



第五届“讯飞杯”中文机器阅读理解评测报告

基于选项理解增强的可解释机器阅读理解

队 伍: Yunxiaomr

成 员: 赵云肖, 乔雪妮, 闫智超

导 师: 李 茹 教授

山西大学 计算机与信息技术学院

Sunday, October 30, 2022



报告提纲

一、任务背景

二、任务分析

三、模型方法

四、实验结果

五、结论展望



报告提纲

一、任务背景

二、任务分析

三、模型方法

四、实验结果

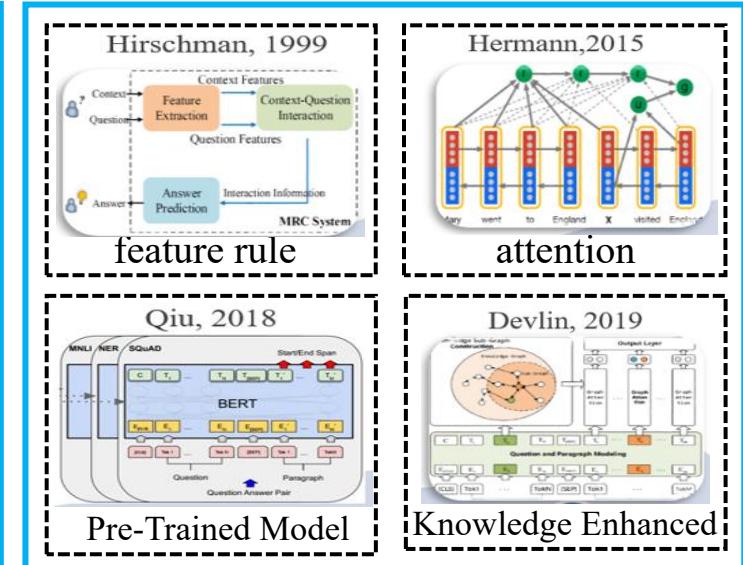
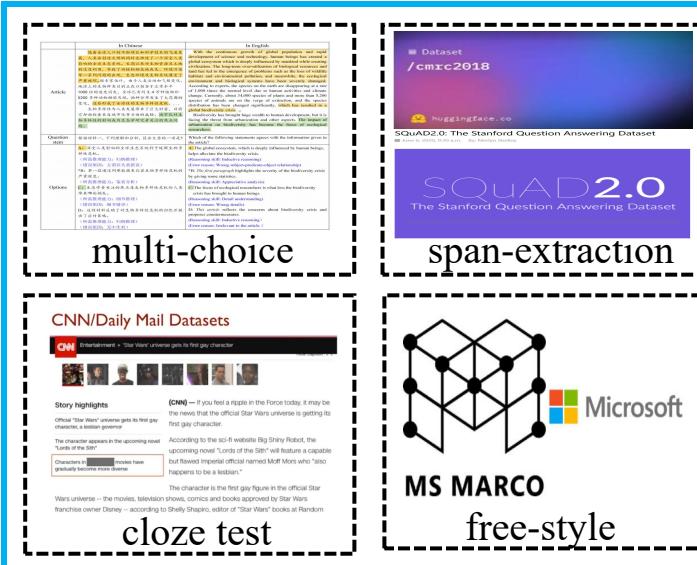
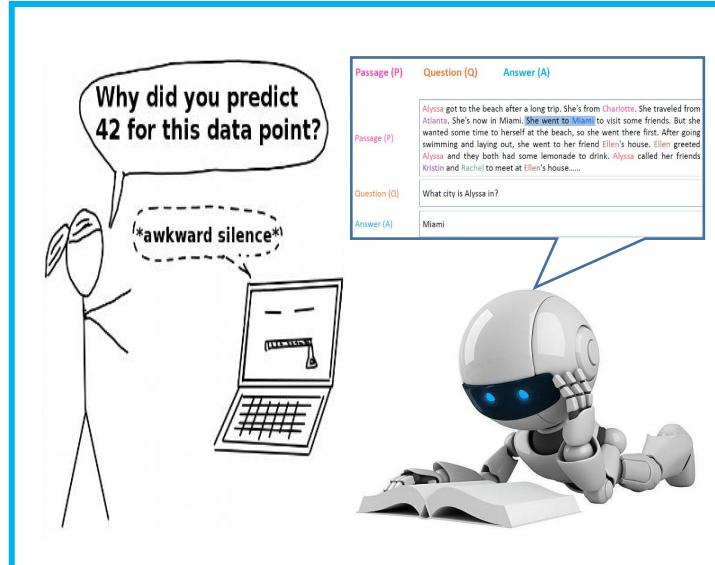
五、结论展望



一、任务背景

■ 机器阅读理解 (Machine Reading Comprehension, MRC)

- 机器阅读理解要求机器阅读和理解给定的材料并回答相关问题.
- 多项选择(C3,GCRC,Race,Dream); 区间抽取(SQuAD, NewsQA,CMRC2018); 完形填空(CNN/Daily Mail); 自由问答(MS MARCO), and so on.
- BiDAF, Attention Sum Reader, Attention-over-Attention Reader, QANet, BERT, RoBERTa, ELECTRA, Ernie, and so on.





一、任务背景

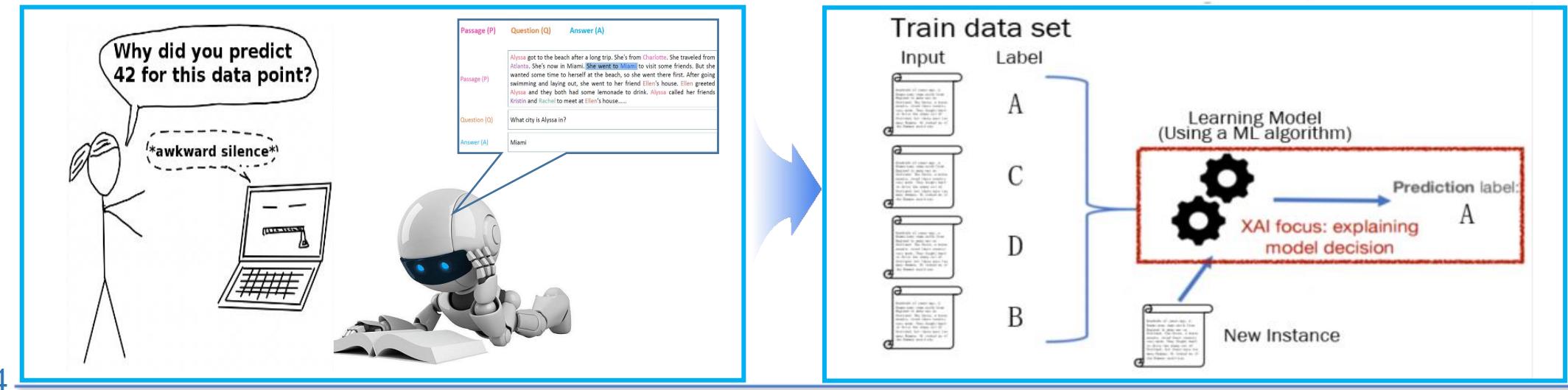
■ 任务背景

□ MRC系统的性能在部分数据集上已经实现接近甚至超过人类的水平

- ✓ (Tan et al. 2021) Numerous MRC datasets have been proposed and facilitate the progress of MRC systems, which have achieved near-human performance on some datasets.
- ✓ (Cui et al. 2022) the state-of-the-art systems could achieve better performance than the average human on some MRC datasets with the help of pre-trained language models.

.....

□ 然而，MRC系统不能同时给出问题的答案和解释，其面临着可解释性差的问题





一、任务背景

■ 任务介绍 (CMRC2022)

□ 可解释机器阅读理解 (Explainable Machine Reading Comprehension, ExpMRC) :

- 给定篇章以及和篇章相关的问题，设计可以同时抽取答案和佐证依据的机器阅读理解系统。
- 两个赛道
 - ✓ 选择型阅读理解赛道：给出选项答案和支撑证据
 - ✓ 抽取型阅读理解赛道：抽取答案片段和支撑证据

□ 评测指标

$$F1_{\text{overall}} = F1_{\text{answer}} \times F1_{\text{evidence}}$$

□ 评测难点

- 不提供带标注的训练集合
- 聚焦于无监督或弱监督的机器阅读理解系统

表1: ExpMRC中的数据样例 (Cui et al. 2022)

Subset	Passage	Question & Answer
SQuAD	<p>... Competition amongst employers tends to drive up wages due to the nature of the job, since there is a relative shortage of workers for the particular position. <u>Professional and labor organizations</u> may limit the supply of workers which results in higher demand and greater incomes for members. Members may also receive higher wages through collective bargaining ...</p>	<p>Q: Who works to get workers higher compensation? A: Professional and labor organizations</p>
CMRC 2018	<p>... 钩盲蛇（学名：“Ramphotyphlops braminus”）是蛇亚目盲蛇科下的一种无毒蛇种，主要分布在非洲及亚洲，不过现在钩盲蛇的分布已推广至世界各地。钩盲蛇是栖息于<u>地洞</u>的蛇种，由于体型细小，加上善于掘洞...</p>	<p>Q: 钩盲蛇一般生活在什么地形中? A: 地洞</p>
RACE ⁺	<p>... My biology teacher, Mr. Clark, divided us into three groups and asked us to play a game about natural selection and how birds find food. He gave the first group one spoon to every student, the second group forks and my group knives. ... When I almost picked a bean, it dropped back to the ground. When I finally picked up several beans, one of my friends ran into me. I fell over. <u>All my beans dropped to the ground!</u> Just at that moment, Mr. Clark called us back. ...</p>	<p>Q: How many beans did the writer get at last? A: None. B: One. C: Several. D: Many.</p>
C ³	<p>... 大学生生活是走上社会的预演，可以说，大学里的处世态度和人际关系的成功与否，直接决定着将来在社会上的成败。人是社会性的动物，生活中的每个人都离不开别人的帮助，同时也在帮助着别人。不管是学习、生活、工作，都要求自己要有良好的处理人际关系的能力。一个人要想有良好的人际关系，就要遵循以下几个原则：一是“主动”。要主动和别人交往，主动帮助别人。二是“诚信”。...</p>	<p>Q: 说话人认为什么因素决定在社会上的成败? A: 工作的态度 B: 朋友的数量 C: 大学里的学习成绩 D: 大学里的人际关系</p>



