

文章编号: 1003-0077(2020)12-0073-09

基于 CFN 和篇章主题的概括型问答题的解答

杨陟卓¹, 李春转¹, 张 虎¹, 钱揖丽¹, 李 茹^{1,2}

(1. 山西大学 计算机与信息技术学院, 山西 太原 030006;

2. 山西大学 计算智能与中文信息处理教育部重点实验室, 山西 太原 030006)

摘 要: 相对于普通阅读理解, 高考语文阅读理解难度更大, 问句更加抽象, 答案候选句的抽取除了注重与问句的相似性分析, 还注重对材料内容以及作者的观点的概括归纳。因此该文提出了利用汉语框架网 (Chinese FrameNet) 抽取与问句语义相似的候选句的方法, 通过识别篇章主题 (段落主题句和作者观点句), 生成与问句相关的内容要点以及作者的观点态度, 最终选取 top 6 作为答案句。在近 12 年北京市高考真题上进行测试, 召回率达到了 68.69%, 验证了该方法的有效性。

关键词: 概括型问答题; 框架语义; 篇章主题; 答案句抽取

中图分类号: TP391

文献标识码: A

Question Answering for Overview Questions Based on CFN and Discourse Topic

YANG Zhizhuo¹, LI Chunzhuang¹, ZHANG Hu¹, QIAN Yili¹, LI Ru^{1,2}

(1. School of Computer and Information Technology, Shanxi University, Taiyuan, Shanxi 030006, China;

2. Key Laboratory of Computation Intelligence and Chinese Information

Processing of Ministry of Education, Shanxi University, Taiyuan, Shanxi 030006, China)

Abstract: Reading comprehension QA for College Entrance Examination on Chinese is much challenging due to the fact that the questions are more abstract. In addition to the question similarity analysis, the extraction of answer candidate sentences should also pay more attention to the topic and opinion sentences. This paper proposes to extract the candidate answer sentences by frame semantic match and frame semantic relation. By identifying the discourse topic sentences, the topic and opinion sentences related to the questions are generated. Then the top-six candidate answers are selected based on ranking results. In the experiment, the recall of the method on the College Entrance Examination of Beijing in recent twelve years is 68.69%, which verifies the effectiveness of the method.

Keywords: general-problems; frame semantic; discourse topic; answer sentence extraction

0 引言

近年来问答技术研究受到了广泛的关注, 目前发展比较成熟的问答系统有麻省理工学院人工智能实验室研发的 Start、密歇根大学的 AnswerBus、苹果公司研发的 Siri 和微软公司开发的 Cortana 系统。国内也推出一些以智能问答技术为核心的机器人, 例如微软小冰、百度小度等, 但是基于阅读理解的问答研究还远远不够。为此, 国家重点研发计划重点

专项课题“文本生成及语言问题求解关键技术与系统”提出研究面向高考的类人答题系统。

根据《2019 年普通高等学校招生统一考试大纲》, 对高考阅读理解的考试范围及要求如下: (1) 理解: ①理解文中重要概念的含义; ②理解文中重要句子的含意。(2) 分析综合: ①筛选并整合文中的信息; ②分析文章结构, 归纳内容要点, 概括中心意思; ③分析论点、论据和论证方法; ④分析概括作者在文中的观点态度。本文主要研究概括型问答题, 问题样例如表 1 所示。

收稿日期: 2019-12-16 定稿日期: 2020-04-07

基金项目: 国家重点研发计划(2018YFB1005103); 国家自然科学基金(61772324)

表 1 高考语文概括型问答题样例

2018 年北京高考语文第 7 题
问句: 根据材料一、材料二, 简要说明人类对人工智能的认识是如何不断深化的。
参考答案对应的答案句(来自背景材料中):
(1) 材料一第二段: 1936 年, 英国数学家阿兰·麦席森·图灵从模拟人类思考和证明的过程入手, 提出利用机器执行逻辑代码来模拟人类的各种计算和逻辑思维过程的设想。
(2) 材料一第三段: 20 世纪 80 年代, 美国哲学家约翰·希尔勒教授用“中文房间”的思维实验, 表达了对“智能”的不同思考。(段落主题句)
(3) 材料一第四段: 在人工智能研究领域中, 不同学派的科学家对“何为智能”的理解不尽相同。(段落主题句)
(4) 材料一第五段: 但是, 正是对“何为智能”这个核心问题的不断思考和解答, 推动了人工智能技术在不同层面的发展。(作者观点句)
(5) 材料二第一段: 2018 年 5 月, 谷歌 Duplex 人工智能语音技术(部分)通过了“图灵测试”。
(6) 材料二第五段: 尽管研究者对“奇点”到来的时间和可能性还有争议, 但是不管“奇点”时刻能否真的到来, 在技术不断完善的过程中, 我们都要小心被人工智能“异化”。(段落主题句)
(7) 材料二第六段: 面对人工智能可能带来的种种冲击, 上世纪 50 年代美国科幻小说家阿西莫夫提出的机器人三大定律, 今天对我们依然有借鉴意义。(作者观点句)

