

## 多跳式机器阅读理解研究进展综述

仇亚进<sup>1,2</sup>, 奚雪峰<sup>1,2,3\*</sup>, 崔志明<sup>1,2</sup>, 盛胜利<sup>4</sup>, 周悦尧<sup>1,2</sup>

(1. 苏州科技大学 电子与信息工程学院, 江苏 苏州 215000;

2. 苏州市虚拟现实智能交互及应用重点实验室, 江苏 苏州 215000;

3. 苏州智慧城市研究院, 江苏 苏州 215000;

4. 德州理工大学, 德克萨斯州 卢伯克市 79401)

**摘要:** 机器阅读理解的目标是使机器更好地理解自然语言文本并在此基础上回答提出的问题, 是自然语言处理领域热门的研究方向之一。早期, 由于受到了数据集的约束, 对机器阅读理解的认识大多仅限于一个单跳式的问答。随着最近几年多跳式机器阅读理解数据集的发展, 多跳式机器阅读理解得到广泛的研究, 极大地推动了机器阅读理解领域的发展。从以下几个方面对基于多跳式的机器阅读理解进行归纳总结: 介绍了机器阅读理解任务定义与发展历程; 阐述了多跳式机器阅读理解任务定义并梳理总结相关数据集; 详细整理了多跳式机器阅读理解基于注意力机制和图神经网络以及问题分解相关模型方法的研究进展; 最后, 对多跳式机器阅读理解未来研究重点和所面临的研究挑战进行展望。

**关键词:** 多跳式机器阅读理解; 注意力机制; 图神经网络; 问题分解; 数据集

中图分类号: TP391

文献标识码: A

文章编号: 1673-629X(2023)02-0009-08

doi: 10.3969/j.issn.1673-629X.2023.02.002

## Review on Research Progress of Multi-hop Machine Reading Comprehension

QIU Ya-jin<sup>1,2</sup>, XI Xue-feng<sup>1,2,3\*</sup>, CUI Zhi-ming<sup>1,2</sup>, SHENG Sheng-li<sup>4</sup>, ZHOU Yue-yao<sup>1,2</sup>

(1. School of Electronic and Information Engineering, Suzhou University of Science and Technology, Suzhou 215000, China;

2. Suzhou Key Laboratory of Virtual Reality Intelligent Interaction and Application Technology, Suzhou 215000, China;

3. Suzhou Smart City Research Institute, Suzhou 215000, China;

4. Texas Institute of Technology, Lubbock 79401, USA)

**Abstract:** The goal of machine reading comprehension is to enable machines to better understand natural language texts and answer questions on this basis, which is one of the hot research directions in the field of natural language processing. In the early days, due to the constraints of data sets, the understanding of machine reading comprehension was mostly limited to a single-hop question and answer. With the development of multi-hop machine reading comprehension data sets in recent years, multi-hop machine reading comprehension has been widely studied, which has greatly promoted the development of the field of machine reading comprehension. We summarize the multi-hop machine reading comprehension from the following aspects: Introduce the definition and development of machine reading comprehension task, expound the definition of multi-hop machine reading comprehension task, and sort out the related data sets. The research progress of multi-hop machine reading comprehension based on attention mechanism, graph neural network and problem decomposition related models is sorted out in detail. Finally, the future research emphases and challenges of multi-hop machine reading comprehension are prospected.

**Key words:** multi-hop machine reading comprehension; attention mechanism; graph neural network; question decomposition; dataset

收稿日期: 2022-04-11

修回日期: 2022-08-17

基金项目: 国家自然科学基金(61876217, 62176175); 江苏省“六大人才高峰”高层次人才项目(XYDXX-086)

作者简介: 仇亚进(1995-), 男, 硕士, CCF 会员(H9807G), 研究方向为机器阅读理解; 通信作者: 奚雪峰(1978-), 男, 博士, 副教授, 硕导, 研究方向为自然语言处理、机器学习; 崔志明(1961-), 男, 博士, 教授, 博导, 研究方向为知识挖掘、人工智能; 盛胜利(1969-), 男, 博士, 教授, 博导, 研究方向为机器学习、知识挖掘。

## 0 引言

机器阅读理解 (Machine Reading Comprehension, MRC) 任务主要是指通过利用现代人工智能等信息技术,使机器在阅读并掌握理解人类自然语言的基础上,从非结构化文档中寻找合适的答案,回答来自文本内容的相关问题,以此度量机器对自然文本的理解能力。近年来,MRC 因其开放性和交互性在学术界和工业界受到越来越多的重视和研究,已经发展成为 NLP 领域重要的研究热点方向。

机器阅读理解有着悠久的历史,图灵早在 20 世纪 50 年代,就曾提出通过人机交互衡量机器的智能水平<sup>[1]</sup>。然后 Terry Winograd 提出构想,认为实现阅读理解的三大要素是:语法、语义和推理<sup>[2]</sup>。Hirschman 等人于 1999 年开始探索 MRC 技术的研究,设计出第一个以小学年级故事为语料库的自动阅读理解基线系统 Deep Reed<sup>[3]</sup>。Riloff 等人<sup>[4]</sup>根据人工制定规则,对问题与文章中候选语句的匹配度打分,然后选择最高分的候选语句当作答案。基于传统特征的 MRC 大都采用模式匹配来提取特征,其鲁棒性差、耗时长、早期发展比较慢。直到 Hermann 等人<sup>[5]</sup>于 2015 年提出使用神经网络模型,该领域才逐渐得到发展。Liu 等人<sup>[6]</sup>于 2019 年发表一篇关于神经机器阅读理解方法的综述,重点归纳了基于深度学习的 MRC 方法。Zhang 等人<sup>[7]</sup>于 2019 年列出并比较了各种 MRC 数据集的具体特征,并描述一些典型的 MRC 模型的主要思想。Dzendorik 等人<sup>[8]</sup>于 2021 年发表一篇关于 MRC 英文数据集的综述,其详细介绍了数据集的数据收集以及创建过程,并对各个数据集从多维度进行分析对比。

