项目编号：201810006002



国家大学生创新创业训练计划

结题报告

项 目 名 称： 基于深度学习的智能问答系统

立 项 人： 李明浩，谢一凡

所 在 单 位： 北京航空航天大学计算机学院

手 机： 13121276176

电 子 信 箱： li\_mh1997@163.com

资 助 金 额： 6500.00（元）

指 导 教 师： 荣文戈

训 练 场 地： 北京航空航天大学

研 究 期 间： 2018.05.04-2019.03.31

北京航空航天大学教务处

项 目 摘 要

（500字）

随着深度学习的发展，机器的智能水平已经在许多领域超越了人类的水平，但是在使用自然语言进行通信交流方面，机器远不及人类的水平，其关键在于机器无法准确的理解蕴含在自然语言中的信息。尽管如此，人们依然提出了许多方法来解决这类问题，并在机器翻译，智能问答等方向研究出了有效的模型策略。

我们首先实现了提取特征的逻辑回归模型，经过SQuAD数据集进行训练后，在官方的测试集上取得了50%的正确率。随后调研了智能问答方向最前沿的模型，并选择在微软亚洲研究院提出的R-Net模型的基础上进行改进与完善，实现一个智能问答应用。R-Net模型适用于针对一个具体段落提出的问题，并选取段落原文进行回答，R-Net模型在该问题上的正确率为71%，性能基本达到人类水平。我们将模型进行部分改进，使其可以处理针对由多个段落构成的目标文本提出的问题，并选取目标文本的语句进行回答。围绕改进后的模型，实现了单段落多段落问答接口，以及智能客服系统和通用型智能问答系统的应用，均取得了不错的表现。

关键词（3～5个）：阅读理解，智能问答，SQuAD，R-Net

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 项目名称 | | | | | 基于文本理解的智能问答系统 | | | | | | | | | | |
| 项目来源 | | | | | 科研 | | | | | | | | | | |
| 审批经费（元） | | | | | 6500.00 | | | | 起止时间 | 2018年 5月 至 2019年 3月 | | | | | |
| 项目组成员(第一人为立项人) | 姓名 | 学号 | | 专业/年级 | | | 学院 | | 项目分工 | | E-mail | | | 联系电话 | |
| 李明浩 | 15061108 | | 计算机科学与技术/大四 | | | 计算机学院 | | 前端 | | li\_mh1997@163.com | | | 13121276176 | |
| 谢一凡 | 15061105 | | 计算机科学与技术/大四 | | | 计算机学院 | | 后端 | | xieyifan@buaa.edu.cn | | | 18009646235 | |
|  |  | |  | | |  | |  | |  | | |  | |
|  |  | |  | | |  | |  | |  | | |  | |
|  |  | |  | | |  | |  | |  | | |  | |
| 导 师 | 姓名 | | 荣文戈 | | | 单位 | | 北京航空航天大学计算机学院 | | 职务 | |  | 职称 | | 副教授 |
| 电话（手机） | | 010-82338199 | | | E-mail | | w.rong@buaa.edu.cn | | 性别 | | 男 | 年龄 | |  |

|  |
| --- |
| **一、**项目背景（不少于400字）  当前，构建具有阅读理解能力或基于真实数据的开放领域问答能力的智能系统是人工智能、特别是自然语言处理领域的一个主要目标。这样的系统能够为社会的方方面面带来巨大价值。目前，已有的产品如Cortana,Siri等均已实现大规模的生产和应用，这得益于近些年深度语音识别技术的快速发展。尽管这些类型的语音助手已经深受欢迎，但消费者们认为如果语音助手能对一个问题给出答案而非只是列出搜索引擎的搜索结果会更加方便实用。同样的，具有阅读理解和问答能力的系统还可以应用到企业的智能客服上面，从而产生巨大的商业效益。  阅读理解本质是一种问答任务的类型，其答案可以从读者不熟悉的特定实体和事件的段落中找到。特别的，这样的实体和事件是结构化数据库中所涉及到的。阅读理解问题旨在衡量系统能否从非结构化文本中直接提取关于实体和关系的语义信息。近些年，许多的神经网络模型被开发出来用于机器阅读。这些模型可以分为两类，即聚合模型和显示参考模型。聚合模型计算一个涉及问题敏感关注的段落矢量表示，然后根据传递向量选择一个答案。显示参考模型基于问题计算段落的注意力权重。相反，他们依赖于一种共同参考注解，即在该位置是否引用候选答案。这些模型计算一个注意力，就像在聚合模型中一样，但是不是计算段落向量。它们只是选择最有注意力的答案。  目前使用人工智能、深度学习方法，解决这类问题，正逐渐成为一种趋势。 |

