

基于深度学习的机器阅读理解综述

李舟军 王昌宝

(北京航空航天大学计算机学院 北京 100191)

摘 要 阅读理解能力是人类智能中最关键的能力之一,而机器阅读理解作为自然语言处理领域皇冠上的明珠,一直是该领域的研究焦点。近年来,随着深度学习方法的快速发展,机器阅读理解技术获得了长足的进步。首先,对基于深度学习的机器阅读理解技术的研究背景和发展历史进行了概述;然后,详细介绍了词向量、注意力机制以及答案预测这三大关键技术的研究进展;在此基础上,分析了目前机器阅读理解研究所面临的问题;最后,对机器阅读理解技术的未来发展趋势进行了展望。

关键词 自然语言处理,机器阅读理解,深度学习,词向量,注意力机制

中图法分类号 TP391 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2019.07.002

Survey on Deep-learning-based Machine Reading Comprehension

LI Zhou-jun WANG Chang-bao

(School of Computer Science and Engineering, Beihang University, Beijing 100191, China)

Abstract Natural language processing is the key to achieving artificial intelligence. Machine reading comprehension, as the crown jewel in the field of natural language processing, has always been the focus of research in the field. With the rapid development of deep learning and neural network in recent years, machine reading comprehension has made great progress. Firstly, the research background and development history of machine reading comprehension were introduced. Then, by reviewing the important progress in the development of word vector, attention mechanism and answer prediction, the problems in recent research related to machine reading comprehension were proposed. Finally, the outlook of machine reading comprehension was discussed.

Keywords Natural language processing, Machine reading comprehension, Deep learning, Word vector, Attention mechanism

1 引言

近年来,深度学习技术快速发展,并已在图像识别、语音识别和围棋等领域超越了人类水平。自然语言处理(Natural Language Processing, NLP)是实现人工智能的核心技术。在自然语言处理领域,深度学习技术也在机器翻译、人机对话等方面取得了重大突破。机器阅读理解(Machine Reading Comprehension, MRC)作为自然语言处理领域的核心技术,借助于深度学习技术,也获得了快速发展。

顾名思义,机器阅读理解就是让机器学会阅读和理解文章,即对于给定的问题,从相关文章中寻找答案。MRC 涉及到语言理解、知识推理、摘要生成等复杂技术,因而面临着诸多挑战。

传统的 MRC 技术往往采用基于人工规则的模式匹配方法,或者借鉴信息抽取的方法构造关系数据库来寻找答案。这些方法效率低下,且准确率不高。随着深度学习技术的日

趋成熟,以及各种大规模阅读理解数据集(如 CNN/Daily^[1], SQuAD^[2], MS MARCO^[3]等)的发布,基于深度学习的 MRC 技术取得了巨大的进步。研究 MRC 技术,对推动智能搜索和智能客服等人工智能应用的发展具有重要的现实意义。

在智能搜索方面,MRC 技术被认为是下一代搜索引擎的核心技术。目前的搜索引擎技术主要依靠用户提供的关键词,从海量的互联网数据中搜索相关性较大的网页,用户需要在返回的诸多网页中进一步遴选答案。为了得到期望的答案,用户在搜索过程中往往需要花费较多的精力,并需要掌握一定的搜索技巧。将 MRC 应用于搜索引擎,将会极大地改善用户体验,引领搜索引擎技术的变革。针对用户输入的问题,基于 MRC 技术的搜索引擎会对整个互联网文档进行阅读理解,直接为用户返回更加精准和智能的答案,为用户节省了从众多网页中寻找答案的精力和时间。

在智能客服方面,MRC 正在取代人工规则,赋予“智能客服”真正的智能。在电商领域,“阿里小蜜”为用户解读交易规

则,“店小蜜”为用户提供商品售后服务咨询。基于阅读理解技术的“智能客服”,可以快速且准确地为用户提供答案,极大地提高服务效率。

2 发展历史

1999 年,Hirschman 等^[4]最早开始了 MRC 技术的研究。Hirschman 使用小学 3 年级到 6 年级的阅读材料构造了一个小规模 MRC 测试集,再用词袋模型分别对问题和文章中的每条语句进行信息抽取和模式匹配,然后从文章中选取与问题匹配度最高的语句作为问题的答案。Riloff 等^[5]则通过人工制定规则,对文章中的候选语句和问题进行匹配度打分,然后选取得分最高的候选语句作为答案。Poon 等^[6]使用关系数据库将海量文本信息进行结构化表示,构建了一个能够进行联合推断和持续学习的阅读理解系统。

