目录

[第1章 绪论 1](#_Toc1944419)

[1.1 选题背景及研究意义 1](#_Toc1944420)

[1.1.1 机器阅读理解的概念 2](#_Toc1944421)

[1.1.2 研究应用价值 3](#_Toc1944422)

[1.2 机器阅读理解的难点 4](#_Toc1944423)

[1.2.1 普遍存在的不确定性 4](#_Toc1944424)

[1.2.2 语言知识处理的复杂性 6](#_Toc1944425)

[1.2.3 输入不规范问题 7](#_Toc1944426)

[1.3 国内外研究现状 8](#_Toc1944427)

[1.4 本文主要工作 10](#_Toc1944428)

[1.5 本文的组织结构 12](#_Toc1944429)

[参考文献 13](#_Toc1944430)

# 绪论

* 1. 选题背景及研究意义

在这个星球上，人类是惟一可以用语言交流的生物。人类会说话使知识的传播成为可能，对人类文明的形成和发展具有重要而深远的影响。进行语言交流，是人类区别于其他动物的最显著的特征。20世纪50年代，数字计算机研制成功，研究者开始探索试图将人类智能简化成符号处理。从那时也诞生了自然语言处理（Natural Language Processing，NLP）这一人工智能和语言学领域的分支学科，力求探讨机器如何处理及运用自然语言达到认知、理解、生成自然语言等目的。1950年，图灵提出现在所谓的“图灵测试[[1]](#footnote-1)”作为判断智能的条件，以测试某机器是否能表现出与人等价或无法区分的智能。然而，时到今日，人类还未突破技术瓶颈研发出能够通过“图灵测试”的机器。自然语言处理作为人工智能分支中最重要的一环之一，还需研究者们投入大量努力。本文将就自然语言处理里面中文机器阅读理解这一个子方向展开探究。

如今，随着科技的发展，人们的生活节奏越来越快，计算机已经在诸如商业、建筑、医疗等很多领域代替大量重复人工劳动。于是在很多场景中，计算机的智能问答已经被开发引进并逐渐走向成熟。其中，Apple在2010年引入了Siri程序，IBM在2011年推出Watsons系统，三星的Bixby语音助手在2012年出现，Amazon的Alexa系统在2014年推出以对用户的问题进行回答，微软小冰主打模范人类感情，并在2014年上市，Google的语音助手在2016年开始成型，阿里巴巴的阿里小蜜也在2018年逐渐接管了大部分商家客服服务，小米的小爱同学在2019年开始结合文字、语音和图像与用户进行深入交流。人机对话系统未来科技发展的趋势，其模仿人类交流的特点使得其成为各种科技的载体。然而，目前智能人机对话系统最大的问题，在于机器对于人类真实想法与意图识别不准确，时常出现回答文不对题、千篇一律，亦或是由于信息缺乏而选择不知道。这个问题很大程度上限制了智能对话系统在商业中的应用。针对这一个问题，科研工作者们提出了让机器先学会小学阶段最基础的阅读理解，这项最基础的任务将会是研究的根基，使得后续机器无论是理解对于人类没有任何难度的常识信息亦或是理解人类基础情绪或者自身培养出性格并随外界输入信息表现出不同的情绪成为可能。如果机器能顺利的完成最阅读理解任务，并且能在各种场景下进行调节适应，那么距离“图灵测试”的破解也就不远了。

* + 1. 机器阅读理解的概念

阅读理解是人类的基本技能之一，自小学起必须系统地学习。它通常包含一篇文章和一些关于特定段落的问题。要回答这些问题，需要首先通过收集文章中与答案相关的句子来收集信息，有时可以直接复制文章中的原始句子作为最终答案。这是一个微不足道的“直觉问题”，然而难点不在回答问题本身，而是要针对问题回答进行断言，推断作者的意图并改进，最后总结出一套完整的体系。这整个推理过程尤其重要，但在复杂的语言交流中，这个能力只是基础。

机器阅读理解（Machine Reading Comprehension）是指让机器阅读文本，然后回答和阅读内容相关的问题。阅读理解是自然语言处理和人工智能领域的重要前沿课题，对于提升机器智能水平、使机器具有持续知识获取能力具有重要价值，近年来受到学术界和工业界的广泛关注。从学术研究的角度来看，这是学术研究的最前沿，代表着机器在当今时代可能达到的最高智能水平。从工业界的视角来看，对搜索引擎、聊天机器人、私人机器助手等重要产品都有直接的应用创新进展，拥有可产生巨大商业价值的发展空间[1]。机器阅读理解任务作为研究和评价机器理解语言文字的重要方式，具有重要的理论研究和实际应用价值。