|  |
| --- |
| **二、**研究内容（不少于600字）  实现在现实世界中具有阅读理解和问答能力的系统是一项非常有挑战性的任务。这其中涉及了自然语言处理领域的多个子任务，如阅读、处理、理解、推理和生成答案。  ­­­­­­­目前，许多开源的大型数据集已经引领了很多在该领域的技术突破。但许多该领域的数据集都存在的一个问题是关于数据文本的问题并不是来自现实世界的用户的，通常的做法是使用众包的方法来收集相关问题，但这样产生的问题与现实世界中人们的提问存在较大的偏差。此外，另一个与实际不太相符的问题是大多数数据集的问题答案都局限于一段文本中。而现实环境下，为了保证问题答案的质量，通常需要从多个段落的信息中拼接得出最佳答案，因为最理想的答案不仅需要直接回答出问题，还需要对答案能够有一定的解释。  基于以上背景，我们选择了斯坦福大学公开的SQuAD数据集来训练具有阅读理解和问答能力的智能系统。  我们期望实现一个智能问答系统，用户输入一个问题后，系统将问题发送至问答搜索引擎，爬取问答搜索引擎的搜索结果，将搜索结果与问题一同发送至智能问答模型中，之后再获取模型的预测答案，将答案返回给用户。我们期望使用网页界面与用户交互。  实现这个系统需要处理很多问题，其中最关键的难点就是智能问答模型的实现，智能问答语言模型的性能直接影响到了本系统的用户体验。SQUAD作为一个公开数据集，目前已经吸引了众多研究团队参与其中，并且许多研究团队也提出了大量优质的模型，我们借鉴了由微软亚洲研究院所提出的R-net模型的结构，并在他的基础上进行了一定程度的改进，从而使其适应我们的系统。 |
| **三、**技术路线及解决方案（可分章节，附图或表，不少于5000字，可续页）  实现本项目所预期的智能问答系统，主要有两个方面的任务。第一个是智能问答模型的构建，本项目需要根据用户所提出的问题，爬取特定的文本内容，然后根据文本内容与用户提出的问题，从文本中选取对应的答案。本项目中我们分别采用了传统的机器学习技术以及比较前沿的深度神经网络技术，来实现我们的智能阅读理解模型，其中核心模型最终选择了深度神经网络模型，并在现有的性能优秀的模型的基础上做了一定程度的改进，以使模型能够满足任务的需要。  另一个工作量比较大的任务，就是实现与用户交互的手段，我们采用了网页界面的方法，并围绕改进后模型的特点，设计了三个方面的应用。分别是智能客服，智能问答以及阅读理解。同时为了尝试解释神经网络进行文本阅读理解的过程，我们还在网页中加入了网络可视化的内容。  下面我们将分三个章节来分别介绍项目的模型搭建，前端网页界面的搭建，以及最终模型的实验效果。  第一章 模型介绍  1.1 数据集介绍与问题描述  SQuAD(Stanford Question Answering Dataset)是一个由来自斯坦福大学的自然语言处理研究小组发布的，阅读理解与问题回答方向的公开数据集。研究小组从维基百科(Wikipedia)中选择了网页排名前10000的文章，并从中随机抽样了536篇文章来构建数据集。这些文章包含有23215个段落，内容涵盖了从知名音乐人到一些抽象概念等多个方向。数据集的构建采用了众包的方式，通过雇佣一些“云工作者”对挑选出来的段落提出适当的问题并为其作答，最终通过一些评估手段选出合格的问题-答案对，需要注意的是，这些答案来源于原文中存在的连续语句，换句话来说，问题的答案抽取自原文中一段连续的文字。表1是一个数据集中的样本样例。  **表 1 一个SQuAD数据集的样本样例**   |  | | --- | | **Passage**: The Broncos took an early lead in Super Bowl 50 and never trailed. Newton was limited by Denver's defense, which sacked him seven times and forced him into three turnovers, including a fumble which they recovered for a touchdown. **Denver linebacker Von Miller was named Super Bowl MVP**, recording five solo tackles, 2½ sacks, and two forced fumbles.  **Question**: Who was the Super Bowl 50 MVP?  **Answer**: Von Miller |   本数据集需要的是一个类似于阅读理解风格的问题回答模型，通过给定一个段落P和一个问题Q，我们需要基于段落P来给出问题Q的答案A，其中A是P的一段连续的子序列。这个问题的挑战在于模型需要同时理解问题与段落，并准确推断出答案所在范围。  