

引文格式:

刘啸, 杨敏. 基于动态图神经网络的会话式机器阅读理解研究 [J]. 集成技术, 2022, 11(2): 67-78.

Liu X, Yang M. Research on conversational machine reading comprehension based on dynamic graph neural network [J]. Journal of Integration Technology, 2022, 11(2): 67-78.

基于动态图神经网络的会话式机器阅读理解研究

刘 啸^{1,2} 杨 敏^{1*}

¹(中国科学院深圳先进技术研究院 深圳 518055)

²(中国科学院大学深圳先进技术学院 深圳 518055)

摘 要 近年来, 基于深度学习的机器阅读理解模型研究取得显著进展, 但这些模型在全局语义关系构建以及较长距离推理上仍有显著缺陷, 在对段落文本进行推理时, 大多只把文本信息看作词的序列, 而没有探索词与词之间丰富的语义关系。为了解决上述问题, 该文提出一种新的基于动态图神经网络的会话式机器阅读理解模型。首先, 提取文本中的实体, 使用句法结构与句子之间的语义关系进行建模; 然后通过语义融合模块, 将基于序列化结构得到的上下文嵌入表示与基于图结构得到的实体节点嵌入表示进行融合; 最终使用图神经网络实现对答案的预测。同时, 该模型可在每轮对话过程中动态地构建问题和会话历史的推理图, 能有效地捕捉对话中的语义结构信息和会话历史流程。实验结果表明, 在两个最近提出的会话挑战 (CoQA 和 QuAC) 上表现了出色的性能。

关键词 会话式机器阅读理解; 会话问答; 图神经网络

中图分类号 TP 399 **文献标志码** A **doi**: 10.12146/j.issn.2095-3135.20211122001

Research on Conversational Machine Reading Comprehension Based on Dynamic Graph Neural Network

LIU Xiao^{1,2} YANG Min^{1*}

¹(Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China)

²(Shenzhen College of Advanced Technology, University of Chinese Academy of Sciences, Shenzhen 518055, China)

*Corresponding Author: min.yang@siat.ac.cn

Abstract Machine reading comprehension models based on deep learning have achieved remarkable success recently. However, these models have significant defects in constructing long-distance and global semantic relationships, which affect their performance in reading comprehension tasks. Moreover, when reasoning over passage text, most of them simply regard it as a word sequence without exploring rich semantic relationships between words. In order to solve this problem, this paper proposes a new system

收稿日期: 2021-11-22 修回日期: 2021-12-10

基金项目: 国家自然科学基金项目 (61906185, 61902385); 广东省自然科学基金项目 (2019A1515011705)

作者简介: 刘啸, 硕士研究生, 研究方向为自然语言处理-对话系统; 杨敏(通讯作者), 副研究员, 研究方向为自然语言处理和推荐系统等, E-mail: min.yang@siat.ac.cn.

effective graph structure named Dynamic Conversational Graph Network (DCGN). Firstly, named entities are extracted from the text, and the semantic relationship between syntactic structure and sentence is used for modeling. Then, the context-embedded representation based on serialization structure and the entity node embedded representation based on graph structure are fused by semantic fusion module. Finally, dynamic graph neural network is used to realize machine reading comprehension. The model dynamically builds inference graphs of questions and session history during each round of conversation, which can effectively capture the semantic structure information and the history flow of the conversation. Experimental results show that the model performs well on two recent session challenges (CoQA and QuAC).

Keywords conversational machine comprehension; conversation question and answer; graph neural network

Funding This work is supported by National Natural Science Foundation of China (61906185, 61902385), Natural Science Foundation of Guangdong (2019A1515011705)

1 引 言

机器阅读理解^[1]是自然语言处理领域中一项重要且富有挑战的任务,该任务通过向机器输入一段非结构化文本及其相应的问题,要求机器能够根据文本回答相关问题。然而,这些问题往往充斥着迷惑性,如果不理解文本内容而根据规则进行回答,由此得到的答案大概率为错误答案。近年来,机器阅读理解得益于其在信息检索和问答任务中的广泛应用而受到越来越多的关注。大量大规模数据集的发布,使其迎来了迅速的发展^[2-4]。当前的大部分阅读理解模型主要聚焦于单轮问答中文章的答案提取,同一篇文章的不同问题和答案之间没有联系。相关模型在 SQuAD^[5]数据集上的性能甚至已经超过了人类。

然而,在现实生活中,人类的交流更多是基于对话式的。在对话行为中,人类会基于已学习的信息,通过不断地询问来获取额外的信息,这是机器智能化的体现。为了探索机器在历史对话中提取有效信息并结合文章内容进行逻辑推理的能力,一种全新的机器阅读理解任务,即会话式机器阅读理解任务被提出。会话式机器阅读理解^[6]是由机器阅读理解与会话式问答交叉

形成的新领域,即在传统机器阅读理解的基础上加入多轮问答,使问答对之间具有一定的承接关系。它要求机器能够模拟人类捕获信息的方式来回答问题。会话式机器阅读理解任务被证实比普通机器阅读理解任务更具有挑战性,因为它不仅需要理解文章段落,还需要理解会话历史中的语言现象,诸如指代关系、主语省略、主题转移等。相比于单轮机器阅读理解,会话式机器阅读理解更符合人类对相关主题的问答习惯。为了探索这一新领域,越来越多的学者尝试构建此类数据集并尝试不同的模型和方法。其中,最经典的数据集为 Reddy 等^[7]在 2018 年创建的多轮机器阅读理解数据集 CoQA,它包含了 127 000 个问题及相应答案,以及 8 000 篇对应的文章。此后,有众多研究者对该数据集展开了大量实验,主要分为以下几种方式:(1)将原训练数据中的多轮对话问题拆解为单轮对话问题,简化任务形式,并通过多次打乱问题顺序来提高模型的泛化能力。对于问题轮数较少的样本而言,这种处理方法可以在一定程度上利用模型的大规模参数来强行拟合数据,达到较好的结果。但是,对于问题轮数较多,且上下文关联性非常强的样本而言,这种方法会破坏上下问题的