* + 1. 研究应用价值

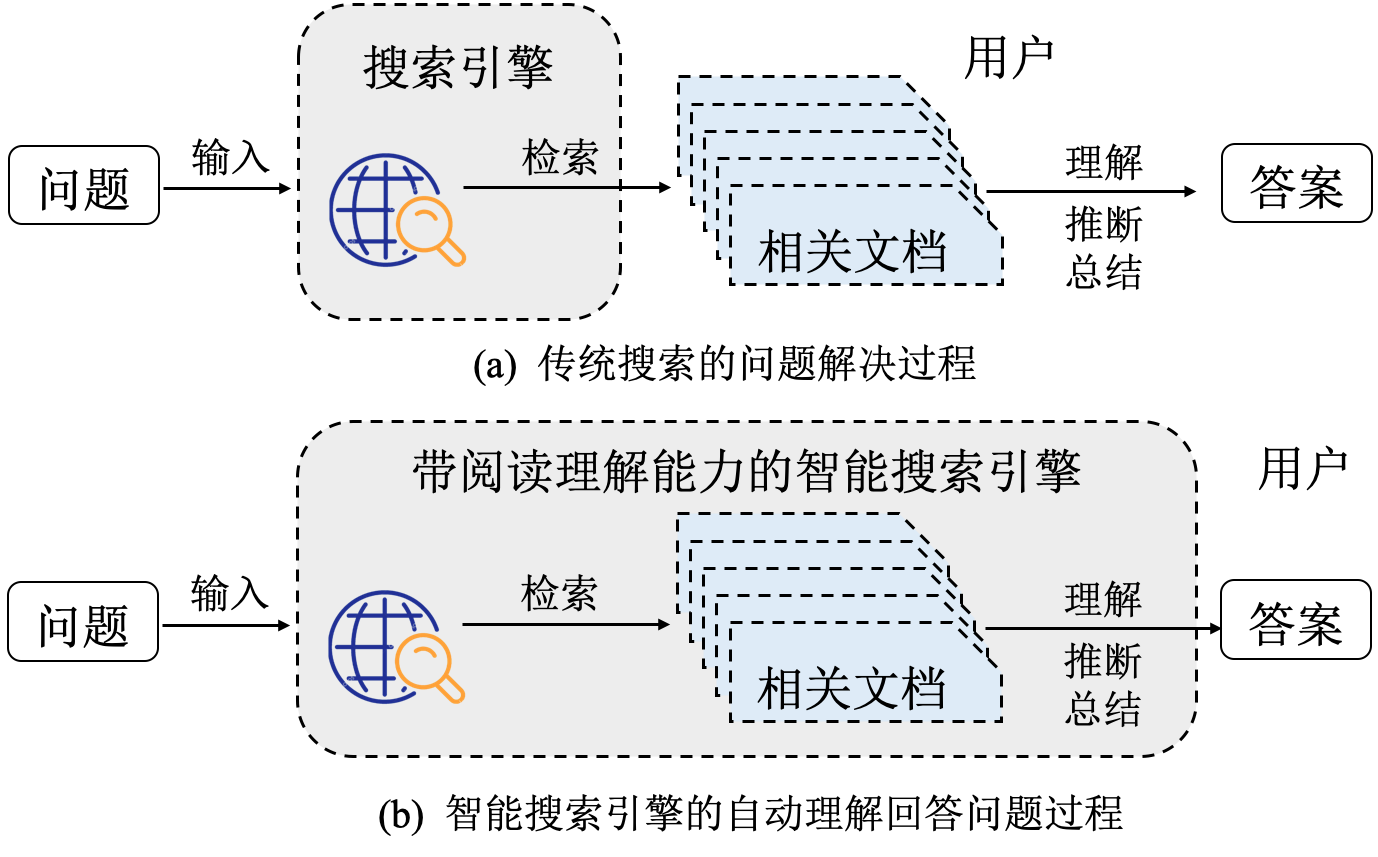


图 1 传统搜索引擎和带智能机器理解的搜索引擎的问答过程

计算机作为一种工具，在各个领域帮助人类解决了各种枯燥繁杂的工作。但目前其能解决的都是基于规则编程的简单任务，若是其能解决复杂的交流任务，将大大提升社会运行效率。以一个常用的寻医需求举例，图 1展示了用户使用外部资源探求问题的过程。传统的方式中，当用户探求一个未知问题时，其可以通过互联网检索出大量与之相关的文档，在花一定时间进行筛选、判别、推断和理解后，用户能得到问题的恰当答案。然而，这个过程在如今的快节奏生活中是费时费力的，而且能否得到合适的答案是非常依赖于用户获取到的相关文档的数目以及质量。举一个例子，当用户提出问题“华西医院有关脊柱侧弯的专家有哪些，哪一个最好”，要得到这个问题的解答，用户需要登上四川省华西医院的官网，找到骨科医生列表，判断哪些医生的专业与“脊柱侧弯”相关。在筛出的医生中，为了判断哪一个最好最合适，用户还得重新搜索这些医生的信息进行判断。这是一个非常繁杂的过程。试想，如果实现了出可以自动实现阅读理解的智能搜索引擎，则用户提问后，这个耗时耗力的搜索以及判断过程将由搜索引擎代劳。由于计算机在计算效率上的优势，这个过程的时间将大大简短，而且计算机可以短时间对成千上万的相关资料进行分析，结果将更全面而有说服力。

另外，机器阅读理解也将是强人工智能[[2]](#footnote-2)实现的毕竟之路。现在的智能程式多是弱人工智能只处理特定的问题，不需要具有人类完整的认知能力，甚至是完全不具有人类所拥有的感官认知能力，只要设计得看起来像有智慧就可以了。这种设计的局限很强，在实现机器阅读理解后，计算机将在更多更广的场景中帮助人类。

* 1. 机器阅读理解的难点

如前文所述，机器阅读理解的难点很大程度上就是自然语言处理的核心难点，其主要可以分为三个：普遍存在的不确定性、语言知识处理的复杂性、输入的不规范性。

* + 1. 普遍存在的不确定性

这个问题包括消除歧义，涵盖词法、句法、语用分析中存在的歧义问题。在每一个层次中或是在下一层次向上一层次转变中都存在着歧义和多义现象，即形式上一样的一段字符串，在不同的场景或不同的语境下，可以理解成不同的词串、词组串等，并有不同的意义。造成歧义出现的原因是，对于机器来说，人类活动非常复杂，而语言的词汇和句法规则又是有限的，它们中的大多数表述需要根据特定的相应的语境和场景的规定而得到解决，这就造成同一种语言形式可能具有多种含义。这个问题在中文阅读理解中更严重。

