

文章编号: 2095-2163(2023)09-0158-06

中图分类号: TP391.1

文献标志码: A

结合主题和位置信息的两阶段文本摘要模型

任淑霞¹, 张靖¹, 赵宗现¹, 饶冬章²

(1 天津工业大学 软件学院, 天津 300387; 2 天津工业大学 计算机科学与技术学院, 天津 300387)

摘要: 预训练模型 BERT 显著提升了文本摘要领域模型的性能, 但是其在探索文档全局语义方面和对句子位置信息的利用方面还存在着不足。为了解决以上问题, 本文提出了一种结合双主题嵌入和句子绝对位置嵌入的两阶段自动摘要生成模型。首先, 文章在两个阶段分别引入主题嵌入, 融合了丰富的语义特征, 以捕获更准确的全局语义; 其次, 在抽取式阶段引入句子绝对位置嵌入将句子位置信息进行完全整合, 获得更全面的摘要抽取辅助信息以进行摘要提取; 在此基础上, 模型采用抽取-生成两阶段混合式摘要框架, 通过抽取阶段对文本重要信息的提取降低生成摘要内容的冗余性, 并进一步提高了模型的性能。在 CNN/Daily Mail 数据集上实验结果表明, 本文提出的模型取得了较好的结果。

关键词: 混合式摘要; BERT; 双主题嵌入; 句子绝对位置嵌入

A two-stage text summarization model combining topic and position information

REN Shuxia¹, ZHANG Jing¹, ZHAO Zongxian¹, RAO Dongzhang²

(1 School of Software, Tianjin University, Tianjin 300387, China;

2 School of Computer Science and Technology, Tianjin University, Tianjin 300387, China)

[Abstract] The pre-trained model BERT can significantly improve the performance of the text summarization task, but it still has some shortcomings in exploring the global semantics of documents and using the sentence position information. In order to solve the above problems, this paper proposes a two-stage automatic summary generation model which combines two-topic embedding and sentence absolute position embedding. Firstly, topic embedding is introduced in two stages, which integrate rich semantic features to capture more accurate global semantics. Secondly, absolute position embedding is introduced in the extraction stage to fully integrate the sentence position information, so as to obtain more comprehensive auxiliary information for abstract extraction. On this basis, the model adopts the extractive-abstractive two-stage hybrid summary framework, which reduces the redundancy of the generated summary content by extracting the important information of the text in the extraction stage and further improves the performance of the model. Experimental results show that the proposed model achieves good results on CNN/Daily Mail dataset.

[Key words] hybrid summarization; BERT; two-topic embedding; sentence absolute position embedding

0 引言

文本摘要生成是从输入文档中有效地压缩和提取信息, 同时保留其关键内容的过程。典型的摘要生成方法包括抽取式和生成式。抽取式摘要技术通过在原文中选择一个句子子集来生成摘要, 方法简单且内容与语法更准确, 但存在着内容冗余、语义不一致以及信息缺失等问题; 生成式摘要通过使用自然语言生成(NLP)方法产生具有理解力的新词或短语生成摘要, 可以更好的理解文档, 生成更符合人类所能理解的摘要。

Rush 等^[1]受到神经机器翻译成功的启发, 在 2015 年首次将 seq2seq 和注意力机制结合起来应用于文本摘要任务中; Chopra 等^[2]提出模型编码器, 同样使用卷积模型构造, 但添加了输入单词的位置信息, 而解码器则基于循环神经网络构造。以上方法相比于传统的方法, 效果有了极大的提升, 但仍存在着未登录词(Out-Of Vocabulary, OOV)、生成重复以及长程依赖的问题。2017 年, See 等^[3]提出了一种带有覆盖的混合指针生成器架构(PTGEN)。首先使用混合指针生成器网络, 通过指向从源文档复制单词, 有助于信息的准确再现, 缓解 OOV 问题, 同

基金项目: 天津市自然科学基金(19JCYBJC18700)。

作者简介: 任淑霞(1973-), 女, 博士, 副教授, 主要研究方向: 自然语言处理; 张靖(1995-), 男, 硕士研究生, 主要研究方向: 自然语言处理; 赵宗现(1997-), 男, 硕士研究生, 主要研究方向: 自然语言处理; 饶冬章(1997-), 男, 硕士研究生, 主要研究方向: 自然语言处理。