联系, 导致效果降低。(2) 另一种处理方式是将样本中的多个问题处理为流机制的形式, 上一个问题的处理结果会以输入参数的形式参与当前问题的处理过程, 这样在一定程度上保留了问题之间的关联性。但是过往的方法在处理问题流时仅关注了问题与文章单一字、词的联系, 忽略了问题流与全文的交互过程, 造成信息损失。(3) 还有一些处理方法是通过对大规模预训练语言模型构建而成, 如基于变换器的双向编码器表示技术 (Bidirectional Encoder Representations from Transformers, BERT) 或基于广义自回归预训练方法 (Generalized Autoregressive Pretraining Method, XLNet) 的阅读理解模型^[8-9]。预训练语言模型利用迁移学习的思想, 从相关任务的大量数据中预先训练出有效的语言模型, 然后迁移到目标任务中加以优化, 可以在一定程度上提高模型的准确率。但是, 这些模型不能有效整合句子的句法结构、句子间长距离的语义关系等信息, 从而造成模型对问题和篇章理解不够充分。

受到图神经网络模型在自然语言处理等领域获得成功的启示, 本文提出了一种用于会话式机器理解的动态会话图神经网络 (Dynamic Conversational Graph Network, DCGN)。首先, 针对当前机器阅读理解模型无法构建全局语义关系以及较长距离推理关系的问题, 该方法在传统的词序列结构基础上添加了文本实体图结构。基于问题和对话语境中的命名实体构建动态文本图, 文本图会随着问题和对话历史的处理过程而发生动态变化。通过这种方式, 系统可以在一系列上下文本图中对语义和时间依赖性进行建模。然后利用动态图神经网络获得上下文中实体的语义嵌入表示, 并与采用 RoBERTa^[10] 预训练模型所提取的序列化结构的上下文嵌入表示进行融合。最终根据问题嵌入与融合后的语义嵌入表示计算每轮的匹配分数并进行答案预测。通过在相关数据集上进行实验, 结果表明本文系统的性能

超过了所有基线模型。

2 会话式机器阅读理解研究现状

2.1 会话式机器阅读理解

近年来, 随着深度学习技术的不断进步, 机器阅读理解成为自然语言处理领域里一个重要且富有挑战的研究课题, 受到了越来越多的关注。传统的机器阅读理解任务由文章段落、问题、答案三要素构成, 主要研究单轮问答的文章的答案提取, 并提出了几种单轮模型, 如在 SQuAD 数据集上的 BiDAF 模型^[11]、QANet 模型^[12]等。然而, 随着数据集逐渐从虚构内容转向真实场景, 问题由简单问题向复杂问题发展, 机器阅读理解的研究开始向更复杂的会话式机器阅读理解领域进行探索。会话式机器阅读理解是由机器阅读理解与会话式问答交叉形成的新领域, 即在传统机器阅读理解的基础上加入多轮问答, 使问答对之间具有一定的承接关系。这更符合真实场景中人们通过多轮问答逐层深入了解相关主题信息的情况。通常来说, 会话式机器阅读理解任务由文章段落、会话、会话历史三要素构成。机器不仅要理解文章段落, 还要理解会话历史中的语言现象, 诸如指代关系、主语省略、主题转移等。相比于单轮机器阅读理解, 会话式机器阅读理解更符合人类对相关主题的问答习惯。

为了将会话历史信息融入模型, 早期的处理方式大多为显性地将过去的问题以及答案直接拼接到当前的问题之前, 这样就可以保留对话中的历史信息。典型的做法有 Choi 等^[13]提出的 BiDAF++w/x-ctx 模型——利用“问题-文章段落”和“文章段落-问题”之间双向的注意力来捕捉回答问题的有效信息。另外, 为了将会话历史信息添加到当前问题中, BiDAF++w/x-ctx 模型将上一个问题的答案在文章中所对应的下标信息拼接到文章的词向量矩阵里, 再进行

文章与问题的双向注意力操作。与 Choi 不同的是, Reddy 等^[7]提出的 DrQA+PGNet 模型,直接将上一个问题以及答案与当前问题进行拼接,并用特殊符号进行标记,再将其以普通机器阅读理解任务的形式进行建模。随后, Zhu 等^[14]提出的 SDNet 模型,将前几轮的问题和答案拼接到当前问题之前,以纳入上下文信息,而且在文章和问题上运用了互相关注和自我关注的方法,深度融合文章与问题的信息,获得了不错的效果。自以 BERT 为代表的大规模模型问世以来,有众多研究者探索这种预训练模型在会话式机器阅读理解任务上的表现。Ohsugi 等^[15]提出了 BERT+Finetune 模型——使用 BERT 模型训练当前问题-文章、上一问题-文章、上一问题的答案-文章,取得文章以及问题的高质量特征数据,再按普通机器阅读理解的形式进行建模,取得了非常好的效果。然而,这些处理方式只是简单地将之前的问题与当前问题的答案拼接起来,从而保持历史问题与当前问题的关联性,本质上并没有解决会话式机器阅读理解任务中模型需要从会话历史中获取与问题相关的信息的初衷,忽略了基于历史信息的会话推理过程。

会话推理是会话式机器阅读理解模型的核心,主要负责将问题、文章段落以及会话历史进一步融合,从而获得三者之间的相互加权关系以及深层次语义信息,并为预测答案提供推理依据。与循环神经网络通过序列传递历史信息类似,流是一种潜在表示序列,它将先前会话推理过程中产生的潜在表示序列传递到当前问答推理中,使得当前模型可以沿着会话进程方向推理。Huang 等^[16]构建的 FlowQA 模型提出了一种沿着会话进程方向以及上下文词方向相互交叉融合信息的集成流机制。它在以往单轮机器阅读理解模型对文章进行双向循环神经网络提取特征的基础上,加入了会话进程方向的单向循环神经网络,将历史问题的推理信息以隐向量的形式传入当前