基于传统特征的 MRC 技术虽然取得了一定的进展,但仍然存在一些问题需要解决。传统的 MRC 技术大多采用模式匹配的方法进行特征提取,因而不能有效处理表达的多样性问题;同时,由于匹配时往往使用固定窗口,因此其无法解决多个句子之间的长期依赖问题。

为了弥补传统 MRC 技术的缺陷,研究人员开始考虑采用深度神经网络进行机器阅读理解的研究。使用深度神经网络的前提是需要大规模的训练语料,早期的数据集,如 MCT-est^[7],数据规模太小,因而只能用于测试,并不能用于训练。Hermann 等^[1]提出了一种高效的生成大规模训练数据的方法,并构造了 CNN 和 Daily Mail 两个大规模数据集。区别于传统方法,Hermann 提出了基于注意力机制的深度神经网络模型,该模型在 CNN 和 Daily Mail 上取得了远超传统方法的结果。在此基础上,Kadlec^[8]使用点积简化了 Hermann 的注意力模型,并获得了更好的效果。Chen^[9]则使用双线性函数作为激活函数来获得更加灵活、有效的注意力模型。区别于只在文章上使用注意力机制的方法,Cui^[10]通过在文章和问题上相互使用注意力机制,增强了模型的特征提取能力。Dhingra^[11]通过加深网络的深度,并在层与层之间增加门限来保留重要信息,以达到多步推理的效果。

2016 年,斯坦福发布了大规模数据集 SQuAD^[2]。不同于 CNN/Daily Mail 的答案是原文中的实体词,SQuAD 数据集的答案是原文的一个连续片段,因而更具有挑战性。SQuAD 发布之后,极大地推动了 MRC 的快速发展。Wang 等^[12-13]通过注意力机制将问题融合到文章表示中,接着利用指针网络^[14]预测答案序列在文章中的开始位置和结束位置。Seo 等^[15]通过多阶段信息对文章和问题进行编码,并通过双向注意力机制来提升模型性能。Xiong 等^[16]提出了动态迭代和双路注意力机制,可以同时为文章和问题使用注意力机制,并通过不断迭代预测结果来提升性能。Wang 等^[17]则提出了一个多层网络结构 R-Net,分别从 4 个层面对 MRC 任务进行建模,率先超越了人类水平。

SQuAD 发布后不久,微软亚洲研究院便发布了来源于真实应用场景的数据集 MS MARCO^[3]。MS MARCO 数据集的所有文章和问题全部来源于必应搜索,每个问题对应 10 篇候选文章,问题的答案由人工生成,而不再是原文的连续片段。类似的数据集还包括 TriviaQA^[18]和百度发布的中文数

据集 DuReader^[19]。这类数据集要求模型能从多篇候选文章中排除干扰项,并具备生成答案的能力。Tan^[20]在 R-Net 的基础上增加了文章排名算法和答案生成模块,性能显著优于答案抽取模型的性能。Clark 等^[21]通过使用跨文章的全局归一化方法,增强了不同候选答案得分之间的可比性。Wang 等^[22]则使用注意力机制对多个候选答案进行交叉验证,以实现跨文章的信息交流。

Kociskyd 等^[23]分析了主流的 MRC 数据集,发现现有数据集的问题大多比较肤浅,并且答案往往来源于局部上下文而非全局信息,于是发布了更具挑战的数据集 NarrativeQA。该数据集的语境不再是一篇文章,而是一整部小说,要求模型在理解小说内容的基础上回答问题,对模型的理解能力和推理能力提出了更大的挑战。

