

融合关键信息与专家网络的生成式文本摘要

魏盼丽, 王红斌

(昆明理工大学 信息工程与自动化学院, 云南省人工智能重点实验室, 昆明 650500)

摘要: 针对现有生成式摘要模型生成过程中存在原文本关键信息缺失和内容难控制的问题, 提出一种结合抽取方法引导的生成式文本摘要方法. 该方法首先通过抽取模型从原文本中获取关键句, 然后采用双编码策略, 分别编码关键句和新闻文本, 使关键信息在解码过程中引导生成摘要, 最后引入专家网络在解码时筛选信息, 以进一步引导摘要生成. 在数据集 CNN/Daily Mail 和 XSum 上的实验结果表明, 该模型可有效改进生成式文本摘要的性能. 该方法在一定程度上提高了生成摘要对原文本关键信息的包含量, 同时缓解了生成内容难控制的问题.

关键词: 生成式文本摘要; 双编码器; 关键信息; 专家网络; 引导感知

中图分类号: TP391 **文献标志码:** A **文章编号:** 1671-5489(2024)04-0951-09

Fusing Key Information and Expert Network for Abstractive Text Summarization

WEI Panli, WANG Hongbin

(Yunnan Key Laboratory of Artificial Intelligence, Faculty of Information Engineering and Automation,
Kunming University of Science and Technology, Kunming 650500, China)

Abstract: Aiming at the problems of missing key information and difficult control of content in the original text during the generation process of existing generative summary models, we proposed a generative text summarization method guided by extraction methods. This method first obtained key sentences from the original text through an extraction model, and then adopted dual encoding strategy to encode key sentences and news text respectively, so that key information was guided to generate a summary during the decoding process. Finally, expert network was introduced to screen information during decoding to further guide the generation of summary. The experimental results on CNN/Daily Mail and XSum datasets show that the proposed model can effectively improve the performance of abstractive text summarization. This method improves the content of key information in the original text for generating summary to a certain extent, while alleviating the problem of difficult control of generated content.

Keywords: abstractive text summarization; double encoder; key information; expert network; guided perception

随着互联网产生的文本数据越来越多, 文本信息过载问题日益严重, 如新闻、微博等, 因此对各

收稿日期: 2023-09-08.

第一作者简介: 魏盼丽(1995—), 女, 汉族, 硕士, 从事自然语言处理的研究, E-mail: 3289480148@qq.com. **通信作者简介:** 王红斌(1983—), 男, 汉族, 博士, 教授, 从事自然语言处理、信息检索和机器学习的研究, E-mail: whbin2007@126.com.

基金项目: 国家自然科学基金(批准号: 61966020)和云南省基础研究计划面上项目(批准号: 202202AT070157).

类文本进行降维处理非常必要,自动文本摘要是一种重要方法.自动文本摘要的主要目的是将一个较长的文档压缩为较短的文本,以便于用户在互联网海量数据中快速获取有用信息.

自动文本摘要旨在从输入文档中生成一个精炼、简洁的摘要,同时保留输入文档的关键信息.自动文本摘要任务主要可分为抽取式摘要和生成式摘要.抽取式摘要是从输入文档中识别出最合适的单词或句子,并将它们连接成摘要.神经网络模型将抽取式摘要视为句子分类问题,首先创建输入文本的适当表示,以方便文本分析,然后进行句子评分,根据输入的文本表示对句子进行排序,最后从文档中选择评分较高的语句,并将其连接形成摘要.生成式摘要能对原文进行理解生成新的句子和单词,可自由地生成摘要,所以可能包含原文本中未出现过的表达.因此,与抽取式摘要相比,生成式摘要更灵活,能产生流畅、连贯的摘要.