问题中并参与推理过程。这一简单的处理极大地提升了原本模型在此任务中的效果,也有力地证明了流机制在会话式机器阅读理解任务中的重要性。此外, Chen 等^[17]提出的 GraphFlow 模型利用基于图神经网络的集成图流机制来保留历史问题的推理信息。GraphFlow 将构建问题和会话历史感知的上下文图作为潜在表示序列传入图神经网络中,并使用 K 最近邻(K-Nearest Neighbor, KNN)分类算法选择 k 个最近邻(包括自身)以及每个上下文节点的注意力值,来保存文章中重要的推理关系。这也是第一个将图神经网络与会话式机器阅读理解任务相结合的模型,取得了不错的效果。由此可见,会话一般具有较强的上下文关联性,因此,可以将会话历史结构作为抽象信息辅助模型推理。然而,现有的基于流机制的方法尚处于萌芽阶段,大部分的模型将文本简单地看作词的序列,没有探索词之间丰富的语义关系。

2.2 图神经网络

图神经网络最早由 Scarselli 等^[18]提出,是一种用于处理图结构数据的神经网络模型,在信息传播、关系归纳偏置上展现了优秀的性能。当信息在图的节点之间传播时,图神经网络中每个节点会学习到它的邻居节点状态表示,并以此来更新自己的节点表示。近年来,图神经网络在自然语言处理的各种任务中取得了较好的成果。越来越多的学者将图神经网络用于机器阅读理解任务,并显示出优于传统方法的优势。其中, Qiu 等^[19]提出的 DFGN 模型基于文本内容构建实体图,并通过图注意力机制传播节点之间的信息。以从问题中提取的实体作为起点,沿着基于文章内容动态构建的实体图进行探索,并逐步找到给定文章中相应的支持实体。同时,添加了 Graph2Doc 模块,将每个实体节点的嵌入表示与该实体对应的单词嵌入表示进行融合,然后送入长短期记忆,将图结构中实体的表示转化为序列

表示来抽取答案。Thayaparan 等^[20]提出的文档图网络 (Document Graph Network, DGN) 模型, 采用预过滤步骤来限制节点的数量, 并直接在提取的文档图上训练门控图神经网络, 从而为文档图结构表示的识别提供证据文本。Ding 等^[21]提出一个多跳阅读理解框架 CogQA, 该框架以认知科学中的双过程理论为基础, 使用 2 个系统来维护 1 张认知图谱: 系统 1 采用 BERT 模型, 隐式地从句子中提取相关实体并对其中的信息进行编码, 然后将它们提供给系统 2; 系统 2 利用图神经网络将这些实体及其编码信息构建为 1 张认知图谱, 通过图来对这些相关信息进行推理计算, 同时指导系统 1 进行实体提取。CogQA 模型通过模拟人类的认知方式来解决机器的阅读理解问题, 利用协同隐式提取模块和显式推理模块, 在迭代过程中逐步构建认知图, 并在认知图基础上解答问题。虽然上述方法在多跳阅读理解任务上测试证明能够取得不错的性能, 但会话式机器阅读理解的核心问题是建模会话推理过程。如何进一步运用图神经网络来完善机器的逻辑推理能力, 使其拥有真正的自然语言理解能力, 是未来一个重要的研究方向。

3 基于动态图神经网络的阅读理解模型

3.1 问题描述

Reddy 等^[7]给出了会话式机器阅读理解任务的定义: 它通过给机器一段文本, 及其具有上下文内关系的一组序列问题, 要求机器能够聚集会话中的信息, 并且给出相应的回答。该任务的数学定义形式为: 输入给定的文章段落 C , 会话历史信息 $H^{(i)} = \{Q^{(i-N)}, A^{(i-N)}, \dots, Q^{(i-1)}, A^{(i-1)}\}$ 。其中, $Q^{(i-1)}, A^{(i-1)}$ 分别为第 $i-1$ 轮的问题和答案, N 代表前 N 轮对话。会话式机器阅读理解的任务是预测当前问题 $Q^{(i)}$ 的答案 $A^{(i)}$, 输出是该答案在文章段落中提取的跨距文本的起始和终止两个词的位置 (W_{start}, W_{end})。同时, 系统也会预测当前问题答案的类型 $T^{(i)}$ 。具体来说, 在 CoQA 任务中, 答案种类有 “Yes” “No” “No Answer” 和 “Answer Span” 4 种, 分别代表肯定回答、否定回答、没有答案, 以及答案是文章中的一段句子或短语。具体例子如图 1 所示。

3.2 动态图神经网络的会话式机器阅读理解模型

本文所提方法的模型框架如图 2 所示, 主要

文章段落	Jessica went to sit in her rocking chair. Today was her birthday and she was turning 80. Her granddaughter Annie was coming over in the afternoon and Jessica was very excited to see her. Her daughter Melanie and Melanie's husband Josh were coming as well. Jessica had...
问题与答案	问题 1: Who had a birthday? 答案 1: Jessica 线索 1: Jessica went to sit in her rocking chair. Today was her birthday and she was turning 80
	问题 2: How old would she be? 答案 2: 80 线索 2: she was turning 80.
	问题 3: Did she plan to have any visitors? 答案 3: Yes 线索 3: Her granddaughter Annie was coming over
	问题 4: How many? 答案 4: Three 线索 4: Her granddaughter Annie was coming over in the afternoon and Jessica was very excited to see her. Her daughter Melanie and Melanie's husband Josh were coming as well.
	问题 5: Who? 答案 5: Annie, Melanie and Josh 线索 5: Her granddaughter Annie was coming over in the afternoon and Jessica was very excited to see her. Her daughter Melanie and Melanie's husband Josh were coming as well.

图 1 CoQA 数据示例

Fig. 1 Example of CoQA dataset

包括信息编码层、信息推理层和答案预测层 3 个模块。

3.2.1 信息编码层

大量研究工作证明, 预训练模型可以有效提升模型性能, 这同样适用于会话式机器阅读理解模型。预训练模型不仅考虑到词的全局统计信息, 也涵盖了上下文信息。通常情况下, 传统的机器阅读理解模型使用预训练模型的最后一层输出值作为文本的嵌入表示。但在会话式机器阅读理解任务中, 上述方法无法处理输入序列中的会话历史信息。由于会话历史包含不同轮次的问答结果, 模型理解当前问题所需的信息可能在会话历史的任意位置。这需要模型可以捕捉不同距离的依赖关系。而 BERT 的每一层 Transformer 会生成不同的权重表示, 该权重表示可以看作模型对当前输入序列的不同阶段的表示。因此, 采用加权平均的方法融合 BERT 不同层 Transformer 的权重, 并将其作为输入序列的上下文嵌入。与 BERT 相比, RoBERTa^[22]的改进主要有以下几点: (1) 使用更多的训练资源和训练数据, 耗时更长。(2) RoBERTa 使用了动态掩码的方式, 每