可以看出, 相对于一般阅读理解任务, 高考阅读理解问答题的问句更具多样性、抽象性, 不可以简单地划分为 who, where, when, what, why 五种类型。同时, 问答题的答案句在字面上与问句相关性不大, 通过词语匹配的方法很难将这些句子召回。但是, 将参考答案对应到材料中的答案句发现: 有些答案句不是分布于某一段落, 而是分布于不同材料的不同段落, 且是从不同方面对整个篇章的概括归纳。如何将这些答案句召回, 是一个巨大的挑战, 也是提高高考阅读理解问答题得分率的关键。

1 相关研究

近年来, 随着人工智能的快速发展, 基于阅读理解的问答任务越来越受到国内外研究者的青睐。首次提出阅读理解系统的是 MITRE 公司, 他们开发了 Deep Read 系统^[1]。之后 Riloff 构建了 Quarc 系统^[2], 该系统是一个基于启发式规则的系统。文献^[3]提出将故事与现实世界背景知识相结合回答问

题的方法。

目前, 主流方法主要包含基于检索模型和基于深度学习两大研究方向^[4]。传统的基于统计特征的检索方法, 例如, 文献^[5]借助 WordNet 获取更多的词汇语义信息。文献^[6]提出将问题和所有候选句的语义角色标注结果表示成树状结构, 用树核的方法计算问题和每个候选句之间的语义结构相似度。文献^[7]提出了基于语义依存树和改进编辑距离相结合的句子相似度计算方法, 文献^[8]提出一种基于同义词词林的句子语义相似度方法, 通过对词形、词序、语义依存相似度加权结合获得句子之间的最终相似度。文献^[9]根据问句和候选答案句的对应关系, 抽取 10 个特征, 采用最大熵模型对中文阅读理解问题回答进行建模。

基于深度学习的方法, 例如, 文献^[10]使用多层 Bi-LSTM 和关键字匹配相结合的模型完成答案选择任务。文献^[11]在 BIDAf 模型的嵌入层和交互层分别融入了问题类型和问题主题、问题焦点三种问题特征, 在 DuReader 数据集上进行测试, ROUGE-L 和 BLEU-4 分别达到了 43.25% 和 40.02%。文献^[12]提出 BERT 模型, 该模型预先训练未标记文本的深层双向表示, 只需一个额外的输出层, 就可以对预先训练的 BERT 模型进行微调, 以创建用于广泛任务(如问答和语言推理)的最新模型, 而无须对特定任务的体系结构进行大量修改。

另外, 文献^[13]提出针对高考语文文意理解选择题, 利用基于词袋模型的词匹配、HowNet 相似度、Word2Vec 相关度及 CFN 的多维度投票算法的句子语义相关度计算方法选取答案的最佳选项。文献^[14]提出基于框架语义相似性、有定零形式线索以及框架关系的答案抽取方法。文献^[15]提出借助框架语义匹配和框架语义关系抽取答案候选句, 在排序时引入流形排序的答案句抽取方法。

目前, 基于检索模型和深度学习的方法在一般阅读理解问题上已经取得了很好的效果^[16], 但是在高考语文阅读理解题上的效果却不尽如人意, 这是由于高考阅读理解问题更加复杂, 需要对问题和背景材料进行深层次的语义理解和推理。此外, 基于深度学习的方法虽然可以自动学习特征, 但是需要足够规模的数据进行训练。本文针对高考问答题的特点, 提出一种弱监督的方法, 首先通过词语相似度匹配的方法抽取浅层语义匹配的答案候选句, 然后通过汉语框架网获取语义场景下的语义信息, 查找

在背景材料中与问题语义场景一致或相关的候选句。最后,通过抽取篇章主题,包括段落主题句和作者观点句,弥补语义信息难以匹配的需要深层次语义理解、推理的答案候选句。本文提出的方法不仅在高考语文阅读理解中具有较好的答题效果,也可以应用在文本蕴含、自动文摘及复杂问题求解等自然语言处理任务中。

2 方法

通过分析试题发现,少数问答题答案句集中在材料中的某一段落,而多数问答题的答案句分布在分散的段落中。例如表 1 中案例的答案分散于多个段落。阅读理解问答题较多地考察考生对文章归纳内容要点、概括中心思想以及分析概括作者在文中的观点态度的能力。对于此类问题,本文对其定义为概括型问答题。它的特点是:①需要筛选并整合文中与问句相关的重要信息;②需要理解并概括材料内容;③需要分析概括作者在文中的观点态度。本文提出一种针对概括型问答题的解题模型,因此需要对问答题的类型进行判别。通过研究大量高考真题后发现,概括型问答题符合以下四种特征:

(1) 问句类型(抽象或具体):概括型问答题的问句类型一般比较抽象。通过统计各省的高考真题及模拟题,发现抽象型问答题包括“意义”“特点”“作用”“认识”“观点”“启发”等关键词。