1. **词法歧义问题**。在词法中，中文存在词汇边界界定的歧义问题，即分词。在口语中，词与词是连贯表述的，在书写上中文也没有词与词之间的边界，词和词之间没有将不同单词切分开来的分隔符，这时候，对于中文的自然语言处理就面临不同于英文等西方语言的另一步工序，分词，即由机器在词与词之间自动加上分隔符，从而将中文文本切分为独立的单词。通俗的说，就是查字典，把句子从左到右扫描一次，遇到字典里有的词标识出来。这种问题能解决多数问题，比如将句子“今天天气炎热”切分，变成“今天/天气/炎热”。但当遇到二义性的分割就无能为力了。例如“乒乓球拍卖完了”，切分为以下两种情况都是合理的，“乒乓球拍/卖/完了”，“乒乓球/拍卖/完了”，但其表达意思完全不一样。正确的词汇切分是后续机器阅读的基础。分词中的第二大难点是新词发现。新词是指在从前的分词词典中没有收录，后来出现并被大家公认的词语。新词发现难的原因在于新词覆盖面广，词语构成没有普遍规律；低频新词数据稀疏，识别难度大。比如“屌丝”、“不明觉厉”、“累觉不爱”等人为创造出的未登录词。除了人名、未登录词以外，地名、机构组织名、事件名、缩略语、派生词、各种专业术语也属于新词。中文分词是中文自然语言处理的基础，在自然语言处理中扮演着重要角色。
2. **句法歧义问题**。除了词法以外，句法也存在着各种歧义问题。例如在短语级别上，结构关系不同和结构层次不同都会造成歧义的出现。比如“修改方案”可以理解为动宾关系（修改了一个方案），也可以理解为偏正关系（一个修改方案），这是结构关系不同所造成的歧义；比如“许多朋友送来的礼物”，既可以理解为礼物多，也可以理解为朋友多，这是结构层次不同造成的歧义。又如在句子级别上，可能出现语义组合歧义。比如“他在汽车上画画”，介词短语“在汽车上”的语义既可以指“他”（他在汽车上上，他画画），也可以指向“画画”（他画画，画在了汽车上）。由此可见，同样一个单词、短语或者句子有多种可能的理解，表示多种可能的语义。如果不能解决好各级语言单位的歧义问题，我们就无法正确理解语言要表达的意思。
3. **语用歧义问题**。当语气不同的时候，表达的意思可能大相径庭。例如问题“你今天钓到鱼了吗？”，回答“运气好极了，一条都没有钓到。”机器需要判定，这里的“运气好极了”不是有好运气的感慨，而是带有反讽的自嘲，表达的恰恰是相反的意思。这对于人类很好理解，但对于机器是一个极其困难的任务。
   * 1. 语言知识处理的复杂性