次向模型输入一个序列时都会随机 mask 不同的 token。动态掩码的方式可以使模型在大量数据输入的过程中, 逐渐适应不同的掩码策略, 学习到不同的语言表征。(3) RoBERTa 使用更大的 byte 级别 BPE 词汇表来训练 BERT, 且没有对输入进行任何模型之外的预处理或分词操作。这可以有效防止 “unknown” 问题。因此, 本文使用 BERT 以及 RoBERTa 模型, 对输入文本进行编码。

模型采用预训练模型 BERT 的输入方式: $input = [CLS] + Q^{(i)} + [SEP] + H^{(i)} + [SEP] + C + [SEP]$ 。其中, $Q^{(i)}$ 为第 i 轮的问题; $H^{(i)}$ 为会话历史信息; C 为篇章; $[SEP]$ 为分隔符; $[CLS]$ 为整个输入的聚合表示。预训练模型 RoBERTa 可以表示为一个变换函数 $F(\cdot)$, 它接受一个训练实例的输入, 并为每个字符生成向量化表示, 具体计算如公式 (1) 所示。

$$T_{Q^{(i)}}, T_{H^{(i)}}, T_C = F(Q^{(i)}, H^{(i)}, C) \quad (1)$$

3.2.2 信息推理层

会话推理是会话式机器阅读理解模型的核心, 主要负责将问题、文章段落以及会话历史进一步融合, 从而获得三者之间的相互加权关

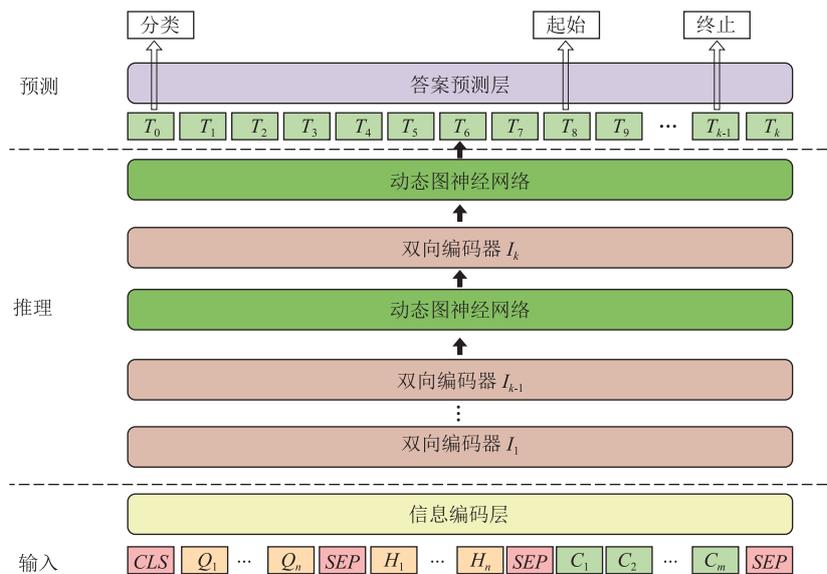


图 2 本文所提模型框架示意图

Fig. 2 The framework of the model proposed in this paper

系以及深层次语义信息，并为预测答案提供推理依据。当前，大部分的模型将文本简单地看作词的序列，没有探索词之间丰富的语义关系。为了解决这一问题，本文将图神经网络应用到会话式机器阅读理解任务中，利用文本实体图结构建立全局语义关系以及长距离推理关系，得到包含了语义结构信息的文本嵌入表示。同时，本文提出了一种动态会话图神经网络(Dynamic Conversational Graph Network, DCGN)，该网络基于问题和对话语境中的实体动态地构建文本图，利用动态图神经网络获得上下文中实体的语义嵌入表示。然后，通过融合机制将前轮的推理过程及结果并入当前上下文图中，提高了当前轮次的推理性能。该模块结构如图 3 所示，它由实体图构造模块、语义融合模块和动态图神经网络模块 3 部分组成。

(1) 实体图构造模块

构建文本实体图结构首先需要对文章进行命名实体抽取，实体是机器阅读理解任务中的关键要素。当前机器阅读理解数据集中，大多数问题都是基于事实的问题。实体在段落的语义信息中

有重要的地位，实体之间的语义关系也是推理过程中的重要线索。因此，通过构造文本中的实体图结构可以辅助获取文章的深层次语义信息，提高模型的推理性能。在当前命名实体抽取任务^[23]中有大量的研究成果，并取得了很好的效果。其中，SpaCy 工具包^[24]由于它的高效而被广泛运用，因此本文采用 SpaCy 工具包进行命名实体抽取。在得到实体后，将文章的所有实体构造为一个无向图，构建的图可表示为 $G(E, A)$ 。其中， E 为图的节点集合，该图中的顶点即为文中所有的实体； A 为图的边集合，本文采用双仿射注意力模型^[25]提取实体之间的依存关系。实体图中边的定义方式如下所示：(1) 基于实体的依存关系，如果一个句子中的两个实体之间存在依存句法关系，则两个实体之间建立边；(2) 基于句子的共现关系，如果两个实体出现在同一个句子中，则认为它们之间存在共现关系；(3) 基于上下文的同义词关系，如果文本中的两个实体具有相同词性时，则两个实体之间建立链接。共同词性边通过词性复现在句子之间建立关联，但是考虑到机器阅读理解任务中有时篇幅比较短，重复词较

