C]//Proceedings of International Conference on Learning Representations (ICLR2018). UAS: ICLR Organizing Committee, 2018: 1-8.
- [7] RAJPURKAR P, ZHANG J, LOPYREV K, et al. SQuAD: 100 000+ Questions for Machine Comprehension of Text[C]//Proceedings of the Conference on Empirical Methods in Natural Language Processing. USA: Association for Computational Linguistics, 2016: 2383-2392.
- [8] DUNN M, SAGUN L, HIGGINS M, et al. SearchQA: A New Q&A Dataset Augmented with Context from a Search Engine [J]. arXiv:1704.05179, 2017.
- [9] TANG H, LI H, LIU J, et al. DuReader_robust: A Chinese Dataset Towards Evaluating Robustness and Generalization of Machine Reading Comprehension in Real-World Applications [J]. arXiv:2004.11142, 2020.
- [10] VELIKOVI P, CUCURULL G, CASANOVA A, et al. Graph Attention Networks[C]//Proceedings of International Conference on Learning Representations (ICLR 2018). USA: ICLR Organizing Committee, 2018: 1-11.
- [11] PENG H, LI J, HE Y, et al. Large-scale hierarchical text classification with recursively regularized deep graph-cnn[C]//Proceedings of the 2018 World Wide Web Conference. 2018: 1063-1072.
- [12] SONG L, WANG Z, YU M, et al. Exploring graph-structured passage representation for multi-hopreading comprehension with graph neural networks [J]. arXiv:1809.02040, 2018.
- [13] DE CAO N, AZIZW, TITOV I. Question answering by reasoning across documents with graph convolutional networks[C]//Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Hu-

- man Language Technologies, Volume 1, 2019:2306-2317.
- [14] FANG Y, SUN S, GAN Z, et al. Hierarchical Graph Network for Multi-hop Question Answering [J]. arXiv: 1911. 03631, 2019.
- [15] XIAO Y, QU Y, QIU L, et al. Dynamically Fused Graph Network for Multi-hop Reasoning[C]//Proceedings of the 57st Annual Meeting of the Association for Computational Linguistics, 2019:6140-6150.
- [16] LI R. Research on Frame Semantic Structure Analysis Technology for Chinese Sentences [D]. Taiyuan: Shanxi University, 2012.
- [17] LIU M. Sentence similarity calculation based on word vector and its application in case-based machinetranslation [D]. Beijing: Beijing Institute of Technology, 2015.
- [18] FILLMORE C J. Frame semantics and the nature of language [J]. Annals of the New York Academy of Sciences; Conference on the Origin and Development of Language and Speech, 1976, 280(1):20-32.
- [19] BAKER C F, FILLMORE C J, LOWE J B. The Berkeley FrameNet Project[C]// Proceedings of the 17th International Conference on Computational Linguistics-Volume 1. Association for Computational Linguistics, 2002:86-90.
- [20] MANNING C D, SURDEANU M, BAUER J, et al. The Stanford CoreNLP Natural LanguageProcessing Toolkit[C]// Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics; System Demonstrations. 2014:55-60.
- [21] LIU Q, LI S J. Vocabulary semantic similarity calculation based on HowNet[J]. Chinese Computational Linguistics, 2002, 7(2): 59-76.
- [22] DEVLIN J, CHANG M W, LEE K, et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding [C]// Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics; Human Language Technologies, USA; Association for Computational Linguistics, 2019:4171-4186
- [23] HOCHREITER S, SCHMIDHUBER J. Long Short-Term Memory[J]. Neural Computation, 1997, 9(8):1735-1780.
- [24] LAN Z, CHEN M, GOODMAN S, et al. ALBERT: a lite BERT for self-supervised learning of language representations [EB/OL]. (2020-02-09) [2021-06-07]. <https://arxiv.org/abs/1909.11942v6>.
- [25] SUN Y, WANG S H, LI Y K, et al. ERNIE: enhanced representation through knowledge integration[C]// Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, USA; Association for Computational Linguistics, 2019: 1441-1451.
- [26] CUI Y M, CHE W X, LIU T, et al. Pre-training with whole word masking for Chinese BERT [EB/OL]. (2019-10-29) [2021-06-07]. <https://arxiv.org/abs/1906.08101v2>.
- [27] LIU Y, OTT M, GOYAL N, et al. RoBERTa: a robustly optimized BERT pretraining approach [EB/OL]. (2019-07-26) [2021-06-07]. <https://arxiv.org/abs/1907.11692v1>.
- [28] YANG Z L, DAI Z H, YANG Y M, et al. XLNet: generalized autoregressive pretraining for language understanding [C] // Proceedings of 33rd Conference on Neural Information Processing Systems (NeurIPS 2019). Canada: Neural Information Processing Systems Foundation, Inc, 2019:5754-5764.



YANG Zhizhuo, born in 1983, Ph.D, associate professor, is a member of China Computer Federation. His main research interests include natural language processing and reading comprehension Q&A.

(责任编辑:何杨)