目前,基于循环神经网络(recurrent neural networks, RNN)的序列到序列(sequence to sequence, seq2seq)模型在机器翻译领域取得了很好的效果.自动文本摘要问题也可视为原文本到目标文本的映射,因此可通过该方法解决.近年来,Wang 等^[1]提出了基于卷积神经网络的 seq2seq 框架,并引入了注意力机制和基于主题信息的强化学习,以帮助模型生成连贯和信息丰富的摘要.为解决摘要生成内容重复的问题,See 等^[2]提出了指针网络和覆盖机制,指针网络在保留新词内容的同时从原文本中抽取内容,以生成更准确的摘要;覆盖机制用于在解码器中记录已经生成的内容,以减少生成内容的重复.Narayan 等^[3]概念化抽取式摘要作为句子排序任务,提出了一种新的训练算法,通过强化学习系统优化 ROUGE 评估指标进行全局训练,以生成含有丰富信息的摘要.随着 BERT (bidirectional encoder representation from transformers)^[4]等一系列预训练模型的发展,预训练模型在自然语言处理领域被广泛应用,可在简化模型的同时取得更好的效果.Liu 等^[5]提出了一种基于对比学习的打分模型,通过训练无参考摘要的打分模型近似需要参考摘要的评价指标,直接学习评价指标本身的打分模式,在一定程度上缓解目标函数与评价指标不一致的问题,从而在候选摘要中选出性能更高的摘要.Su 等^[6]提出了一种两阶段的变长生成文本摘要方法,其由一个文本分割模块和一个基于两级转换器的摘要模块组成,在捕捉句子之间的关系方面取得了良好的效果.Zhong 等^[7]将文档摘要作为语义匹配任务,并使用 Siamese-BERT 作为匹配模型.Jin 等^[8]和 Zhu 等^[9]从原文档中提取(主题、关系、对象)形式的关系三元组,并用图神经网络进行表示,然后由解码器负责处理所提取的关系,以生成更贴合原文档的摘要.Dou 等^[10]提出了一种基于 BERT 的双编码器引导模型,通过使用各种类型的引导信号更好地处理摘要生成内容难控制的问题,同时也使摘要内容与原文档偏离更小,更具可控性.Jiang 等^[11]提出了基于图的主题感知生成文本框架 GTASum,无缝地集成了一个神经主题模型寻找文本中潜在的主题信息,通过维护文档级特征生成摘要.Cui 等^[12]提出了一种混合文本摘要模型,该模型使用神经主题模型(neural topic model, NTM)推断潜在主题作为一种全局信息,以此提高生成摘要的准确度.由于生成式摘要比较灵活,因此如何确保生成的摘要相对原文档的事实一致性至关重要.Ravaut 等^[13]提出了基于专家混合架构的多任务学习框架 SummaReranker,其是第一个用于生成摘要的多任务重新排序框架,可在多个措施上进行联合优化.Zhou 等^[14]提出了一种选择性门控网络,旨在使生成的文本摘要中保留更多的关键信息.但由输入文本表示控制的选择性门控网络只控制一次从编码器到解码器的信息流,如果某些关键信息未通过网络,则很难出现在摘要中,从而导致生成的摘要缺乏关键内容,甚至还可能导致事实性错误.

由于序列到序列模型的广泛应用,生成式摘要模型可生成较高 ROUGE 分数的摘要,虽然这些模型已被证明可以捕捉到自动文本摘要的规律性,但对摘要的内容在生成过程中却很难控制,有时生成摘要的内容不符合原文事实,易导致事实性错误.例如,对原文本“The classic video game ‘Space Invaders’ was developed in Japan back in the late 1970’s”,生成的摘要却为“Video game ‘Space Invaders’ was developed in Japan back in 1970”.错误地将原文本中所表示的“在 20 世纪 70 年代”表达为“在 1970 年”.针对这种问题,如果除编码原文档外,再加以外部的引导信号,不仅能获取输入原文本的上下文内容,还可以得到外部知识,在摘要生成过程中对生成内容进行控制,减少事实性错误,从而缓解生成摘要准确率较低的问题.虽然现有的注意力机制会注意到文本中的一些信息,但对部分

关键信息的识别不足, 仍会在解码过程中丢失关键信息, 使生成摘要的内容不准确, 不能完全概括原文的主旨.

为解决上述问题, 本文提出一种融合关键信息与专家网络的生成式文本摘要方法, 使用抽取式摘要模型抽取的语句作为关键信息引导生成摘要, 使模型输出摘要内容与原文档的偏差更小, 并可使生成摘要的内容存在一定的可解释性. 本文模型将预先选择的关键信息添加到注意力机制中, 使模型更关注原文本的关键内容, 在解码生成输出时用关键句引导模型倾向于关注原文本中的主旨内容, 然后使用专家选择网络进一步筛选信息, 以提高生成摘要的准确性和可读性, 有效减少冗余的生成. 本文的主要贡献如下:

- 1) 基于改进的 Transformer 架构, 采用双编码策略, 引入丰富的文本语义表征; 在解码端引入专家网络筛选信息, 减少冗余内容的生成.
- 2) 关键信息作为一种附加知识融入到解码过程中, 约束文本摘要的生成过程, 使输出内容忠于原文, 有效保持摘要和原文档事实一致.
- 3) 在数据集 CNN/DM 和 XSum 上的实验结果表明, 本文模型改进有效, 该方法在 ROUGE-1, ROUGE-2, ROUGE-L 评估指标上准确率均有提升.