3 基于深度学习的机器阅读理解

如图 1 所示,基于深度学习的阅读理解模型主要包括词向量模块、编码模块、注意力模块与答案预测模块。

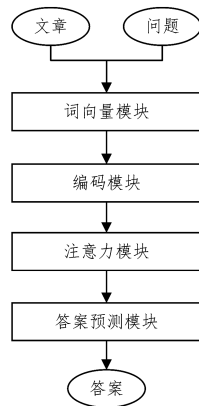


图 1 深度机器阅读理解模型的基本结构

Fig. 1 Basic architecture of deep MRC model

1)词向量模块:与其他自然语言处理任务一样,在 MRC 任务中,第一步就是要将所有单词映射到一个向量空间,使得每个单词具有唯一的向量表示。该映射要能反映单词的语法和语义信息,以及词与词之间(诸如同义、反义等)的关系。

2)编码模块:词向量模块得到的是每个单词的独立表示,并不包含上下文信息。编码模块则以词向量表示的文本序列作为输入,通过单层或多层的双向循环神经网络(Bidirectional Recurrent Neural Network, BiRNN)和双向长短期记忆(Bidirectional Long Short Term Memory, BiLSTM)对输入的文本序列进行特征提取,提取的特征具有更丰富的上下文信息和语义信息。

3)注意力模块:对于一篇文章而言,并非每个位置的单词都同等重要,与问题相关的往往是文章的一部分。注意力机制可以从文章中挑选出与问题关联度最大的部分内容,排除不相关信息,使模型将注意力放在文章的关键部分。

4)答案预测模块:随着 MRC 技术的发展,越来越多的数据集被发布出来,这些数据集也越来越具有挑战性。答案从早先的抽取式(答案是文章的一个子片段)演变为生成式(答案是由人工生成的任意语句),问题的答案从由单篇文章提供发展到由多篇文章综合生成。如何从给定的输入数据中提取答案,也成为 MRC 技术的重点研究内容。

当前方法中的编码模块广泛使用 BiLSTM 进行编码,结构比较单一;下文将主要从词向量模块、注意力模块和答案预测模块角度介绍 MRC 技术的重要进展。

3.1 词向量模块

几乎所有 NLP 任务的第一步都是将单词进行向量表示,词的向量表示问题一直都是 NLP 领域研究的热点问题。在 MRC 任务中,词向量的表示也尤其重要,使用高质量的词向量能极大提升 MRC 模型的性能。目前的词向量研究主要基于词的相似性分布理论,即一个词可以通过其所在的上下文表示出来。根据词向量的生成方式,可以将词向量的研究方法分为:矩阵分解法,参数学习法和上下文学习法。

3.1.1 矩阵分解法

从相似性分布理论出发,共现频率矩阵 M 是表示一个词最简单直观的方式,其中 $M_{i,j}$ 表示单词 i 在上下文 j 中出现的次数。根据需要,上下文可以是整篇文档,也可以是任意长度的窗口。但直接用共现矩阵表示词向量存在诸如维数灾难和稀疏性等问题,因此更有效的做法是对共现矩阵进行奇异值分解,并合成低维度矩阵用于词向量表示。这种低维度、高密度的词向量表示有效解决了维数灾难和稀疏性问题,被广泛应用。Deerwester 等^[24]将“词-文档”共现矩阵分解,用于潜在语义分析。Lund 等^[25]则利用“词-词”共现矩阵构建语义存储模型。Rohde 等^[26]提出了基于熵的归一化方法,进一步提高了词向量的可靠性和稳定性。

但基于矩阵分解的表示方法仍然存在明显的缺点:对于大规模语料,对其共现矩阵进行分解时需要耗费庞大的计算量;另外,对语料进行增量式更新时,共现矩阵的奇异值分解要重新计算。

3.1.2 参数学习法

矩阵分解法简单直观,但是存在计算量大等诸多缺点,于是研究人员提出了直接学习词向量表示的方法。

Bengio 等^[27]最早构建了基于神经网络的概率语言模型,并将词向量作为语言模型的参数进行学习。Bengio 将每个单词表示为一个低维度的特征向量,再利用词的特征向量构造语言模型的联合概率密度,然后通过优化联合概率密度函数不断学习词的特征向量。Mikolov^[28-29]延续了这一工作,并提出了更加简单高效的 word2vec 模型,具体包括连续词袋(Continuous Bag-of-Word, CBOW)模型和 Skip-gram 模型。word2vec 模型没有使用神经网络,而是借鉴了 N-gram 的思想,对固定窗口的局部上下文进行建模。word2vec 由于训练高效,并且得到的词向量具有线性相关等优秀的性质而被广泛应用;但其并没有考虑全局信息。Pennington 等^[30]将共现矩阵引入联合概率函数,有效地将全局的统计信息和局部上下文信息进行结合。