(2) 段落匹配数目:对问句与背景材料句子进

行匹配,统计每段的匹配情况。如果匹配的段落数目超过预定的阈值,则符合概括型问答题的特征。如表 1 中间句内的关键词在背景材料的很多段落中出现。

(3) 问句出处段^[15]句子中的数目:根据关键字匹配定位问句出处段,并统计该段的句子数目,如果超过预定的阈值,则符合概括型问答题的特征。

(4) 背景材料的段落和句子数目:若段落数或者句子数超过预定的阈值,则符合概括型问答题的特征。例如 2018 年北京高考题段落和句子数目较多,分别是 11 段和 48 句。

最后,将各个特征通过线性插值的方法进行计算,超过预定阈值,判断为概括型问答题。

阅读理解问答题解题过程可以形式化为:给定问题 Q 与背景材料 D ,本文首先进行预处理,由于背景材料是一则或多则,因此先将背景材料按照所属材料、段落、句子进行形式化处理,即 $D = \langle S_{d_i p_j s_k} \rangle$,其中 d_i 表示所属第 i 则材料, p_j 表示所属第 j 段落, s_k 表示所属第 k 句;对问句进行分词去停用词,得到问句关键词,即 $Q = \langle K_1, K_2, \dots, K_m \rangle$, K_i 表示第 i 个关键词,然后根据以上四种特征判断问题的类型。如果该问题是概括型问答题,根据基于词语相似度匹配、框架匹配、篇章主题的方法计算各个候选句 S 的分数;如果该问题不是概括型问答题,则根据词语相似度匹配和框架匹配的方法计算答案候选句的分数;最后根据候选句的分数高低进行排序,选取前六句作为答案句集 S_A 。技术路线图如图 1 所示。

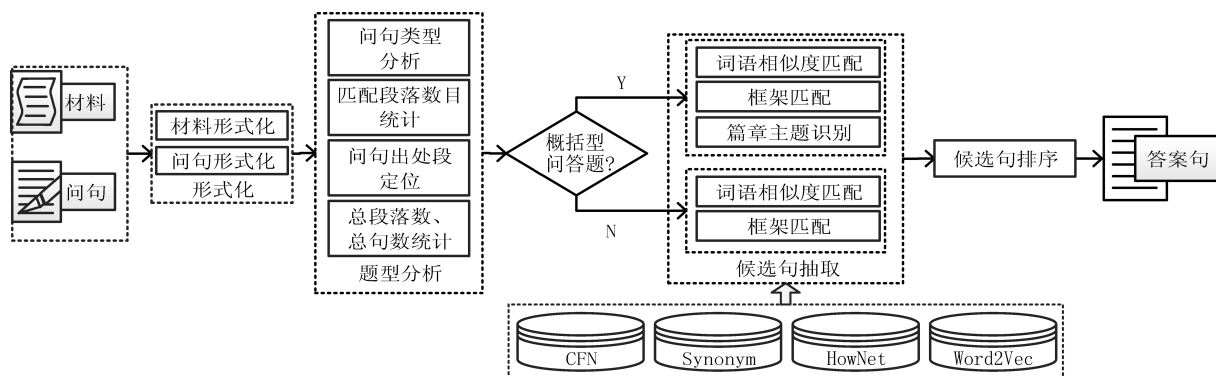


图 1 技术路线图

2.1 基于词语相似度匹配的候选句抽取

基于词语相似度匹配的候选句抽取是从较为浅层的语义上抽取候选句,本文融合 HowNet 和

Word2Vec 词向量计算词汇间相似度,如算法 1 所示。

算法 1 基于 HowNet 和 Word2Vec 的词语相似度匹配

```

输入: 问句 Q、背景材料句子 S、词语相似度阈值
输出: 句子 S 与问句 Q 的词语匹配数 sumword
1. 对问句、句子分词去停用词  $Q = \{K_1, K_2, K_m\}$ 、 $S = \{W_1, W_2, \dots, W_n\}$ , 并初始化匹配数  $\text{sumword} = 0$ 、各关键字与问句关键字相似度  $\text{simword} = 0$ 
2. FOR  $K_i$  IN  $Q$ 
3.   FOR  $W_j$  IN  $S$ 
4.     IF  $K_i = W_j$ 
5.        $\text{sumword}++$ ;
6.     ELSE
7.        $\text{simword} = 0.4 \times \text{simHowNet}(K_i, W_j) + 0.6 \times \text{Cos}(K_i^v, W_j^v)$ ;
8.     END IF
9.   IF  $\text{simword} > \text{词语相似度阈值}$ 
10.     $\text{sumword}++$ ;
11.    break;
12.   END IF
13. END FOR
14. END FOR
    返回 sumword
  
```

其中, $\text{simHowNet}(K_i, W_j)$ 表示通过 HowNet^[17] 计算问句关键词 K_i 和句子关键词 W_j 的相似度。利用 Word2Vec 计算词向量的余弦相似度^[18], 如式(1)所示。

$$\cos(K_i^v, W_j^v) = \frac{K_i^v \cdot W_j^v}{\|K_i^v\| \times \|W_j^v\|} \quad (1)$$

