



天津科技大学学报

Journal of Tianjin University of Science & Technology

ISSN 1672-6510, CN 12-1355/N

《天津科技大学学报》网络首发论文

题目： 基于句子平衡策略的动态遮盖文本摘要方法
作者： 雷罡，梁琨，黄越洋，张贤坤
DOI： 10.13364/j.issn.1672-6510.20230199
收稿日期： 2023-10-22
网络首发日期： 2024-07-12
引用格式： 雷罡，梁琨，黄越洋，张贤坤. 基于句子平衡策略的动态遮盖文本摘要方法[J/OL]. 天津科技大学学报. <https://doi.org/10.13364/j.issn.1672-6510.20230199>



网络首发：在编辑部工作流程中，稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定，且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式（包括网络呈现版式）排版后的稿件，可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定；学术研究成果具有创新性、科学性和先进性，符合编辑部对刊文的录用要求，不存在学术不端行为及其他侵权行为；稿件内容应基本符合国家有关书刊编辑、出版的技术标准，正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性，录用定稿一经发布，不得修改论文题目、作者、机构名称和学术内容，只可基于编辑规范进行少量文字的修改。

出版确认：纸质期刊编辑部通过与《中国学术期刊（光盘版）》电子杂志社有限公司签约，在《中国学术期刊（网络版）》出版传播平台上创办与纸质期刊内容一致的网络版，以单篇或整期出版形式，在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊（网络版）》是国家新闻出版广电总局批准的网络连续型出版物（ISSN 2096-4188，CN 11-6037/Z），所以签约期刊的网络版上网络首发论文视为正式出版。



DOI: 10.13364/j.issn.1672-6510.20230199

基于句子平衡策略的动态遮盖文本摘要方法

雷 罡, 梁 琨, 黄越洋, 张贤坤
(天津科技大学人工智能学院, 天津 300457)

摘要: 自动文本摘要技术是自然语言处理领域的焦点研究方向之一。针对现有特征提取方法捕捉语义能力较弱以及摘要结果中存在大量信息冗余的情况, 本文通过调整句子嵌入阶段生成词向量的方式和遮盖策略, 同时在句子选择阶段引入了冗余性、关键性联合评分机制, 分别在中文短文本和长文本数据集上进行抽取式摘要的实验。实验结果表明, 通过在句子嵌入和选择阶段进行一系列改进, 本文模型的性能明显优于 TextRank+BERT、BERTSUM 等其他基准模型。

关键词: 自然语言处理; 文本摘要; 抽取式摘要; BERT

中图分类号: TP391.1 **文献标志码:** A **文章编号:** 1672-6510 (2024)00-0000-00

Dynamic Masked Text Summarization Method Based on Sentence Balancing Strategy

LEI Gang, LIANG Kun, HUANG Yueyang, ZHANG Xiankun
(College of Artificial Intelligence, Tianjin University of Science & Technology, Tianjin 300457, China)

Abstract: Automatic text summarization technology is one of the focal research directions in the field of natural language processing. In response to the weak ability of existing feature extraction methods to capture semantics and the presence of a large amount of information redundancy in summary results, in our present study we adjusted the way word vectors were generated during the sentence embedding stage and the masking strategy. At the same time, a joint evaluation mechanism of redundancy and criticality was introduced in the sentence selection stage, and experiments on extracted summaries were conducted on Chinese short text and long text datasets, respectively. The experimental results showed that by making a series of improvements in the sentence embedding and selection stages, the performance of our model was significantly better than other benchmark models such as TextRank+BERT, BERTSUM, etc.

Key words: natural language processing; text summary; extract summary; BERT

在“信息过载”时代, 面对网络文本资源指数级的膨胀, 迅速且精确地抽取其中的关键信息成了一个迫切需要解决的问题。自动文本摘要技术通过计算机的强大处理能力, 能够从长篇幅的文本中筛选出核心内容, 生成简明扼要、流畅易读的摘要, 使用户能够快速掌握文本的要点。