在陈丹琦博士的学位论文中,根据答案形式将 MRC 任务分为完形填空式、多选题式、片段抽取式、自由生成式四种类型,这四类见证了 MRC 技术的发展<sup>[9]</sup>。近年来,随着数据集的不断丰富,问题从简单的单询问问题转向更复杂的综合性问题,文本中的答案分布也随之变化,由单段落单片段答案到多片段多答案,从而对 MRC 的研究也面向更为复杂的多跳式阅读理解领域探索。

面对多种多样阅读理解任务的出现,近年国内学者包玥等人<sup>[10]</sup>对抽取式 MRC 任务作了总结;随着深度学习的较快发展,李舟军等人<sup>[11]</sup>对基于深度学习的 MRC 任务作了总结,认为高质量的词向量表示仍是 MRC 任务的研究重点;张超然等人<sup>[12]</sup>对基于预训练模型的 MRC 任务作了总结,总结了使用预训练模型的阅读理解模型在相关数据集上的表现;徐霄玲等人<sup>[13]</sup>对 MRC 任务的研究技术作了总结,比较了采用不同技术的 MRC 模型在不同类型数据集上的表现。

但目前还没有学者对多跳式 MRC 任务进行系统地介绍,因此,该文对多跳式阅读理解的典型代表数据集进行介绍分析,对相关模型方法进行分类总结,希望能够对多跳式阅读理解任务的研究起到一定的推动作用。

## 1 多跳式机器阅读理解

多跳式机器阅读理解 (Multi-Hop Machine Reading Comprehension, MHMRC) 是指由机器阅读理解与多跳式推理回答交叉形成的新领域,即在传统 MRC 抽取答案的基础上引入多跳式问答,是 MRC 任务中重要的一类。多跳推理式阅读理解中的问题一般都是人工提出的复杂的综合问题,在文中不易直接找到答案。面对复杂问题,有可能需要从文本中多次推理才能对答案实现精准定位。具体来说,给定一个问题,系统只通过一个文档或段落是无法正确回答问题的,需要系统逐步结合多处信息来回答,所以需要多跳推理,如图 1 所示<sup>[14]</sup>。

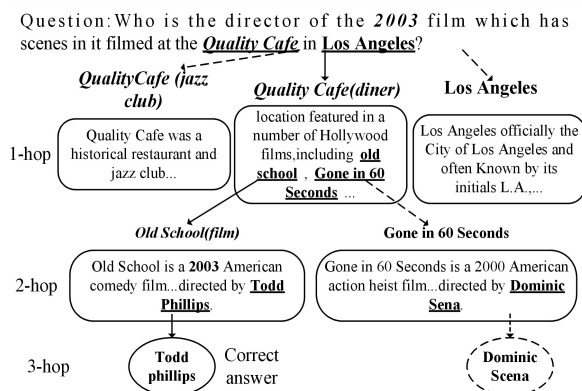


图 1 多跳阅读理解认知

对于上例问题,由于问题是复杂的组合问题,给出的信息都是一些间接的信息,因此没有办法直接抽取到答案。对于此类复杂问题,需要经历多次“跳转”才能找到答案,所以被称为 MHMRC 问题,可以看出,多跳问题比一般的单跳问题要复杂得多。

## 2 多跳式 MRC 相关数据集

数据集是支撑神经网络模型得以快速发展的基础,大规模 MRC 数据集的出现促进了神经机器阅读模型更好地发展。传统的单跳式阅读理解数据集中一直聚焦于“文章段落-问题-答案”三者之间的联系,从而忽略了真实情况中多跳问答具有的承前启后、逐步推理的性质。随着研究者不断深入探索多跳式 MRC,各种具有挑战性的数据集被提出,其数据集规模不断增加并且答案类型不断丰富,越来越接近真实场景。多跳式数据集中的问题,系统需要根据多篇文档来回答,即需要多跳推理来锁定答案。目前,具有代表性的多跳式 MRC 数据集主要包括 QAngaroo<sup>[15]</sup>、

HotpotQA<sup>[16]</sup>等。QAngaroo 与 HotpotQA 的发布,不仅推动了该类问题的研究,也推动了 MRC 的发展。

## 2.1 QAngaroo

QAngaroo 是 2017 年伦敦大学学院推出的多文档推理阅读理解数据集。它由两个数据集组成: WikiHop 和 MedHop,它需要多个推理步骤,将来自多个文档的事实结合起来。

WikiHop 是开放域,来源于维基百科文章,是一个多跳问答数据集。WikiHop 的查询由来自 WikiData<sup>[17]</sup>的实体和关系构成,而支持文档来自 WikiReading<sup>[18]</sup>。与 WikiHop 格式相同,MedHop 数据集基于医疗库 PubMed 的研究论文摘要,查询是关于药物对之间的相互作用。必须通过结合来自药物和蛋白质的一系列反应的信息来推断出正确的答案。