1 方法设计

与基于 RNN 的体系架构相比, Transformer 的编码器和解码器采用注意力机制作为其主要架构, 能更好地编码文本的上下文语义信息, 提取文本语义表征, 所以本文采用 Transformer 模型作为文本摘要模型的基本架构. 首先使用微调的 BERT 预训练模型分别编码输入的文档 $x = \{x_1, x_2, \dots, x_n\}$ 和预先选择的关键信息文本 $h = \{h_1, h_2, \dots, h_n\}$, 然后映射到序列连续表示 $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n)$ 和 $\mathbf{H} = (\mathbf{H}_1, \mathbf{H}_2, \dots, \mathbf{H}_n)$ 中. 为使模型在解码过程中更好地获取到原文本的关键信息, 本文将关键信息向量 $\mathbf{H} = (\mathbf{H}_1, \mathbf{H}_2, \dots, \mathbf{H}_n)$ 输入到 Transformer 解码器的底层编解码注意力层中, 注意力机制首先关注关键信息, 通知解码器应该关注原文档的哪一部分, 然后解码器基于引导感知的表示处理整个原文档, 在生成目标摘要 $y = \{y_1, y_2, \dots, y_n\}$ 的标记过程中通过专家网络筛选信息, 最后模型以一种自回归的方式建模条件概率 $p = (y_1, y_2, \dots, y_n | x_1, x_2, \dots, x_n)$, 根据词表生成相对应的摘要. 模型总体框架如图 1 所示.

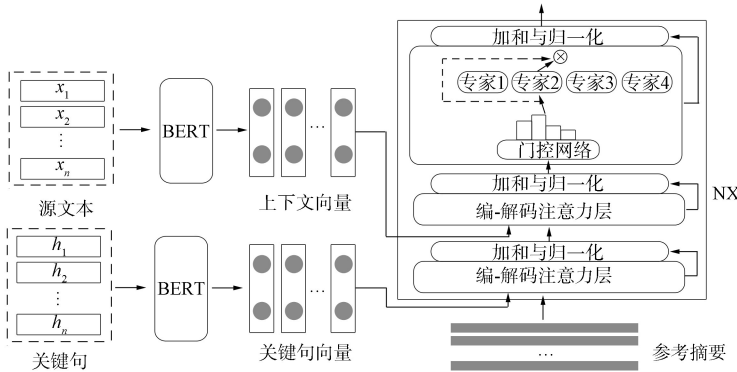


图 1 模型总体框架

Fig. 1 Overall frame of model

1.1 关键信息表示模块

仅编码原文本的模型局限于关注文本的上下文信息, 很难凸显原文本的显著特征, 即关键信息, 因此本文引入关键信息作为一种外部知识, 增强语义特征, 使生成摘要内容忠于原文事实.

抽取式文本摘要方法从原文档中抽取文本的子集形成摘要, 因此, 原文档中的重要语句对生成式摘要会起关键的引导作用, 故本文选择使用抽取的语句作为关键信息确切地告知模型应该重点关注原文档的哪一部分, 使生成内容更贴合文本主旨. 与输入关键字作为辅助信息相比, 关键字可能会丢失上下文信息之间的联系, 例如实体之间的关联情况等, 关键句抽取模块采用抽取摘要模型 Match-

Sum^[7]或 Bertext^[15]实现自动预测.

由于 BERT 通过联合调节上下文生成文本的双向表示,可更好地表示关键信息,因此本文采用微调的 BERT 预训练模型对关键句编码:

$$\mathbf{H} = \text{BERT}(h_1, h_2, \dots, h_n), \quad (1)$$

其中 h_i 表示预选择的第 i 个关键句.

在训练时,为使模型更密切地关注关键信息,本文使用贪婪搜索算法训练模型在原文档中找到一组 ROUGE 得分最高的句子,视为能较好传达文本主旨的语句,将其作为关键引导句.在测试时,本文使用抽取摘要模型 Match-Sum^[7]或 Bertext^[15]实现关键句的自动预测,以约束模型的输出,如图 2 所示.在解码阶段模型会首先关注关键信息并产生相应的表示,关键信息再通知解码器应该重点关注原文档的哪一部分.

1.2 原文档编码表示模块

对原文档的编码表示,本文同样采用微调的 BERT 预训练模型编码,与 Transformer 模型类似,BERT 预训练模型的每层包含一个多头自注意力模块和一个前向反馈层,两个子层之间用残差连接,然后进行层归一化,用公式表示为

$$\mathbf{X} = \text{LN}(\mathbf{X} + \text{SELFATTN}(\mathbf{X})), \quad (2)$$

$$\mathbf{X} = \text{LN}(\mathbf{X} + \text{FEEDFORWARD}(\mathbf{X})), \quad (3)$$

其中 LN 为层归一化表示.

在多头自注意力模块中,使用放缩点积注意力函数,输入由维度为 d_k 的查询向量和键向量以及维度为 d_v 的值向量组成,然后计算所有键查询的点积,用 Softmax 函数输出分布在值向量 \mathbf{V} 上的注意力权重,输出矩阵为

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{Softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}}\right)\mathbf{V}, \quad (4)$$

其中 $\mathbf{Q}, \mathbf{K}, \mathbf{V}$ 分别表示查询向量、键向量和值向量, d_k 表示键向量 \mathbf{K} 的维度. $\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V})$ 是 \mathbf{V} 的加权和向量,表示当前的上下文信息.对 Transformer 中的注意力机制采用多头实现,但 $\mathbf{Q}, \mathbf{K}, \mathbf{V}$ 较小,其维数分别为原始维数的 $1/h$,来自 h 头的注意力拼接在一起,通过线性投影形成最终的注意力,这样多头注意力机制提供了一个有利于最终性能的注意行为的多个视角,用公式表示为

$$\text{MultiHead}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{Concat}(\text{head}_1, \dots, \text{head}_h), \quad (5)$$

其中 $\text{head}_i = \text{Attention}(\mathbf{Q}\mathbf{W}_i^Q, \mathbf{K}\mathbf{W}_i^K, \mathbf{V}\mathbf{W}_i^V)$, $\mathbf{W}_i^Q, \mathbf{W}_i^K, \mathbf{W}_i^V$ 是可学习的参数矩阵.