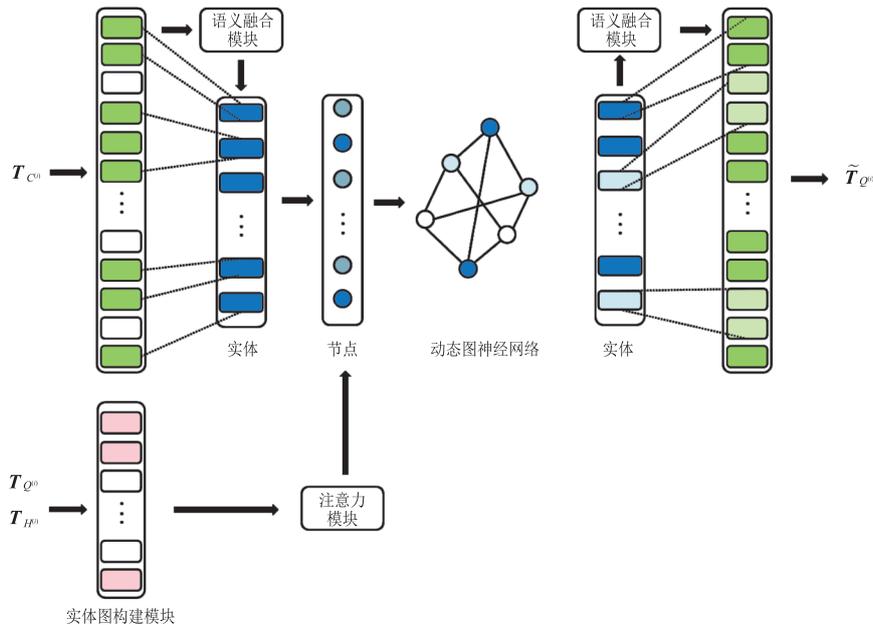


图 3 DCGN 模型框架示意图

Fig. 3 Overall architecture of the DCGN model

少。因此,本文引入词性关系边,通过词性的重复将不同句子关联起来,同时也克服了图稀疏的缺陷。具体的,满足以下规则之一时,两个实体节点 E_1 和 E_2 是相似的:① E_1 等于 E_2 ;② E_1 包含 E_2 ;③ E_1 和 E_2 之间重叠的单词数大于 E_1 和 E_2 最小单词数的一半。据此可以得到一个基于依存关系的具有多个节点和边缘类型的语义图——捕获了文章中实体之间的核心语义关系。此外,为了模仿人类的一些推理行为,实体图在推理过程中会发生动态变化。

(2) 语义融合模块

由于文本具备序列化结构的特点,因此在自然语言处理任务中,通常使用序列化结构进行文本嵌入表示。在构建了命名实体图之后,本文使用与实体关联的文本嵌入表示来计算实体图节点的向量表示。首先,构造一个大小为 $m \times m$ 的二进制矩阵 \mathbf{M} ,用于选择与实体关联的上下文跨度。其中, m 为上下文中的单词数; n 为实体数。如果上下文中第 i 个单词在第 j 个实体的范围内,则 $M_{i,j}=1$ 。由于每个实体节点中可能包含的单词数量不止一个,所以对每个节点中所包含的单词向量都进行最大池化与平均池化操作,再将得到的特征向量进行拼接,得到最终的每个节点信息表示向量如公式(2)所示:

$$\mathbf{E}^{(i)} = [\mathbf{e}_1^{(i)}, \mathbf{e}_2^{(i)}, \dots, \mathbf{e}_n^{(i)}] \quad (2)$$

其中, $\mathbf{E}^{(i)}$ 是大小为 $2d \times n$ 的矩阵; n 为实体数; $2d$ 为每个节点中所包含的单词向量进行最大池化与平均池化操作拼接之后的维度大小。

同时,为了模仿人类基于关键词的推理行为,在每轮对话过程中,本文使用当前问题嵌入向量 $\mathbf{T}_{Q^{(i)}}$ 和对话历史嵌入向量 $\mathbf{T}_{H^{(i)}}$ 来指导实体图的构造。通过在实体之间创建一个注意力网络来识别与问题相关的节点。节点与问题和会话历史信息相关性越大,从邻居节点接收的信息越多,具有更高的权重。本文使用问题嵌入、会话历史

嵌入和实体嵌入之间的注意网络来预测高权重节点。具体计算公式如下所示:

$$\mathbf{T}_{QH^{(i)}} = \text{MeanPooling} \left(\left[\mathbf{T}_{Q^{(i)}}; \mathbf{T}_{H^{(i)}} \right] \right) \quad (3)$$

$$\gamma_t^{(i)} = \frac{\mathbf{T}_{QH^{(i)}} \mathbf{V}^{(i)} \mathbf{e}_t^{(i)}}{\sqrt{d}} \quad (4)$$

$$\text{Att}_t^{(i)} = \sigma \left(\left[\gamma_1^{(i)}, \gamma_2^{(i)}, \dots, \gamma_n^{(i)} \right] \right) \quad (5)$$

$$\tilde{\mathbf{E}}^{(i)} = \left[\text{Att}_1^{(i)} \mathbf{e}_1^{(i)}, \text{Att}_2^{(i)} \mathbf{e}_2^{(i)}, \dots, \text{Att}_n^{(i)} \mathbf{e}_n^{(i)} \right] \quad (6)$$

其中, $\mathbf{V}^{(i)}$ 为线性投射矩阵; σ 为 sigmoid 函数; $\text{Att}_t^{(i)}$ 是注意力向量; $\tilde{\mathbf{E}}^{(i)}$ 是实体节点的动态嵌入向量。

(3) 动态图神经网络模块

将经过上下文语义融合模块编码后的特征向量输入图神经网络,可以得到图中节点的原始向量。由于每个节点都会与多个节点相连,因此要求节点有选择性地获取相邻节点的信息,并在实体图中传递最为相关的信息,所以模型采用门控图神经网络来更新图节点的表示,通过聚合来自邻居的信息来更新当前图节点的表示。在每轮对话中,通过门机制使当前节点 $\tilde{\mathbf{e}}_k^{(i)}$ 在获取其邻居节点 $\tilde{\mathbf{e}}_l^{(i)}$ 信息的时候更有选择性。具体公式如下所示:

$$\mathbf{h}_{IN_k}^{(i)} = \sum_{\tilde{\mathbf{e}}_k^{(i)} \in IN_k} \alpha_{kl}^{(i)} \tilde{\mathbf{e}}_k^{(i)} \quad (7)$$

$$\mathbf{h}_{OUT_k}^{(i)} = \sum_{\tilde{\mathbf{e}}_k^{(i)} \in OUT_k} \alpha_{kl}^{(i)} \tilde{\mathbf{e}}_k^{(i)} \quad (8)$$