根据不同类型的任务以及所使用的数据集,对模型的推理能力要求各不一样。QAngaroo 最大的特点是要从多篇文章中找相关文档,结合多个文档的多个推断步骤,联合文本多处进行综合推理,问题的答案不能从一个段落中单独得出。随着推理跳数的增加,统计距离逐渐变远,与问题的相关性逐渐变小,因此, QAngaroo 对于算法设计和分析多段落的能力提出了很高的要求。

## 2.2 HotpotQA

2018 年由卡耐基·梅隆大学、斯坦福大学、蒙特利尔大学和 Google 公司等联合推出多段落推理阅读理解数据集 HotpotQA<sup>[16]</sup>。HotpotQA 拥有 11 万个基于维基百科问答对段落,是一个以自然的、多跳的问题为特征的问答数据集。HotpotQA 不仅要模型需要读的文本范围从单段落扩展到多段落或多篇章段落,还要求模型可以对支持答案的证据之间的逻辑关系构建至少两步的推理链,这比单跳式 MRC 更有挑战性。

HotpotQA 包含比较问题和是/否问题,涵盖了围绕实体、位置、事件、日期和数值的各种问题。HotpotQA 作者从数据集中抽取了 100 个例子,发现 HotpotQA 涵盖了广泛的答案类型,这与笔者对问题类型的初步分析相匹配,其中大多数问题都是关于文章中的实体(68%)和包括各种属性(如日期(9%))以及其他描述性属性(如数字(8%)和形容词(4%))的不可忽略的问题。

与 QAngaroo 数据集相比,HotpotQA 数据集无论是在训练集还是在测试集的数据量上都有成倍的增加,且增加了带干扰设置的数据集,更为复杂,如表 1 所示。

另外,为了证明模型的确有利用原文中的相关证据进行推理并提升模型的可解释性,HotpotQA 不仅要求模型给出最终答案,还要给出推理所用到的支持性

事实。

表 1 数据集对比

数据集	Train	Dev	Test	来源
WikiHop	43.7K	5.1K	2.4K	维基文章
MedHop	1.6K	0.3K	0.5K	医疗论文库
HotpotQA	90.5K	7.4K	7.4K	带干扰文章 全维基文章

## 2.3 数据集评估标准

不同的 MRC 任务有不同的评估指标,可信的评估标准是衡量模型理解自然语言能力的关键参考指标。对于 MHMRC 任务,需要模型对比预测所得答案和真实答案。通常使用精确匹配(EM 值)和模糊匹配(F1 值)进行评价。

EM 值指的是模型的预测答案和给出的正确答案完全相同的比例。F1 值指的是模型的预测答案和正确答案的覆盖率,通过计算准确率(Precision)与召回率(Recall)得到,即使模型的预测答案和正确答案不完全相同也可得分,是一种模糊匹配计算。其计算公式如下:

(1) EM 值。

$$EM = \frac{N_{pa}}{N_{ta}} \quad (1)$$

其中,  $N_{pa}$  表示模型的预测答案和真实答案完全匹配的数量,  $N_{ta}$  表示真实答案的总数量。

(2) F1 值。

$$Precision = \frac{N_{ptw}}{N_{paw}} \quad (2)$$

$$Recall = \frac{N_{ptw}}{N_{taw}} \quad (3)$$

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} \quad (4)$$

其中,  $N_{ptw}$  表示预测正确的单词/字符数目,  $N_{paw}$  表示所预测答案的所有单词/字符数目,  $N_{taw}$  表示真实答案的所有单词/字符数目。

## 3 多跳式 MRC 相关模型方法

传统的基于机器学习的 MRC 模型采用模式匹配进行特征提取,性能提升有限,已不能满足多文档多答案以及复杂问题的需求。随着深度学习的研究发展,基于神经网络的 MRC 模型发展很快。循环神经网络(RNN)、卷积神经网络(CNN)、长短时记忆网络(LSTM)、门控循环单元(GRU)、注意力机制等深度学习方法和图神经网络的出现大大推动了 MRC 的发展。近年来有很多学者基于注意力机制来研究 MHMRC,与此同时,随着图神经网络技术的不断发展,也有学者就此展开对 MHMRC 的研究,考虑到有



些复杂的问题其实可以拆分成简单子问题,而问题的答案就在这些子问题中,也有学者提出将复杂问题分解成若干子问题来解决 MHMRC 问题。

因此,目前对于 MHMRC 任务的解决方案根据方法主要分为基于注意力机制方法、基于图神经网络方法以及基于问题分解方法三种。

### 3.1 基于注意力机制方法

由于受到人类视觉神经系统可以快速发现较远处视野中的重点关注区域的启发,计算机视觉领域早在 2000 年就提出注意力机制概念<sup>[19]</sup>。自然语言处理模型在读取文本时也可以效仿图像处理,重点关注和任务相关的文本内容。基于此, Bahdanau 等人<sup>[20]</sup>于 2014 年在机器翻译(NMT)任务上应用注意力机制,同时进行翻译和对齐工作,显著提升了译文质量,这是首次在 NLP 领域中使用注意力机制。Hermann 等人<sup>[5]</sup>将注意力机制应用于 MRC 模型中,大大的提升了模型准确率。自此,注意力机制很快被推广应用到各种自然语言处理任务中。