其中, K_i^v 与 W_j^v 分别为 K_i 与 W_j 的词向量。

随后对 sumword_i 进行归一化处理, 最终基于词语相似度匹配的计算如式(2)所示。

$$\text{score}_{\text{sumword}} = \frac{\text{sumword}_i}{\{\max_{1 \leq i \leq n}(\text{sumword}_i) - \min_{1 \leq i \leq n}(\text{sumword}_i)\}} \quad (2)$$

2.2 基于框架匹配的答案句抽取

由于基于词语相似度的方法不能挖掘到句子的深层语义信息, 本文利用汉语框架网(CFN)^[19] 捕捉语义场景下的语义信息。CFN 是山西大学以 Fillmore 的框架语义学^[20] 为基础, 以加州大学伯克利分校的 FrameNet^[21] 为参照建立的汉语词汇语义知识库。其中, 框架^[19] 是与一些激活性语境相一致的结构化范畴, 它是存储在人类认知经验中的图示化情景, 是理解词语的背景和动因。框架元素^[19] 是由某个目标词激起的框架下的一类词语, 这些词语具有共同的认知结构, 支配相同的语义角色。目标词^[19] 是句子中能激起框架的词元。框架关系^[19] 是指基于框架语义学建立的语义场景下的框架与框架

之间的关系, 主要包括继承、透视、使用、总分、起始、参照和因果 7 种关系。

首先定位问句出处句, 然后根据问句出处句扩展问句目标词, 从而扩展目标词激起的框架, 最后借助框架语义匹配和框架语义关系抽取候选句^[15]。

(1) 框架语义匹配: 当问句目标词激起的框架与句子 $S_{d_i p_j s_k}$ 的目标词激起的框架相同时, 匹配数加一。

(2) 框架语义关系匹配: 当问句目标词激起的框架与句子 $S_{d_i p_j s_k}$ 的目标词激起的框架之间的距离小于等于 2 时, 匹配数加一。最终得到句子 $S_{d_i p_j s_k}$ 与问句 Q 的框架匹配数 sumframe_i 。然后对其归一化处理, 最终基于框架匹配计算如式(3)所示。

$$\text{score}_{\text{sumframe}} = \frac{\text{sumframe}_i}{\{\max_{1 \leq i \leq n}(\text{sumframe}_i) - \min_{1 \leq i \leq n}(\text{sumframe}_i)\}} \quad (3)$$

基于框架匹配的候选句抽取示例如图 2 所示。问句 Q 中目标词“启发”激起的框架与候选句 S 中目标词“激发”激起的框架相同; 问句 Q 中目标词“关系”激起框架与候选句 S 中的目标词“有利于”激起的框架存在继承关系。问句 Q 与候选句 S 涉及到的语义场景非常相似。因此, 基于框架匹配的方法将该句抽取出来, 作为答案候选句。

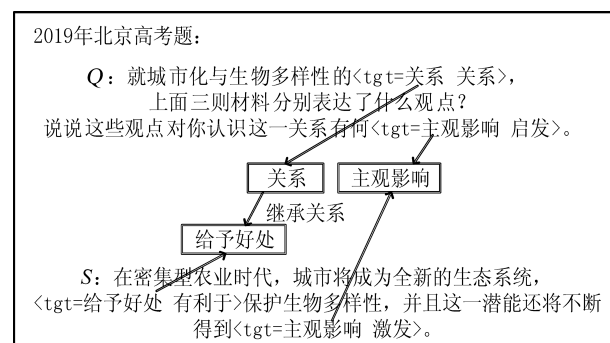


图 2 基于框架匹配的候选句抽取示例

2.3 基于篇章主题的答案句抽取

通过研究考试大纲发现, 高考语文经常考察考生归纳内容要点、概括中心意思的能力, 本文针对高考语文问答题的特点, 提出了基于篇章主题的候选句抽取方法, 该方法可以将一些与问句关键字和框架不匹配的答案候选句召回。篇章主题包括段落主题句和作者观点句。

2.3.1 段落主题句的抽取

例 1 【2017 年北京高考题】

Q: 请结合上述三则材料, 简述让文物“活”起来的含义与作用。

S: 移动应用程序(APP)以新鲜时尚的方式提供给观众审美与求知、娱乐与鉴赏的多元文化体验。

在例 1 中, S 是所处段落的段落主题句。通过研究大量的科技文材料发现, 主题句通常位于此段的段首或段尾, 同时该句子与此段落中的其他句子主题相关, 因此本文通过位置信息和句子之间的语义相似性计算各个句子的主题句分数。

(1) 位置信息

段落主题句是对段落内容的概括总结, 是反映段落主旨的句子, 一般分布在段落的开头或者结尾, 因此根据句子的位置信息对段落中的各句子进行计算, 如式(4)所示。

$$\text{score}_i = \begin{cases} 1 & i = 1, n \\ 1 - \frac{\log i}{\log n} & \text{其他} \end{cases} \quad (4)$$

其中 i 为句子编号, n 为每段的总句数。

对于不同段落, 一般来说, 材料的首段和尾段可以反映材料的主题, 因此首尾段落的权重应该更大, 按式(5)根据段落的位置对各段落的主题句进行计算:

$$\text{score}_{\text{loc}} = \begin{cases} 0.7 \times \text{score}_i & i = 1 \text{ 或 } i = m \\ 0.3 \times \text{score}_i & \text{其他} \end{cases} \quad (5)$$