报告提纲

一、任务背景

二、任务分析

三、模型方法

四、实验结果

五、结论展望



二、任务分析

■ 数据样例（一）：子句证据

Article: "大学生活是走上社会的预演，可以说，**大学里的处世态度和人际关系的成功与否，直接决定着将来在社会上的成败**。人是社会性的动物，生活中的每个人都离不开别人的帮助，同时也在帮助着别人。不管是学习、生活、工作，都要求自己要有良好的处理人际关系的能力。一个人要想有良好的人际关系，就要遵循以下几个原则：一是“主动”。要主动和别人交往，主动帮助别人。二是“诚信”。与人交往，贵在一个“诚”字。以真心换真心，以诚信换诚信。要小聪明只能蒙得了一时，但“日久见人心”。三是“包容”。大千世界，芸芸众生，人的脾气千差万别，与人交往时不要斤斤计较，要能容得下别人的缺点，只要这个缺点无伤大雅。四是“感恩”。“受人滴水之恩，当涌泉相报。”对别人给予自己的帮助，时刻不能忘怀，哪怕是很小的一件事。",

Question: 说话人认为**什么因素决定在社会上的成败**？

Options: ["工作的态度","朋友的数量","大学里的学习成绩","**大学里的人际关系**"]

Answer: D

Evidence: "**大学里的处世态度和人际关系的成功与否，直接决定着将来在社会上的成败。**"

基于给定材料，给出相关问题的作答结果，并给出支撑答案的相关支撑依据。



二、任务分析

■ 数据样例（二）：子句证据 & 句子证据

Article: "春秋战国时期，秦国的商鞅实行变法，发布了新的政令。为了让老百姓相信自己，商鞅下令在都城南门外竖起一根三丈长的木头，并当众许下诺言，**谁能把这根木头搬到北门，赏金10两。大家不相信如此简单的事能得到如此高的奖赏，结果没人肯出来试一试。于是，商鞅将赏金提高到50两。**终于有人禁不住奖赏的诱惑，将木头扛到了北门。商鞅立即赏了他50两黄金。商鞅这一举动在百姓心中树立了威信，于是商鞅接下来发布的法令就很快在秦国推广开了。商鞅变法使秦国渐渐强盛，最终统一了中国。而就在商鞅“立木取信”的地方，在商鞅之前400年，却曾发生过一场令人哭笑不得的闹剧。周幽王有个美丽的妃子，从来不笑。………幽王再次点起烽火，但是将士们谁也不愿再上当了，结果幽王被逼自杀。一个“立木取信”，一诺千金；一个帝王无信，玩弄将士。可见，“信”对一个国家的兴衰存亡起着非常重要的作用。"

Question: 商鞅为什么要提高赏金？

Options: ["**因为没有人去搬木头**","**因为商鞅很有钱**","**因为他想看看人们是否贪财**","**因为人们想看商鞅的笑话**"]

Answer: A

Evidence: "**谁能把这根木头搬到北门，赏金10两。大家不相信如此简单的事能得到如此高的奖赏，结果没人肯出来试一试。于是，商鞅将赏金提高到50两。**"