MRC 中最核心的任务就是上下文的语义推理,一般是将单层网络模型经过多次迭代计算语义从而包含更多相关语义信息,在每一层的更新词语计算中加入注意力机制,注意力机制应用在 MRC 模型中读取文本时重点关注与问题相关的文本部分。面对使用注意力机制的 MRC 模型无法充分利用上下文与问题之间的关联信息,只是单向的应用以及只应用于文本中部分内容, Seo 等人<sup>[21]</sup>提出双向注意力流网络结构 BiDAF,这是一个根据不同粒度分成多阶段来获得问

题感知的上下文表示结构。在此基础上, Liu 等人<sup>[22]</sup>提出一种新的基于注意力的机器理解任务神经网络模型—双重交互模型(DIM Reader),用于研究文本的理解能力,该模型构建了多跳的双重迭代交替注意力机制。Li 等人<sup>[23]</sup>提出一种基于双向注意流(BiDAF)模型和词向量结合的方法,该模型使用双向注意流机制来捕捉上下文和问题之间的注意力。Xu 等人<sup>[24]</sup>提出一个由粗到细的注意力网络,设计了一个多阶段分层模型 BiDMF,其上下文和问题由双向 LSTM RNN 编码;然后通过注意力机制的多次迭代,得到更准确的交互信息。

自注意力(Self-Attention)<sup>[25]</sup>是注意力机制的一种改进,即查询来自源文本序列自身,用于建模源文本序列内部元素间的依赖关系,以加强对源文本语义的理解。Wang 等人<sup>[26]</sup>提出了门控自匹配网络 R-Net,首次将自注意力机制应用于 MRC 模型。MHMRC 需要结合多篇文章的信息和推理来推断答案的能力。Zhuang 等人<sup>[27]</sup>引入动态自注意力网络 DynSAN 来完成多篇文章的阅读理解任务,它在标记级处理跨篇文章的信息,同时避免大量的计算开销。针对现实生活中很多问题的答案往往由多个片段组成,苏立新等人<sup>[28]</sup>提出面向多片段答案抽取的阅读理解模型 Bert-Boundary,该模型采用预训练 Bert 作为底层结构来理解问题和文本,再利用序列标注建模多个答案片段。面对具有挑战性的多跳生成任务(叙事任务), Bauer 等人<sup>[29]</sup>提出多跳指针生成器模型 MHPGM,如图 2 所示。

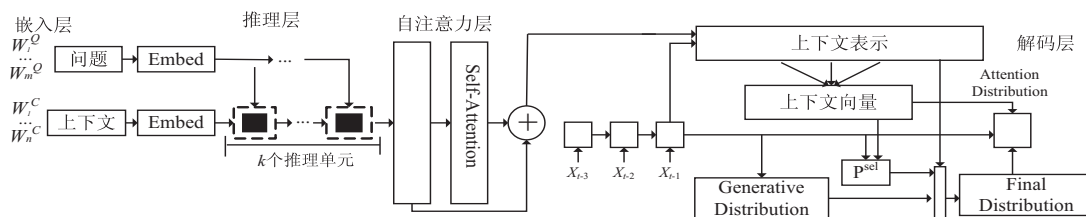


图 2 多跳指针生成器模型框架

MHPGM 编码器使用双向注意力结合自注意力的混合注意力机制来执行多跳推理,并使用指针生成解码器来有效地读取和推理长文章,合成与问题一致的答案。MHPGM 模型框架由 Embedding Layer、Reasoning Layer、Self-Attention Layer、Pointer-Generator Decoding Layer 四部分组成。各部分功能结构列举如下:

- 嵌入层:用一个学习过的  $d$  维嵌入空间嵌入上下文和问题中的每个单词,通过来自语言模型(ELMo)的预嵌入获得每个单词的上下文感知嵌入。
- 推理层:推理层由  $k$  个推理单元组成,使嵌入的上下文通过  $k$  个推理单元传递,使用 BiDAF 模拟多跳

推理过程中的单个推理步骤。

- 自注意力层:作为答案生成之前的最后一层,利用剩余静态自我注意机制来帮助模型处理具有长期依赖性的长上下文。
- 指针生成解码层:使用一个注意力指针生成解码器并从上下文中复制来创建答案。

随着注意力机制的快速发展, Song 等人<sup>[30]</sup>提出一种新的全局注意力推理(GAI)神经网络结构,它通过动态终止的多跳推理机制从结构知识中学习有用的线索来回答完形填空式的问题。Duan 等人<sup>[31]</sup>提出一种端到端的深度学习模型来回答多选题。该模型采用双 GRU 对文章和问题进行上下文编码,并通过六种注

意力函数对给定文章和问题之间的复杂交互进行建模,然后利用多层次的注意力转移推理机制,进一步获得更准确的综合语义。针对现实生活中的问题多样性,谭红叶等人<sup>[32]</sup>对此展开研究,提出一种基于 Bert 的多任务阅读理解模型,该模型先对问题分类预测,然后利用双向注意力机制来捕获问题和篇章的关系,从而回答多样性问题。Zhong 等人<sup>[33]</sup>提出一种新的问答模型—粗粒度细粒度知识网络(CFC),CFC 应用协同注意力和自我注意力来学习候选、文档和实体的查询感知节点表示。由于 RNN 的顺序性,基于注意力机制的循环神经网络阅读理解模型在训练和推理方面较