将词向量作为参数进行学习的方法有效解决了维数灾难和计算量的问题,但仍然没有解决一词多义的问题。例如,“我喜欢吃苹果”和“我喜欢用苹果手机”两个语句中,“苹果”具有不同的含义。产生这一问题的原因在于没有考虑词汇所处的上下文信息。对上下文信息的有效利用,是解决一词多义问题的关键。

3.1.3 上下文学习法

参数学习法是词向量的一种静态的表示方法,即同一个

词汇不论处于何种语境都具有相同的词向量表示,因而无法解决一词多义等问题。解决一词多义问题的关键在于有效利用上下文语境,即输入不再是单独的某个词,而是该词所处的上下文语境,然后获得该词在当前语境下的向量表示。这是一种动态的词向量表示方法,词向量不再是静态的模型参数,而是动态的模型输出。Melamud 等^[31]用 BiLSTM 取代 word2vec 模型中的词向量矩阵,提出用 BiLSTM 对上下文进行编码,从而获得词向量的表示。McCann 等^[32]则先预训练一个神经机器翻译模型,然后用编码器模块来生成词向量。该方法对包括 MRC 在内的诸多任务带来了显著的性能提升,但受平行语料规模的限制。Peters 等^[33]则提出了 ELMo (Embeddings from Language Models) 词向量表示法,通过使用双向语言模型进行预训练,然后将中间结果融合作为词向量的表示,摆脱了平行语料的限制。如表 1 所列,通过在基线模型上使用 ELMo 词向量,模型的 $F1$ 值提升了 4.7%,效果显著。

表 1 SQuAD 测试集上 ELMo 方法与基线模型的对比

Table 1 Comparison of baseline and ELMo method on SQuAD test set

Model	F1
BIDAF ^[15] +self-attention	81.1
BIDAF+self-attention+ELMo ^[33]	85.8

3.2 注意力机制

注意力机制很早就被提出^[34],但最早的成功应用是在视觉图像领域,Mnih 等^[35]在循环神经网络(Recurrent Neural Network, RNN)模型上使用了注意力机制来进行图像分类。随后,Bahdanau 等^[36]使用注意力机制在机器翻译任务上将翻译和对齐同时进行,第一次将注意力机制应用到 NLP 领域中。

在处理阅读理解问题时,基于神经网络的模型几乎都直接或者间接使用了注意力机制。简单来说,就是在编码或者解码的阶段,为序列的每个位置生成一个权重。权重大小代表了当前位置对最终预测结果的重要性。权重的计算有各种形式,但需要综合各方面的信息。

在 MRC 任务中,根据注意力机制的结构,可以将注意力模型分为单路注意力模型、双路注意力模型和自匹配注意力模型。

3.2.1 单路注意力模型

单路注意力模型被用来模拟人做阅读理解时所经历的过程——带着问题去阅读,即先看问题,再去文本中有目标地阅读以寻找答案。通过结合问题和文本段落二者的信息,生成一个关于文本段落各部分的注意力权重,对文本信息进行加权,该注意力机制可以更好地捕捉文本段落中与问题相关的信息。

Hermann 等^[1]最先将基于注意力机制的深度神经网络模型用于 MRC 任务,即以问题为上下文,计算文章中每个单词的权重,并生成文章的最终表示。在此基础上,Kadlec^[8]和 Chen^[9]等分别对注意力函数做了一些改进,但它们在本质上并没有摆脱单路注意力模型的框架。

3.2.2 双路注意力模型

不同于单路注意力模型只在文章上使用注意力机制,双路注意力模型同时在文章和问题上使用注意力机制。如图 2

所示,这类方法一般先通过编码器分别获得文章序列 d 和问题序列 q ,然后生成二维相似度矩阵 $A^{|d| \times |q|}$,其中 $|d|$ 表示序列 d 的长度。按行计算注意力,得到的是文章单词 d_i 和问题 q 中每个单词的相似度;按列计算注意力,得到的是问题单词 q_i 和文章 d 中每个单词的相似度。双路注意力模型主要研究如何从二维相似度矩阵中获取文章的整体表达。Cui 等^[37-38]按列进行注意力计算,获得“文章-问题”注意力矩阵;然后采用启发式规则,按行进行合并,得到文章的注意力。Seo 等^[15]则是首先对相似度矩阵按列使用注意力机制,得到问题的注意力表示,再按行使用注意力机制,得到文章的注意力表示,然后将二者进行融合,得到文章的最终表示。