通讯作者: 任淑霞 Email: r_sx@126.com

收稿日期: 2022-10-13

时保留通过生成器生成新单词的能力;其次,使用一个覆盖机制来跟踪总结的内容,从而避免生成重复问题。同年,Paulus等^[4]提出了一种用于生成式摘要的深度强化模型(DRM),该模型是对PTGEN的改进,通过一个内注意力机制和新的学习目标来进一步解决生成重复的问题;2018年,Narayan等^[5]为了解决长程依赖问题,提出了一个完全基于卷积神经网络(CNN)的主题条件神经模型,与循环神经网络(RNN)相比,CNN能够捕获文档中的长距离依赖关系并识别相关内容。

随着预训练语言模型在NLP领域的快速发展,其在文本摘要任务中的应用也有极大的进展。Liu等^[6]在2019年提出了一种新的基于BERT的文档级编码器(BERTSUM),该编码器能够对文档进行编码,并获得其句子的表示,同时提出了一种通用的抽取式摘要和生成式摘要框架,性能有了极大的提升。然而,其模型编码器只嵌入了句子的相对位置信息,而没有嵌入绝对位置信息,对抽取摘要的结果造成影响。除此之外,BERTSUM模型更善于探索局部标记之间的关系,而对于更高层次(如句子、主题)的语义理解并不充分。近年来,预训练模型在文本摘要领域的发展主要在提升缺少标签数据情况下的模型性能,而对于如何解决预训练框架中的全局语义问题并没有过多的研究。

为了进一步提升文本摘要生成的性能,本文采用了抽取-生成两阶段混合式摘要生成方法。首先,对源文档的重要信息进行选择、合并、压缩或删除;其次,将提取得到的重要信息当作新的源文档,利用生成式摘要技术进行摘要生成。混合式方法结构非常清晰,只需将抽取阶段的结果当作生成阶段的输入,而不需要从根本上改变其架构。

本文构建了抽取-生成两阶段的摘要框架,对抽取的句子进行摘要生成,在两阶段间传递信息的同时大大减少句子冗余。首先,在两个阶段同时引入主题嵌入形成双主题嵌入,融合了丰富的语义特征,以捕获更准确的全局语义表示,从而提升生成摘要的质量;其次,在抽取式阶段引入句子绝对位置嵌入,将句子位置信息进行完全整合,获得更全面的摘要抽取辅助信息,从而提取更加重要的句子以进行摘要生成。实验结果表明,本文提出的模型在CNN/Daily Mail数据集上取得了较好的结果。本文还进行了人工评估实验,其结果也优于其他对比模型。

1 本文模型研究

本文模型的体系结构如图1所示。本文的模型

是抽取-生成两阶段的微调模型,通过将抽取式摘要生成和生成式摘要生成技术结合起来以帮助生成更高质量的摘要。模型两个阶段都在编码器中添加主题嵌入以捕获更丰富的全局语义信息,并且在抽取式阶段添加句子绝对位置嵌入来完全利用句子位置信息,提取高质量的摘要信息。

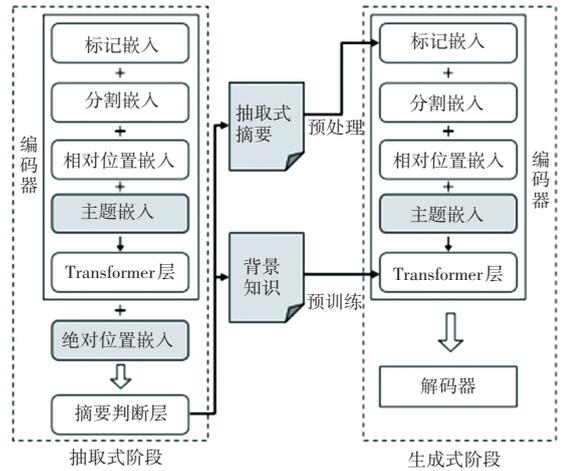


图1 模型的体系结构

Fig. 1 The architecture of model

模型将源文档 D 表示成一个包含 n 个句子的文档 $\{s_1, s_2, \dots, s_n\}$,其中 s_i 表示第 i 句在文档中的文本顺序。两个阶段的编码器都是采用类似BERT的结构。