自动摘要技术通常分为两大类: 抽取式摘要方法和生成式摘要方法。抽取式摘要方法从原文中选

择关键句子或短语构建摘要, 它依赖统计特征(如术语频率、句子位置、标题词和提示短语等)确定哪些句子最适合构建摘要。这种方法的优势在于易于理解和实施, 因为该方法直接提取原文的信息, 确保了摘要的语法正确性和事实准确性。与之相反, 生成式摘要方法模仿人类写作的方式, 使用自然语言生成技术改写原文句子来构建摘要。这种方法会引入原文中不存在的词汇和更复杂的句子结构, 因此

收稿日期: 2023-10-22; 修回日期: 2024-05-17

基金项目: 国家自然科学基金项目(62377036)

作者简介: 雷罡(1996—), 男, 安徽安庆人, 硕士研究生; 通信作者: 梁琨, 副教授, liangkun@tust.edu.cn

生成的摘要通常更具创意和新颖性,然而也可能出现偏离原文事实、引入冗余或重复等问题。

抽取式摘要方法的历史可追溯到 20 世纪 50 年代, Luhn^[1]首次提出自动文本摘要(ATS)的概念,随后经历了多次关键性的发展。在 1961 年,Edmundson 等^[2]通过整合统计特征(包括单词频率、句中标题词频率、提示短语以及句子在输入文档中的位置),对抽取式摘要方法进行拓展。这些早期工作为自动文本摘要研究打下了坚实的基础。随着时间的推移,研究者们持续不断地探索新方法和技术,以提高抽取式摘要方法的性能。2014 年, Padmalahari 等^[3]提出 text_pronouns 算法,该算法借助构建句子和词语级别的文本特征,实现了抽取式自动文本摘要。同时, Fattah^[4]提出一种多文档文本摘要方案,该方案结合了统计特征,如术语频率、标题相似性、句子位置和提示短语等,采用朴素贝叶斯、最大熵和支持向量机 (support vector machine, SVM)等技术,取得了有竞争力的结果。随着机器学习领域的不断进步,研究人员开始将机器学习算法应用于文本摘要任务。在 2017 年, Nallapati 等^[5]提出 SummaRuNNer 方法,采用序列标注方法抽取关键句子。相较于传统的统计方法, SummaRuNNer 更能够从语义角度控制生成文本摘要的质量,然而仍然存在一些问题。首先,在特征提取方面,该方法无法充分捕捉词向量在当前上下文中的具体语义;其次,在使用神经序列模型进行评分时, SummaRuNNer 在学习给定部分输出摘要的情况下如何以最大化相对增益预测句子,会导致信息过度冗余。针对特征提取,本研究采用动态遮盖模型捕获文本的语义特征,克服传统特征方法的词义歧义问题,并借助 BERT(bidirectional encoder representations from transformers)模型^[6]的双向 Transformer 结构更好地提取文本的语义特征。针对生成摘要中信息冗余的问题,本研究引入了一个额外的句子选择模型,以平衡构建摘要的关键性和冗余性。该模型整合了表面特征(如 n -gram 重叠比和语义匹配分数)用于衡量多样性,通过减少冗余提升了模型性能。

1 相关工作

1.1 句子嵌入

在抽取式文本摘要领域,近几年的模型一直依赖于基于 BERT 模型结构的方法。BERT 模型是一种自然语言处理(NLP)领域的深度学习模型,它于 2018 年由 Google 提出并开创了一种全新的预训练文本表

示方法。它是一种基于 Transformer 架构的模型,其主要创新在于预训练方式。传统的自然语言处理(natural language processing, NLP)模型通常使用从左到右或从右到左的单向语言模型进行预训练,而 BERT 模型是采用双向的语言模型,能够同时考虑上下文中的所有单词,从而更好地捕捉单词之间的语义关系。通过在大规模文本语料库上进行无监督的预训练,学习到了丰富的语言表示,然后可以在各种下游任务中进行微调,以适应特定任务的需求。

在进行抽取式摘要生成时,关键步骤涉及输出每个句子的词元表示。BERT 模型作为一种掩码语言模型,能够输出句子的词元向量,这些向量是构建抽取式摘要的基础。此外, BERT 模型还利用分段嵌入来区分每个独立的句子,通常通过在每个句子的开头放置[CLS],并在句子结尾放置[SEP],这样可以让模型理解句子之间的独立性。同时,为了区分在抽取式摘要中的句子, BERT 模型使用了两种标签,即 E_A 和 E_B 。BERT 模型中的 3 种 Embedding 如图 1