1.2 Logistic Regression(逻辑回归)模型  我们首先考虑使用传统的机器学习方法来解决这个问题，我们参考了Rajpurkar等[6]的相关工作，首先对段落与问题进行了特征工程，针对文本样本提取一些特征，之后将提取出特征向量输入到逻辑回归模型来对答案进行预测。具体的做法是，先根据问题对段落进行预处理，将段落原文分割成多个字串作为答案候选集，之后对每一条候选答案进行特征提取，提取的特征主要有4类，如表2所示。之后将这些特征组成一个特征向量，输入到逻辑回归模型中对候选答案的可能性进行计算，最后针对候选答案集选出置信度最高的候选答案作为该问题的结果输出。  **表 2 特征列表**   |  |  | | --- | --- | | 特征名称 | 描述 | | 词匹配率 | 对问题与答案都出现的词语使用TF-IDF计算出它们的频率并进行求和，来评估问题与候选答案的词相似程度。 | | 二元词语匹配率 | 与上述方法类似，但是使用的是二元计算方法，主要区别在于TF-IDF的计算方式不同，我们采用的是Shirakawa等[7]提出的方法。 | | 长度 | 对于在句子中提取出的候选答案序列，原句的长度、候选答案序列的长度、位于候选答案序列左侧的句子子序列长度以及位于候选答案序列右侧的句子子序列长度。 | | 候选答案序列词频 | 整个候选答案序列每个单词的TF-IDF频率和。 |   作为传统机器学习模型，逻辑回归模型的性能十分依赖于特征的选取，由于我们挑选的特征并不足够典型，模型最后的效果不太令人满意。  1.3 R-Net模型  为了提高原模型的表现效果，我们决定使用深度神经网络模型来处理此问题。通过调研，我们发现微软亚洲研究院NLC研究小组提出的R-Net模型在此问题上表现出色，我们对该模型进行了复现，并进行了一定程度上的改进以满足更一般的应用场景。  1.3.1 原模型描述  模型首先使用了一个Bi-RNN (bidirectional recurrent neural network,双向循环神经网络)来对段落和问题进行编码，随后使用了一个基于门-注意力机制的循环神经网络来对问题和段落进行匹配。第三步是本文的一个亮点，NLC小组使用一个自匹配注意力模型，来进一步提取段落特征并对段落进行提纯精炼，最后使用了Pointer Network (指针神经网络)来根据问题对答案在段落的的起始与终止位置进行预测。图1展示了网络的结构。下面将详细介绍每一层网络的结构细节。  **图 1 R-Net网络结构**  (1) 编码层  编码层的作用是让网络将文本和问题读入到模型当中。考虑一个问题和段落，首先使用词向量转换工具将文字转换为词级别的向量表示形式和，以及字符级别的向量表示形式和，并将它们分别输入到两个双向循环网络中，获得他们的编码:  这里的RNN单元采用的是GRU (Gated Recurrent Unit，门循环单元)方法。  (2) 门-注意力机制循环网络层  这一个网络层的作用，是让模型聚焦到文本中信息量较多的位置，从而捕获文本与问题的核心。通过使用第一层得到的文本信息和，通过一个加入了注意力机制[12]的RNN网络可以得到一个问题-段落匹配向量:  其中为从中使用注意力机制获得的结果，计算方法如下:  这与其他的注意力模型基本一致。  (3) 自匹配注意力层  这层网络是本模型的亮点所在，这一个网络层的作用是让模型将问题与文本段落结合起来，进行一个综合的信息提取，以便确定答案所在的大致区域与具体位置。作者在使用注意力机制获得的问题-段落匹配向量后，对此向量又使用了一次注意力机制，用于原文之间的自匹配，最终获得自匹配向量:  其中为从中使用注意力机制获得的结果，计算方法如下:  通过自匹配层，模型从原段落提取到了根据当前的文章和问题信息获得的特征。  (4) 输出层  输出层是将模型所预测的最终答案所在文本的位置进行输出，从而得到最终答案。本文采用了指针网络的形式来对答案在原文的区间进行预测，具体做法为使用上一层获得的向量来选择答案的开始位置与结束位置，可以用如下的公式表示:  这里的为RNN网络的隐藏层状态，计算方法如下:  这个循环神经网络使用一个与原问题相关的向量作为了隐藏层初始状态，计算方法为:  这依然是一个基于注意力模型的计算方法。  由于网络输出为问题的开始以及结束的位置，在训练时损失函数设计为输出结果与真实结果的差别。  2.3.2 模型改进  由于SQUAD的数据集的目标是在单独的段落中获取答案，而我们的系统所规划的应用场景是多段落的环境，因此我们需要让针对SQUAD数据集构建的模型符合我们的应用场景。我们在R-net网络的基础上做出了一定的改进，使它可以处理多段落的输入。我们我们首先计算出每个段落与问题的匹配程度:  将输入到softmax层以获得其概率分布:  最后将概率分布结合使用R-Net得到每个段落对问题所得到的指针网络结果，可以得出这些答案的置信度:  上述改进的主要思路是，针对用户提出的问题，我们将之施加于篇章的每一个段落之上，从每一个段落中都获得一个答案，然后比较这些答案与问题的文本相似度，剔除那些答非所问的句子，保留评分最高的答案，作为最终的输出答案。