其中 m 为材料的总段落数。

(2) 基于段落的句间语义相似性

句子 A 的关键词为 A_i , 共 p 个, 句子 B 的关键词为 B_j , 共 q 个。

在计算句间相似度时, 采用 HowNet 分别计算句子相似度。两个词基于 HowNet 的相似度为 $S(A_i, B_j)$, $a_i = \max\{S(A_i, B_1), S(A_i, B_2), \dots, S(A_i, B_q)\}$, $b_j = \max\{S(B_j, A_1), S(B_j, A_2), \dots, S(B_j, A_p)\}$, 则基于 HowNet 的句子相似度^[17]如式(6)所示。

$$\text{sim}(A, B) = \frac{\frac{\sum_{i=1}^p a_i}{p} + \frac{\sum_{j=1}^q b_j}{q}}{2} \quad (6)$$

则基于段落的句子 A 语义相似性如式(7)所示。

$$\text{score}_{\text{sim}} = \frac{\sum_{x=1}^n \text{sim}(A, B_x)}{n} \quad (7)$$

其中 n 为每段的总句数。

最终对上述两种因素进行加权, 得到各个句子段落主题句计算, 如式(8)所示。

$$\text{score}_{\text{topic}} = \alpha_1 * \text{score}_{\text{loc}} + \alpha_2 * \text{score}_{\text{sim}} \quad (8)$$

其中, $\alpha_1 + \alpha_2 = 1$ 。

2.3.2 作者观点句的抽取

作者在文中的观点态度经常出现在参考答案中, 观点句主要表明作者在文中的观点态度, 是对材料内容的总体把握, 是整个篇章的主题。位置信息、句间语义相似性以及提示性词是观点态度句的重要特征。例如表 1 中材料二第六段, 首先该句是文章尾段落的首句, 其次该句子与其他句子主题相似, 表明作者在整篇文章中的观点态度。

(1) 位置信息

通过分析试题发现, 句子在文中出现的位置很重要。作者的观点一般分布在材料的尾段, 因此将尾段作为候选句段落, 同时根据尾段中句子的不同位置进行计算, 计算公式如式(4)所示, 记为 score_i 。

(2) 句间语义相似性

为了避免只根据位置信息带来的主观性, 在抽取尾段作者观点句时, 进行句间语义相似度计算, 计算公式如式(7)所示。

(3) 基于提示性词的启发式规则

根据句子是否包含提示性词抽取候选句, 若包含 $\text{score}_{\text{word}} = 1$, 否则 $\text{score}_{\text{word}} = 0$ 。本文通过同义词词林扩充提示性词表, 提示性词示例如表 2 所示。

表 2 提示性词示例

看来、由此可见、由此看来、可见、无论如何、不管怎样、综上所述、由上述可知、如上所述、总的来看、总的来说、总的说来、总之、总而言之、总体而言、首先、其次、表明、所以

最后对上述三种因素进行加权, 得到作者观点句计算, 如式(9)所示。

$$\text{score}_{\text{opinion}} = \beta_1 * \text{score}_i + \beta_2 * \text{score}_{\text{sim}} + \beta_3 * \text{score}_{\text{word}} \quad (9)$$

其中, $\beta_1 + \beta_2 + \beta_3 = 1$ 。

2.4 根据问答题类型采取不同策略

如果问答题是概括型问答题, 本文利用基于词语相似度匹配、框架匹配、篇章主题的方法对各个句子分数进行计算, 候选句最终计算如式(10)所示。

$$S = \lambda_1 * \text{score}_{\text{sumword}} + \lambda_2 * \text{score}_{\text{sumframe}} + \lambda_3 * \text{score}_{\text{topic}} + \lambda_4 * \text{score}_{\text{opinion}} \quad (10)$$

否则,只利用词语相似度匹配和框架匹配的方法进行计算,如式(11)所示。

$$S = \lambda_1 * \text{score}_{\text{sumword}} + \lambda_2 * \text{score}_{\text{sumframe}} \quad (11)$$

其中, λ_k 为第 k 个维度的权重, $k \in [1, K]$, 且

$$0 \leq \lambda_k \leq 1, \sum_{k=1}^K \lambda_k = 1。$$

3 实验及结果分析

3.1 实验数据

实验采用哈尔滨工业大学的语言技术平台 LTP^[22] 进行分词和词性标注,使用山西大学的 CFN^[19] 知识库,以及知网平台提供的词汇相似度计算工具^[17]。

由于高考问答题在高考阅读理解中所占比例较小,本文采用的数据集包括各省的高考真题、模拟题以及选择题改造成的问答题。最终在各省近 12 年高考真题上形成 132 个问句,在各省高考模拟题上形成 511 个问句。

3.2 实验结果及分析

3.2.1 不同实验方法结果比较

目前高考问答题按答题要点给分,本文采取准确率 P 、召回率 R 和 F_1 值作为评价指标;采用五倍交叉实验,将语料平均分成五份,使用其中一份作为测试集,其他四份作为训练集,重复五次实验,取平均值作为最终结果。