基于给定材料，给出相关问题的作答结果，并给出支撑答案的相关支撑依据。



二、任务分析

■ 数据分析（一）：文章长度

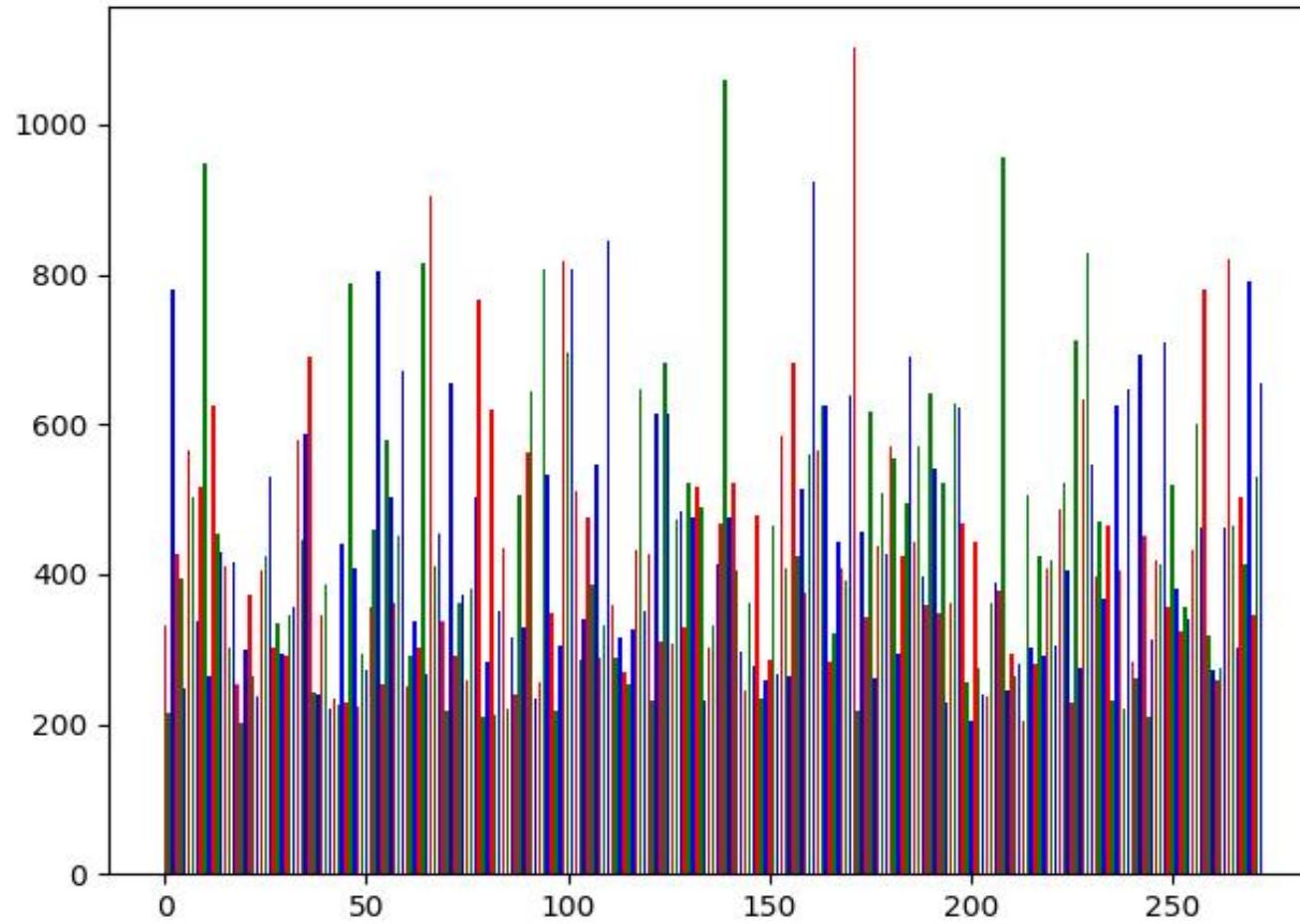


图1: dev dataset的文章长度分布

9

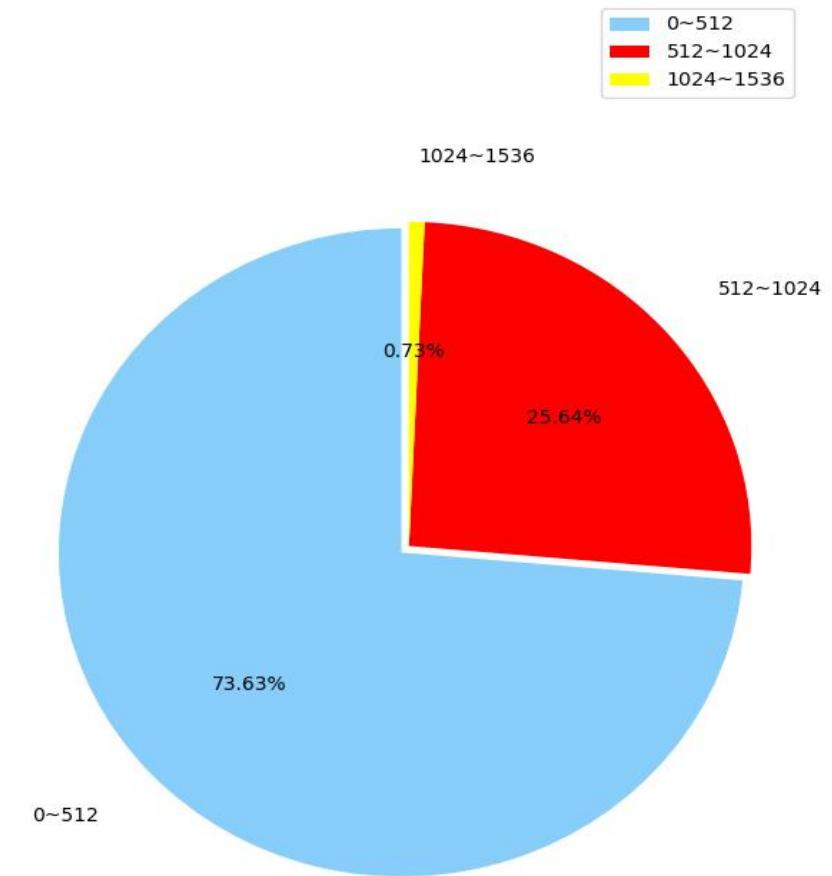


图2: dev dataset的文章长度比例分布



二、任务分析

■ 数据分析（二）：问题长度

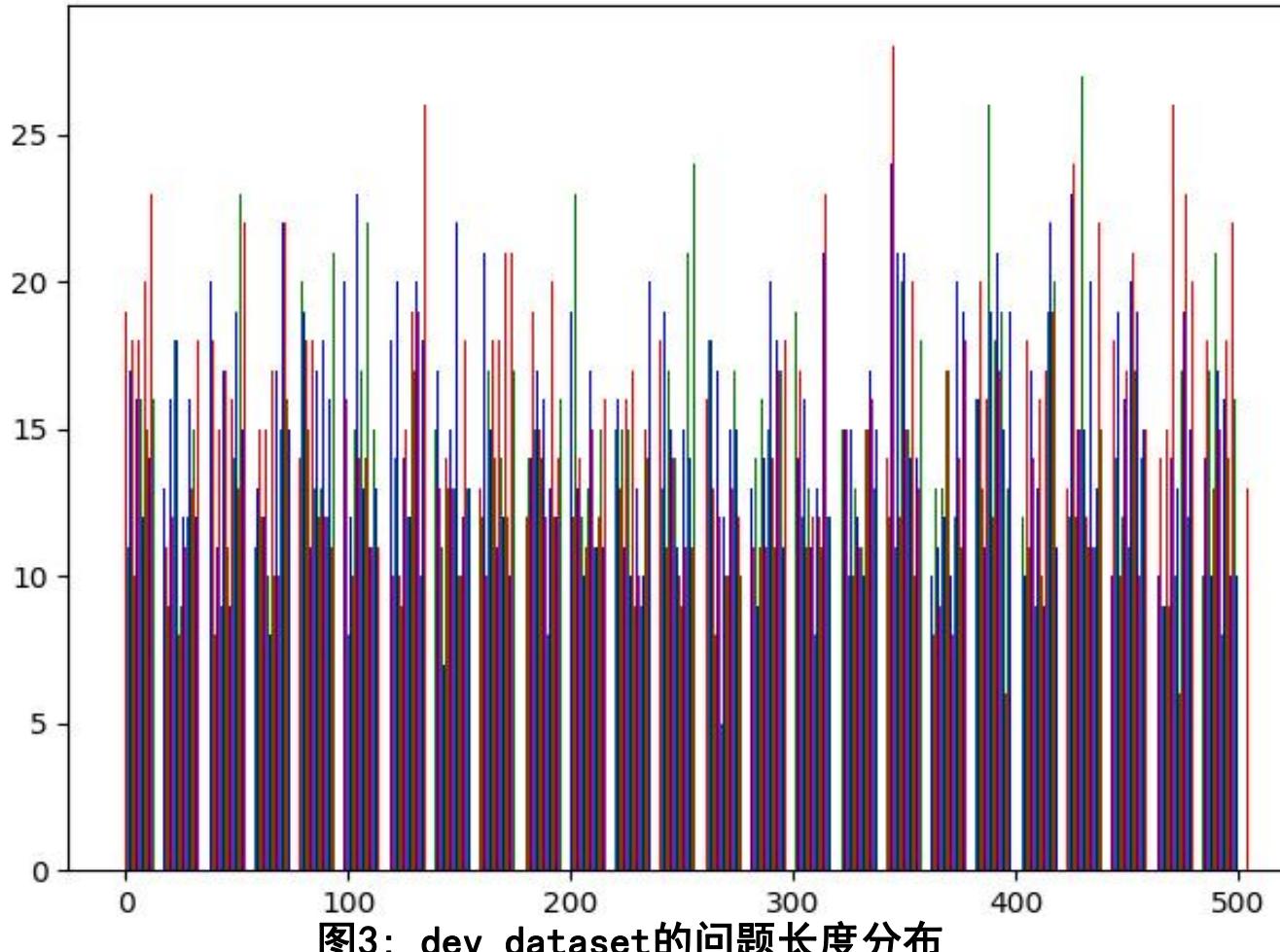


图3: dev dataset的问题长度分布

10

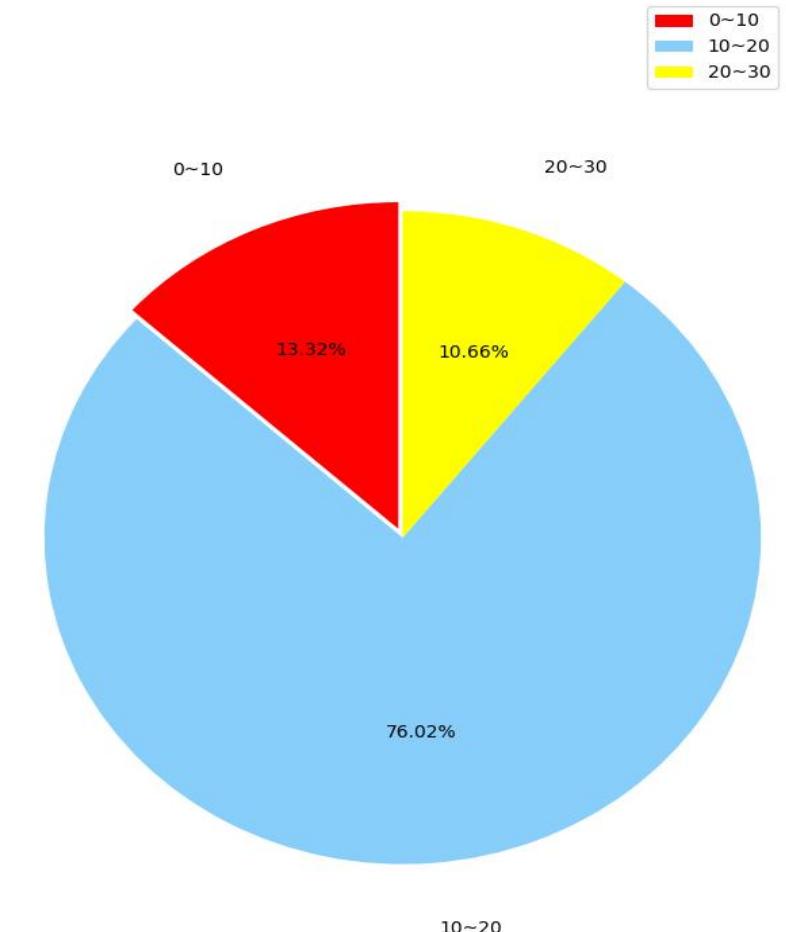


图4: dev dataset的问题长度比例分布



二、任务分析

■ 数据分析（三）：选项长度

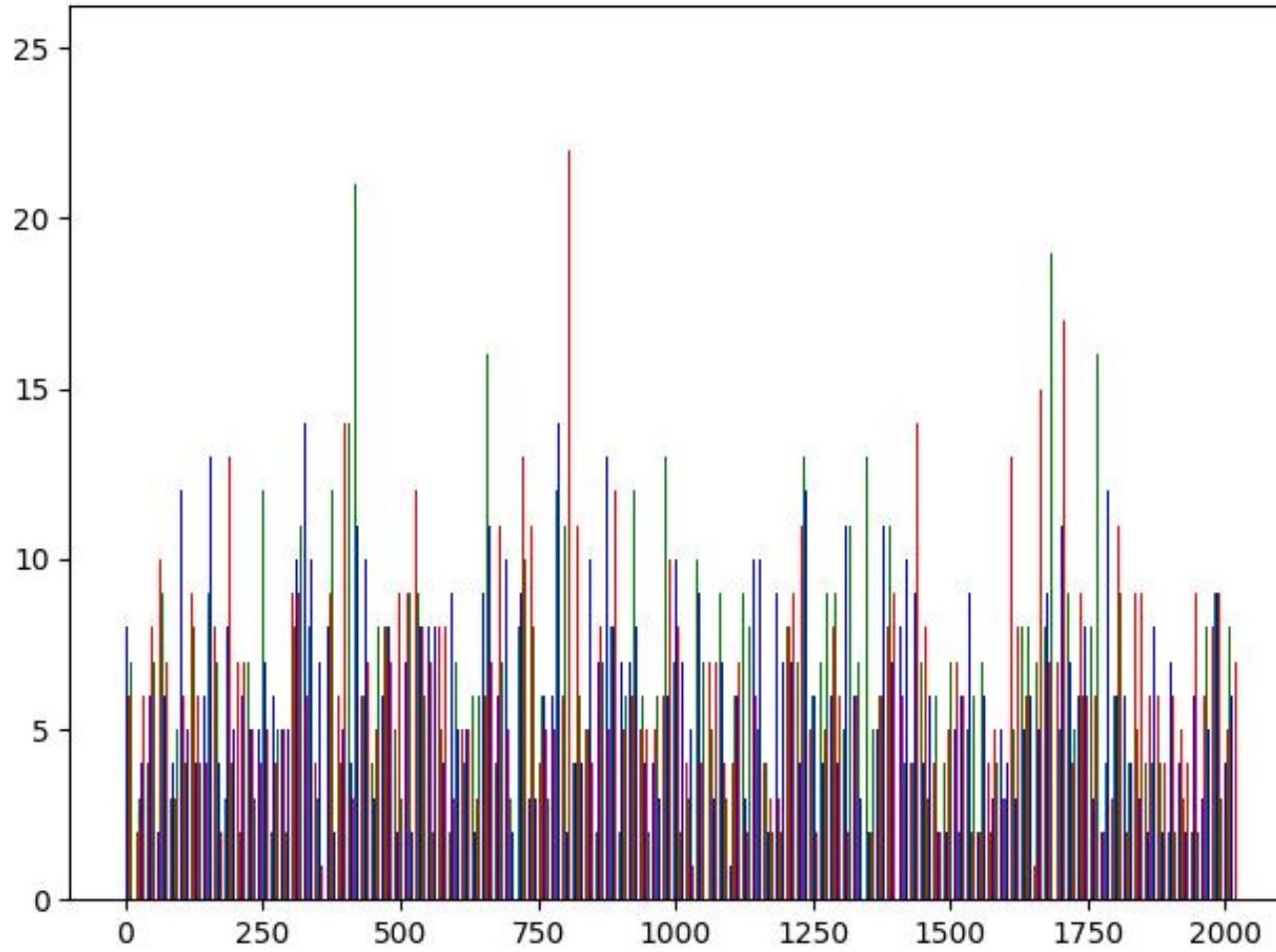


图5: dev dataset的选项长度分布

11

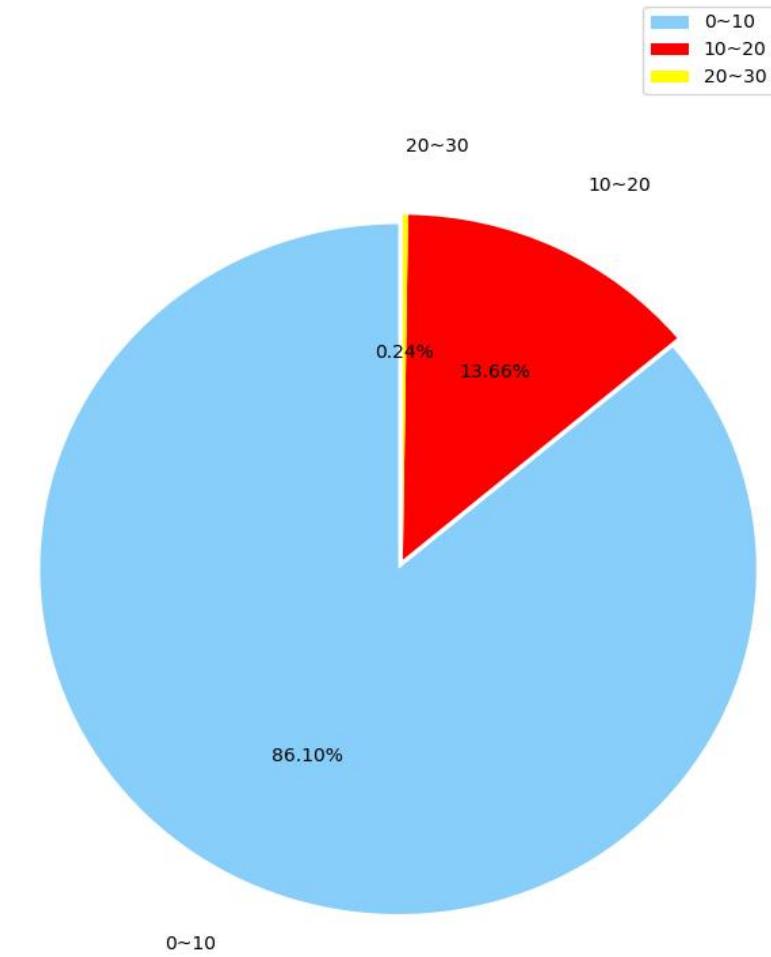


图6: dev dataset的选项长度比例分布

二、任务分析