双路注意力模型由于对文章和问题之间的相似度进行了更加细粒度的建模,因此包含更加丰富的细节信息,其性能往往比单路注意力模型好。

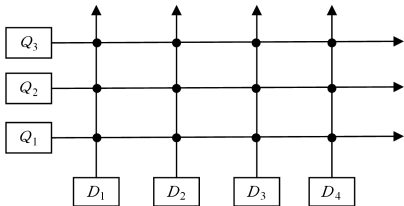


图2 问题-文章相似度矩阵
Fig. 2 Question-document similarity matrix

3.2.3 自匹配注意力模型

上述两种注意力机制可以被认为是带着问题去阅读文章,或是带着文章去阅读问题。这种注意力机制过于依赖经验信息,导致只重视文章中与问题相关度较高的局部信息,丢失了文章本身强调的重点信息。自匹配注意力模型则是反复阅读文章并找出重点。

基于这一点,Wang 等^[17]提出了单路注意力模型和自匹配注意力模型结合的网络结构,Clark 等^[21]则是通过在双路注意力模型之后加入自匹配注意力模块,使模型性能获得了进一步提升。

注意力模块是机器阅读理解模型的核心模块,对提升模型的性能至关重要。如表 2 所列,从单路注意力机制到双路注意力机制,再发展到自匹配注意力机制,机器阅读理解模型的性能获得了持续提升。

表2 SQuAD 测试集上不同注意力模型的对比
Table 2 Comparison of different attentive models on SQuAD test set

Model	EM	F1
Match-LSTM ^[12]	64.7	73.7
BIDAF ^[15]	68.0	77.3
R-Net ^[17]	72.3	80.7

3.3 答案预测

从早期的从原文中抽取一个单词或者子片段,发展到阅读多篇候选文章生成答案,再发展到阅读整篇小说生成答案,MRC 任务越来越具有挑战性。根据答案的不同产生方式,答案预测方法可分为答案抽取和答案生成。

3.3.1 答案抽取

答案抽取是指从给定的文章中挑选出关键片段,然后生成答案。在数据集 SQuAD^[2]中,每个问题的答案均是原文的一个子片段,因而是典型的答案抽取式数据集。由于输入的

文章长度是变化的,大部分阅读理解模型广泛地使用了指针网络进行答案抽取。如图 3 所示,指针网络的输出是输入序列的某个位置,因而可以用来预测答案中每个词语在文章中的位置。基于指针网络,Wang 等^[39]提出了两种答案抽取模型:序列模型和边界模型。

1)序列模型:利用指针网络预测答案中的每个词在原文中的位置,然后将对应位置的词汇取出并按顺序组合成答案。同时,通过引入停止符来控制生成答案的长度。

2)边界模型:边界模型假定答案是原文的一个连续子序列,因而可以通过预测答案的开始位置和结束位置对答案进行定位。

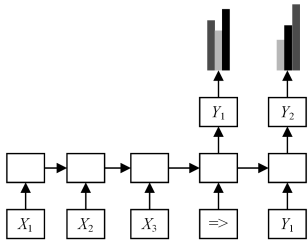


图3 指针网络的基本结构
Fig. 3 Basic architecture of Pointer Network

相比较而言,序列模型比较灵活,不要求答案是原文的一个连续子序列。但是,由于序列模型的答案是由多个词汇组合而成,因此其在语法规范和语义表达方面明显弱于边界模型。

3.3.2 答案生成

不同于 SQuAD 数据集,微软亚洲研究院发布的数据集 MS MARCO^[3]由人工阅读多篇候选文章之后进行总结,然后生成答案。类似的数据集还有百度提出的中文数据集 Du-Reader。这类数据集的答案不再受限于文章片段,要求模型具有生成答案的能力。Tan 等^[20]尝试通过答案抽取模型从候选文章中提取片段作为线索,然后用机器翻译模型将线索翻译为答案,并在 MARCO 数据集上获得了比答案抽取模型更优的性能,如表 3 所列。