本文按照试题所给的参考答案人工找到其在阅读材料中对应的多个句子,标记为答案句的集合 A , S_A 是使用本文方法,根据分数从高到低排序的前 6 句的集合。

$$P(\text{准确率}) = \frac{S_A \text{ 中答对的句子总数}}{S_A \text{ 的句子总数}} \times 100\% \quad (12)$$

$$R(\text{召回率}) = \frac{S_A \text{ 中答对的句子总数}}{A \text{ 的句子总数}} \times 100\% \quad (13)$$

$$F_1 = \frac{2PR}{P+R} \quad (14)$$

为了验证本文方法的有效性,使用文献^[15]的方法(框架匹配)作为 baseline1,该方法在本文 2.2 节进行阐述。同时为了与国际上阅读理解任务中效果较好的方法相比较,本文将 BERT 模型^[12] 作为 baseline2。由于原始的 BERT 模型只能输出连续的答案区间,而高考问答题的答案区间通常是不连续的,因此本文利用 BERT 模型对候选句进行二分

类,即判断背景材料中每个候选句是否为答案句。利用除北京外各省高考真题和模拟题(包括 122 道高考真题和 511 道高考模拟题,根据标准答案人工标记背景材料中的答案句)训练模型。各种方法在北京近 12 年真题上进行对比,实验结果如表 3 所示。

表 3 不同方法答案句抽取结果比较

方法	$P/\%$	$R/\%$	$F_1/\%$
baseline1(框架匹配)	33.33	50.48	40.15
baseline2(BERT)	35.30	39.50	37.28
词语相似度匹配	33.33	48.57	39.53
段落主题句	33.33	39.41	36.12
作者观点句	31.67	38.45	34.73
词语相似度匹配+框架匹配	36.67	52.74	43.26
词语相似度匹配+框架匹配+段落主题句+作者观点句	50.00	63.69	56.02
根据问答题类型采取不同策略	51.67	68.69	58.98

在实验中,算法 1 中输入的词语相似度阈值为 0.7;框架匹配方法抽取答案句根据式(3);词语相似度匹配方法抽取答案句根据式(2);段落主题句方法抽取答案句根据式(8),其中 $\alpha_1=0.7, \alpha_2=0.3$;作者观点句方法抽取答案句根据式(9),其中 $\beta_1=0.3, \beta_2=0.1, \beta_3=0.6$;根据问答题类型采取不同策略:如果是概括型问答题采用式(10), λ_k 设置为 0.3、0.2、0.3、0.2,否则采用式(11), λ_k 设置为 0.6、0.4。

可以发现 BERT 模型的召回率仅仅为 39.50%,说明 BERT 模型应用在高考题上效果欠佳,高考题相对普通阅读理解问答题难度更大,在目前的训练数据规模上,不足以训练一个高效的高考答题模型。

当使用词语相似度匹配+框架匹配+段落主题句+作者观点句方法时,抽取前 6 个答案句的准确率、召回率以及 F_1 值分别达到了 50.00%、63.69% 和 56.02%;当判断问答题类型而采取不同策略时,准确率、召回率和 F_1 值有了进一步的提升。实验结果表明,本文提出的方法确实可以提高高考语文阅读理解问答题的得分率。

3.2.2 问答题题型判断实验结果

本文将问题是概括型问答题的情况作为正例,否则作为负例,采用式(14)~式(16)作为评价指标,在训练集和测试集上进行判别,实验结果如表 4 所示。

$$P(\text{准确率}) = \frac{TP}{TP+FP} \quad (15)$$

$$R(\text{召回率}) = \frac{TP}{TP + FN} \quad (16)$$

其中, TP 代表预测值为 1, 真实值为 1; FP 代表预测值为 1, 真实值为 0; FN 代表预测值为 0, 真实值为 1。

表 4 概括型问答题判别结果

数据集	P/%	R/%	F ₁ /%
训练集	80.95	83.95	82.42
测试集	86.87	77.25	81.77

由表 4 可见, 本文的方法可以有效区分问答题的类型。通过分析错误原因发现: 本文提出的四种特征都有可能造成错误判断, 比如特点类问句在大多数情况下属于概括型问答题, 但有一小部分却是其他类型问答题; 问句的抽象表达会造成问句出处段定位的不准确; 对各个特征阈值的设定在大多数情况下比较准确, 在个别试题上存在判断错误的情况。

3.2.3 抽取段落主题句和作者观点句的准确率

对北京 12 年高考真题中的段落主题句和作者观点句进行标注, 共 19 则材料, 89 个段落主题句, 26 个作者观点句。利用本文方法对北京 12 年高考真题进行测试, 实验结果如表 5 所示。

表 5 段落主题句和作者观点句实验结果

方法	P/%
段落主题句识别	80.62
作者观点句识别	75.00

通过分析试题发现, 北京高考科技文与一般新闻类型的文章相比, 段落主题句的抽取难度要大, 如例 2 所示: 该段的主题句是“演唱昆曲是厅堂里的事情”, 是段落的精简概括。但这个句子与其他句子的相似度较小, 需要更深层的语义推理技术。作者观点句抽取难度在于: 有部分阅读材料没有明确的作者观点句。例如, 例 2 中全文共四段, 首段引出昆曲, 接下来三段从不同角度说明昆曲的长处与局限, 但是没有明显的概括性的观点态度。