在前向反馈层中,由两个线性转换和中间的 ReLU 激活函数组成,其作用是增加模型的非线性拟合能力,用公式表示为

$$\text{FFN}(x) = \max\{0, x\mathbf{W}_2 + b_2\}\mathbf{W}_3 + b_3, \quad (6)$$

其中 $\mathbf{W}_2, \mathbf{W}_3$ 为线性转换, b_2, b_3 为偏置.

1.3 带有专家网络的解码器

传统的 Transformer 解码器每层末尾都有一个前馈网络层,用于聚合来自注意力模块多个头的输出,即所有的参数全部参与计算,但并非模型中存储的所有信息都与特定的输入有关,为筛选模型中与特定输入有关的信息,本文使用专家网络层,如图 3 所示.该层从专家集合 $\{E_i(\mathbf{z}_t)\}_{i=1}^N$ 中选择将特定输入路由给最优的专家,进一步筛选信息.

本文将解码器中编-解码注意力模块的输出 \mathbf{z}_t 作为专家网络层的输入,在每个步骤 t 中,经过编-解码注意力模块输出,用公式表示为

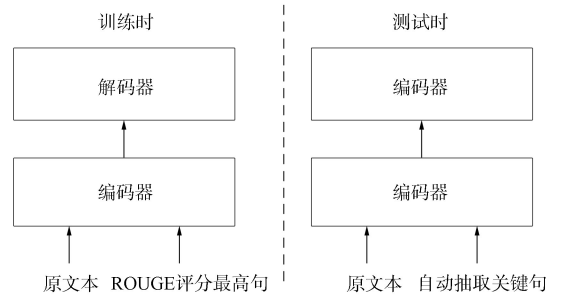


图 2 模型两阶段训练示意图

Fig. 2 Schematic diagram of two-stage training of model

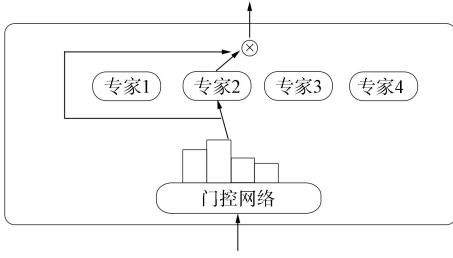


图 3 专家网络

Fig. 3 Expert network

$$z_t = \text{LN}(z_t + \text{SELFATTN}(z_t)), \quad (7)$$

$$z_t = \text{LN}(z_t + \text{CROSSATTN}(z_t, H)), \quad (8)$$

$$z_t = \text{LN}(z_t + \text{CROSSRATTN}(z_t, X)), \quad (9)$$

其中 LN 为层归一化表示。

编-解码注意力模块输出 z_t 进入专家网络层后先乘以路由矩阵 W_r , 得到各专家对于特定输入的得分, 即 $w = W_r \cdot z_t$, 然后由 Softmax 函数归一化为概率分布, 专家 i 的门控值计算如下:

$$p_i(z_t) = \frac{e^{w_i}}{\sum_{j=1}^N e^{w_j}}. \quad (10)$$

参考 Fedus 等^[16]的工作, 与选择多个专家相比, 本文采用一种简化策略, 在每层对特定的输入, 模型只激活一个专家, 因此输入 z_t 会选择最高的概率通过专家网络, 被选择的专家 i 的概率分布如下:

$$P(z_t) = \max\{p_i(z_t)\}. \quad (11)$$

专家网络层的输出由专家产生激活, 并通过其概率得分加权, 计算公式如下:

$$Y = P(z_t)E_i(z_t), \quad (12)$$

其中 $E_i(z_t)$ 表示对给定输入 z_t 第 i 个专家的输出. 在模型解码器的顶部, 用 Softmax 层将解码器的输出转换为摘要字生成概率。

1.4 损失函数

在训练过程中采用交叉熵损失函数, 用标签平滑策略, 从而提高神经网络的泛化能力和学习速度, 防止模型过拟合. 标签平滑可降低经过集束搜索后的单词错误率, 经过标签平滑过后的样本交叉熵损失不仅考虑了训练样本中 one-hot 标签为 1 位置的损失, 也考虑了到 one-hot 标签为 0 位置的损失, 在一定程度上通过标签平滑策略可缓解模型“盲目自信”的问题, 从而提高模型的学习能力. 经过标签平滑后的交叉熵损失可表示为

$$\text{Loss} = - \sum_{i=1}^N \text{LS}(\text{one_hot}) \times \log(\text{Softmax}(\text{logit}_i)), \quad (13)$$