■ 数据分析（四）：证据长度

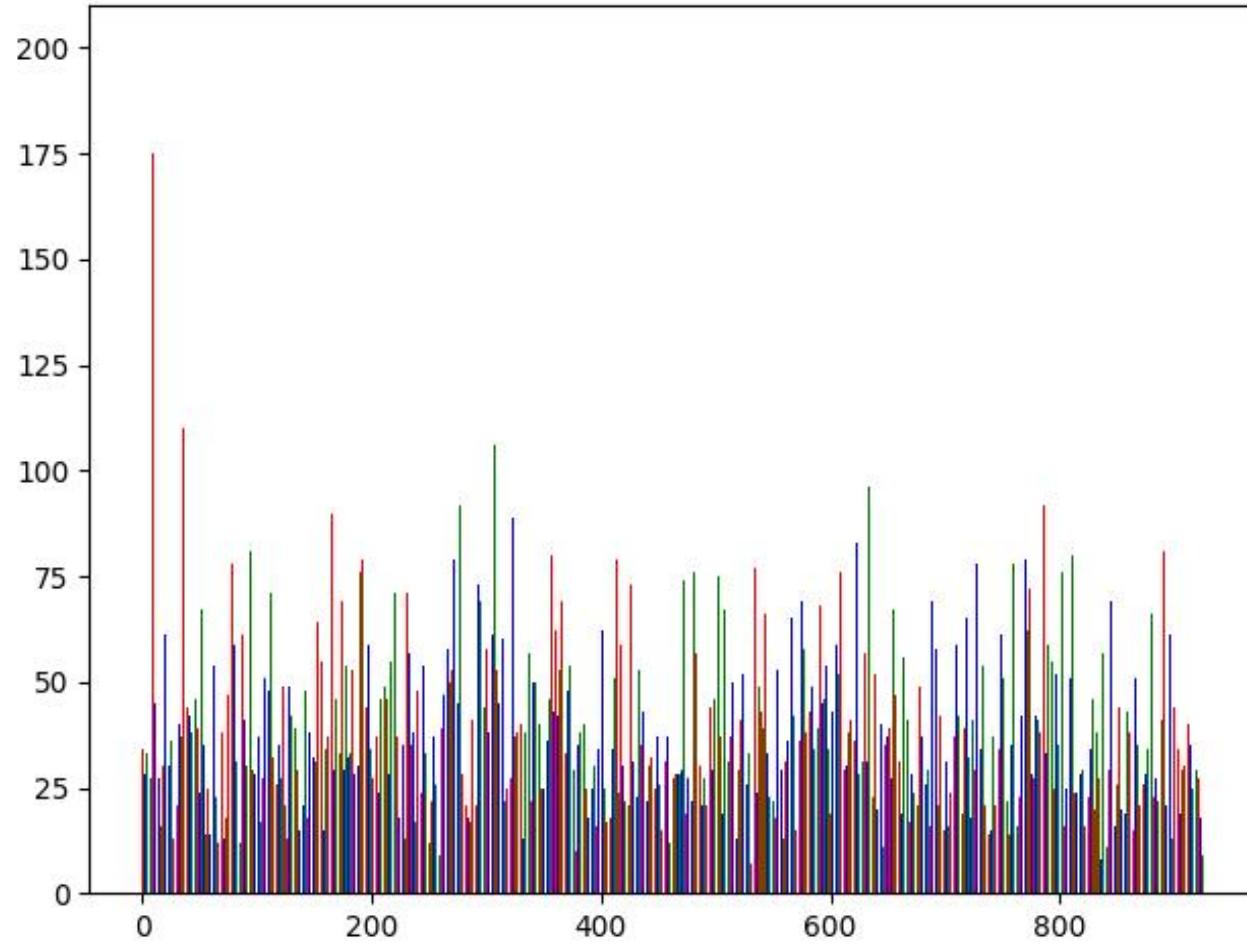


图7：dev dataset的证据长度分布

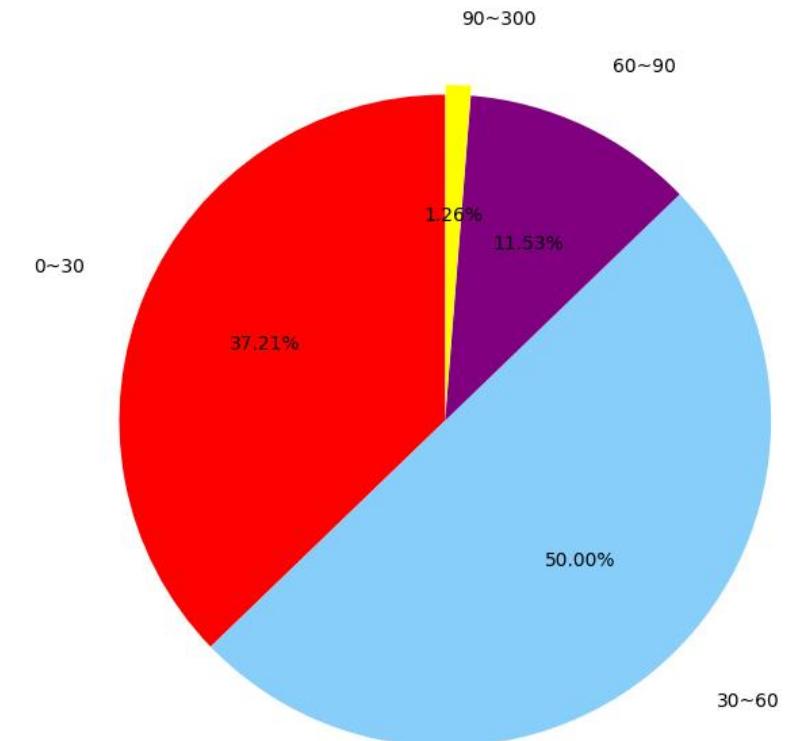


图8：dev dataset的证据长度比例分布

二、任务分析

■ 数据分析（五）：证据句数

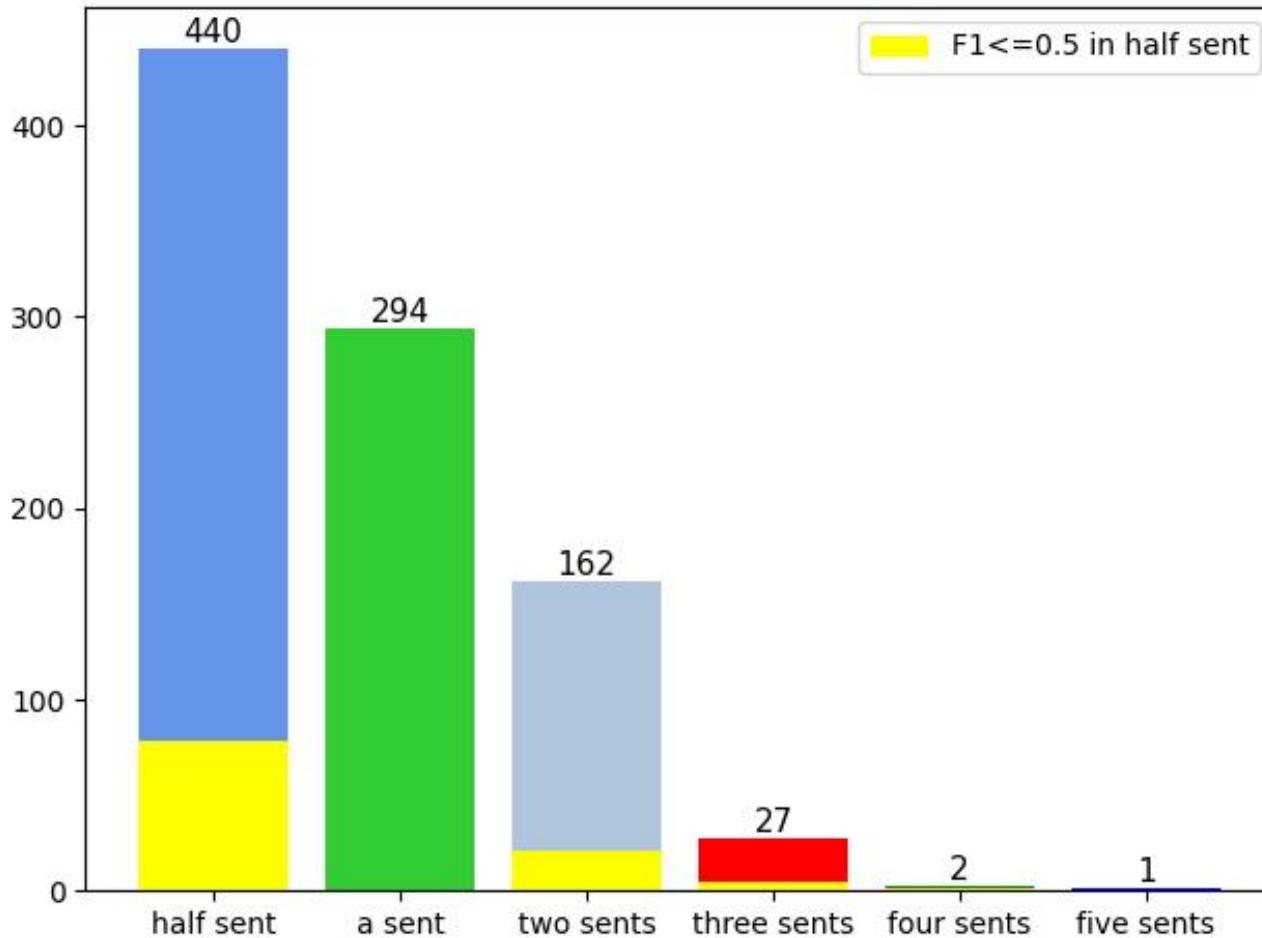


图9: dev dataset的证据句数分布

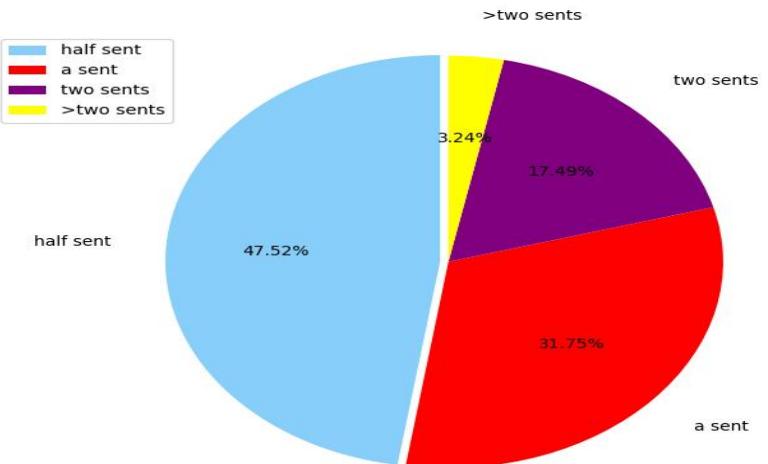


图10: dev dataset的证据句数比例分布

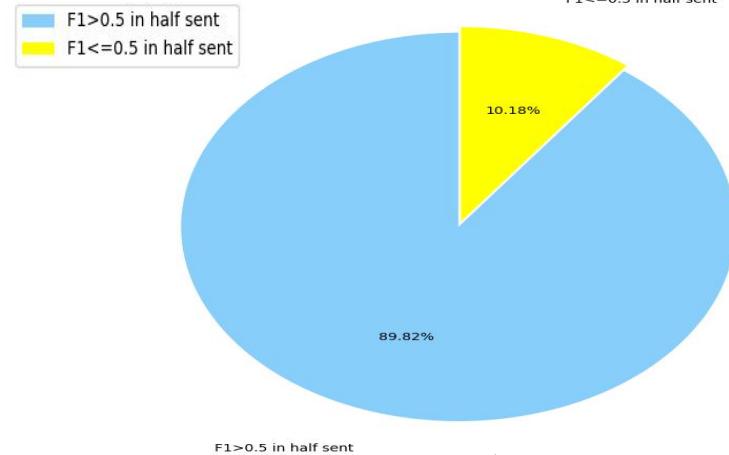


图11: dev dataset的子句证据F1比例分布



二、任务分析

■ 证据分析（一）：证据的多样性

Article: "很早以前，.....在风的大家族中，还有一种人们比较陌生的风，它就是火风。火风又称“干旱风”，是一种气象灾害。它多出现在春末夏初，会使谷物或果实早熟，严重时可引起农作物枯萎，甚至造成山林火灾。可见，风与火有着密切的关系。风，就是流动的空气。燃烧离不开氧气，而空气里约有五分之一是氧气。