从而让原本针对但段落而设计的问答模型，可以应用到多段落的场景里。整个网络的结构如图2所示。  **图 2 模型改进部分的结构**  第二章 模型应用及前端配置  2.1 模型应用  在前一章介绍的模型中，Logistics模型和R-Net模型可以接收的输入为一个不限领域的段落和针对该段落提出的一个问题，模型返回的是从段落原文中截取的部分单词词组或短句作为问题的答案；我们的模型可以接受的输入为来自同一篇文章或针对同一话题的若干个段落和根据这个文章或话题提出的一个问题。在本章节中的模型应用均基于以上两种模式。  单段落阅读理解接口。接受两个长度不等的字符串，一个为段落，一个为问题。将段落和问题解析后，生成每个单词的词向量并输入到模型中。模型返回预测的答案区间的起始下标和结束下标，并将区间对应的段落原文以字符串的形式返回给接口调用端。同时返回基于问题的段落单词权重、基于上下文的段落单词权重、段落每个下标的指针网络得分，为后续的深度学习模型可视化作准备。  例：维基百科中加拿大简介的第一段与问题“加拿大在哪里？”  输入段落：Canada (/ˈkænədə/ (About this sound listen); French: [kanadɑ]) is a country located in the northern part of North America. Its ten provinces and three territories extend from the Atlantic to the Pacific and northward into the Arctic Ocean, covering 9.98 million square kilometres (3.85 million square miles), making it the world's second-largest country by total area. Canada's southern border with the United States is the world's longest bi-national land border. Canada is sparsely populated, the majority of its land territory being dominated by forest and tundra and the Rocky Mountains. It is highly urbanized, with 82 percent of the 35.15 million people concentrated in large and medium-sized cities, many near the southern border. Its capital is Ottawa, and its three largest metropolitan areas are Toronto, Montreal, and Vancouver. Canada's climate varies widely across its vast area, ranging from arctic weather in the north, to hot summers in the southern regions, with four distinct seasons.  输入问题：Where is Canada?  返回结果：northern part of North America  多段落阅读理解接口。接受一个字符串型数组和一个独立字符串，分别为多个段落和问题。将段落和问题解析后，生成每个单词的词向量并输入到模型中。模型会生成数目等于段落数目个候选答案，并为所有候选答案评分，最后返回得分最高的候选答案作为最终答案。  智能客服系统。接受一个机构（包括但不限于学校，公司，政府等）官方网站介绍页的网址和一个针对该机构的问题。将自动爬取该页面的重点内容，输入到多段落阅读理解接口中，得到答案并返回给用户。具体实现为：大多介绍页内容比较多，往往是多段落的形式。在Html5语言中，段落最常使用的标签是p标签。后端获取到介绍页的网址，读取内容并使用BeautifulSoup进行解析，获取所有p标签的内容，选取长度超过某个阈值的内容，过滤掉某些不包含重要信息的标签，存储为字符串型数组，与问题一起输入到模型中。  例：麻省理工大学的介绍页与问题：“麻省理工大学是什么时候建立的？“  输入网址：<http://web.mit.edu/aboutmit/>  输入问题：When was MIT built?  