

朱运昌<sup>1,2</sup> 庞亮<sup>1,2</sup> 兰艳艳<sup>1,2</sup> 程学旗<sup>1,2</sup>

1. 中国科学院 计算技术研究所 中国科学院网络数据科学与技术重点实验室; 2. 中国科学院大学

## 论文摘要

将机器开放域问答技术应用在政策咨询领域,有助于各行业及时掌握相关政策,尤其是在新冠疫情期间,能方便企业快速从大量繁琐的政策中找到有价值的信息。现有的抽取式机器阅读理解模型往往更适用于诸如实体类的短答案抽取,而面对长文档长答案的政策型场景,模型往往会遇到输入长度超越极限、长答案抽取不完整的挑战。然而,相比于答案的边界,答案内部冗长的内容对于答案的准确抽取并不是那么关键,为此,该文加强对答案边界的理解并忽略答案内部信息,进而提出了答案边界整合(BoundaryQA)框架,该框架包括局部片段合并、全局边界重选以及结合局部和全局信息的跳跃式重阅读三种边界整合策略。为了验证框架的有效性,该文在真实的疫情政策问答数据上进行了实验,实验结果表明提出的整合策略能够在总体性能提升的基础上,明显提升长答案抽取的性能。

## 问题概述

现有的抽取式阅读理解模型和数据集大多只关注短答案,但政策型问答中的答案却通常很长,并且往往也对应着更长的文档。面对长文档长答案的场景,现有模型往往会遇到输入长度超越极限和长答案抽取不完整的挑战。

任何模型实际中都存在一个输入文本长度的极限,当输入文本长度超过这个极限后,模型就无法正常的工作。

图1所示的预处理方法虽然可以通过滑动窗口的方式将长文档切分成多个文档块来保证长文档以多次输入的形式输入到阅读理解模型之中,但是文档中的潜在答案仍会不可避免地被切分到不同文档块。

- **问题:** 疫情期间,如何安排经费预算?
- **答案长度:** 446
- **长文档切分后的文档块:**

[1, 501]: 关于进一步做好……坚决打赢疫情防控阻

[502, 986]: 现就进一步做好疫情防控经费保障工作……近期,按照党中央、国务院统一……切实履行好经费保障职能。

[987, 1483]: 年初卫生防疫经费预算安排……“三保”经费的保障力度也要切实加强。三、……疫情防控资金拨付。

[2346, 2581]: 七、切实做好……财政部 2020年1月31日

图1 分块后的长答案问答实例。

下划线文本为参考答案, [start, end]表示文档块所处的区间

## 核心想法

为了克服这两个挑战,本文注意到抽取式问答的目标只是找到精准的答案边界位置,其对与问题无关的片段和答案内部冗长的内容其实并不特别关心。因此,在初步对整个文档进行过分块阅读之后,问答系统可以自信地排除与问题无关的片段,只考虑候选答案的边界,之后对它们进行整合就可以得到一个更完整精确的答案。

## BoundaryQA

BoundaryQA提供了三种答案边界整合策略:

- **局部片段合并:** 利用启发式规则来局部地合并那些因文档块切分而被分开的两个邻近候选答案片段,从而得到更完整的答案

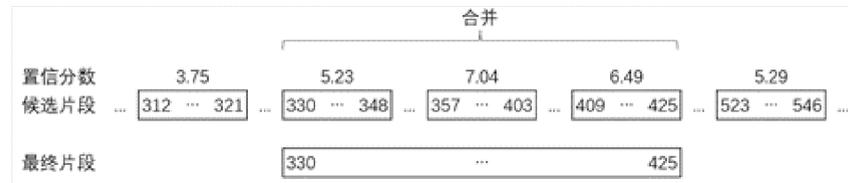


图2 局部片段合并示意图

- **全局边界重选:** 额外训练一个答案边界重选模型,让其自己学习从所有候选答案边界中全局地重新选择出一对最好的答案开始和结束边界

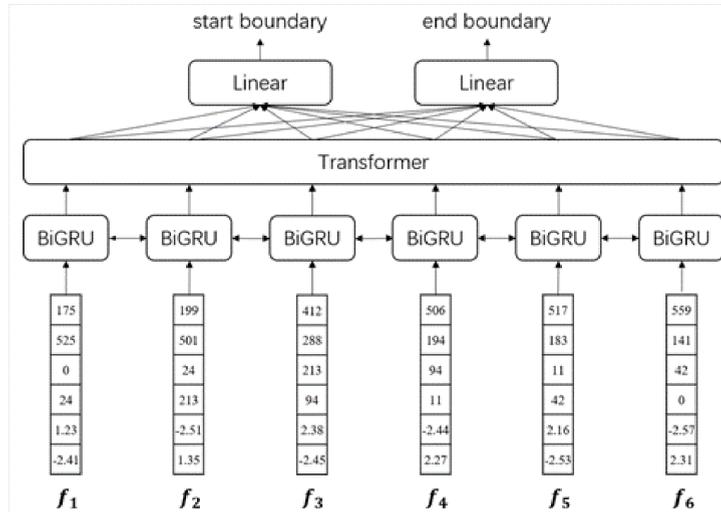


图3 全局重选模型的网络结构

- **跳跃式重阅读:** 先利用粗读模型预测出的所有候选答案生成一段仅包含候选答案边界附近上下文的摘要,然后训练一个精读模型让其重读这段简短摘要并抽取更准确的答案