有风时，空气流动较快，氧气供应充足，火就烧得旺。风的大小和方向对火势的蔓延往往起着决定性作用.....众所周知，可燃物燃烧需要同时满足两个条件：一是可燃物要与氧气接触，二是可燃物的温度必须达到着火点。例如，往煤炉内吹风，供给了煤燃烧所需要的足量氧气。冷空气虽能使炉内燃料降温，却不能让炉内温度降至煤的着火点以下，这样炉火仍可持续燃烧，所以会越扇越旺；而用扇子扇烛火，会使环境温度低于蜡烛的着火点，所以蜡烛一扇即灭。因此，风能助火也能灭火。当风无法使可燃物温度降低到着火点以下时，风便助火；当风大到使可燃物的温度低于着火点时，风便灭火。油井着火时，消防人员在油井附近引爆炸药，用爆炸引起的强大气流来扑灭油井大火，就是利用了这个原理。我们相信，随着科学技术的不断发展，人们一定能认识和掌握风的“脾气”和特点，从而让它更好地为人类服务。",

Question: 根据上文，下列哪项正确？

Options: ["火风多伴随降雨","火风会对农作物造成危害","往煤炉内吹风易扑灭炉火","可燃物在无氧情况下也可燃烧"]

Answer: B

Evidence: "火风又称“干旱风”，是一种气象灾害。它多出现在春末夏初，会使谷物或果实早熟，严重时可引起农作物枯萎，甚至造成山林火灾。"