返回结果：1861  通用型智能问答系统。接受一个任意领域通用知识性问题，例如：“How old is Barack Obama? ”将自动在多种知识问答社区搜索回答，并把社区中的回答作为段落，与问题一起输入到模型中，再次对答案进行理解提炼，返回给用户。具体实现为：以用户提出的问题作为关键字，在Yahoo Answers等英文问答社区上进行搜索。选取第一个或前若干个问题的最佳答案，和问题一起输入到单段落或多段落阅读理解接口中，并将提取后的答案返回给用户。  例：2014年FIFA世界杯在哪个国家举办？  输入问题：Which country did the FIFA World Cup 2014 hold in?  返回结果：BRAZIL  模型可视化。由于R-Net模型属于深度学习模型，含有大量权重，因此内部处于黑盒状态，无法直观看到内部的运行机制及作用原理。为此，在单段落阅读理解接口实现的过程中，本团队将R-Net模型的二、三、四层中涉及到的权重或指针网络分数，即基于问题的段落单词权重、基于上下文的段落单词权重、段落每个下标的指针网络得分导出，使用echarts前端库中的热力图进行可视化。在第一部分的热力图中可以看出，单词where与located，单词where与in等组合具有极高的权重值，由此可以看出模型对问题和段落的语义理解是十分准确的。在第二部分的热力图中可以看出，单词and，以及用于连接的标点符号具有较高的权重值，由此可以看出模型对于回答具有两个部分的情况有良好的表现，而目前问答领域其他的模型倾向于只回答一个部分而忽视连词连接的另一个部分。在第三部分的热力图可以看出，答案的开始和结束位置的单词有着最高的分数，可以很好地理解指针网络的原理。  2.2 前端配置  本项目的前端由Html5，JS，Jquery实现，由两部分组成。第一部分是对两个模型在测试集上的表现，第二部分是API的展示页面。  第一部分是在模型训练后，输入测试集，并将所有的问题的预测结果作为字典的值，问题的ID作为字典的键存储在字典里并导出为Json文件。读取Json文件，以及SQuAD测试集的内容，每篇文章生成一个页面，页面上罗列该文章所有的段落，以及根据段落生成的问题的标准答案和预测答案。使用SQuAD主页提供的实例模板和JS库，可实现在原文中显示标准答案和预测答案的位置。  第二部分是API的实现。使用文本框和输入框接受输入的段落和问题，使用Jquery将段落和问题POST给后端服务器，并接受返回的答案显示在页面上。可视化部分接受权重和指针网络分数数组，传入echarts对象，生成对应热力图。  第三章 实验结果  由于阅读理解模型所得到结果的好坏，评价起来是一件较为主观的事情，为了尽可能地给出一个客观量化的评价数据，我们借助了SQUAD数据集官方所提供的测试数据集以及测试方法。SQuAD官方提供了两类测评分数:EM与F1，其中EM分数是用来评价预测结果与其中一个真实结果相匹配的答案的比值，F1分数衡量预测结果与真实结果的重叠率。本文使用的两类模型与其他的部分模型对比如表3所示。  **表 3 模型性能比较**   |  |  |  | | --- | --- | --- | | 模型 | 验证集(EM/F1) | 测试集(EM/F1) | | **Logistic Regression** | **40.0/51.0** | **40.4/51.0** | | Iterative Coattention Network (Fudan University) | - / - | 67.5 / 76.8 | | T-gating (Peking University) | - / - | 68.1 / 77.6 | | Mnemonic Reader (NUDT & Fudan University) | - / - | 69.9 / 79.2 | | **R-Net** | **72.3/80.6** | **72.3/80.7** |   从上表中的结果可以看出，逻辑回归作为一种传统机器学习技术的算法，在类似于阅读理解这类大量本本数据，以及复杂逻辑的任务中丧失了竞争力。而在基于深度学习技术的方法中，R-net的表现能力超过了其余三个主流模型，最终我们采用R-net网络作为系统的核心模型。 |
| 四**、**项目特色与创新点（不少于300字）  由于SQUAD的数据集的目标是在单独的段落中获取答案，而我们的系统所规划的应用场景是多段落的环境，因此我们需要让针对SQUAD数据集构建的模型符合我们的应用场景。我们在R-net网络的基础上做出了一定的改进，使它可以处理多段落的输入。我们我们首先计算出每个段落与问题的匹配程度:  将输入到softmax层以获得其概率分布:  最后将概率分布结合使用R-Net得到每个段落对问题所得到的指针网络结果，可以得出这些答案的置信度:  上述改进的主要思路是，针对用户提出的问题，我们将之施加于篇章的每一个段落之上，从每一个段落中都获得一个答案，然后比较这些答案与问题的文本相似度，剔除那些答非所问的句子，保留评分最高的答案，作为最终的输出答案。从而让原本针对但段落而设计的问答模型，可以应用到多段落的场景里。 |