表3 MARCO 测试集上答案抽取模型 R-Net 与答案生成模型 S-Net 的对比

Table 3 Comparison of answer extraction model R-Net and answer generation model S-Net on MARCO test set

Model	ROUGE-L	BLEU-1
R-Net ^[17]	42.89	42.22
S-Net ^[20]	46.65	44.78

当前 MRC 模型广泛采用边界模型进行答案抽取,答案生成技术只作为一种辅助手段产生答案。如何生成高质量的答案仍是一个亟需解决的问题。

4 MRC 面临的主要问题

通过分析 MRC 关键技术的研究进展可以看出,当前的 MRC 技术在词向量表示、注意力机制和答案预测方面取得了较大的进步;尤其是陆续发布的大规模数据集,在推动 MRC 技术快速发展的同时,也带来了更大的挑战。

4.1 词向量模块仍需改善

目前,使用预训练的语言模型取代预训练的词向量对 MRC 模型带来了显著的性能提升,但基于语言模型的词向量

表示依然有局限性。例如,在 MRC 任务中,为了回答人物轨迹的问题,需要模型具有指代消解的能力,语言模型不具备这种能力。另外,语言模型仅能捕捉到其所见过的信息,对于常识类的信息,仍需借助其他技术。

4.2 模型缺乏推理能力

当前的阅读理解模型并没有形成深度的阅读理解,而是注意某些线索以执行粗浅的模式匹配。无论是语言模型还是注意力模型,都不具备推理能力。语言模型更多地是对语法和语义的建模,注意力模型则是基于浅层语义的相似度计算,而推理能力要求模型能够对高层抽象的语义进行建模。

4.3 模型缺乏外部知识

目前的阅读理解模型所用的信息均来源于文章,而在阅读文章的过程中,常需要结合外部知识来加深对文章的理解。

4.4 答案生成技术研究的不足

当前阅读理解技术广泛地采用了边界模型进行答案预测。边界模型在 SQuAD^[2]数据集上获得了成功,但对 MS MARCO^[3]和 DuReader^[19]等贴近真实应用场景的数据集并不适用。答案生成技术已有一些初步的研究成果,但如何生成高质量的答案仍是一个需要解决的问题。

5 MRC 的发展趋势

通过分析 MRC 关键技术的发展现状,总结当前面临的主要问题,可以看出 MRC 技术需要更深入的研究。

5.1 构建端到端的高效模型

目前,预训练语言模型只是作为独立的特征提取器用于生成上下文相关的词向量表示,并且这一过程无论是训练还是推断都需要耗费较多的时间。考虑到语言模型和阅读理解模型在局部结构和功能上具有相似性,可以将二者进行参数共享,构建一个高效的端到端的阅读理解模型。

5.2 深层结构探索,提高推理能力

为了能够回答复杂的问题,设计有效的推理机制是突破阅读理解能力的关键技术。目前,阅读理解模型的深度较浅,只能提取浅层的语义特征,缺乏推理能力。为了提高模型的推理能力,需要设计更深、更有效的阅读理解模型,利用深层网络提取抽象的高层语义信息。

5.3 与其他 NLP 技术进行结合

MRC 任务涉及到 NLP 领域的许多关键技术,MRC 任务中的一些难题可以通过引入其他技术进行解决。例如,通过引入指代消解技术将文章中的代词关联到真正的实体,以降低推理难度;通过知识库引入外部知识,提供文章相关的背景知识,加强模型对文章的理解。

5.4 答案生成技术的深入研究

目前答案生成技术主要采用机器翻译模型生成答案,模型结构单一,依然需要更加深入的研究。一方面,可以通过构建更大规模的数据集,为生成模型提供更加丰富的语料;另一方面,可以在阅读理解模型中增加语言模型分支,进行多任务训练,增强模型的语言建模能力。

5.5 对多候选文章进行排名

在部分阅读理解数据集中,问题的答案需要从多篇候选文章中综合生成,这些候选文章中有一部分是与问题无关的

噪音。一种简单的解决思路是将多篇文章拼接在一起,然后转化为单篇文章。这一解决思路的问题在于扩大了单篇文章的长度,同时也扩大了答案的解空间,增加了答案预测的难度。另一种解决思路是对多篇候选文章进行排名,从中选出最可能包含答案的文章,然后进行答案预测。