为了消除歧义，机器需要获取更多的知识，而消除歧义所需要的知识在获取、表达以及运用上存在困难，主要包含上下文知识和背景知识的问题。所谓的“上下文”指的是当前所说这句话所处的语言环境，例如说话人所处的环境，人说话时指代的对象，人说话时省略的部分是什么，或者是这句话的前几句话或者后几句话等等。在试图理解一句话的时候，上下文的影响十分重要。由于上下文对于当前句子的暗示形式是多种多样的，因此如何考虑上下文知识获取是语言知识处理复杂性的一大体现。

上下文知识获取的问题有不同的表现。第一，在日常对话中，人们时常会在表述中做省略，让表达更简洁。因此，上下文知识获取中存在对省略部分进行补充的一步，称为“省略恢复”。 比如用户向客服系统发问：“发票快递费怎么收，在线支付。”未等客服系统做出回答，他继续说道：“到付呢？”这是自然语言中典型的省略问题，要求机器能够根据上下文的相关信息，对可能的语法空位和语义空位进行填充，实现对省略部分的恢复。在这个例子中，机器应在“到付呢”前恢复问题的陈述部分“发票快递费怎么收”，并给出合适的答案。第二，在人的交流过程中，时常会有指代，在表达中适当地使用指代会让表达更加简练却不影响本意的阐述。文本中是否存在指代现象，准确识别指代的是什么内容，机器遇到的这一不容小觑的难点称为“指代消解”。比如识别并理解文中普遍出现的“他/她/它”，“上次”，“另外一个”这些指示代词具体是什么。这是一个难以归纳出规律的问题。第三，在某些以任务驱动的对话中，用户通过使用某类产品会在某一领域获得相对完整的服务，解决一个复杂问题，或者获得某种方向性的引导。在这种情况下，整个对话的场景被限定，需要确定一些关键的信息才能完成任务，但用户一般不会再一次表述中就表达完所有需要提供的信息。这时候，机器需要能够主动向用户询问，并且会根据上下文来判断继续提出什么样的问题，这就是“槽位填充”。比如餐厅定位，涉及到时间、地点、餐厅名、用餐人数等属性。槽位填充就是在对话中识别出用户提供的餐厅名、用餐人数这些属性，这些属性的状态就是这个对话的状态，对话状态跟踪就是记录当前用户所有动作以及对应的槽位的状态并维护整个对话的状态。

此外，正确理解人类语言还要有足够的背景知识。机器也需要了解约定俗成的东西。比如某打车软件司机向客服系统询问：“都十点了，怎么还没奖励？”这就需要系统知道该打车软件在晚间9点以后就有夜间奖励。如果机器不知道这个政策，往往会给出令人啼笑皆非的答案。对俗语、谚语等的深入理解、运用也同样重要。在英语中“The spirit is willing but the flesh is weak.”，意思是“心有余而力不足”。但是当时的某个机器翻译系统将这句英文翻译到俄语，然后再翻译回英语的时候，却变成了“The Voltka is strong but the meat is rotten.”，意思是“伏特加酒是浓的，但肉却腐烂了”。从字面意义上看，“spirit”与“Voltka”对译似无问题,而“flesh”和“meat”也都有肉的意思，但两句翻译的意思却南辕北辙。关键的问题就在于在翻译的过程中，机器翻译系统是从字面上进行翻译，却没有理解句子真正的寓意。