不同于其他类型的阅读理解任务，多项选择任务中不同选项关注的证据语义信息不同。



二、任务分析

■ 证据分析（二）：证据的多视角性

Article: "有一块小石头不停地哭泣，哭声惊动了身边的一块大石头。大石头问：“小石头，你为什么不停地哭啊？”小石头伤心地说：“我不是小石头，我是会发光的金子，我想发挥我的光和热。”大石头安慰它：“你不要伤心，是金子在哪儿都会发光的。总有一天，你会大展才华的。”事实上，**小石头在这儿已经数百年，由于地处偏僻，未能被人发现。**小石头也想：“**是金子在哪儿都发光。**”于是他日夜吸收天地之精华，把自己冲涤得闪闪发亮，终究又被脏东西所掩盖。如今眼看着周围的石块风化消失，它怎能不着急。了解了小石头的情况后，大石块只能长叹一声，却又无可奈何。**不久，山洪暴发，小石头和大石块都被洪水冲进大河，深深地埋在泥沙之中，只是偶尔在河水中会听到极其弱小的哭泣声，那一定是来自小石头的。"**

Question: 最后小石头的命运怎么样？

Options: ["消失了","继续发出亮光","被人们发现了","被掩盖在泥沙之中"]

Answer: D

Evidence: "所以，预防感冒的关键措施是：不用手乱摸东西；勤洗手帕或者使用一次性的面巾纸；感冒患者不要与别人握手等等。"

选项的证据存在共现现象，即来自同一证据的语义信息对不同选项判别可能存在语义贡献。



二、任务分析

■ 证据分析（三）：证据的关联性

Article: "春秋战国时期，秦国的商鞅实行变法，发布了新的政令。为了让老百姓相信自己，商鞅下令在都城南门外竖起一根三丈长的木头，并当众许下诺言，谁能把这根木头搬到北门，赏金10两。大家不相信如此简单的事能得到如此高的奖赏，结果没人肯出来试一试。于是，商鞅将赏金提高到50两。终于有人禁不住奖赏的诱惑，将木头扛到了北门。商鞅立即赏了他50两黄金。商鞅这一举动在百姓心中树立了威信，于是商鞅接下来发布的法令就很快在秦国推广开了。商鞅变法使秦国渐渐强盛，最终统一了中国。而就在商鞅“立木取信”的地方，在商鞅之前400年，却曾发生过一场令人哭笑不得的闹剧。周幽王有个美丽的妃子，从来不笑。………幽王再次点起烽火，但是将士们谁也不愿再上当了，结果幽王被逼自杀。一个“立木取信”，一诺千金；一个帝王无信，玩弄将士。可见，“信”对一个国家的兴衰存亡起着非常重要的作用。"

Question: 商鞅为什么要提高赏金？

Options: ["因为没有人去搬木头","因为商鞅很有钱","因为他想看看人们是否贪财","因为人们想看商鞅的笑话"]

Answer: A

Evidence: "谁能把这根木头搬到北门，赏金10两。大家不相信如此简单的事能得到如此高的奖赏，结果没人肯出来试一试。于是，商鞅将赏金提高到50两。"

支撑答案的证据核心语义关联着当前片段上下文多种粒度（子句和句子）的语义。



报告提纲

一、任务背景

二、任务分析

三、模型方法

四、实验结果

五、结论展望



三、模型方法

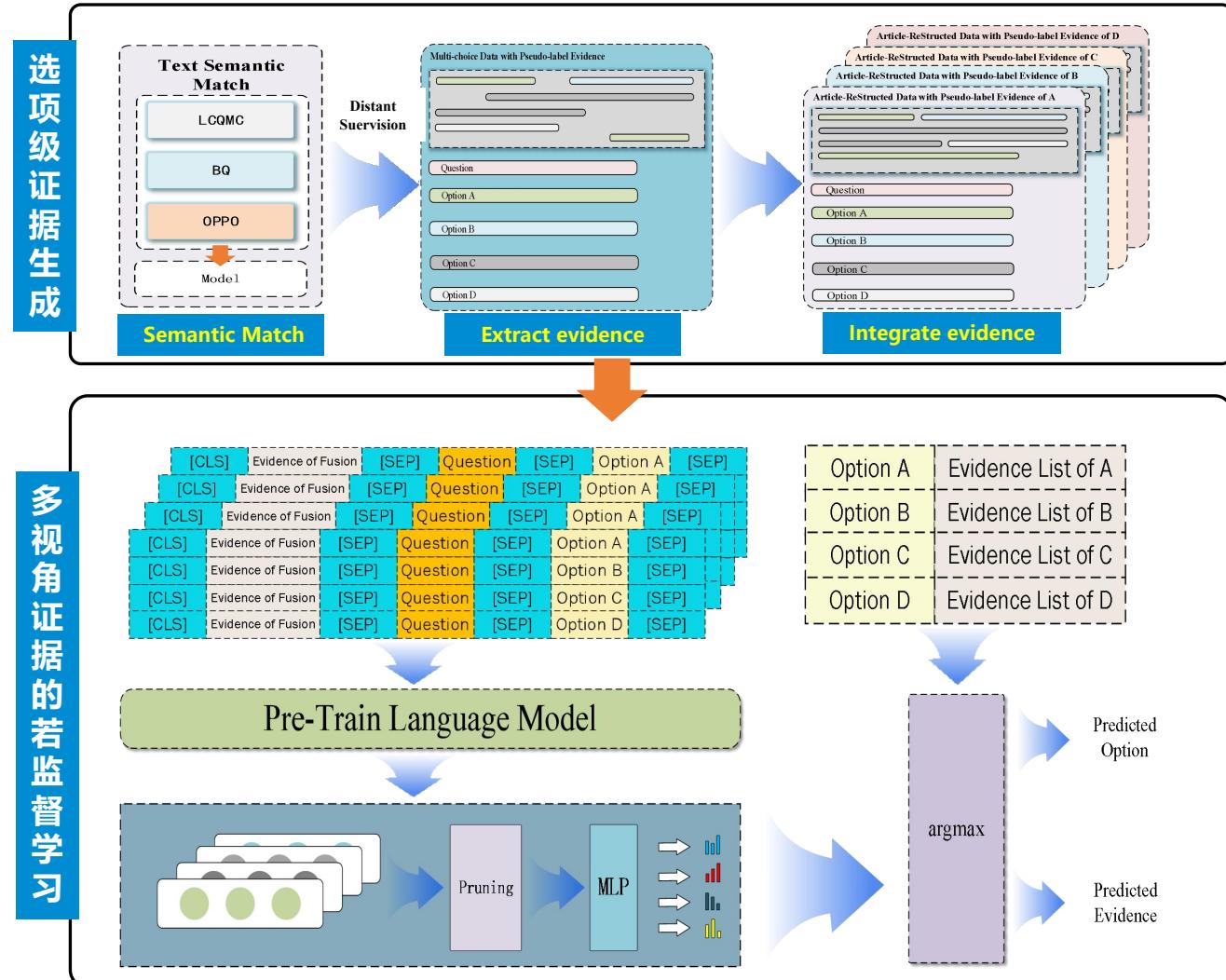
■ 整体架构

□ 选项级的证据生成方法

- 语义匹配
- 远程监督
- 证据抽取
- 证据整合

□ 多视角证据学习的弱监督方法

- 输入层
- 表示层
- 剪枝层
- 分类层
- 输出层



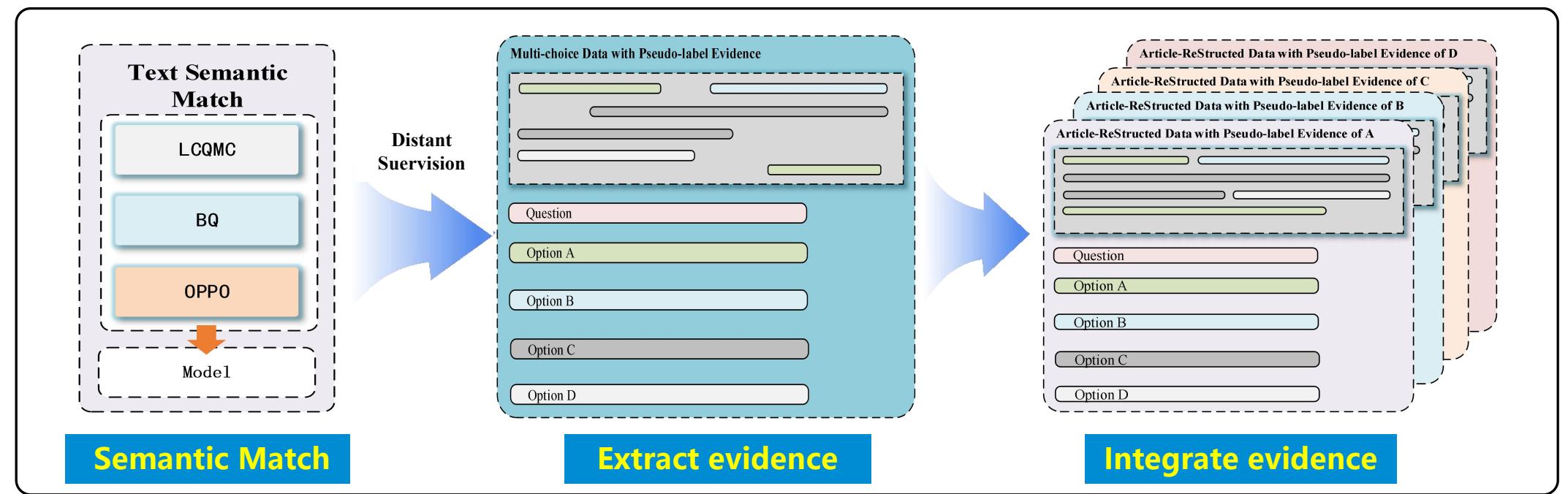
基于上述任务分析，研究构建了一种选项理解增强的可解释阅读理解方法。



三、模型方法

■ 选项语义理解：选项级的证据生成方法

- ✓ 针对证据的多样性问题提出选项级的证据生成方法，充分提升系统对选项信息的理解能力；
- ✓ 针对证据的关联性问题构建相似度计算策略，通过特定阈值分析，进行核心证据的证据关联；
- ✓ 具体地，构建文本语义匹配模型；远程监督的核心证据生成；相关证据关联；证据整合。