图 1 BERT 模型的 Embedding

Fig. 1 Embedding of BERT model

与传统的 BERT 模型有所不同,动态遮盖模型改变了输入编码方式,摒弃了传统的段向量的使用,采用<s>和</s>标记明确标识句子的起始和结束,从而简化了输入编码的过程。这种新的方式需要在每个句子前加入<s>标识符,并在每个句子的末尾添加</s>标识符,以区分不同句子的界限。此外,为了更好地区分同一文档中的不同句子,引入了区间向量,如 s_i 代表句子 I。对包含多个句子的序列,会分配相应的区间段向量,以提供给动态遮盖模型作为输入,如图 2 所示。

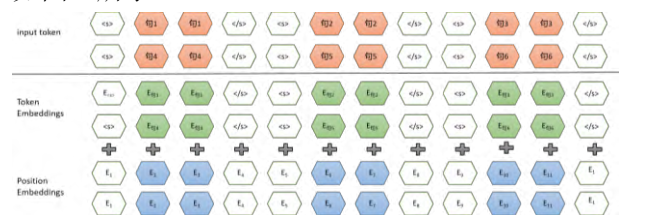


图 2 动态遮盖模型的 Embedding

Fig. 2 Embedding of dynamically masked models

掩码语言建模(masked language model, MLM)是 BERT 模型的预训练任务之一, 可以理解完形填空任务。在原始 BERT 模型中, 随机遮盖其中 15% 的词(在一个滑动窗口里取一个词, 10%~15% 的比例都可以), 用其上下文进行预测。动态遮盖模型中, 不再对 15% 的词元采取随机遮盖, 而是选择遮盖连续的词元以达到遮盖序列的效果。

综上所述, 动态遮盖模型能够更精确地捕捉文本中的语义信息, 提高文本摘要模型的性能和效率, 提升文本摘要生成的质量和准确性。

1.2 句子选择

在先前的研究中, 句子选择方法的性能改进显得较为有限。Zhong 等^[7]将句子选择任务转换为语义文本匹配问题, 对原始文档和候选摘要进行语义空间匹配, 找到最佳句子。Roit 等^[8]结合强化学习, 设计合适的奖励函数, 在每一步生成摘要的过程中, 根据当前状态和领域知识, 逐步优化选择句子的质量。McDonald^[9]引入了一种基于整数线性规划的方法, 将句子选择问题构建为在满足摘要长度限制的前提下的优化问题。另外, Lin 等^[10]提出一种方法, 旨在寻找具有子模块功能的句子的最佳子集。Ren 等^[11]采用两组手工特征, 分别捕获句子选择中的信息性和冗余性。与单独训练句子选择模型不同, Zhou 等^[12]提出一种共同学习句子评分和选择的序列生成模型。Liu 等^[13]在关键性评分后, 采用启发式方法选择不同的句子, 他提出的 BERTSUMEXT 模型是一种先进的三角块模型, 通过在每个时间步过滤掉与先前选择的句子存在三角重叠的句子减少冗余。三角块的基本思想是监测和干预生成模型生成文本中的三元组(3 个相邻的词)。在文本生成的过程中, 当模型生成一个新词时, 三角块会检查生成的词与前面两个词组成的三元组是否已经在之前的文本中出现过。如果检测到生成的三元组已经存在于之前生成的文本中, 三角块会采取相应措施(例如重新生成该词或者修改生成的词), 确保生成文本的多样性和低冗余性。三角块的引入有效地解决了文本生成中的一些问题, 如重复性和生成的文本不自然, 有助于改善生成文本的质量, 提高了自然性和可读性。尽管该方法在某些数据集上表现出色, 但是在其他数据集上可能表现欠佳, 这是因为它对所有文档都应用相同的规则。本文提出一种信息冗余性感知的提取摘要自适应学习过程, 该过程通过在句子评分时综合考虑冗余性和关键性, 或者通过引入额外的选择模型扩展 BERTSUMEXT 模型, 如图 3 所示。