其中, IN_k 以及 OUT_k 分别为节点 $\tilde{\mathbf{e}}_k^{(i)}$ 出度与入度的集合; $\alpha_{kl}^{(i)}$ 为节点 $\tilde{\mathbf{e}}_k^{(i)}$ 与邻居节点 $\tilde{\mathbf{e}}_l^{(i)}$ 的注意力系数。 $\text{Attn}(\cdot)$ 为注意力函数,其计算方式如公式(9)所示:

$$\alpha_{kl}^{(i)} = \frac{\exp \left(\text{Attn} \left(\tilde{\mathbf{e}}_k^{(i)}, \tilde{\mathbf{e}}_l^{(i)} \right) \right)}{\sum_t \exp \left(\text{Attn} \left(\tilde{\mathbf{e}}_k^{(i)}, \tilde{\mathbf{e}}_t^{(i)} \right) \right)} \quad (9)$$

然后,利用门控循环神经网络(Gate Recurrent

Unit, GRU) 对节点状态进行更新:

$$\mathbf{h}_k^{(i)} = GRU\left(\tilde{\mathbf{e}}_k^{(i)}, \left[\mathbf{h}_{IN_k}^{(i)}; \mathbf{h}_{OUT_k}^{(i)}\right]\right) \quad (10)$$

至此, 得到更新后的实体图节点向量表示:

$$\mathbf{h}_E^{(i)} = \left[\mathbf{h}_1^{(i)}, \mathbf{h}_2^{(i)}, \dots, \mathbf{h}_n^{(i)}\right] \quad (11)$$

此外, 会话式机器理解任务中, 不同轮次的问答对之间具有一定的承接关系。而图作为一种描述数据节点之间关系的数据结构, 本身并不含有图之间的顺序依赖关系。因此, 本文设计了类似循环神经网络 (Recurrent Neural Network, RNN) 的信息融合机制, 将之前的推理过程的结果融入当前的推理过程中, 补充了回答当前问题所需要的额外信息。具体计算方式如下所示:

$$\tilde{\mathbf{h}}_E^{(i)} = Fuse\left(\mathbf{h}_E^{(i)}, \mathbf{h}_E^{(i-1)}\right) \quad (12)$$

$$Fuse\left(\mathbf{h}_E^{(i)}, \mathbf{h}_E^{(i-1)}\right) = \mathbf{z} \times \mathbf{h}_E^{(i)} + (1 - \mathbf{z}) \times \mathbf{h}_E^{(i-1)} \quad (13)$$

$$\mathbf{z} = \sigma\left(\mathbf{W}_z \left[\mathbf{h}_E^{(i)}; \mathbf{h}_E^{(i-1)}; \mathbf{h}_E^{(i)} \times \mathbf{h}_E^{(i-1)}; \mathbf{h}_E^{(i)} - \mathbf{h}_E^{(i-1)}\right] + \mathbf{b}_z\right) \quad (14)$$

其中, 融合函数是两个信息源的门控和; σ 为 Sigmoid 激活函数; \mathbf{W}_z 是可学习参数矩阵; \mathbf{z} 为门控函数向量; \mathbf{b}_z 为偏置向量。最后, 将更新后的图节点向量 $\tilde{\mathbf{h}}_E^{(i)}$ 和文本表示向量 $\mathbf{T}_C^{(i)}$ 进行融合, 得到一个包含序列化结构特征和图结构特征的统一嵌入表示 $\tilde{\mathbf{T}}_C^{(i)}$ 。具体是通过语义融合模块所描述的相同的二进制矩阵 \mathbf{M} 将对应位置的嵌入表示向量进行融合, 如公式 (15) 所示:

$$\tilde{\mathbf{T}}_C^{(i)} = Fuse\left(\mathbf{T}_C^{(i)}, \mathbf{M}^T \tilde{\mathbf{h}}_E^{(i)}\right) \quad (15)$$

3.2.3 答案预测层

在问题和文章充分交互融合之后, 将最后的编码表示输入一个全连接层中来预测答案区间, 输出为序列中每个位置作为开始位置或者结束位置的概率。本文采用与 FlowQA 模型^[16]相同的答案预测模块, 先计算答案的起始位置 $P_{i,j}^S$, 再计算答案的终止位置 $P_{i,j}^E$, 这种方式可以更好地联合

答案开始与结尾的关系。同时, 由于答案中存在 “Yes” “No” “No Answer” 和 “Answer Span” 4 种类型, 因此需设计一个答案类型分类器来对答案的类型进行预测。具体计算公式如下所示:

$$P_{i,j}^S \propto \exp\left(\left(\tilde{\mathbf{T}}_{C_{i,j}}\right)^T \mathbf{W}_S \mathbf{T}_{Q_i}\right) \quad (16)$$

$$\tilde{\mathbf{T}}_{Q_i} = GRU\left(\mathbf{T}_{Q_i}, \sum_{i,j} P_{i,j}^S \tilde{\mathbf{T}}_{C_{i,j}}\right) \quad (17)$$

$$P_{i,j}^E \propto \exp\left(\left(\tilde{\mathbf{T}}_{C_{i,j}}\right)^T \mathbf{W}_E \tilde{\mathbf{T}}_{Q_i}\right) \quad (18)$$

$$P_i^T \propto \exp\left(\left[\frac{1}{m} \sum_j \tilde{\mathbf{T}}_{C_{i,j}}; \max_j \left(\tilde{\mathbf{T}}_{C_{i,j}}\right)\right]^T \mathbf{W}_T \tilde{\mathbf{T}}_{Q_i}\right) \quad (19)$$

其中, $\mathbf{W}_S, \mathbf{W}_E, \mathbf{W}_T$ 是 $d \times d$ 的可学习参数矩阵; $\tilde{\mathbf{T}}_{C_{i,j}}$ 为第 i 轮问题中文章的第 j 个位置的嵌入表示向量; $P_{i,j}^S$ 为第 i 轮问题中文章的第 j 个位置为起始位置的概率; $P_{i,j}^E$ 为第 i 轮问题中文章的第 j 个位置为终止位置的概率; P_i^T 为第 i 轮问题的答案类型; $GRU(\cdot)$ 为门控循环神经网络; $\exp(\cdot)$ 是以自然常数 e 为底的指数函数。