首先,插入两个特殊标记对输入文本进行预处理,在每个句子的开头插入标记[CLS],利用该标记计算出的输出来整合每个序列中的信息,在每个句子的结尾插入标记[SEP],作为句子边界的指示符;其次,利用分段嵌入来区分句子,达到在不同层次学习相邻句子特征或段落句子特征的目的。由于Transformer的自我注意层处于不同位置的相同单词具有相同的输出表示,因此需要引入位置嵌入来恢复位置信息,本文使用正弦位置嵌入来允许模型学习给定序列中标记的相对位置。为了全面准确地表达每个序列的上下文信息,本文添加了主题嵌入。

对于每个输入句子,本文通过4种嵌入类型即标记嵌入、分割嵌入、位置嵌入和主题嵌入的总和输出,以充分表达句子的语义特征;将前面的输出结果 h^0 通过一个多层的双向transformer进行处理,得到上下文嵌入向量,式(1)和式(2):

$$\tilde{h}^l = \text{LN}(h^{l-1} + \text{MHAtt}(h^{l-1})) \quad (1)$$

$$h^l = \text{LN}(h^l + \text{FFN}(\tilde{h}^l)) \quad (2)$$

其中,LN为层归一化操作;MHAtt为多头注意

操作;FFN 为两层前馈网络操作;上标 l 表示堆叠层的深度。

最终输出的上下文嵌入表示为 $c \triangleq [c_1, c_2, \dots, c_n]$, 其中 c_i 为第 i 个 [CLS] 标记对应的向量, 可以被看作是第 i 个句子 s_i 的向量表式。

1.1 双主题嵌入

预训练语言模型主要针对词或句子级别进行设计, 因此在捕获整个文档的语义方面表现不佳。比如, BERTSUM 模型更善于探索局部标记之间的关系, 而在更高层次, 如句子、主题、语义理解方面做的并不好。因此, 本文添加主题信息嵌入来解决全局语义缺失问题, 模型采用双主题嵌入, 即在抽取和生成两个阶段同时引入主题嵌入, 双主题嵌入降低了源文档无关信息的干扰, 更准确的捕获全局语义信息。主题模型擅长捕获文本的全局语义, 本文使用潜在的狄利克雷分配 (LDA) 来获得文档和单词的主题分布。

让 \bar{D} 表示一个由一系列文档 $\{D_1, D_2, \dots, D_m\}$ 组成的文档集, 每个文档 D 由一系列单词 $\{w_1, w_2, \dots, w_n\}$ 组成。LDA 的生成过程:

$$\theta_D \sim Dir(\alpha_0)$$

$$z_n \sim Multi(\theta_D)$$

$$w_n \sim Multi(\beta_{z_n})$$

其中, α_0 是狄利克雷先验的超参数, β_{z_n} 表示给定主题分配的单词的主题分布 z_n 。

抽取阶段引入的主题嵌入 T 主要依赖源文档的文档主题分布 z_n 和单词主题分布 w_n , 而在生成阶段引入的主题嵌入 T' 主要依赖源文档的文档主题分布 z_n 以及抽取后文档单词的主题分布 w'_n 。式(3)和式(4):

$$T = z_n \otimes w_n \quad (3)$$

$$T' = z_n \otimes w'_n \quad (4)$$

其中, \otimes 表示点乘。单词 w_n 与 w'_n 的主题分布本质上捕捉了单词本身的主题程度(局部上下文), 而文档的主题分布 z_n 表示文档的整体主题(全局上下文)。

本文通过双主题嵌入降低了文档冗余信息的影响, 更准确的捕获文档的全局语义信息, 进一步提高了生成摘要的质量。

1.2 绝对位置嵌入

为了捕获文档级特征以提取摘要, BERTSUM 模型在 BERT 输出的顶部构建了句子间的 Transformer 层, 但 Transformer 并没有对其输入句子