结束语 随着深度学习技术的快速发展,基于深度神经网络的机器阅读理解技术也获得了长足的进步。本文通过回顾 MRC 关键技术的相关进展,分析了 MRC 面临的主要问题,并提出了可能的解决办法,最后对未来的发展趋势进行了展望。

参 考 文 献

[1] HERMANN K M,KOCISKY T,GREFENSTETTE E,et al. Teaching machines to read and comprehend[C]// Advances in Neural Information Processing Systems. Cambridge:MIT Press, 2015:1693-1701.

[2] RAJPURKAR P,ZHANG J,LOPYREV K,et al. SQuAD: 100000+ Questions for Machine Comprehension of Text[C]// Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. Stroudsburg: ACL, 2016: 2383-2392.

[3] NGUYEN T,ROSENBERG M,SONG X,et al. MS MARCO: A human generated machine reading comprehension dataset [DB/OL]. [2018-08-28]. <https://arxiv.org/abs/1611.09268>.

[4] HIRSCHMAN L,LIGHT M,BRECK E,et al. Deep read: A reading comprehension system[C]// Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics. Stroudsburg: ACL, 1999: 325-332.

[5] RILOFF E,THELEN M. A rule-based question answering system for reading comprehension tests[C]// Proceedings of the 2000 ANLP/NAACL Workshop on Reading comprehension tests as evaluation for computer-based language understanding systems-Volume 6. Stroudsburg: ACL,2000:13-19.

[6] POON H,CHRISTENSEN J,DOMINGOS P,et al. Machine reading at the university of washington[C]// Proceedings of the NAACL HLT 2010 First International Workshop on Formalisms and Methodology for Learning by Reading. Association for Computational Linguistics. Stroudsburg: ACL,2010:87-95.

[7] RICHARDSON M,BURGES C J C,RENSHAW E. Mctest: A challenge dataset for the open-domain machine comprehension of text[C]// Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing. Stroudsburg: ACL, 2013:193-203.

[8] KADLEC R,SCHMID M,BAJGAR O,et al. Text understanding with the attention sum reader network [DB/OL]. [2018-08-28]. <https://arxiv.org/abs/1603.01547>.

[9] CHEN D,BOLTON J,MANNING C D. A Thorough Examination of the CNN/Daily Mail Reading Comprehension Task[C]// Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Stroudsburg: ACL,2016:2358-2367.

[10] CUI Y,CHEN Z,WEI S,et al. Attention-over-Attention Neural Networks for Reading Comprehension[C]// Proceedings of the