其中 one_hot 表示样本标签转化的独热向量, $\text{LS}(\text{one_hot})$ 表示标签平滑操作后的样本标签, logit_i 表示解码器经过全连接层后的输出。

2 实验与分析

2.1 实验数据集

实验采用新闻自动文本摘要数据集 CNN/DM 和 XSum, 各数据集信息列于表 1. 数据集 CNN/DM 包含从美国有限新闻网(CNN)和每日邮报网(Daily Mail)上收集的新闻文章及对应的摘要. 本文采用 Hermann 等^[17]处理后的版本, 其中包含 287 226 组数据用于训练, 13 368 组数据用于验证, 11 490 组数据用于测试, 用于训练的原文档每条新闻平均包含 760 个单词, 共 29.74 句组成; 其对应的摘要由 53 个单词, 共 3.72 句组成. 数据集 XSum 中每篇新闻文章对应的摘要仅为一句话, 其中新单词占 83.71%, 因此是高度抽象的. 本文使用 Stanford CoreNLP 工具包分割文本语句, 并对实验数据集进行预处理, 输入文档被截断为 512 个令牌。

表 1 各数据集信息

Table 1 Information of each dataset

数据集	数据对(训练/验证/测试)	文档平均长度		摘要平均长度		参考摘要中 新单词占比/%
		单词数	句子数	单词数	句子数	
CNN	90 266/1 220/1 093	760.50	33.98	45.70	3.59	52.90
DailyMail	196 961/12 148/10 397	653.33	29.33	54.65	3.86	52.16
XSum	203 028/11 273/11 332	431.07	19.77	23.26	1.00	83.71

2.2 评价指标

自动文本摘要的评估方法可分为内部评估方法和外部评估方法两类. 内部评估方法提供摘要, 并基于参考摘要评估生成摘要的质量; 外部评估方法并未提供参考摘要. 本文使用 Lin^[18] 提出的内部评价方法 ROUGE 指标评估生成摘要的质量. ROUGE 指标主要评估生成摘要与参考摘要之间的共现信息, 共现信息量越多, 评价指标分数越高, 则表明模型生成的摘要质量越高. ROUGE 评价指标按信息共现量分为 ROUGE-1, ROUGE-2 和 ROUGE-L, 其中 ROUGE-1, ROUGE-2 分别表示生成摘要与参考摘要中词和二元词语的重合程度, ROUGE-L 表示生成摘要与参考摘要中最长公共子序列的重合程度, 其计算方法如下:

$$\text{ROUGE-N} = \frac{\sum_{S \in \{\text{Ref}\}} \sum_{n\text{-grams} \in S} \text{Count}_{\text{match}}(n\text{-gram})}{\sum_{S \in \{\text{Ref}\}} \sum_{n\text{-grams} \in S} \text{Count}(n\text{-gram})}, \quad (14)$$

其中 $n\text{-gram}$ 表示 n 个单词, $\{\text{Ref}\}$ 表示参考摘要, $\text{Count}_{\text{match}}(n\text{-gram})$ 表示生成摘要与参考摘要中同时出现 $n\text{-gram}$ 的数量, $\text{Count}(n\text{-gram})$ 表示参考摘要中出现 $n\text{-gram}$ 的数量.

2.3 参数设置

本文用预训练 BERT-base-uncased 模型初始化文档编码器, 并随机初始化解码器. 由于随机梯度下降在整个训练过程中保持单一学习率更新所有权重, 而 Adam 优化算法通过计算梯度的一阶矩估计和二阶矩估计为不同的参数设计独立的自适应学习率, 因此在解码阶段使用 Adam 优化算法默认学习率设为 2×10^{-3} , 动量参数 $\beta_1 = 0.9$, $\beta_2 = 0.999$. 为解决编码器与解码器不匹配的问题, 采用新的微调计划, 分离编码器和解码器的优化器:

$$lr_{\text{ENC}} = 2 \times 10^{-3} \cdot \min\{\text{step}^{-0.5}, \text{step} \cdot \text{warmup}_{\text{ENC}}^{-1.5}\}, \quad (15)$$

$$lr_{\text{DEC}} = 0.1 \cdot \min\{\text{step}^{-0.5}, \text{step} \cdot \text{warmup}_{\text{DEC}}^{-1.5}\}, \quad (16)$$

其中: 编码器学习率为 0.002, $\text{warmup}_{\text{ENC}} = 20\ 000$; 解码器学习率为 0.1, $\text{warmup}_{\text{DEC}} = 10\ 000$. 为防止模型过拟合, 将非专家网络层的 Dropout 设为 0.2, 专家网络层的 Dropout 设为 0.5, 专家层个数设为 4. 在测试阶段使用集束搜索算法进行解码, 大小设为 5. 实验采用单张 GTX3090Ti(GPU)进行训练.