的绝对位置信息进行编码, 且 BERT 中的位置嵌入层也不能提供文档级的句子位置信息。但对抽取式摘要来说, 句子位置信息对于文档重要信息的提取十分重要。因此, 本文通过句子的绝对位置嵌入来合并句子位置信息, 解决了 BERT 中的位置嵌入层不能在文档级别提供句子位置信息的问题。

假设摘要生成句代表源文档的主要内容, 抽取式摘要可以定义为源文档中的每个句子分配标签 $y_i \in \{0, 1\}$ 的任务, 0 或 1 代表着当前句子是否作为摘要提取的句子。抽取式阶段的流程如图 2 所示。

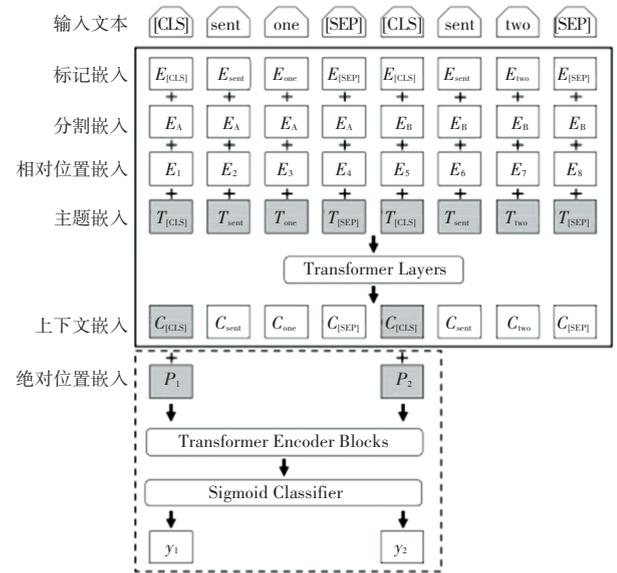


图 2 抽取式阶段流程

Fig. 2 The process in extractive stage

通过编码器编码获得句子向量 c_i 后, 引入句子绝对位置嵌入 $p \triangleq [p_1, p_2, \dots, p_n]$, p_i 代表着第 i 个句子在文档中的绝对位置, 最后通过摘要判断层来获得最终的标签以提取摘要。本文的摘要判断层使用两个句子间 Transformer 层和简单分类器的组合, 即式(5)和式(6):

$$z^0 = \text{PosEmb}(C) \quad (5)$$

$$z^l = \text{Transformer}(x^{l-1}), l = 1, \dots, L \quad (6)$$

其中, $\text{PosEmb}()$ 表示绝对位置嵌入, $\text{Transformer}()$ 表示句子间 Transformer 层。

在 Transformer 的输出上添加一个线性全连接层, 并使用一个 sigmoid 函数进行二分类, 获得预测标签 Y_i , 式(7):

$$Y_i = \sigma(W_0 h_i^L + b_0) \quad (7)$$

其中, σ 表示 sigmoid 函数; h_i^L 表示 Transformer 的顶层的第 i 个句子 s_i 的向量; L 表示二分类之前堆叠了 Transformer 层数。

如果标签结果为0,则将该句子归为背景知识,用于生成式阶段编码器的预训练任务;如果标签结果为1,则将该句作为抽取式阶段的抽取结果,用于生成式阶段的摘要生成。

本文通过句子的绝对位置嵌入来将句子位置信息进行完全整合,解决了BERT中的位置嵌入层不能在文档级别提供句子位置信息的问题,获得更全面的摘要抽取辅助信息。

1.3 抽取-生成两阶段模型

抽取-生成两阶段混合式摘要方法首先通过抽取式技术对源文档中的重要信息进行提取,然后将其作为新的文档通过生成式技术进行摘要生成。其将两种方法结合起来生成摘要,取长补短,进一步提升模型的性能。但传统的混合式摘要方法只是简单的将抽取式方法和生成式方法结合起来,对于输入的源文档内容的利用并不是很充分。