4 结果分析与评估

4.1 数据集与评价指标

本文使用多轮机器阅读理解数据集 CoQA^[7] 和 QuAC^[13] 作为实验数据。CoQA 是第一个高质量会话式机器理解数据集——来自 7 个不同的领域, 共计 8 399 个文章段落, 12.7 万个对话。每个文章段落平均约有 15.2 轮问答, 文章平均长度达 271 个词。QuAC 数据集提取自维基百科, 包含 13 594 个会话以及 98 407 个问题, 答案平均长度约达 15 个词且为跨距文本。与 CoQA 不同, QuAC 考虑了现实中师生问答的情况。学生试图通过一系列的问题来了解隐藏的知识, 而教师用简短的段落和所有的会话来回答相应问题。

对于会话式机器阅读理解，本文沿用机器阅读理解中使用的 F_1 值来作答案评价。相较于精确率、召回率， F_1 值更能准确评价一个模型的好坏。 F_1 值的计算如公式(20)所示：

$$F_1 = \frac{2 \times \text{精确率} \times \text{召回率}}{\text{精确率} + \text{召回率}} \quad (20)$$

QuAC 的评价方法除了 F_1 值以外，还引入了人类等效分数 (HEQ) 以计算模型 F_1 值超过或等于人类 F_1 值的样本数百分比。 HEQ 包括 $HEQ-Q$ 和 $HEQ-D$ ： $HEQ-Q$ 计算所有会话中正确回答问题的百分比； $HEQ-D$ 计算所有会话中每一组会话中正确回答问题的百分比。

4.2 实验结果

本文将所提出的 DCGN 模型分别基于 BERT 预训练模型与基于 RoBERTa 预训练模型进行实验，选取目前主流的会话式阅读理解模型作为基线模型与本文所提模型进行对比，对比的基线模型有 FlowQA^[16]、FlowDelta^[26]、SDNet^[27] 和 GraphFlow^[17]。此外，本文亦对其他基于 BERT 的模型进行了比较，如 BERT+HAE^[15]、RoBERTa^[22] 等。实验结果如表 1、表 2 所示，本文模型优于各种先进的基线模型。

4.3 讨论与分析

本文将对模型分为基于 BERT 预训练模型与基于 RoBERTa 预训练模型。从表 1 可以观察

表 2 模型在 QuAC 测试集的评价值

Table 2 Evaluation results on the QuAC test set

模型	F_1 (%)	$HEQ-Q$ (%)	$HEQ-D$ (%)
BERT+HAE	62.4	57.8	5.1
FlowQA	64.1	59.6	5.8
GraphFlow	64.9	60.3	5.1
FlowDelta	65.5	61.0	6.9
RoBERTa	67.1	62.8	7.8
RoBERTa-DCGN	69.1	64.3	9.0

到，DCGN 模型在 CoQA 数据集上显著超过了最先进的基于注意力的模型。与同样基于图网络思想的 GraphFlow 模型相比，本文模型可以更好地学习实体之间的语义关系与句子之前的句法结构等信息，从而解决模型对问题和文章理解不充分的问题。同样，从表 2 可以看出，DCGN 模型在 QuAC 数据集上优于其他方法。此外，本文将 DCGN 模型与基于 RoBERTa 预训练模型进行了比较，证明了本文模型与基于 RoBERTa 预训练模型同样具有竞争力，且更容易集成到现有模型中。为了证明模型各模块对提升实验效果的有效性，在验证集上进行模型消融实验来验证不同模块对于模型效果的影响。本文对 DCGN 模型中重要模块进行删除或者替换实验，结果如表 3 所示。

由表 3 结果可知，删除实体图融合模块，

表 1 模型在 CoQA 测试集上各领域的 F_1 值

Table 1 F_1 values of the model in each field on the CoQA test set

模型	儿童故事 (%)	文学 (%)	中学英语 (%)	新闻 (%)	维基 (%)	社交 (%)	科学文章 (%)	总分 (%)
FlowQA	73.7	71.6	76.8	79.0	80.2	67.8	76.1	75.0
SDNet	75.4	73.9	77.1	80.3	83.1	69.8	76.8	76.6
GraphFlow	77.1	75.6	77.5	79.1	82.5	70.8	78.4	77.3
FlowDelta	75.9	75.6	80.1	82.1	82.3	69.8	78.8	77.7
BERT w/2-contex	76.0	77.0	80.5	82.1	83.0	72.5	79.6	78.7
BERT-DCGN	78.5	76.2	80.1	82.8	83.3	71.5	78.9	79.2
RoBERTa	80.3	80.8	83.3	84.1	82.6	75.5	80.6	81.6
RoBERTa-DCGN	82.2	82.7	85.6	85.9	85.1	77.6	83.1	84.3

表 3 模型在 CoQA 测试集上消融实验的 F_1 值Table 3 The model ablation experimental F_1 values on the

CoQA test set

模型	F_1 (%)
DCGN	84.3
去掉实体图融合模块	82.1
去掉语义融合模块	80.9
去掉动态图神经网络模块	81.6

将文章中的单词作为图节点, 模型的效果下降了 2.2%。由此可见, 使用实体构造的图结构能够提取文本中的有效信息, 避免冗杂信息的引入。删除语义融合模块, 利用图嵌入模型获得每个图节点的向量表示, 模型的效果下降了 3.4%, 究其原因是无法匹配节点与文章信息, 造成图信息与文本信息的不匹配问题。删除动态图神经网络模块, 模型的效果下降了 2.7%, 只利用 BERT 预训练模型对答案进行预测, 缺少基于图结构的机器阅读理解能够利用图结构构建全局语义关系以及较长距离推理关系的优势。