慢,Zheng 等人<sup>[34]</sup>提出一个新的多粒度 MRC 模型,该模型利用图注意网络来获得不同层次的表示,对两个粒度答案之间的依赖关系进行建模,从而为彼此提供证据。最近,Wu 等人<sup>[35]</sup>利用选择引导策略(S2G)由粗到细的方式精确检索证据段落,并结合两种新颖的注意机制,设计出符合多跳推理本质的 MRC 模型。

综上所述,可知在 MRC 任务中结合注意力机制的模型层出不穷,表 2 以时间顺序梳理了近五年来不同注意力机制应用在 MRC 中的典型模型方法及主要贡献。

表 2 基于注意力机制方法的典型 MRC 模型及主要贡献

时间	模型	主要贡献
2016	BiDAF <sup>[21]</sup>	根据不同粒度分成多阶段来获得问题感知的上下文表示结构
2017	DIM Reader <sup>[22]</sup>	构建基于多跳的双重迭代交替注意力机制
2017	R-Net <sup>[26]</sup>	首次在 MRC 模型上引入自注意力机制
2018	MHPGM <sup>[29]</sup>	使用混合注意力机制(双向注意力加自注意力)来执行多跳推理
2018	GAI <sup>[30]</sup>	提出一种新的全局注意力推理神经网络结构
2019	DynSAN <sup>[27]</sup>	在 MRC 任务中,引入动态自注意力网络
2019	CFC <sup>[33]</sup>	使用协同注意力和自注意力来设计粗粒度-细粒度协同网络问答模型
2020	Zheng 等人提出的模型 <sup>[34]</sup>	提出基于图注意力网络结合 Bert 的多粒度 MRC 模型,采用联合训练策略
2021	graph-free S2G model <sup>[35]</sup>	引入两种特定任务注意机制,根据每个特定任务的性质来限制每个表征的作用域

### 3.2 基于图神经网络方法

图神经网络的概念最初由 Gori 等人<sup>[36]</sup>于 2005 年提出,并在 Scarselli<sup>[37]</sup>等人的文章中进一步阐述。最近几年,深度学习方法在图上得到了广泛的探索与研究。在 CNN、RNN 和深度自动编码器的基础上,衍生出很多处理图数据的神经网络结构,即图神经网络(GNN)。按照其建模方式的不同,将图神经网络划分为图卷积网络(GCN)、图注意力网络(GAN)、图自编码器(Graph Auto-encoders)、图生成网络(GGN)和图时空网络五大类别。

目前,图的深度学习和 NLP 的交叉研究浪潮影响着很多 NLP 任务,人们应用和开发不同的 GNNs 变体并在诸多 NLP 任务中取得相当大的成功。陈雨龙等人<sup>[38]</sup>对图神经网络在自然语言处理中的不同应用进行系统论述,认为如何用图来建模不同任务中的关键信息是未来工作中的重点。Song 等人<sup>[39]</sup>探索过基于图结构的 MRC 任务,主要解决 MRC 中的多跳语义关

联问题。Su 等人<sup>[40]</sup>提出用于问题生成的多跳编码融合网络(MulQG),MulQG 使用图卷积网络在多跳中进行上下文编码,并通过编码器推理门进行编码融合。受人类推理过程的启发,Tang 等人<sup>[41]</sup>从支持文档中构建了基于路径的推理图,该图结合基于图的方法和基于路径的方法,更适合多跳推理。

对于需要通过多次跳转一组文档才能得到正确答案的问题,模型需要:(1)选出和问题相关的段落;(2)找出段落中有力的支持证据;(3)根据收集到的证据推理出正确答案。基于图神经网络其固有的消息传递机制,可以通过图传播多跳信息,Fang 等人<sup>[42]</sup>提出一种用于多跳问题回答的层次图网络(HGN),该模型通过层次框架中的多级细粒度图来支持联合回答/证据预测,对每个问题,HGN 不是只使用实体作为节点,而是构建一个层次图来从不同粒度级别(问题、段落、句子、实体)中获取线索,HGN 模型各部分如图 3 所示。

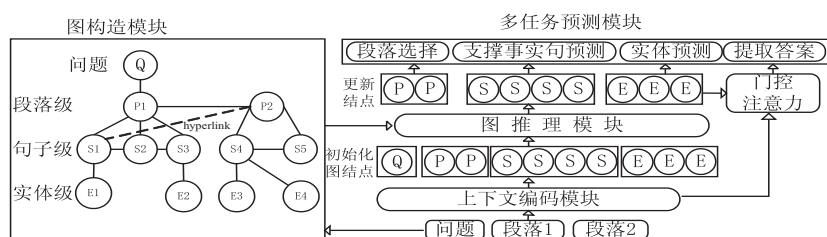


图 3 HGN 模型架构

HGN 主要由四部分组成,图构造模块:用于构造分层图以连接不同粒度信息,其中分层图需要分两步构建,首先要识别出和问题相关联的各段落,然后再添加所选段落中各级实体(句子/实体)之间联系的边;上下文编码模块:也就是通过基于 RoBERTa 的模型编码器得到图中各个图节点的初始表示;图推理模块:在上下文编码完成后,HGN 对分成图推理,通过图神经网络将上下文表示的图节点转换成更高级的特征,并利用基于图注意力的方法完成节点的表示更新;多任务预测模块:图推理完成后,更新的节点表示被用于不同的子任务(选择段落+寻找支持证据+实体预测+提取答案),完成这些子任务后才得到答案。