阅读问题

查看选项

寻找证据

语义推导

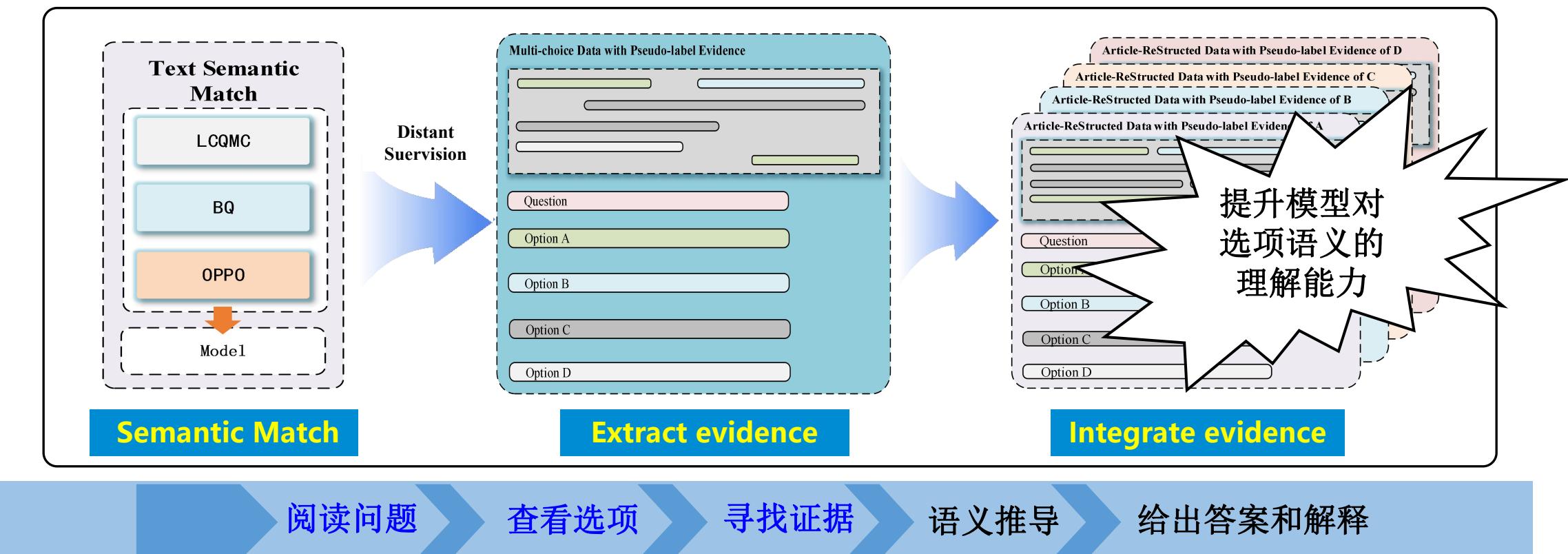
给出答案和解释



三、模型方法

■ 选项语义理解：选项级的证据生成方法

- ✓ 针对证据的多样性问题提出选项级的证据生成方法，充分提升系统对选项信息的理解能力；
- ✓ 针对证据的关联性问题构建相似度计算策略，通过特定阈值分析，进行核心证据的证据关联；
- ✓ 具体地，构建文本语义匹配模型；远程监督的核心证据生成；相关证据关联；证据整合。

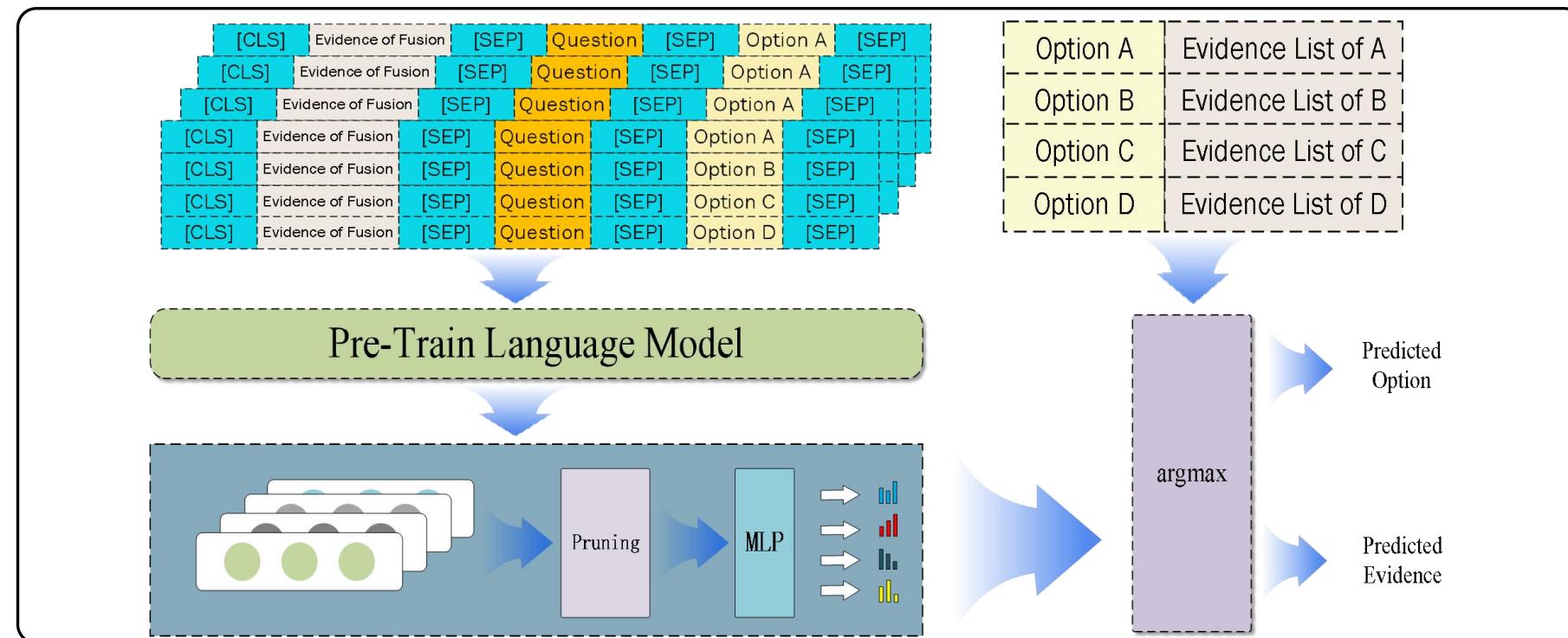




三、模型方法

■ 选项信息利用：多视角证据学习的弱监督方法

- ✓ 针对证据的多视角性问题提出多视角证据学习的弱监督方法，充分利用选项固有的语义信息；
- ✓ 首先，证据输入；其次，编码表示；然后，模型剪枝，线性分类；最后，答案和证据输出。



阅读问题

查看选项

寻找证据

语义推导

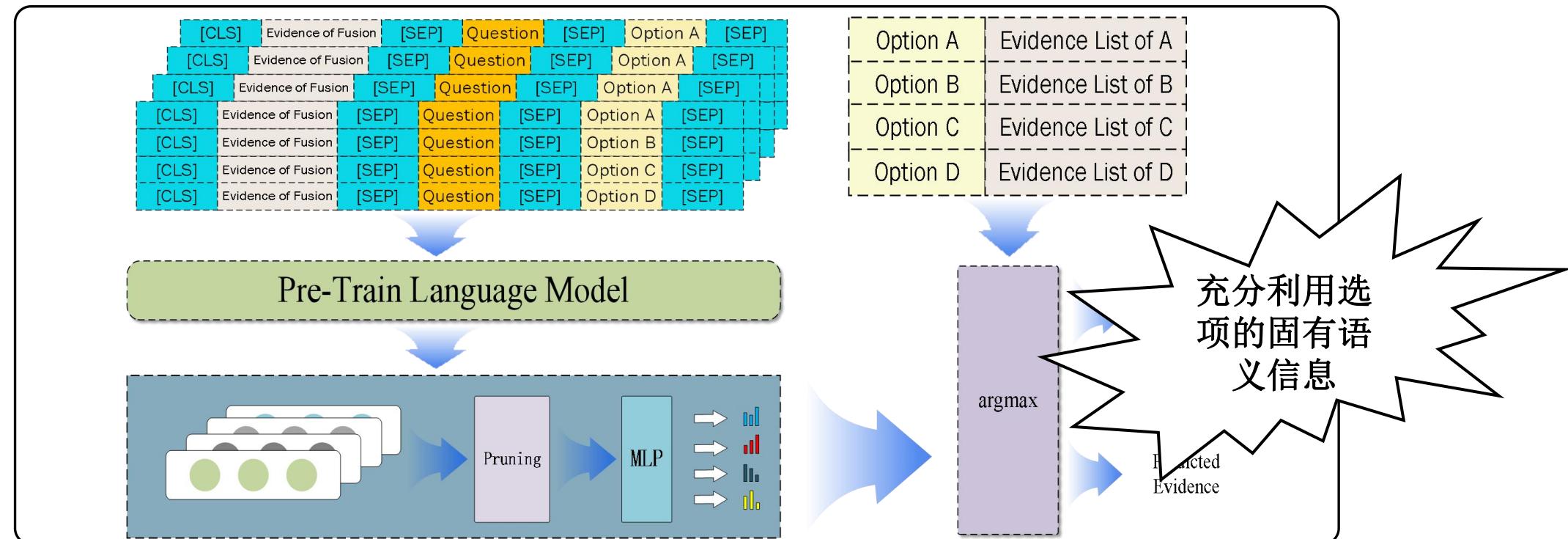
给出答案和解释



三、模型方法

■ 选项信息利用：多视角证据学习的弱监督方法

- ✓ 针对证据的多视角性问题提出多视角证据学习的弱监督方法，充分利用选项固有的语义信息；
- ✓ 首先，证据输入；其次，编码表示；然后，模型剪枝，线性分类；最后，答案和证据输出。