2.4 实验结果分析

2.4.1 对比实验

为证明本文方法的有效性, 将对以下基准模型并直接从原文献中抽取实验结果.

1) RNN-Ext-Abs+RL^[19]: 基于强化学习的句子级的文本摘要生成模型. 先对选取的句子进行重写, 并根据强化学习中的梯度策略, 提出将句子选择与句子摘要连接起来的方法, 在一定程度上减少了冗余内容的生成.

2) Bert-Abs^[15]: 先通过 BERT 预训练模型编码原文档, 再将原文档的上下文表示输入到 Transformer 解码器中生成摘要.

3) Bert-Hybrid^[20]: 其为一种新的抽取与生成混合框架, 先由抽取模型选择语句, 生成模型根据所选句子重写摘要, 再通过联合学习选择语句和重写摘要完成文本摘要任务.

4) Bert-Ext-Abs+RL^[21]: 基于 BERT 的抽取体系架构生成摘要, 直接最大化通过强化学习获得的概要级 ROUGE 评分, 优化获取的摘要.

5) Bert-Ext-Abs^[15]: 先由抽取式摘要模型 BertExt 抽取摘要, 再将抽取的摘要作为唯一输入到生成式摘要模型 BertAbs 中, 重新编码生成摘要.

6) ESCA-BERT^[22]: 其为一种新的抽取-生成框架, 该框架侧重于可解释性, 配备了成对排序抽取模型, 与配备句子级注意指针的生成式摘要模型无缝连接.

7) Bert-Copy/Rewrite+HRL^[23]: 基于分层强化学习, 提出一种端到端的强化方法, 将抽取模块和重写模块连接在一起, 根据冗余度灵活地在复制和重写句子之间切换, 提高摘要性能.

8) T-BERTSum^[24]: 先通过神经主题模型(NTM)将编码的潜在主题表示与嵌入的 BERT 表示进行匹配, 指导主题的生成, 然后通过 Transformer 网络学习长期依赖关系, 以端到端的方式共同探索

主题推理和文本摘要.

上述各自动文本摘要模型在数据集 CNN/DM 上的实验结果列于表 2.

表 2 不同自动文本摘要模型在数据集 CNN/DM 上的实验结果

Table 2 Experimental results of different automatic text summarization models on CNN/DM dataset

模型	ROUGE-1	ROUGE-2	ROUGE-L
RNN-Ext-Abs+RL ^[19]	40.18	17.80	38.54
Bert-Abs ^[15]	41.72	19.39	38.76
Bert-Hybrid ^[20]	41.76	19.31	38.86
Bert-Ext-Abs+RL ^[21]	41.90	19.08	39.64
Bert-Ext-Abs ^[15]	42.13	19.60	39.18
ESCA-BERT ^[22]	42.01	19.52	39.07
Bert-Copy/Rewrite+HRL ^[23]	42.92	19.43	39.35
T-BERTSum ^[24]	43.06	19.76	39.43
本文	43.26	20.30	40.14

由表 2 可见, 本文方法在 ROUGE-1, ROUGE-2 和 ROUGE-L 评价指标上优于其他对比方法. 对比 RNN-Ext-Abs+RL, Bert-Ext-Abs+RL, Bert-Copy/Rewrite+HRL 模型等使用强化学习或复制机制的摘要生成方法, 本文在各项评价指标上的实验效果仍有所提升, 表明在不使用复杂算法的情况下, 用文本关键信息有效引导, 模型仍可学习到文本的主旨内容, 提高了生成摘要的精确度. 与使用 RNN 编码的模型 RNN-Ext-Abs+RL 相比, 本文模型的实验效果也有一定提升, 表明使用 BERT 预训练模型编码, 不仅可提高模型的并行能力, 而且可提高模型的文本语义特征提取能力. 本文方法优于其他方法的关键原因是首先使用关键句作为引导信息, 可在一定程度上引导解码器关注原文档的重要内容, 使生成的摘要与原文档的内容偏差较小; 另一方面, 本文在注意力层后加入了专家网络层进一步筛选信息, 可有效减少冗余内容的生成. 实验结果表明, 本文方法在自动文本摘要生成任务上有效. 此外, 本文也在相对较抽象的数据集 XSum 上进行了实验验证, 实验结果列于表 3. 由表 3 可见, 本文模型性能优异, 但相对于偏抽取的数据集 CNN/DM 性能欠佳.