* + 1. 输入不规范问题

在输入时，可能会出现错误。常见的不规范输入主要是错别字、口语化、语法不对这三方面。首先，就客服系统来说，用户时常会有一些有瑕疵的输入，例如：“打车卷怎么用啊，不会用”、“伺机居然绕路，气死我了”，用户错把“打车券”打成了“打车卷”、把“司机”打成了“伺机”，这要求机器具备“纠错”功能。其次在口语化问题上，比如，有时候客服系统还会遇到这样的情况：“跟他说了我赶时间他还不麻溜儿地过来，这不能够啊。”如果不能具备理解口语化的能力，机器无法理解这是一个投诉司机速度慢的信息。值得一提的是，这对于不熟悉本地方言的人类都尚且是一个问题，对于机器则更加困难。最后在语法问题上，正确的句式应该是主、谓、宾的格式，但人的的表达往往没那么多条条框框：“顺风车没有到目的地就强制让我下车，他自己找找错误我然后还要就不按我给她说的路线走他自己绕路，还要让我强制交高速费”。可以看到，用户可能因为情绪激动语序颠倒。机器需要修正这些错误的语法，理解用户投诉的三个要点：未达到强制用户下车、绕路、强制交高速费，投诉对象是顺风车车主。

上面三个方面的主要困难可以看到机器阅读理解乃至自然语言处理这个难题的根源就是人类语言的复杂性和语言描述的外部世界的复杂性。人类语言承担着人类表达情感、交流思想、传播知识等重要功能，因而自然语言处理也要求机器需要具备强大的灵活性和表达能力。理解语言所需要的知识又是无止境的，机器阅读理解任务道阻且长。

* 1. 国内外研究现状

近几年来，有不少研究者们在机器阅读理解问题上展开了研究，并取得了一定的成果。值得一提的是，机器阅读理解这一领域不同于其他领域，其模型强烈地依赖于数据集，不同的数据集有着迥然不同的处理方式。

* + 1. 国外机器阅读理解研究现状

目前国际上的机器阅读理解形式有很多，可基本分为单项填空（cloze-style），多项填空，开放域（世界知识）问答和多项选择几种类型。每一种类型的问题都有对应数据集，每个数据集都有很多研究工作进行攻坚。下面将分别介绍几种类型的阅读理解以及研究现状。

1. **单项填空**。让计算机阅读并理解一篇文章内容后，对机器发出问题，问题往往是抽掉某个单词或者实体词的一个句子，而机器回答问题的过程就是将问题句子中被抽掉的单词或者实体词预测补全出来，一般要求这个被抽掉的单词或者实体词是在文章中出现过的。最出名的数据集有来自Hermann 等人[7]的CNN/Daily News。对于该数据集，其答案一定为出现在原文中的一个词，根据这一特点，Kadlec[8]受Pointer Network[9]启发提出了AS Reader模型，直接将原文中每个词的注意力相加作为其成为答案的概率，进而输出答案。这一模型非常简单，但取得了当时最好的效果。受ASReader的启发，Cui[10]提出了AoA Reader，他的创新点是在注意力的基础上再计算注意力的注意力，借此指示每个注意力的重要性。同样受AS Reader启发的还有Dhingra[11]，他提出GA Reader来引入一种计算注意力的新机制。Chen[12]提出基于人工构建的特征的传统分类模型和基于Herman[7]提出的模型改进后的深度神经网络模型，主要是计算注意力时使用Bilinear函数代替Tanh函数，并去掉最后的非线性层，实验表明Bilinear函数能够显著提升模型效果，并且最后的非线性层对结果影响不大。更重要的是作者对数据进行了分析，通过将100个采样问题进行分类分析，由于数据集本身的问题，得出当前模型已经达到了该数据集上的最佳效果，几乎没有进一步提升的空间。
2. **多项填空**。区别于单项填空，其标准答案不仅限于一个单词，可能是原文中的一个短语或者句子。这其中，最出名的SQuAD是Rajpurkar等人[13]发布的阅读理解数据集，采用众包的方式构造包含10000个问题的大规模数据集。SQuAD是问答题而非选择题，它的答案不是文中的一个词，而是原文中的一段连续文本片段，在难度和数量上均要明显高于之前的任何数据集。Yang[14]提出通过Gating机制一定程度上缓解了未登录词所带来的问题。Seo[15]提出BiDAF模型，同时使用问题到原文的注意力和原文到问题的注意力，取得了很好的效果。Wang[16]结合Match-LSTM[17]进行实验，发现直接预测答案边界的模型要优于直接预测答案的模型。受Match[16]启发，微软提出R-NET[18]模型，它是一个基于四个层面的神经网络的端到端系统，在该数据集上取得了很好的效果。
3. **单项选择**。就是阅读完一篇文章后，给出问题，正确答案是从几个选项中选择出来的，典型的任务比如托福的听力测试，目前也有研究使用机器来回答托福的听力测试，这本质上也是一种阅读理解任务。Ricardson等人提出了MCTest数据，其总体形式类似于英语考试中的单项选择阅读理解题。但是不同点在于MCTest严格限制的文档本身是儿童能够理解的故事内容，从而保证阅读理解的过程中不需要过量的外部知识。MCTest作为机器阅读理解数据集，只包含660篇文章，共2640个选择题型问题，其较小的数据规模导致大多数工作都是基于特征工程的工作。事实上，Richardson在发布数据集时提出了两种基于词汇匹配的非监督基线方法。第一种使用滑动窗口同问题与答案组合中词汇匹配分数的方法，第二种方法首先将问题以及答案按照启发式规则进行拼接，然后计算上述拼接结果与原文信息之间的关系。这两种方法都比较简答，但是在数据集上都取得了不错的效果。Smith等人[3]对这种方法进行了扩展，使用变窗口大小的方式改进了模型。Trischler等人[4]利用词嵌入，充分考虑了不同级别窗口以及n-gram级别的匹配信息，并使用一个MLP来结合上述匹配效果，最终实现了端到端的训练。Wang等人[5]使用了线性模型融合多种框架、句法以及语义级别的特征，使模型结果有了大量的提高。Sachan[6]使用了修辞结构理论（RST）和事件实体共指方法对多对句子进行建模，获取和问题相关部分，并通过有机组合来判断文档和问题之间的蕴含关系。