由于层次图可以应用于不同的多跳问答数据集,为了推广到其他需要使用较长文档的数据集,可使用基于滑动窗口的方法将长序列分成短序列,或者用能够处理长序列的其他基于变压器的模型代替基于 Bert 的主干。Song 等人<sup>[39]</sup>基于共同参考全局信息或滑动窗口构建实体图,提出一种新的模型 MHQA-GRN 来更好地连接全局证据,形成了更复杂的图。Tu 等人<sup>[43]</sup>通过添加文档节点以及在文档、实体和候选答案之间创建交互来丰富实体图中的信息,引入一种具有不同类型节点和边的异构图,称为异构文档实体图

(HDE),HDE 使用基于图神经网络(GNN)的消息传递算法来积累证据。

可解释的多跳问题回答需要对多个文档进行逐步推理,并找到分散的支持事实来回答问题。已有工作提出用实体图方法来聚合实体信息,提高推理能力。然而,实体图有时也会丢失一些对理解语义也很重要非实体信息。此外,分布在噪声句子中的实体可能会误导推理过程。为此,Zhang 等人<sup>[44]</sup>提出一种粗粒度和细粒度图网络(CFGGN),CFGGN 是一种结合句子信息和实体信息来回答多跳问题的新的可解释模型,由执行句子级推理的粗粒度模块和进行实体级引用的细粒度模块组成。为解决多跳问答模型在没有正确推理的情况下反而得到正确答案的偏差问题,Lee 等人<sup>[45]</sup>提出一种学习证据的新方法,决定答案预测是否得到正确证据的支持。为处理动态和复杂的问题,Xu 等人<sup>[46]</sup>提出一种新的方法,通过动态地构建语义图并对其进行推理,在为多跳问答提供可解释性证据的同时,发掘更多的有效事实。

通过上述分析可知,近年来基于图神经网络方法来研究 MRC 任务逐渐兴起,其相关典型模型方法及主要贡献如表 3 所示。

表 3 基于图神经网络方法的典型 MRC 模型及主要贡献

时间	模型	主要贡献
2018	MHQA-GRN <sup>[39]</sup>	介绍一种新的方法来解决多跳阅读理解(MHMR)与证据整合过程
2019	HGN <sup>[42]</sup>	提出一个用于多跳问答的层次图网络,其中异构节点被编织成一个完整的层次图
2019	HDE <sup>[43]</sup>	提出一种新的基于 GNN 的跨文档的多跳 RC 方法,引入异构图,用于在不同粒度上进行多跳推理
2020	MulQG <sup>[40]</sup>	使用图卷积网络(GCN)在多跳 MRC 中进行上下文编码,迈出在多跳方面评估模型性能的第一步
2021	Xu 等人 <sup>[46]</sup> 提出的方法	提出一个新的框架,结合 AMR( Abstract Meaning Representation)来为多跳问答系统进行可解释的知识检索和推理,以反映相关事实的内在关系

### 3.3 基于问题分解方法

MHMR 具有挑战性,因为它需要跨多个段落收集证据来回答一个问题,并且需要对两个证据如何相互关联进行更复杂的推理。多跳问题大都为复杂的问题,然而复杂问题可以分解为若干单跳问题。在单跳 MRC 中,系统可以使用单个句子获得良好的性能。基于此,Min 等人<sup>[47]</sup>于 2019 年提出一个用于多跳 RC 的系统 DECOMPRC,其学会使用原始问题的跨度将组合多跳问题分解成简单的单跳子问题。此外,他们还提出一种重排序方法,从不同的可能分解中获得答案,并用答案重排序每个分解,以决定最终答案,而不是一开始就决定分解。同年,Min 等人<sup>[48]</sup>认为问题组合性不是多跳推理的充分条件,即使是高度复合的问题,如果这些问题针对的是特定实体类型,或者回答这些问题所需的事实是多余的,也可以用一跳来回答。为此,他

们提出一个基于单跳 Bert 的阅读理解模型,如图 4 所示。

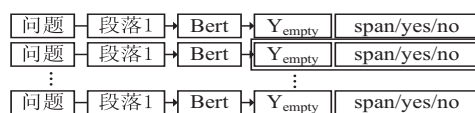


图 4 单段落 Bert 模型

单段落 Bert 模型分别对每段进行评分和回答,然后从  $y_{empty}$  得分最低的段落中得到答案。其中  $y_{span}$ 、 $y_{yes}$ 、 $y_{no}$  和  $y_{empty}$  分别表示答案是有跨度区间、是类型、非类型或没有答案类型。如表 4 所示,与代表性模型相比单段落 Bert 模型在多跳数据集 HotpotQA 上 F1 值达到 67.08%。这表明大多数 HotpotQA 问题都可以使用单跳模型加干扰物设置来回答。最后,他们还设计一个评估环境,在这个环境中,人类没有看到预期的多跳推理所需的所有段落,但仍然可以回答 80% 以



上的问题,同样表明大多数组合问题都可以分解成单跳问题。

表4 单段落 Bert 模型在 HotpotQA 上的 F1 值 %

模型	有干扰环境	开放域环境
	F1 值	F1 值
Single-paragraph Bert *	67.08	38.04
BiDAF <sup>[21]</sup>	58.99	32.89
GRN <sup>[39]</sup>	66.71	36.48
QFE <sup>[49]</sup>	68.06	38.06
DecompRC <sup>[47]</sup>	69.63	40.65
CogQA <sup>[14]</sup>	-	48.87