5 结 论

本文针对会话式机器阅读理解任务提出了一种新的基于动态图神经网络的机器阅读理解模型 (DCGN), 该模型首先识别上下文中的实体, 使用句法结构与句子之间的语义关系进行建模; 然后将通过文本图结构得到的上下文实体嵌入表示融合到基于序列化结构得到的上下文嵌入表示中; 最终使用图神经网络在每轮对话过程中动态地构建问题和会话历史的推理图, 从而实现会话式机器阅读理解问答。与之前的方法相比, 本文提出的模型获得了较好的结果。然而, 会话式机器阅读理解领域仍处于最初的研究探索阶段, 模型的推理能力和会话理解能力还有巨大的提升空间。未来, 在机器阅读理解可解释方法上进行研究, 从会话历史筛选出与当前问题最相关

的历史信息, 并且生成可解释的推理路径。

参 考 文 献

- [1] 顾迎捷, 桂小林, 李德福, 等. 基于神经网络的机器阅读理解综述 [J]. 软件学报, 2020, 31(7): 2095-2126.
Gu YJ, Gui XL, Li DF, et al. Survey of machine reading comprehension based on neural network [J]. Journal of Software, 2020, 31(7): 2095-2126.
- [2] Hermann KM, Kocisky T, Grefenstette E, et al. Teaching machines to read and comprehend [C] // Proceedings of the 28th International Conference on Neural Information Processing Systems, 2015: 1693-1701.
- [3] Hill F, Bordes A, Chopra S, et al. The goldilocks principle: reading children's books with explicit memory representations [Z/OL]. arXiv Preprint, arXiv: 1511.02301, 2015.
- [4] Lai GK, Xie QZ, Liu HX, et al. Race: large-scale reading comprehension dataset from examinations [Z/OL]. arXiv Preprint, arXiv: 1704.04683, 2017.
- [5] Rajpurkar P, Zhang J, Lopyrev K, et al. Squad: 100,000+ questions for machine comprehension of text [Z/OL]. arXiv Preprint, arXiv: 1606.05250, 2016.
- [6] 李堃, 李艳玲, 林民. 会话式机器阅读理解综述 [J]. 计算机科学与探索, 2021, 15(9): 1607-1618.
Li K, Li YL, Li M. Review of conversational machine reading comprehension [J]. Journal of Frontiers of Computer Science and Technology, 2021, 15(9): 1607-1618.
- [7] Reddy S, Chen D, Manning CD. CoQA: a conversational question answering challenge [J]. Transactions of the Association for Computational Linguistics, 2019, 7: 249-266.
- [8] Devlin J, Chang MW, Lee K, et al. Bert: pre-training of deep bidirectional Transformers for language understanding [Z/OL]. arXiv Preprint, arXiv: 1810.04805, 2018.
- [9] Yang ZL, Dai ZH, Yang YM, et al. X1Net: generalized autoregressive pretraining for language understanding [C] // Proceedings of the 33rd

- International Conference on Neural Information Processing Systems, 2019: 5753-5763.
- [10] Liu YH, Ott M, Goyal N, et al. RoBERTa: a robustly optimized bert pretraining approach [Z/OL]. arXiv Preprint, arXiv: 1907.11692, 2019.
- [11] Seo M, Kembhavi A, Farhadi A, et al. Bidirectional attention flow for machine comprehension [Z/OL]. arXiv Preprint, arXiv: 1611.01603, 2016.
- [12] Yu AW, Dohan D, Luong MT, et al. QANet: combining local convolution with global self-attention for reading comprehension [Z/OL]. arXiv Preprint, arXiv: 1804.09541, 2018.
- [13] Choi E, He H, Iyyer M, et al. QuAC: question answering in context [Z/OL]. arXiv Preprint, arXiv: 1808.07036, 2018.
- [14] Zhu C, Zeng M, Huang X. SDNet: contextualized attention-based deep network for conversational question answering [Z/OL]. arXiv Preprint, arXiv: 1812.03593, 2018.
- [15] Ohsugi Y, Saito I, Nishida K, et al. A simple but effective method to incorporate multi-turn context with BERT for conversational machine comprehension [Z/OL]. arXiv Preprint, arXiv: 1905.12848, 2019.
- [16] Huang HY, Choi E, Yih W. FlowQA: grasping flow in history for conversational machine comprehension [Z/OL]. arXiv Preprint, arXiv: 1810.06683, 2018.
- [17] Chen Y, Wu L, Zaki MJ. GraphFlow: exploiting conversation flow with graph neural networks for conversational machine comprehension [Z/OL]. arXiv Preprint, arXiv: 1908.00059, 2019.
- [18] Scarselli F, Gori M, Tsoi AC, et al. The graph neural network model [J]. IEEE Transactions on Neural Networks, 2008, 20(1): 61-80.
- [19] Qiu L, Xiao YX, Qu YR, et al. Dynamically fused graph network for multi-hop reasoning [C] // Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 2019: 6140-6150.
- [20] Thayaparan M, Valentino M, Schlegel V, et al. Identifying supporting facts for multi-hop question answering with document graph networks [Z/OL]. arXiv Preprint, arXiv: 1910.00290, 2019.
- [21] Ding M, Zhou C, Chen QB, et al. Cognitive graph for multi-hop reading comprehension at scale [Z/OL]. arXiv Preprint, arXiv: 1905.05460, 2019.
- [22] Liu Y, Ott M, Goyal N, et al. RoBERTa: a robustly optimized bert pretraining approach [Z/OL]. arXiv Preprint, arXiv: 1907.11692, 2019.
- [23] Huang Z, Xu W, Yu K. Bidirectional LSTM-CRF models for sequence tagging [Z/OL]. arXiv Preprint, arXiv: 1508.01991, 2015.
- [24] Schmitt X, Kubler S, Robert J, et al. A replicable comparison study of NER software: StanfordNLP, NLTK, OpenNLP, SpaCy, Gate [C] // Proceedings of the 2019 Sixth International Conference on Social Networks Analysis, Management and Security, 2019: 338-343.
- [25] Dozat T, Manning CD. Deep biaffine attention for neural dependency parsing [Z/OL]. arXiv Preprint, arXiv: 1611.01734, 2016.
- [26] Yeh YT, Chen YN. FlowDelta: modeling flow information gain in reasoning for conversational machine comprehension [Z/OL]. arXiv Preprint, arXiv: 1908.05117, 2019.
- [27] Zhu C, Zeng M, Huang X. SDNet: contextualized attention-based deep network for conversational question answering [Z/OL]. arXiv preprint, arXiv: 1812.03593, 2018.