| 五、项目总结（不少于500字）  本项目期望实现一个智能问答系统，用户输入一个问题后，系统将问题发送至问答搜索引擎，爬取问答搜索引擎的搜索结果，将搜索结果与问题一同发送至智能问答模型中，之后再获取模型的预测答案，将答案返回给用户。我们期望使用网页界面与用户交互。  我们首先实现了提取特征的逻辑回归模型，经过SQuAD数据集进行训练后，在官方的测试集上取得了50%的正确率。随后调研了智能问答方向最前沿的模型，并选择在微软亚洲研究院提出的R-Net模型的基础上进行改进与完善，实现一个智能问答应用。R-Net模型适用于针对一个具体段落提出的问题，并选取段落原文进行回答，R-Net模型在该问题上的正确率为71%，性能基本达到人类水平。我们将模型进行部分改进，使其可以处理针对由多个段落构成的目标文本提出的问题，并选取目标文本的语句进行回答。  之后我们围绕改进后的模型，实现了应用的功能。我们搭建了前端界面，并且根据改进后模型的一些特性，实现了单段落多段落问答接口，以及智能客服系统和通用型智能问答系统的应用，以及模型的可视化——尝试用于解释深度神经网络理解文本的过程。这些应用均取得了不错的表现。  本项目借助深度学习与人工智能技术，实现了以往传统技术无法达到的应用与效果，最终完成了一个通用型智能问答系统。 |
| 六、项目进度安排 （从项目提出、论证、开展工作、到答辩验收全部过程）  2018.5-2018.7 项目提出，调研相关工作，论证项目可行性  2018.7-2018.8 阅读相关论文，寻找最优的智能客服系统解决方案  2018.8-2018.10 确定使用MSRA的R-Net网络作为最终网络，进行网络  2018.10-2018.12 使用SQuAD数据集训练网络，调整网络参数  2018.12-2019.1 扩展单段落匹配为多段落匹配  2018.1-2019.2 设计前端，结合深度学习模型与现实应用  2019.2-2019.3 填写结题报告，准备答辩 |
| 七、具体成果（包括研制的样机、实物，发表的论文或提交的报告，取得专利的名称、专利人姓名、专利号 、专利批准时间 等详细信息）  成型的智能问答系统，后端使用TensorFlow框架+Flask进行阅读理解，前端使用html+js |
| 八、项目经费明细（不少于200字）  1167.82 办公用具及软件费用  机械硬盘，存储训练数据和模型  显示器，调试网页界面，前端，排版网页内容  拓展坞，外接显示器  正版office word，PowerPoint，Excel软件，Dreamviewer，SQL Server等购买费用    2199 外设  GTX 1060显卡，用于运行深度学习模型训练任务  300 众包测试  众包网站上提交任务，众包测试智能问答系统前端，吞吐量，运行时延等性能  285.51 交通费  往返微软中国研发集团调研模型训练相关事宜  前往北京语言大学语言研究所调研基于规则的智能问答系统发展现状    400 校内网络费和电费  合计 4352.33（元） |
| 九、应用前景评价 （不少于500字）  随着深度学习的发展，机器的智能水平已经在许多领域超越了人类的水平，但是在使用自然语言进行通信交流方面，机器远不及人类的水平，其关键在于机器无法准确的理解蕴含在自然语言中的信息。尽管如此，人们依然提出了许多方法来解决这类问题，并在机器翻译，智能问答等方向研究出了有效的模型策略。  当前，构建具有阅读理解能力或基于真实数据的开放领域问答能力的只能系统是人工智能，特别是自然语言处理领域的一个主要目标。这样的系统能够为社会的方方面面带来巨大价值。目前，已有的产品如Cortana,Siri等均已实现大规模的生产和应用，这得益于近些年深度语音识别技术的快速发展。尽管这些类型的语音助手已经深受欢迎，但消费者们认为如果语音助手能对一个问题给出答案而非只是列出搜索引擎的搜索结果会更加方便实用。同样的，具有阅读理解和问答能力的系统还可以应用到企业的智能客服上面，从而产生巨大的商业效益。  本项目在传统搜索引擎的基础上，使用人工智能技术，把获得搜索结果后信息处理的步骤也进行智能化，直接返回一个确定的结果，这样就为用户省去了面对众多搜所结果自己思考的步骤，提高了搜索效率以及用户的使用效率，因此我们相信我们提出的这种搜索系统比现有的一些搜索引擎更有应用价值，因而具有广泛的应用市场。 |