阅读问题

查看选项

寻找证据

语义推导

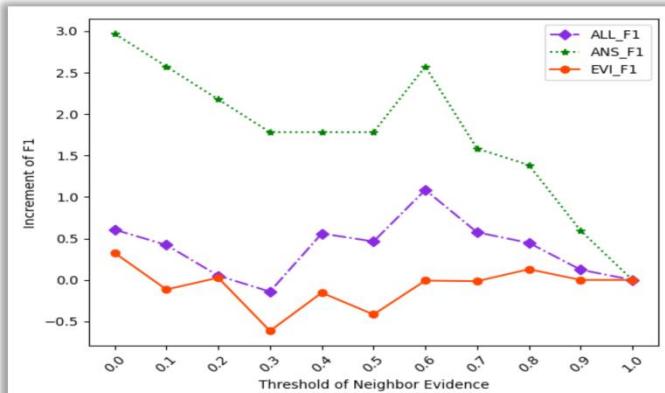
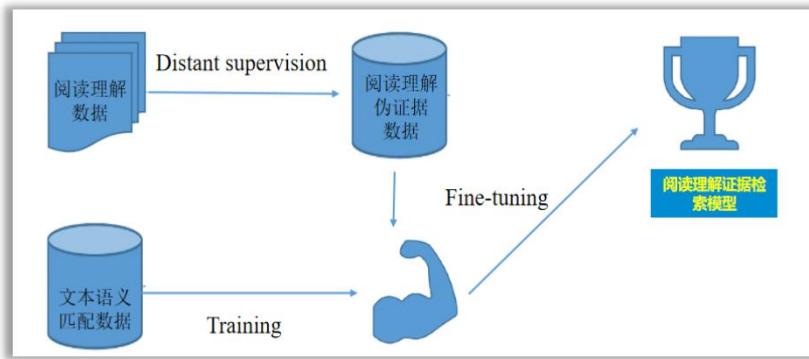
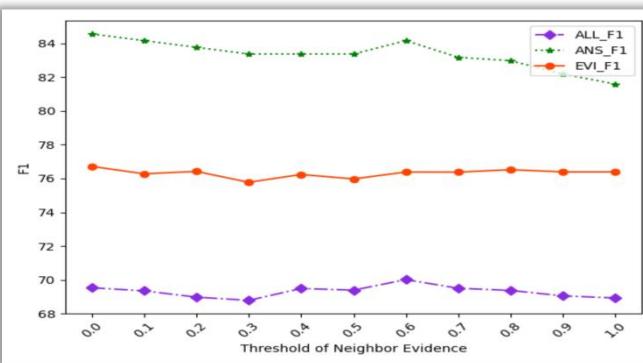
给出答案和解释



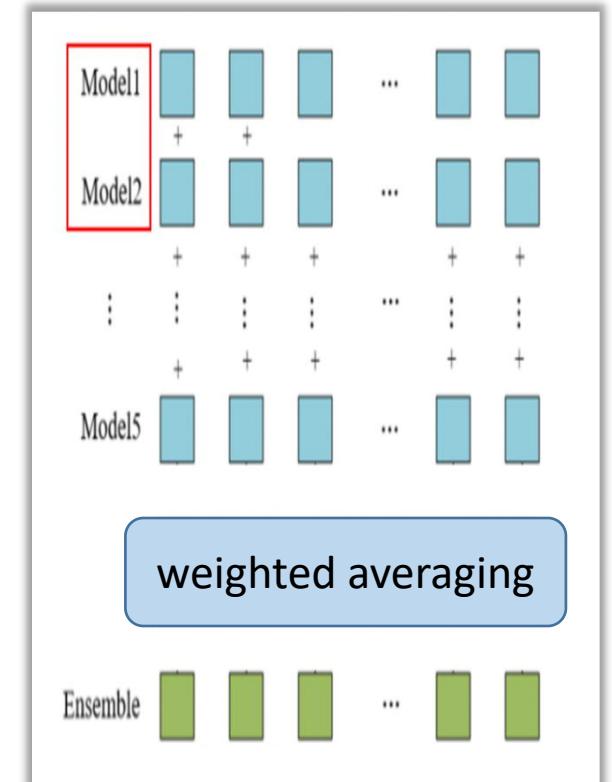
三、模型方法

■ 其他策略

- ✓ 相关证据关联：相似度阈值分析，取综合表现最佳的参数。
- ✓ 以文本语义匹配模型为基模型，进行阅读理解证据数据增量后进行再训练。
- ✓ 数据增强：对话型数据转换材料型数据。
- ✓ 模型集成：设置不同的随机种子，训练证据检索模型；加权平均法，融合模型。



This diagram illustrates the conversion of dialog-based data to material-based data. It shows two examples: one for '对话型数据' (Dialog-based data) and one for '材料型数据' (Material-based data). The dialog-based data example involves a conversation about flight delays due to foggy weather, followed by a question and multiple-choice answers. The material-based data example shows the same information in a more structured, document-like format.





报告提纲

一、任务背景

二、任务分析

三、模型方法

四、实验结果

五、结论展望



四、实验结果

■ 主要实验结果

□ 基于选项理解增强的可解释阅读理解方法在选择型阅读理解数据集上的对比结果

Model	Dev			Test		
	Ans_F1	Evi_F1	ALL_F1	Ans_F1	Evi_F1	ALL_F1
Most Similar Sent.	73.1	59.4	49.9	72.0	52.7	43.9
Most Similar Sent. w/Ques	73.1	63.2	50.9	72.0	58.4	46.0
Pseudo-data training	76.4	64.3	50.7	74.4	59.9	47.3
Most Similar Sent w/Ques.(Normalization)	-	-	51.2	-	-	-
Evidence(Ques.Op) & Answer(Article)	70.1	67.8	53.2	-	-	-
Evidence(Ques.Op) & Answer(Evidence)	67.9	68.3	54.1	-	-	-
Evidence(Ques.Op) & Answer(Evidence with FGS Neighbor)	76.2	68.6	60.1	-	-	-
Evidence(Ques.Op) & Answer(Evidence with FGS Neighbor)	79.8	70.9	62.9	-	-	-
Most Similar Sent Using Semantic Match	79.2	73.7	65.8	-	-	-
Evidence(Ques.Op) & Answer(Evidence with FGS Neighbor) & Neighbor Evidence Contact & Train & Rule	80.6	75.8	67.7	78.4	70.2	59.2
Evidence(Ques.Op) & Answer(Evidence with FGS Neighbor) & Evidence Interaction Using Semantic Match & Train	82.4	75.9	68.5	-	-	-
Evidence(Ques.Op) & Answer(Evidence with FGS Neighbor Threshold) & Evidence Interaction Using Semantic Match & Train	84.1	76..3	70.0	-	-	-
Evidence(Ques.Op) & Answer(Evidence with FGS Neighbor Threshold) & Evidence Interaction Using Semantic Match & Train & Aug & emsemble	86.1	77.6	70.7	81.0	70.9	61.2



四、实验结果

■ 评测结果



选择型阅读理解赛道

排名	队伍名	答案准确率	证据F1	综合F1↓
1	yunxiaomr 山西大学	81.000	70.902	61.196
2	jerome 阳光出行	77.000	71.241	60.133
3	hsz779 北京理工大学	75.400	69.857	57.784
4	Smile	74.000	61.526	50.952
5	Two	69.200	61.605	47.833
6	leishu	66.400	58.646	41.458



报告提纲

一、任务背景

二、任务分析

三、模型方法

四、实验结果

五、结论展望



五、结论展望

■ 结论

- ✓ 针对证据多样性、多视角性和关联性三个问题，围绕选项语义理解和选项信息利用两个角度提出一种选项理解增强的可解释阅读理解方法。
 - ✓ 在选项语义理解方面，针对证据的多样性和关联性问题，提出了一种选项级的证据生成方法，构建了一种相似度计算策略。
 - ✓ 在选项信息利用方面，针对证据的多视角性问题，提出了一种多视角证据学习的弱监督方法。
-
- ✓ 该方法考虑了答案与证据的交互，在给出答案的同时，可给出与答案相关的事后解释证据；同时，该方法模拟人类答题过程，具有一定的自解释能力。
 - ✓ 选项证据生成仅考虑了句子级的证据抽取，在多片段或多粒度文本抽取的场景下尚存在缺陷。

■ 展望

- ✓ 从事后解释的角度，研究阅读理解模型的可解释；
- ✓ 从自解释的角度，构建具有自解释能力的阅读理解模型。



Thanks!



<https://github.com/yunxiaomr>



yunxiaomr@163.com