通过研究多跳问题的最佳表现模型是否像人类一样理解潜在的子问题。2020 年, Tang 等人<sup>[50]</sup>采用神经分解模型为多跳复杂问题生成子问题,然后提取相应的子答案。此外还提出一个新的模型来提高回答子问题的性能,这项工作朝着构建一个更易于解释的多跳质量保证系统迈出了重要一步。2021 年, Cao 等人<sup>[51]</sup>引入粗粒度复杂问题分解 (CGDe) 策略,在没有任何附加注释的情况下将复杂问题分解为简单问题,并结合细粒度交互 (FGIn) 策略,以更好地表示文档中的每个单词,提取与推理路径相关的更全面准确的句子。

## 4 结束语

该文归纳总结了目前多跳式机器阅读理解的概念和多跳领域的数据集,以及常用的模型方法,即基于注意力机制、图神经网络、问题分解的三种多跳式 MRC 模型。

针对当前多跳式 MRC 任务发展中存在的难点和挑战,提出以下几点未来研究方向:

(1) 目前针对多跳式 MRC 中文数据集很少,中文数据集与英文数据集在语法结构、语义分析、语序等多方面存在较大差异,如何构建一个好的面向多跳领域的中文数据集仍然是一个巨大的挑战。另外,面向各专业领域数据集相对较少,各行各业如何结合行业发展趋势,推出各自领域的相关数据集,这也是 NLP 各项任务未来必不可少的研究。

(2) 从上述研究方法中,不难发现能否设计出高效准确的推理方法至关重要,如何结合注意力机制以及图神经网络设计出合理高效的推理方法仍是研究热点之一。

(3) 在实际情况下 MRC 模型的输入是很复杂的,对于一些夹杂着噪声的输入常常影响模型的泛化性能,因此如何增强模型鲁棒性的研究也是一个巨大的挑战。

(4) 人类在解答问题时通常会联想到问题的相关背景以及有关联的信息等,而阅读理解模型在寻找答案时所考虑的信息都来自于给定文档,没有融合外部知识来理解文章,可想而知,效率以及准确度大大下降。因此,如何引入外部知识并与 MRC 模型相结合来提升模型准确率,也需要深入探讨研究。

## 参考文献:

- [1] TURING A M, HAUGELAND J. Computing machinery and intelligence[J]. Mind, 1950, 59(236): 433-460.
- [2] TERRY W. Understanding natural language[J]. Cognitive psychology, 1972, 3(1): 1-191.
- [3] HIRSCHMAN L, LIGHT M, BRECK E, et al. Deep read: a reading comprehension system[C]//Proceedings of the 37th ACL. Maryland: ACL, 1999: 325-332.
- [4] RILOFF E, MICHAEL T. A rule-based question answering system for reading comprehension tests[C]//Proceedings of the ANLP/NAACL workshop on reading comprehension tests as evaluation for computer-based language understanding systems. Seattle: [s. n.], 2000: 13-19.
- [5] HERMANN K M, KOCISKY T, EDWARD G, et al. Teaching machines to read and comprehend[C]//Advances in neural information processing systems. Montreal: NIPS, 2015: 1693-1701.
- [6] LIU S S, ZHANG X, ZHANG S, et al. Neural machine reading comprehension: methods and trends[J]. Applied Sciences, 2019, 9(18): 3698.
- [7] ZHANG X, YANG A, LI S J, et al. Machine reading comprehension: a literature review[J]. arXiv: 1907.01686, 2019.
- [8] DZENDZIK D, CARL V, JENNIFER F. English machine reading comprehension datasets: a survey[J]. arXiv: 2101.10421, 2021.
- [9] CHEN D Q. Neural reading comprehension and beyond[D]. San Francisco: Stanford University, 2018.
- [10] 包 玥, 李艳玲, 林 民. 抽取式机器阅读理解研究综述[J]. 计算机工程与应用, 2021, 57(12): 25-36.
- [11] 李舟军, 王昌宝. 基于深度学习的机器阅读理解综述[J]. 计算机科学, 2019, 46(7): 7-12.
- [12] 张超然, 袁杭萍, 孙 毅, 等. 基于预训练模型的机器阅读理解研究综述[J]. 计算机工程与应用, 2020, 56(11): 17-25.
- [13] 徐霄玲, 郑建立, 尹梓名. 机器阅读理解的技术研究综述[J]. 小型微型计算机系统, 2020, 41(3): 464-470.
- [14] DING M, ZHOU C, CHEN Q B, et al. Cognitive graph for multi-hop reading comprehension at scale[C]//Proceedings of the 57th ACL. Florence: ACL, 2019: 2694-2703.
- [15] JOHANNES W, PONTUS S, SEBASTIAN R. Constructing datasets for multi-hop reading comprehension across documents[J]. Transactions of the Association for Computational Linguistics, 2018, 6: 287-302.