除此之外，还有开放问题的阅读理解，比如日常聊天中的问题，比如课堂上的专业问题。这些问题的回答没有固定的答案，本文不予讨论。

* + 1. 中文机器阅读理解研究现状

相比起英文的机器阅读理解，中文机器阅读理解更加困难，主要原因在于上一章节提到的中文分词问题。目前，中文机器阅读理解材料不多。针对这一问题，郝晓燕[19]等构建了包含121篇文章，涵盖14个领域的中文阅读理解数据集，并对其构造方法加以详细阐述，为后续中文阅读理解研究提供基础。

基于山西大学的中文阅读理解语料库CRCC v1.1，李继宏[20]使用最大熵模型，根据问句和候选答案的关系，在词法层面和句法层面构造了35个人工特征，最终达到了80.18%的HumSent准确率。

李济洪[21]针对中文阅读理解中的时间、任务、地点、数值、实体、描述共六类问题，使用词法和句法两个层次的规则来构造规则集，在山西大学CRCC v1.1语料库上进行实验，比在相同数据集上基于最大熵的方法高约1%。

国内早期在中文阅读理解上面的研究，受限于语料规模的限制，使用的方法依赖于人工构造特征的机器学习模型和基于规则的方法为主。

HFL-RC[22]是哈工大讯飞联合实验室2016年发布的大规模填空型中文阅读理解数据集，填补了大规模中文阅读理解数据集的空白。该数据集使用人民日报和儿童读物，通过机器自动挖词的方式构造，并且在文中提出了一种基于注意力机制的神经网络模型，取得了良好的效果。

哈工大讯飞联合实验室(HFL)[23]在2017年发布了新的大规模中文阅读理解数据集，这个数据集为第一届“讯飞杯”中文机器阅读理解评测(CMRC2017)比赛所采用。主要任务是“填空型阅读理解”，根据问题的形式，主要分为填空型任务和用户提问类任务两个评测任务。数据主要来源于儿童读物，通过机器和人工联合构造，结果更准确，样式更丰富，为中文机器阅读理解的研究提供了更多的选择。哈工大讯飞联合实验室(HFL)在2018年发布了首个人工标注的中文篇章片段抽取型阅读理解数据集[24]，聚焦基于篇章片段抽取的阅读理解，作为填空型阅读理解任务的进一步延伸，以弥补中文资源在此方面的空白。

* + 1. 国内外现状简析

在国内外学者的共同努力下，随着大规模数据集的不断发布，计算能力的不断提高，新的机器学习方法的进步，机器阅读理解任务已经取得了令人瞩目的进展，同时也提炼出很多关键的技术。但机器阅读理解仍处于非常初级的阶段，其中的关键技术如推理、背景知识等均没有出现里程碑式的技术。在这一领域，仍然有非常多，非常有意义的工作值得去研究。从已有的工作中可以看出，值的研究的方向罗列如下：