- **问题:** 疫情期间,如何安排经费预算?
- **摘要:** …: 一、明确经费保障目标 各级财政部门要按照习近…好疫情防控经费保障工作。确保人民群众不因担心费…统筹安排经费预算 近期,按照党中央、国务院统一部署…时,切实履行好经费保障职能。年初卫生防疫经费预算安排不…经费的保障力度也要切实加强。三、加快调度拨付资金…

图4 针对问题的摘要示例。

“…”表示候选答案内部省略,“…”表示问题无关内容省略,绿表示候选答案开始边界,红表示候选答案结束边界

## 实验结果

表1 疫情政务数据集参考答案长度统计

答案长度区间	[1, 4)	[4, 16)	[16, 64)	[64, 256)	[256, +∞)	[1, +∞)
比例(%)	3.76	23.71	42.55	28.01	1.97	100.00
平均长度	2.67	8.33	35.56	114.56	456.78	58.43

表2 不同答案整合策略在测试集上的表现

答案长度区间	方法	指标	[1, 4)	[4, 16)	[16, 64)	[64, 256)	[256, +∞)	[1, +∞)
基线	ROUGE-L		66.25	69.64	69.96	64.08	51.85	67.76
	$R_{lcs}$		78.92	82.73	77.74	62.55	41.19	74.07
局部片段合并	ROUGE-L		66.25	<b>69.67</b>	69.90	64.64	51.37	67.89
	$R_{lcs}$		78.92	83.10	77.91	63.85	42.03	74.62
全局边界重选	ROUGE-L		68.75	67.10	<b>72.00</b>	69.08	44.81	69.32
	$R_{lcs}$		85.78	<b>83.60</b>	79.57	<b>67.16</b>	35.83	76.48
跳跃式重阅读	ROUGE-L		<b>74.65</b>	69.40	71.68	<b>69.20</b>	<b>52.70</b>	<b>70.15</b>
	$R_{lcs}$		<b>86.76</b>	82.92	<b>80.56</b>	67.04	<b>42.12</b>	<b>77.19</b>

表3 跳跃式重阅读策略采用不同上下文窗口大小在验证集上的表现

答案长度区间	答案长度区间 w	指标	[1, 4)	[4, 16)	[16, 64)	[64, 256)	[256, +∞)	[1, +∞)
8	ROUGE-L		69.31	<b>72.29</b>	72.33	67.11	53.21	70.48
	$R_{lcs}$		88.62	86.54	78.69	65.64	42.28	76.89
16	ROUGE-L		67.10	70.40	71.97	<b>68.68</b>	<b>61.08</b>	70.32
	$R_{lcs}$		91.06	<b>87.24</b>	80.40	67.40	50.31	78.48
32	ROUGE-L		66.80	70.42	72.74	68.65	60.49	70.63
	$R_{lcs}$		91.46	85.62	81.16	<b>67.64</b>	48.65	78.47
64	ROUGE-L		69.66	71.35	72.90	68.04	57.21	<b>70.82</b>
	$R_{lcs}$		90.65	86.82	79.96	65.74	44.89	77.64
128	ROUGE-L		65.53	66.98	<b>73.01</b>	66.91	64.10	69.42
	$R_{lcs}$		91.46	83.82	81.37	66.73	<b>51.15</b>	77.90
256	ROUGE-L		<b>73.37</b>	68.63	72.51	66.83	60.51	69.85
	$R_{lcs}$		<b>93.09</b>	86.94	<b>81.75</b>	66.01	47.16	<b>78.64</b>

## 论文总结

针对政策型问答场景中遇到的输入长度超越极限、长答案抽取不完整的挑战,本文提出了答案边界整合框架BoundaryQA,该框架提供了三种边界整合策略,其中局部片段合并策略局部地合并那些因文档块切分而被分开的两个邻近候选答案片段;全局边界重选策略训练一个答案边界重选模型来全局地从所有候选答案边界中重新选择出一对最好的开始和结束边界;跳跃式重阅读策略则重新阅读一段包含全局候选答案和边界局部上下文信息的简短摘要并从中抽取更准确的答案。本文最后在真实疫情政策问答数据上进行了实验,验证了BoundaryQA框架的有效性。