而本文使用的抽取-生成两阶段的框架,将源文档内容分为两大类,一类是抽取式阶段作为摘要结果的句子,用于下一阶段的摘要生成任务;另一类是其余不作为抽取摘要结果的句子,用于下一阶段编码器的预训练任务。这种改进降低生成摘要冗余的同时不丢失源文档的其他背景知识,使得最终生成的摘要对于源文档的概括更加全面。

本文将源文档 D 通过抽取式阶段得到每个句子的预测标签 y_i ,再将标签为1的句子提取出来,作为一个新的源文档 D' ,输入至生成式阶段。对新的源文档 D' 进行预处理,再次通过4种嵌入来捕获丰富的语义特征,得到嵌入后的结果 E' ,其过程如式(8):

$$E' = \text{concat}(E'_i, E'_s, E'_p, T') \quad (8)$$

将词嵌入的结果通过Transformer层得到新的上下文嵌入向量 C' ,再将其传入解码器进行解码,最终输出生成的摘要。

本文的编码器是预处理的BERT模型的变体,解码器是随机初始化的6层Transformer。为了避免编码器和解码器之间存在的匹配问题,本文编码器和解码器使用两种不同的Adam优化器。两个优化器的 β_1 和 β_2 设置相同,都使用 $\beta_1 = 0.9$ 和 $\beta_2 = 0.999$,而warmup和学习率则不同,式(9)和式(10):

$$lr_E = \widetilde{lr}_E \min(\text{step}^{-0.5}, \text{step warmup}_E^{-1.5}) \quad (9)$$

$$lr_D = \widetilde{lr}_D \min(\text{step}^{-0.5}, \text{step warmup}_D^{-1.5}) \quad (10)$$

其中, \widetilde{lr}_E 和 \widetilde{lr}_D 分别是每个step的编码器和解

码器的学习率。

因为经过预处理的编码器应该以较小的学习速率和更平滑的衰减进行微调。这样,当解码器变得稳定时,编码器可以以更精确的梯度进行训练。因此本文的参数设置:编码器的学习率为0.002, warmup为20 000,解码器的学习率为0.1, warmup为10 000。

本文使用抽取-生成两阶段框架,不仅降低了生成摘要的冗余,而且利用文档的背景知识减少了摘要内容的缺失,进一步提升了模型的性能。

2 实验测试与分析

2.1 数据集

本文实验使用的数据集是CNN/Daily Mail,由CNN和Daily Mail的新闻文章和文章的简要概述组成。本文使用常用的标准分割方法进行训练、验证和测试。具体将数据集分割为90,266/1,220/1,093个CNN文档和196,961/12,148/10,397个Daily Mail文档。在CNN的参考摘要中有52.90%的元组,在Daily Mail中有52.16%的元组。该数据集由于语料库大、文本长而被广泛应用于自动文本摘要任务,适用于抽取式模型和生成式模型。

2.2 实验设置

所有模型都在OpenNMT的PyTorch版本上实现,同时本文选择了“BERTbase”对模型进行微调,共有110 M的参数。词汇量为30 522,词嵌入维度为768。对于主题的数量,本文设置为1。当主题数量为1时,摘要生成的指导效果最好。本文的Transformer解码器有768个隐藏单元,所有前馈层的隐藏大小为2 048,模型在Tesla P100 GPU上训练200 000步,每5步进行梯度积累,模型检查点每隔5 000步在验证集上保存并评估。本文根据验证集上的评估损失来选择前3个检查点,并报告测试集上的平均结果。本文使用Adam优化器,并以0.002和0.1的学习率来训练编码器和解码器。此外还对编码器和解码器设置了两个Adam优化器,分别为 $\beta_1 = 0.9, \beta_2 = 0.999$,摘要的最大句子长度设置为512。为了正则化,本文使用dropout并将dropout率设置为0.1。

2.3 实验结果及分析

本文将提出的模型与下面几种模型进行了比较:

(1) TransformerABS:其编码器是有6层的Transformer,具有768个隐藏单元,前馈层的隐藏大

小为 2 048, 而其解码器与 BERTSUM 模型相同。本文将此模型作为基准。

(2) PTGEN-COV: 该模型通过指针直接从原始文本中复制单词, 并保留通过生成器生成新单词的能力^[5]。

(3) BERTSum: 该模型使用预训练语言模型进行摘要生成, 是一个经过微调的 BERT 变体^[6]。

(4) T-BERTSum: 该模型基于 BERT 强大的体系结构和额外的主题嵌入信息来指导上下文信息的获取^[7]。

本文对于实验结果的评估分为两种方式。为了确保公平的比较, 本文使用完全相同的实验设置来进行对比模型的结果呈现。

实验结果的评估分为自动评估和人工评估两种方式。

2.3.1 自动评估

本文使用 ROUGE (Recall - Oriented Understanding for Gisting Evaluation) 度量来自动评估生成摘要的质量。ROUGE 是一组用于评估自动摘要的度量标准, 其将自动生成的摘要与参考版本 (通常手动生成) 之间的重叠数量进行比较, 以计算相应的分数^[8]。其基本思想是将模型生成的摘要与参考摘要的 n 元组贡献统计量作为评判依据。本文使用 ROUGE-1 和 ROUGE-2 作为评估信息量的标准, ROUGE-L 作为评估流利性的标准。本文将每个 ROUGE 评分的精确度和召回率的调和平均值 (即 $F1$ 分) 作为评价标准。本文的模型和比较模型在 CNN/Daily Mail 数据集上的自动评价结果见表 1。

表 1 在 CNN/DailyMail 数据集上的 ROUGE 评分

Tab. 1 ROUGE scores on the CNN/DailyMail dataset

模型	ROUGE-1	ROUGE-2	ROUGE-L
TransformerABS	40.21	17.76	37.09
PTGEN-COV	39.53	17.28	36.38
BERTSum	42.13	19.6	39.18
T-BERTSum	43.06	19.76	39.43
Ours (LPE)	42.74	19.68	39.22
Ours (TE)	43.18	19.84	39.25
Ours (LPE + TE)	43.54	19.89	39.73

注: Ours (LPE) 是指模型只包含了学习位置嵌入模块, 没有加入主题嵌入模块; Ours (TE) 是指模型只包含了主题嵌入模块, 没有加入学习位置嵌入模块。

由表 1 可知基于 BERT 的预训练模型在很大程度上优于其他模型, 这说明使用预训练模型对于模

型的评估结果会有显著的提升。除此之外, 本文的模型评分总体上优于实验对比的其他模型, 这说明本文提出的模型对于生成摘要的质量有一定的提升。

同时, 本文对模型的学习位置嵌入和主题模型两个模块分别进行了消融实验来检验是否对模型具有优化效果。由表 1 可知, 添加了学习位置嵌入模块没有加入主题嵌入模块的效果好, 这主要是因为学习位置嵌入模块主要是对抽取式摘要的结果进行了优化, 相比于加入主题信息提升全局语义来说效果没有那么明显。但两者相结合之后, 最终结果都比两者单独使用效果要好。

2.3.2 人工评估

本文还通过人工判断来评估生成的摘要。从测试集中选取 100 篇文章, 对模型进行匿名实验, 并随机选取 3 名志愿者, 将实验模型生成的摘要进行匿名评分, 评分范围从 1~5, 分数越高, 说明模型能力越强。

评估包括以下内容:

(1) 突出性: 生成的摘要再现原始信息或观点的能力;

(2) 连贯性: 生成的摘要与文章主题是否一致, 即生成的内容是否混乱;

(3) 冗余性: 生成的摘要中冗余词多少。

对 CNN/Daily Mail 数据集进行人工评估的结果, 见表 2。与自动评价结果一致, 本文的模型取得了较好的成绩, 进一步证明本文的模型能够很好地对原始文档进行建模, 并准确地捕获关键信息。在突出性和连贯性方面, 本文模型都比其他模型分数高。在冗余度评价上, PTGEN-COV 与本文的模型获得了相同的分数。PTGEN-COV 引入了基于指针网络的复制机制, 避免了重叠词的产生。而本文的模型结合了主题嵌入和绝对句子位置嵌入, 更好的利用了上下文, 同时抽取-生成模型相结合共享信息, 减少冗余。