| 十、存在的问题及建议  目前本系统还存在由两个问题，第一个问题是无法支持中文。引文神经网络训练所使用的数据集是英文的，因此训练获得的模型无法对中文进行很好的适配。解决方法比较简单，只需花费人力标注构建一个中文的问答数据集即可。  另一个问题是模型能力有限，无法应对文本量过大、问题逻辑模糊等较为复杂的情况，这个问题就需要全世界的学者共同努力，将深度学习推向更深更高的层次。目前SQUAD数据集已经发布了2.0版本，并涌现了一批效果优秀的模型，相信随着研究者的努力，人工智能的智能水平会得到长足的提升。 |
| 十一、参考资料（整个工作中参考和查阅的文件、书籍等，按照文献引用的顺序和标准规范列出）  Sainbayar Sukhbaatar, Arthur Szlam, Jason Weston, and Rob Fergus. End-to-end memory networks. In Advances in neural information processing systems, pp. 2440–2448, 2015.  Karm Moritz Hermann, Tom Kocisk, Edward Grefenstette, Lasse Espeholt, Will Kay, Mustafa Suleyman, and Phil Blunsom. Teaching machines to read and comprehend. In Proceedings of the Advances in Neural Information Processing Systems (NIPS), 2015.  Rudolf Kadlec, Martin Schmid, Ondrej Bajgar, and Jan Kleindienst. Text understanding with the attention sum reader network. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, 1:908–918, 2016.  Bhuwan Dhingra, Hanxiao Liu, William W. Cohen, and Ruslan Salakhutdinov. Gated-attention readers for text comprehension. Arxiv, 2016  Yiming Cui, Zhipeng Chen, Si Wei, Shijin Wang, Ting Liu, and Guoping Hu. Attention-overattention neural networks for reading comprehension. Arxiv, 2016  Rajpurkar P, Zhang J, Lopyrev K, et al. SQuAD: 100,000+ Questions for Machine Comprehension of Text[J]. 2016:2383-2392.  Shirakawa M, Hara T, Nishio S. N-gram IDF: A Global Term Weighting Scheme Based on Information Distance[J]. 2015:960-970.  R-NET: Machine Reading Comprehension with Self-matching Networks  Mikolov T, Karafiát M, Burget L, et al. Recurrent neural network based language model[C]// INTERSPEECH 2010, Conference of the International Speech Communication Association, Makuhari, Chiba, Japan, September. DBLP, 2010:1045-1048.  Vinyals O, Fortunato M, Jaitly N. Pointer Networks[J]. Computer Science, 2015.  Cho K, Van Merrienboer B, Gulcehre C, et al. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation[J]. Computer Science, 2014.  Rocktäschel T, Grefenstette E, Hermann K M, et al. Reasoning about Entailment with Neural Attention[J]. 2015. |
| 十二、指导教师意见  （学生工作的意义、项目的难度和工作量，本项目的特点以及学生对项目的理解与独创性等方面的评价）  签名：  年 月 日 |
| 十三、专家组评审意见      签名：  年 月 日 |
| 十四、学校意见  负责人（签名、公章）：  年 月 日 |