例 2 【2009 年北京高考题】

演唱昆曲是厅堂里的事情。地上铺了一方红地毯, 就算是剧中的境界; 唱的时候, 笛子是主要的乐器, 声音当然不会怎么响, 但是在一个厅堂里, 也就各处听得见了。搬上旧式的戏台去, 即使在一个并不宽广的戏院子里, 就不及平剧那样容易叫全体观众听清。如果搬上新式的舞台去, 那简直没有法子听, 大概坐在第五六排的人就只看见演员拂袖按鬓了。

3.2.4 不同省份真题及模拟题实验结果比较

本文在不同省份真题及模拟题上同样采用五倍交叉实验。不同省份的概括型问答题存在差异: 北京高考概括型问答题的背景材料为科技文, 而其他省份的概括型问答题的背景材料多是议论文、学术论文、时评、书评、新闻、传记、报告、科普文等。此外, 北京高考问答题大多考查学生筛选并整合文中信息的能力, 别的省份大多考查学生对重要词句的理解、把握文章结构与整体思路的能力。因此在不同省份上得到的实验结果差异较大。不同省份真题及模拟题上的召回率如图 3 所示。

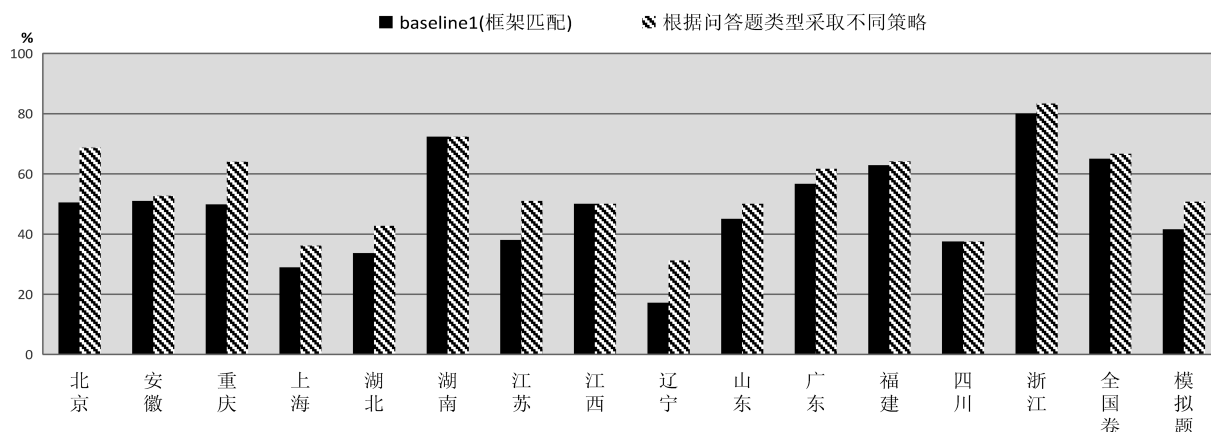


图 3 不同省份真题及模拟题上的召回率

从图 3 可以发现, 利用本文方法在不同省份真题及模拟题上召回率均有不同程度的提高, 同时可以发现有些省份的召回率非常低, 例如例 3。

例 3 【2007 年上海高考题】

问句: 包豪斯艺术能“在人们的生活中留下更长久的印记。”的原因是什么?

答案句：学校强调自由创造，反对模仿因袭。

在设计中注重实用需求，摒弃虚浮奢华；讲究材料自身的质地与色彩，反对附加的修饰与包装。

他们注重发挥结果本身的形式美，采用不对称的构图法，灵活多样、造型简洁。

他的任务，是要在一个普通的水杯上探索和体现美的极致，让美贯穿到普通民众生活的每一个平凡细节当中。

分析其原因发现：①背景材料几乎每句都在围绕“包豪斯”展开，根据关键字“包豪斯”“印记”进行匹配会带来严重的干扰，很多与该问句相关的候选句都不是答案句，需要深层次的语义理解和推理技术。②利用框架匹配发现问句中的“原因”与答案句中的“强调”“注重”“反对”“摒弃”等具体原因词存在着较大的语义鸿沟，利用现有的汉语框架网无法进行语义匹配。③此题不是概括型问答题，无法利用基于篇章主题的方法抽取答案候选句。

4 结论

本文针对高考语文阅读理解问答题，提出了有效的解题策略。首先判断问答题的类型，如果是概括型问答题，则采取基于词语相似度匹配、框架匹配、篇章主题的方法同时抽取候选句，如果是其他类型的问答题，则采取基于词语相似度匹配和框架匹配的方法抽取候选句。然后根据分数对候选句排序，最终选取 top6 的候选句作为答案句。在接下来的工作中，一方面从材料的句间关系入手，对背景材料进行深层语义理解以及推理，研究更加适合的答案句抽取方法；另一方面，由于现有的方法是抽取式模型，而有些答案句是根据材料生成的，因此要研究答案生成模型，进一步提高答题效果。