1. **构建更具备难度的大规模阅读理解数据集**。区手工构建大规模的阅读理解训练数据需要花费极大成本，因此目前的不少数据集都存在一定问题。某些数据集比如MCTest经过人的精心构建，但是规模过小，很难用来有效训练复杂模型。另外一类是采用一定的启发规则自动构建的数据集，这类数据集数据规模可以做得足够大，但是很多数据集失之于太过简单。不少实验以及对应的数据分析结果证明常用的大规模数据集比如CNN和Daily Mail 相对简单，正确回答问题所需要的上下文很短（5个单词窗口大小范围），只要采取相对简单的模型就可以达到较好的性能，复杂技术发挥不出优势，这从某种角度说明数据集难度偏小，这也极大限制了新技术的探索。为了能够促进领域技术的进一步快速发展，需要一些大规模的人工构建的阅读理解数据集合，这样既能满足规模要求，又能具备相当难度。
2. **引入更丰富的世界知识**：对于人来说，如果需要真正理解一篇新的文章，除了文章本身提供的上下文外，往往需要结合世界知识，也就是一些常识或者文章相关的背景知识，才能真正理解内容。目前机器阅读理解任务为了降低任务难度，往往将世界知识排除在任务之外，使得阅读理解任务更单纯，这在该领域研究初期毫无疑问是正确的思路，但是随着技术发展越来越完善，会逐渐引入世界知识来增强阅读理解任务的难度，也和人类的阅读理解过程越来越相似。
3. **发展更为完善的推理机制**：在阅读完一篇文章后，为了能够回答复杂的问题，设计合理的推理机制是突破的关键技术，如上文所述，目前的推理机制大多数还是采用注意力焦点转移的机制来实现，随着复杂数据集的逐渐出现，推理机制会在阅读理解中起到越来越重要的作用，而今后需要提出更加丰富的推理机制。
   1. 本文主要工作

本文基于上文所提到的机器阅读理解任务的难点，以及现有工作的局限性，在前人工作的基础上，首先针对以上所提到任务对R-NET网络进行改进，提出了更加适合机器阅读阅读理解的模型，并在中文数据集上做了适应和推广。在这个过程中，本文将R-NET的参数进行了进一步的调整，使其在实验数据集上得到了更好的结果。本文在机器阅读理解的数据集上，用很多国际主流的模型和本文模型进行对比，并通过对比验证了本文提出模型的性能的优越性。

* 1. 本文的组织结构

本文共分为五章，各章节的内容安排如下：

第一章是绪论部分。主要阐明了本文研究工作的相关背景和机器阅读理解的概念，并介绍了该任务的研究意义、研究难点，然后对国内外的研究现状做了总结，分析了现有工作的存在的不足和局限性。最后介绍了本文的主要工作和组织结构。

第二章介绍了一些相关的研究概念。首先介绍了循环网络中RNN，LSTM，BiLSTM以及GRU的概念；再介绍词向量表示，给出了word2vec，glove和双向Skip-Gram的模型简介；紧接着本文介绍了目前NLP问题中热门的attention机制，和其拓展：self-attention和multi-head attention。为了更好地诠释本文方法，pointer网络以及hard sigmoid概念也在此章中被介绍。

第三章研究探讨了基于R-NET模型的方法。首先对研究问题进行了定义，然后详细介绍了模型的步骤以及每个步骤的意义，包括整体的网络框架，每个子模块的功能和结构，损失函数的选用等。最后是实验部分，介绍了数据集和评价指标，并对实验结果进行分析、与其他算法进行比较。

第四章在第三章的基础上，对R-NET模型进行拓展。首先介绍了研究的问题，然后详细介绍了模型在中文数据集上的难点，然后是实验部分，介绍了数据集和评价指标，并进行实验，最后分别对国际上主流的方法的结果进行分析和总结。

第五章是实验结果总结、分析与展望部分。主要回顾和总结了本文的研究工作，并对未来可以研究的内容进行了展望。

# 

# 参考文献

[1] 朱海潮, 刘铭, 秦兵. 基于指针的深度学习机器阅读理解[J]. 智能计算机与应用, 2017, 7(6):157-159.

[2] Richardson, Matthew, Christopher JC Burges, and Erin Renshaw. 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. 2013:193-203.