- [16] YANG Z, QI P, ZHANG S, et al. HotpotQA: a dataset for diverse, explainable multi-hop question answering [C]//Proceedings of the 2018 EMNLP. Brussels: ACL, 2018; 2369–2380.
- [17] VRANDEĆIĆ D. Wikidata: a new platform for collaborative data collection [C]//Proceedings of the 21st international conference on world wide web. Lyon: ACM, 2012; 1063–1064.
- [18] DANIEL H, ALEXANDRE L, LLION J, et al. WikiReading: a novel large-scale language understanding task over Wikipedia [C]//Proceedings of the 54th ACL. Berlin: ACL, 2016; 1535–1545.
- [19] KASTNER S, UNGERLEIDER, LESLIE G. Mechanisms of visual attention in the human cortex [J]. Annual Review of Neuroscience, 2000, 23(1): 315–341.
- [20] BAHDANAU D, CHO K, BENGIO Y. Neural machine translation by jointly learning to align and translate [J]. arXiv; 1409.0473, 2014.
- [21] SEO M, KEMBHAVI A, FARHADI A, et al. Bidirectional attention flow for machine comprehension [J]. arXiv; 1611.01603, 2016.
- [22] LIU Z, HUANG D G, HUANG K Y, et al. DIM reader: dual interaction model for machine comprehension [C]//Proceedings of the 2017 CCL and NLP–NABD. Nanjing: CCL, 2017; 387–397.
- [23] LI J Q, LI B A, LV X Q. Machine reading comprehension based on the combination of BiDAF model and word vectors [C]//Proceedings of the 2nd international conference on computer science and application engineering. Hohhot: ACM, 2018; 1–4.
- [24] XU H, ZHANG S C, JIANG J. Machine reading comprehension based on multi-headed attention model [C]//MATEC web of conferences. EDP sciences. [s. l.]; EI, 2018; 02047.
- [25] LIN Z H, FENG M W, SANTOS C N, et al. A structured self-attentive sentence embedding [J]. arXiv; 1703.03130, 2017.
- [26] WANG W H, YANG N, WEI F, et al. Gated self-matching networks for reading comprehension and question answering [C]//Proceedings of the 55th ACL. Vancouver: ACL, 2017; 189–198.
- [27] ZHUANG Y M, WANG H D. Token-level dynamic self-attention network for multi-passages reading comprehension [C]//Proceedings of the 57th ACL. Florence: ACL, 2019; 2252–2262.
- [28] 苏立新, 郭嘉丰, 范意兴, 等. 面向多片段答案的抽取式阅读理解模型 [J]. 计算机学报, 2020, 43(5): 856–867.
- [29] BAUER L, WANG Y, BANSAL M. Commonsense for generative multi-hop question answering tasks [C]//Proceedings of the 2018 EMNLP. Brussels: ACL, 2018; 4220–4230.
- [30] SONG J, TANG S L, QIAN T C, et al. Reading document and answering question via global attentional inference [C]//Pacific rim conference on multimedia. Hefei: Springer, 2018; 335–345.
- [31] DUAN L, GAO J, LI A. A multi-attention matching model for multiple-choice reading comprehension [J]. International Journal of Cooperative Information Systems, 2020, 29(01n02): 1950010.
- [32] 谭红叶, 屈保兴. 面向多类型问题的阅读理解方法研究 [J]. 中文信息学报, 2020, 34(6): 81–88.
- [33] ZHONG V, XIONG C M, KESKAR N S, et al. Coarse-grain fine-grain coattention network for multi-evidence question answering [J]. arXiv; 1901.00603, 2019.
- [34] ZHENG B, WEN H Y, LIANG Y B, et al. Document modeling with graph attention networks for multi-grained machine reading comprehension [J]. arXiv; 2005.05806, 2020.
- [35] WU B H, ZHANG Z S, ZHAO H. Graph-free multi-hop reading comprehension: a select-to-guide strategy [J]. arXiv; 2107.11823, 2021.
- [36] GORI M, GABRIELE M, FRANCO S. A new model for learning in graph domains [C]//Proceedings 2005 IEEE international joint conference on neural networks. Montreal: IEEE, 2005; 729–734.
- [37] SCARSELLI F, GORI M, TSOI A C, et al. The graph neural network model [J]. IEEE Transactions on Neural Networks, 2008, 20(1): 61–80.
- [38] 陈雨龙, 付乾坤, 张 岳. 图神经网络在自然语言处理中的应用 [J]. 中文信息学报, 2021, 35(3): 1–23.
- [39] SONG L F, WANG Z G, YU M, et al. Exploring graph-structured passage representation for multi-hop reading comprehension with graph neural networks [J]. arXiv; 1809.02040, 2018.
- [40] SU D, XU Y, DAI W L, et al. Multi-hop question generation with graph convolutional network [C]//Findings of the ACL; EMNLP 2020. [s. l.]; ACL, 2020; 4636–4647.
- [41] TANG Z Y, SHEN Y L, MA X Y, et al. Multi-hop reading comprehension across documents with path-based graph convolutional network [J]. arXiv; 2006.06478, 2020.
- [42] FANG Y W, SUN S Q, GAN Z, et al. Hierarchical graph network for multi-hop question answering [J]. arXiv; 1911.03631, 2019.
- [43] TU M, WANG G T, HUANG J, et al. Multi-hop reading comprehension across multiple documents by reasoning over heterogeneous graphs [C]//Proceedings of the 57th ACL. Florence: ACL, 2019; 2704–2713.
- [44] ZHANG M, LI F, WANG Y, et al. Coarse and fine granularity graph reasoning for interpretable multi-hop question answering [J]. IEEE Access, 2020, 8: 56755–56765.
- [45] LEE K, HWANG S W, HAN S E, et al. Robustifying multi-hop QA through pseudo-evidentiality training [J]. arXiv; 2107.03242, 2021.
- [46] XU W W, ZHANG H H, CAI D, et al. Dynamic semantic graph construction and reasoning for explainable multi-hop