参考文献

- [1] 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. Association for Computational Linguistics, 1999: 325-332.
- [2] 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. As-

- sociation for Computational Linguistics, 2000: 13-19.
- [3] Charnick E. Toward a model of children's story comprehension[D]. PhD Diss., Cambridge! Massachusetts Int of Tech, 1972.
- [4] 安炜杰. 融合知识表示的自动问答系统关键技术研究[D]. 上海: 华东师范大学硕士学位论文, 2018.
- [5] Yih W, Chang M W, Meek C, et al. Question answering using enhanced lexical semantic models[C]//Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2013: 1744-1753.
- [6] 张志昌, 张宇, 刘挺, 等. 基于浅层语义树核的阅读理解答案句抽取[J]. 中文信息学报, 2008, 22(1): 80-86.
- [7] 李旭峰. 中文问答系统中问句理解和相似度计算的研究与实现[D]. 广州: 华南理工大学硕士学位论文, 2010.
- [8] 周艳平, 李金鹏, 蔡素. 基于同义词词林的句子语义相似度方法及其在问答系统中的应用[J]. 计算机应用与软件, 2019, 36(8): 65-68, 81.
- [9] 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.
- [10] Wang D, Nyberg E. A long short-term memory model for answer sentence selection in question answering [C]//Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers), 2015: 707-712.
- [11] 谭红叶, 刘蓓, 王元龙. 基于 QU-NNs 的阅读理解描述类问题的解答[J]. 中文信息学报, 2019, 33(3): 102-109.
- [12] Devlin J, Chang M W, Lee K, et al. BERT: Pre-training of deep bidirectional transformers for language understanding[C]//Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long and Short Papers), 2019: 4171-4186.
- [13] 郭少茹, 张虎, 钱揖丽, 等. 面向高考阅读理解的句子语义相关度[J]. 清华大学学报(自然科学版), 2017, 57(6): 575-579.
- [14] 王智强, 李茹, 梁吉业, 等. 基于汉语篇章框架语义分析的阅读理解问答研究[J]. 计算机学报, 2016, 39(04): 795-807.
- [15] 李国臣, 刘姝林, 杨陟卓, 等. 基于框架语义的高考语文阅读理解答案句抽取[J]. 中文信息学报, 2016, 30(6): 164-172.
- [16] Wang Y, Liu K, Liu J, et al. Multi-passage machine reading comprehension with cross-passage answer verification [C]//Proceedings of the 56th Annual Meeting of the Association for Computational Lin-

- guistics(Volume 1: Long Papers),2018: 1918-1927.
- [17] 刘群,李素建.基于《知网》的词汇语义相似度[J].中文计算语言学,2002,7(2): 59-76.
- [18] 刘敏.基于词向量的句子相似度计算及其在基于实例的机器翻译中的应用[D].北京:北京理工大学硕士学位论文,2015.
- [19] 李茹.汉语句子框架语义结构分析技术研究[D].太原:山西大学博士学位论文,2012.
- [20] Fillmore C J.Frame semantics and the nature of language[J].Annals of the New York Academy of Sciences: Conference on the Origin and Development of Language and Speech,1976, 280(1): 20-32.
- [21] Baker C F, Fillmore C J, Lowe J B. The Berkeley framenet project[C]//Proceedings of the 17th International Conference on Computational Linguistics-Volume 1.Association for Computational Linguistics, 1998: 86-90.
- [22] Che W X, Li Z, Liu T. LTP: A Chinese language technology platform[C]//Proceedings of the 23rd International Conference on Computational Linguistics: Demonstrations. Association for Computational Linguistics,2010: 13-16.



杨陟卓(1983—),博士,副教授,主要研究领域为自然语言处理、阅读理解问答。

E-mail: yangzhizhuo@sxu.edu.cn



李春转(1996—),硕士研究生,主要研究领域为自然语言处理。

E-mail: lcz_124@sina.com



张虎(1979—),博士,副教授,主要研究领域为自然语言处理、大数据和信息安全。

E-mail: zhanghu@sxu.edu.cn

“中国中文信息学会首批会士”名单公布

12月27日举行的中国中文信息学会2020学术年会上,中国中文信息学会首批会士名单对外公布,倪光南等6位院士专家榜上有名。

中国中文信息学会会士的设立,旨在表彰在中文信息处理领域做出突出贡献并取得卓越成就的科学家。本次年会上,中国科协学会学术部杨书宣副部长宣布了“中国中文信息学会首批会士”名单,分别是中国中文信息学会前理事长、中国科学院计算技术研究所倪光南院士;中国中文信息学会前理事长、名誉理事长、前党委书记李生教授;中国中文信息学会前常务理事、海军计算技术研究所沈昌祥院士;中国中文信息学会前常务理事、北京大学何新贵院士;中国中文信息学会理事长、中国电子信息产业集团方滨兴院士;中国中文信息学会前副理事长、新疆大学吾守尔·斯拉木院士获此殊荣。