[3] Smith E, Greco N, Bosnjak M, et al. A Strong Lexical Matching Method for the Machine Comprehension Test[C]. Conference on Empirical Methods in Natural Language Processing. 2015:1693-1698.

[4] Trischler A, Ye Z, Yuan X, et al. A Parallel-Hierarchical Model for Machine Comprehension on Sparse Data[J]. 2016:432-441.

[5] Wang H, Bansal M, Gimpel K, et al. Machine Comprehension with Syntax, Frames, and Semantics[C]. Meeting of the Association for Computational Linguistics and the, International Joint Conference on Natural Language Processing. 2015:700-706.

[6] Sachan M, Dubey K, Xing E, et al. Learning Answer-Entailing Structures for Machine Comprehension[C]. Meeting of the Association for Computational Linguistics and the, International Joint Conference on Natural Language Processing. 2015:239-249.

[7] Hermann K M, Kočiský T, Grefenstette E, et al. Teaching machines to read and comprehend[C]. International Conference on Neural Information Processing Systems. MIT Press, 2015:1693-1701.

[8] Kadlec R, Schmid M, Bajgar O, et al. Text Understanding with the Attention Sum Reader Network[J]. 2016:908-918.

[9] Vinyals O, Fortunato M, Jaitly N. Pointer Networks[J]. Advances in Neural Information Processing Systems 28, 2015: 2692-2700.

[10] Cui Y, Liu T, Chen Z, et al. Consensus Attention-based Neural Networks for Chinese Reading Comprehension[J]. 2016.

[11] Dhingra B, Liu H, Yang Z, et al. Gated-Attention Readers for Text Comprehension[J]. 2016.

[12] Chen D, Bolton J, Manning C D. A Thorough Examination of the CNN/Daily Mail Reading Comprehension Task[C]. Meeting of the Association for Computational Linguistics. 2016:2358-2367.

[13] Rajpurkar P, Zhang J, Lopyrev K, et al. SQuAD: 100,000+ Questions for Machine Comprehension of Text[J]. 2016:2383-2392.

[14] Yang Z, Dhingra B, Yuan Y, et al. Words or Characters? Fine-grained Gating for Reading Comprehension[C]. International Conference on Learning Representations, 2017.

[15] Seo M, Kembhavi A, Farhadi A, et al. Bidirectional Attention Flow for Machine Comprehension[C]. International Conference on Learning Representations,2017.

[16] Wang S, Jiang J. Machine Comprehension Using Match-LSTM and Answer Pointer[C]. International Conference on Learning Representations, 2017.

[17] Wang S, Jiang J. Learning Natural Language Inference with LSTM[C]. Conference of the North American Chapter of the Association for Computational Linguistics. 2016:1442-1451.

[18] Microsoft Research Asia.2017.R-NET: Machine Reading Comprehension with Self-matching Networks. 2017

[19] 郝晓燕, 李济洪, 由丽萍,等. 中文阅读理解语料库构建技术研究[J]. 中文信息学报, 2007, 21(6):29-35.

[20] 李济洪, 王瑞波, 王凯华,等. 基于最大熵模型的中文阅读理解问题回答技术研究[J]. 中文信息学报, 2008, 22(6):55-62.

[21] 李济洪, 杨杏丽, 王瑞波,等. 基于规则的中文阅读理解问题回答技术研究[J]. 中文信息学报, 2009, 23(4):3-9.

[22] Cui Y, Liu T, Chen Z, et al. Consensus Attention-based Neural Networks for Chinese Reading Comprehension[J]. 2016.

[23] Cui Y, Liu T, Chen Z, et al. Dataset for the First Evaluation on Chinese Machine Reading Comprehension[J]. 2017.

[24] [Yiming Cui](https://arxiv.org/search/cs?searchtype=author&query=Cui%2C+Y), [Ting Liu](https://arxiv.org/search/cs?searchtype=author&query=Liu%2C+T), [Li Xiao](https://arxiv.org/search/cs?searchtype=author&query=Xiao%2C+L). A Span-Extraction Dataset for Chinese Machine Reading Comprehension. arXiv preprint arXiv:1810.07366. 2018

1. 图灵测试，维基百科链接：https://en.wikipedia.org/wiki/Turing\_test [↑](#footnote-ref-1)
2. 强人工智能，维基百科链接：https://en.wikipedia.org/wiki/Artificial\_general\_intelligence [↑](#footnote-ref-2)