- 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Stroudsburg: ACL, 2017: 593-602.
- [11] DHINGRA B, LIU H, YANG Z, et al. Gated-Attention Readers for Text Comprehension[C]//Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Stroudsburg: ACL, 2017: 1832-1846.
- [12] WANG S, JIANG J. Machine comprehension using match-lstm and answer pointer [DB/OL]. [2018-08-28]. <https://arxiv.org/abs/1608.07905>.
- [13] WANG S, JIANG J. Learning Natural Language Inference with LSTM[C]//Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Stroudsburg: ACL, 2016: 1442-1451.
- [14] VINYALS O, FORTUNATO M, JAITLY N. Pointer networks [C]//Advances in Neural Information Processing Systems. Cambridge: MIT Press, 2015: 2692-2700.
- [15] SEO, MINJOON, et al. Bidirectional attention flow for machine comprehension [DB/OL]. [2018-08-28]. <https://arxiv.org/abs/1611.01603>.
- [16] XIONG C, ZHONG V, SOCHER R. Dynamic coattention network for question answering; U. S. Patent Application 15/421, 193[P]. 2018-05-10.
- [17] 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). Stroudsburg: ACL, 2017: 189-198.
- [18] JOSHI M, CHOI E, WELD D, et al. TriviaQA: A Large Scale Distantly Supervised Challenge Dataset for Reading Comprehension[C]//Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Stroudsburg: ACL, 2017: 1601-1611.
- [19] HE W, LIU K, LYU Y, et al. DuReader: a Chinese Machine Reading Comprehension Dataset from Real-world Applications [DB/OL]. [2018-08-28]. <https://arxiv.org/abs/1711.05073>.
- [20] TAN C, WEI F, YANG N, et al. S-net: From answer extraction to answer generation for machine reading comprehension [DB/OL]. [2018-08-28]. <https://arxiv.org/abs/1706.04815>.
- [21] CLARK C, GARDNER M. Simple and Effective Multi-Paragraph Reading Comprehension[C]//Meeting of the Association for Computational Linguistics. Stroudsburg: ACL, 2018: 845-855.
- [22] WANG Y, LIU K, LIU J, et al. Multi-Passage Machine Reading Comprehension with Cross-Passage Answer Verification[C]//Meeting of the Association for Computational Linguistics. Stroudsburg: ACL, 2018: 1918-1927.
- [23] KOCISKY T, SCHWARZ J, BLUNSOM P, et al. The NarrativeQA Reading Comprehension Challenge[J]. Transactions of the Association for Computational Linguistics, 2018, 6: 317-328.
- [24] DEERWESTER S, DUMAIS S T, FURNAS G W, et al. Indexing by latent semantic analysis[J]. Journal of the American society for Information Science, 1990, 41(6): 391-407.
- [25] LUND K, BURGESS C. Producing high-dimensional semantic spaces from lexical co-occurrence[J]. Behavior Research Methods, Instruments, & Computers, 1996, 28(2): 203-208.
- [26] ROHDE D L T, GONNERMAN L M, PLAUT D C. An improved model of semantic similarity based on lexical co-occurrence[J]. Communications of the ACM, 2006, 8(627-633): 116.
- [27] BENGIO Y, DUCHARME R, VINCENT P, et al. A neural probabilistic language model[J]. Journal of Machine Learning Research, 2003, 3: 1137-1155.
- [28] MIKOLOV T, SUTSKEVER I, CHEN K, et al. Distributed representations of words and phrases and their compositionality[C]//Advances in Neural Information Processing Systems. Cambridge: MIT Press, 2013: 3111-3119.
- [29] MIKOLOV T, CHEN K, CORRADO G, et al. Efficient estimation of word representations in vector space[J]. arXiv: 1301.3781, 2013.
- [30] PENNINGTON J, SOCHER R, MANNING C. Glove: Global vectors for word representation[C]//Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). Stroudsburg: ACL, 2014: 1532-1543.
- [31] MELAMUD O, GOLDBERGER J, DAGAN I. context2vec: Learning generic context embedding with bidirectional lstm[C]//Proceedings of The 20th SIGNLL Conference on Computational Natural Language Learning. Stroudsburg: ACL, 2016: 51-61.
- [32] MCCANN B, BRADBURY J, XIONG C, et al. Learned in translation: Contextualized word vectors[C]//Advances in Neural Information Processing Systems. Cambridge: MIT Press, 2017: 6294-6305.
- [33] PETERS M, NEUMANN M, IYYER M, et al. Deep Contextualized Word Representations[C]//Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers). Stroudsburg: ACL, 2018: 2227-2237.
- [34] RENSINK R A. The dynamic representation of scenes [J]. Visual cognition, 2000, 7(1/2/3): 17-42.
- [35] MNH V, HEES N, GRAVES A. Recurrent models of visual attention[C]//Advances in Neural Information Processing Systems. Stroudsburg: ACL, 2014: 2204-2212.
- [36] BAHDANAU D, CHO K, BENGIO Y. Neural machine translation by jointly learning to align and translate [DB/OL]. [2018-08-28]. <https://arxiv.org/abs/1409.0473>.
- [37] CUI Y, LIU T, CHEN Z, et al. Consensus Attention-based Neural Networks for Chinese Reading Comprehension[C]//Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers. Pisa: ACM, 2016: 1777-1786.
- [38] CUI Y, CHEN Z, WEI S, et al. Attention-over-Attention Neural Networks for Reading Comprehension[C]//Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Stroudsburg: ACL, 2017: 593-602.
- [39] WANG S, JIANG J. Machine comprehension using match-lstm and answer pointer [DB/OL]. [2018-08-28]. <https://arxiv.org/abs/1608.07905>.