表 3 不同自动文本摘要模型在数据集 XSum 上的实验效果

Table 3 Experimental results of different automatic text summarization models on XSum dataset

模型	ROUGE-1	ROUGE-2	ROUGE-L
PTGEN ^[2]	29.70	9.21	23.24
PTGEN+COV ^[2]	28.10	8.02	21.72
TCONVS2S ^[3]	31.89	11.54	25.75
本文	37.13	14.71	29.45

2.4.2 显著性分析

本文选择从原文档中提取的信息作为引导信号, 但尚不清楚模型是否会过度拟合或者生成新的表达式. 为此, 计算数据集 CNN/DM 中生成文本摘要的重复率, 结果如图 4 所示. 由图 4 可见, 本文模型相比于基线模型重复率明显降低. 结果表明, 本文使用原文档中的重要句子作为关键信息引导模型可以生成新的表达式, 并且没有过于依赖输入的引导信息.

2.4.3 消融实验

为验证本文方法中各模块的重要性, 在数据集 CNN/DM 上进行消融实验, 实验结果列于表 4. 由

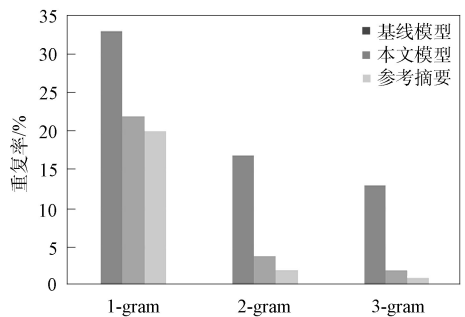


图 4 不同模型生成摘要的重复率

Fig. 4 Repetition rate of generated summarization by different models

表 4 可见, 加入关键句后的模型与基线模型相比, 在评价指标 ROUGE 上分别提升 1.64, 1.21,

1.53 个百分点,表明关键句的确可以有效引导模型,本文模型学会了如何依赖关键引导信号获取文本的主旨内容.在加入专家网络后,本文模型又取得了更高的准确率,表明了各模块在模型中的重要性.

表 4 消融实验结果对比

Table 4 Comparison of ablation experiment results

%

模型	ROUGE-1	ROUGE-2	ROUGE-L
基线模型	41.58	19.02	38.53
+关键信息	43.22	20.23	40.06
+专家网络	43.26	20.30	40.14

2.4.4 案例分析

表 5 列出了数据集 CNN/DM 中不同模型的输出结果.由表 5 可见,本文模型在数据集 CNN/DM 输出的摘要准确概括了文本所表达的主旨,如示例中文本的主要思想是“田纳西州默弗里斯伯勒弗韦尔高中勇士队教练罗恩·艾德洛特在袭击中面部严重受伤”.生成摘要的话语表达与关键句的表达相似,表明本文模型学会了依赖本文所提供的关键句信息,关键句确实在模型解码生成摘要时对生成的内容起到了引导作用.此外,本文生成的摘要相对较简洁,表明专家网络在生成摘要字时进一步筛选信息,在一定程度上减少了冗余内容的生成,提高了生成摘要的简洁性.

表 5 数据集 CNN/DM 中不同模型的输出结果

Table 5 Output results of different models in CNN/DM dataset

参考摘要: Ron aydelott, coach of the riverdale high school warriors in murfreesboro, tennessee, sustained serious facial injuries in the attack.
关键句: Ron aydelott, head coach for the riverdale high school warriors in murfreesboro for nearly ten years, suffered serious facial injuries in the attack, which will require surgery. Witnesses said aydelott in no way provoked the attack but that the 17-year-old alleged attacker became violent after he felt ‘disrespected’, reports news channel five network. the boy was reportedly in the coach’s office in order to turn in paperwork concerning his trying out for the team.
基线模型: Ron aydelott, the head coach of the Warriors at riverdale high school in murfreesboro, tennessee, suffered severe facial injuries in the attack that required surgery. Witnesses said aydelott did not provoke the attack, but the 17-year-old attacker turned violent after feeling ‘disrespected’.
本文模型: Ron aydelott, head coach for the riverdale high school warriors in murfreesboro, tennessee, suffered serious facial injuries in the attack, which will require surgery.

综上所述,针对现有生成式摘要模型生成过程中存在原文本关键信息缺失和内容难控制的问题,本文提出了一种融合关键信息与专家网络的生成式文本摘要模型,该模型采用双编码策略,使用 BERT 预训练模型分别编码原文档和关键句,以更好地获得文本语义表征.用关键信息在解码阶段引导生成摘要的内容,并采用专家网络进一步筛选信息,不仅使生成的摘要更精炼、简洁地概括文本的中心要义,而且使自动文本摘要的生成过程有一定的可解释性.在数据集 CNN/DM 和 XSum 上的实验结果表明,本文模型对生成式摘要的准确度有明显提升.

参 考 文 献

[1] WANG L, YAO J L, TAO Y Z, et al. A Reinforced Topic-Aware Convolutional Sequence-to-Sequence Model for Abstractive Text Summarization [C]//Proceedings of the International Joint Conference on Artificial Intelligence. New York: ACM, 2018: 4453-4460.