表 2 模型人工评价的得分

Tab. 2 The score of manual evaluation of model

模型	突出性	连贯性	冗余性
TransformerABS	2.40	2.87	2.27
PTGEN-COV	2.93	3.10	3.37
BERTSum	2.97	3.47	3.10
T-BERTSum	3.13	3.47	3.17
Ours (ExtAbs)	3.20	3.57	3.37

2.3.3 其他实验

本文对于不同主题数量对于模型结果的影响在 CNN/DM 数据集上进行了实验,实验结果如图 3 所示。可以看出,当主题数量为 1 时,摘要生成的效果最好;当主题数量大于 1 时,模型生成摘要效果会逐渐降低,说明过多的主题对语义的表达是多余的。

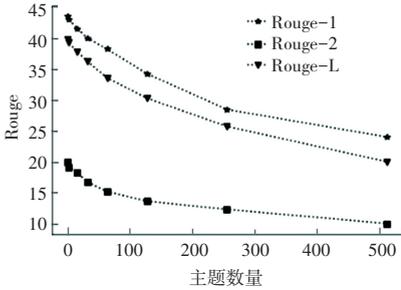


图 3 在 CNN/DailyMail 数据集上不同主题数量的 ROUGE 评分
Fig. 3 ROUGE scores for different number of topics on the CNN/DailyMail dataset

除此之外,本文还进行了超参数的相关实验,对不同编码器和解码器学习率组合方案下的困惑度 (Perplexity) 进行评估,实验结果见表 3,可见当组合方案为 $lr_E = 0.002, lr_D = 0.1$ 时,困惑度最小,因此选取此方案为模型参数。

表 3 在不同编码器和解码器的学习率组合下的模型困惑度

Tab. 3 Perplexity under different combinations of encoder and decoder learning rates

$lr_E lr_D$	1	0.1	0.01	0.001
0.02	48.37	8.45	9.62	17.25
0.002	35.41	6.69	9.13	15.21

3 结束语

本文提出了一种结合双主题嵌入和句子位置嵌

入的抽取-生成两阶段自动摘要生成模型。该模型通过引入双主题嵌入来捕获更准确的全局语义信息,并且通过在抽取阶段引入句子绝对位置嵌入将句子位置信息进行完全整合,获得更全面的摘要抽取辅助信息。同时两阶段框架的使用不仅降低了生成摘要的冗余,还利用文档的背景知识减少了摘要内容的缺失,进一步提升了模型的性能。实验结果表明,本文提出的模型在 CNN/Daily Mail 数据集上取得了较好的结果。

参考文献

- [1] RUSH A M, CHOPRA S, WESTON J. A neural attention model for abstractive sentence summarization [J]. arXiv preprint arXiv: 1509.00685, 2015.
- [2] CHOPRA S, AULI M, RUSH A M. Abstractive sentence summarization with attentive recurrent neural networks [C]// Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies. 2016; 93-98.
- [3] SEE A, LIU P J, MANNING C D. Get to the point: Summarization with pointer-generator networks [J]. arXiv preprint arXiv: 1704.04368, 2017.
- [4] PAULUS R, XIONG C, SOCHER R. A deep reinforced model for abstractive summarization [J]. arXiv preprint arXiv: 1705.04304, 2017.
- [5] NARAYAN S, COHEN S B, LAPATA M. Don't give me the details, just the summary! topic-aware convolutional neural networks for extreme summarization [J]. arXiv preprint arXiv: 1808.08745, 2018.
- [6] LIU Y, LAPATA M. Text summarization with pretrained encoders [J]. arXiv preprint arXiv: 1908.08345, 2019.
- [7] MA T, PAN Q, RONG H, et al. T-bertsum: Topic-aware text summarization based on bert [J]. IEEE Transactions on Computational Social Systems, 2021, 9(3): 879-890.
- [8] LIN C Y. Rouge: A package for automatic evaluation of summaries [C]//Text summarization branches out. 2004: 74-81.