[2] SEE A, LIU P, MANNING C. Get to the Point: Summarization with Pointer-Generator Networks [C]//Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. [S. l.]: ACL, 2017: 1073-1083.

[3] NARAYAN S, COHEN S, LAPATA M. Ranking Sentences for Extractive Summarization with Reinforcement Learning [C]//Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics. [S. l.]: ACL, 2018: 1747-1759.

[4] 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. [S.l.]: ACL, 2019: 4171-4186.
- [5] LIU Y X, LIN P F. Simcls: A Simple Framework for Contrastive Learning of Abstractive Summarization [C]// Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing. [S.l.]: ACL, 2021: 1065-1072.
- [6] SU M H, WU C H, CHENG H T. A Two-Stage Transformer-Based Approach for Variable-Length Abstractive Summarization [J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2020, 28: 2061-2072.
- [7] ZHONG M, LIU P F, CHEN Y R, et al. Extractive Summarization as Text Matching [C]//Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. [S.l.]: ACL, 2020: 6197-6208.
- [8] JIN H Q, WANG T M, WAN X J. Semsun: Semantic Dependency Guided Neural Abstractive Summarization [C]//Proceedings of the AAAI Conference on Artificial Intelligence. Palo Alto: AAAI, 2020: 8026-8033.
- [9] ZHU C G, HINTHORN W, XU R C, et al. Boosting Factual Correctness of Abstractive Summarization with Knowledge Graph [EB/OL]. (2020-03-19)[2023-03-23]. <https://arxiv.org/abs/2003.08612>.
- [10] DOU Z Y, LIU P F, HAYASHI H, et al. GSum: A General Framework for Guided Neural Abstractive Summarization [C]//Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics. [S.l.]: ACL, 2021: 4830-4842.
- [11] JIANG M, ZOU Y F, XU J, et al. GATSum: Graph-Based Topic-Aware Abstract Text Summarization [J]. Information Technology and Control, 2022, 51(2): 345-355.
- [12] CUI P, HU L, LIU Y C. Enhancing Extractive Text Summarization with Topic-Aware Graph Neural Networks [C]// Proceedings of the 28th International Conference on Computational Linguistics. [S.l.]: ACL, 2020: 5360-5371.
- [13] RAVAUT M, JOTY S, CHEN N. SummaReranker: A Multi-task Mixture-of-Experts Re-ranking Framework for Abstractive Summarization [C]//Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics. [S.l.]: ACL, 2022: 4504-4524.
- [14] ZHOU Q Y, YANG N, WEI F R, et al. Selective Encoding for Abstractive Sentence Summarization [C]// Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. [S.l.]: ACL, 2017: 1095-1104.
- [15] LIU Y, LAPATA M. Text Summarization with Pretrained Encoders [C]//Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing. [S.l.]: ACL, 2019: 3730-3740.
- [16] FEDUS W, ZOPH B, SHAZEER N. Switch Transformers: Scaling to Trillion Parameter Models with Simple and Efficient Sparsity [J]. The Journal of Machine Learning Research, 2021, 23(1): 5232-5270.
- [17] HERMANN K M, KOCISKY T, GREFFENSTETTE E, et al. Teaching Machines to Read and Comprehend [C]//Proceedings of the 28th International Conference on Neural Information Processing Systems. New York: ACM, 2015: 1693-1701.
- [18] LIN C Y. Rouge: A Package for Automatic Evaluation of Summaries [C]//Proceedings of the ACL Workshop: Text Summarization Braches Out. [S.l.]: ACL, 2004: 74-81.
- [19] CHEN Y C, BANSAL M. Fast Abstractive Summarization with Reinforce-Selected Sentence Rewriting [C]// Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. [S.l.]: ACL, 2018: 675-686.
- [20] WEI R, HUANG H Y, GAO Y. Sharing Pre-trained BERT Decoder for a Hybrid Summarization [C]//Proceedings of Chinese Computational Linguistics; 18th China National Conference. New York: ACM, 2019: 169-180.
- [21] BAE S, KIM T, KIM J, et al. Summary Level Training of Sentence Rewriting for Abstractive Summarization [C]//Proceedings of the 2nd Workshop on New Frontiers in Summarization. [S.l.]: ACL, 2019: 10-20.
- [22] WANG H N, GAO Y, BAI Y, et al. Exploring Explainable Selection to Control Abstractive Summarization [C]//Proceedings of the AAAI Conference on Artificial Intelligence. Palo Alto: AAAI, 2021: 13933-13941.
- [23] XIAO L Q, WANG L, HE H, et al. Copy or Rewrite: Hybrid Summarization with Hierarchical Reinforcement Learning [C]//Proceedings of the AAAI Conference on Artificial Intelligence. Palo Alto: AAAI, 2020: 9306-9313.
- [24] MA T H, PAN Q, RONG H, et al. T-BERTSum: Topic-Aware Text Summarization Based on BERT [J]. IEEE Transactions on Computational Social Systems, 2022, 9(3): 879-890.