2 模型架构

本文提出一种中文文本抽取式摘要方法。利用动态遮盖预训练模型获取更精确的词语在上下文中的具体语义表示, 包括词向量和句向量。将这些词向量输入 Transformer 编码器层, 该编码器层处理句子级别信息, 并生成相应的隐藏状态表示。这一隐藏状态表示有助于捕捉文档级别特征, 重点关注摘要任务。利用综合考虑关键性和冗余性的评分方法, 并应用 Sigmoid 函数为每个句子生成标签分数。按顺序选择其中适合作为摘要标签的前 m 个句子构建最终的摘要。模型的整体结构如图 4 所示。

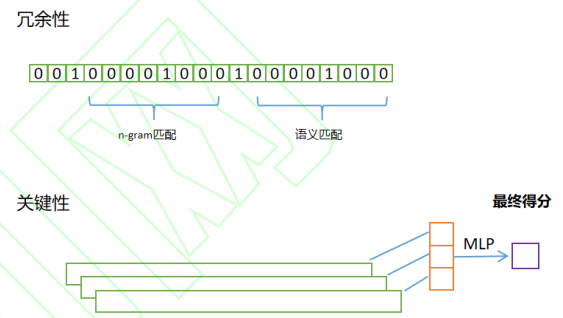


图 3 冗余性与关键性感知

Fig. 3 Perception of redundancy and criticality

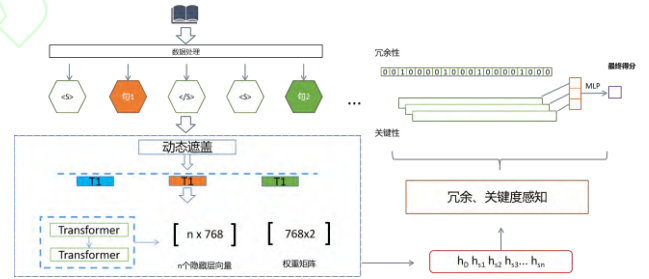


图 4 模型整体结构

Fig. 4 Overall structure of the model

采用句子级特征编码文档特征后, 得到的向量被传递至提取层, 这个提取层由双层 Transformer 构成。这一双层 Transformer 的关键任务是进一步增强文档的特征表示, 以便更好地捕获文档内部不同句子之间的长距离信息, 并提高句子之间的区分能力。

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

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

式中: 上标 l 表示堆叠层深度; $h^0 = \text{PosEmb}(T)$, h^0 表示初始层, T 为 BERT 模型输出的句子向量, PosEmb 为在 T 上添加位置嵌入(表示每个句子的位

置)的函数; LN 为层归一化运算; MHAtt 是多头注意力操作; FFN 为前馈网络; \tilde{h}^l 表示在第 l 层应用多头注意力(MHAtt)到上一层输出 h^{l-1} 后, 进行层归一化(LN)的结果; h^l 为第 l 层的最终输出。

通过多层 Transformer 编码器的训练, 得到文档 A 和每个句子 B_i 的文档级特征向量 h_A 和 h_{B_i} 。

在下游句子选择过程中, 本文采用两步评分和选择学习句子的过程构建摘要。在关键性排序步骤中, 模型专注于学习句子的关键性, 而在句子选择排序步骤中, 模型通过表面特征明确表示冗余, 并使用排名器根据句子的关键性和冗余度的联合程度决定句子的提升或降级。

(1)关键性

假设句子的关键性与之前选择的句子无关, 本文将关键性排序设计为一个单步过程, 而不是一个迭代过程。使用基于 h_A 和 h_{B_i} 之间的双线性匹配的评

分函数 $F_k(B_i)$ 得到句子的重要性评分。

$$F_k(B_i) = \frac{\exp h_A W_{ab} h_{B_i}}{\sum_{j=1}^L \exp h_A W_{ab} h_{B_j}} \quad (3)$$

式中: W_{ab} 表示一个在文档 A 和句子 B_i 之间进行得分计算时使用的参数矩阵; L 表示文档中句子的索引, 用来遍历文档中的每一个句子。

(2)冗余性

在选择步骤中, 对冗余进行明确表示, 让模型专注于学习如何平衡关键性和冗余性。在每个第 m 步提取 n -gram 匹配和语义匹配特征, 用于指示候选句子 B_i 在考虑前文已选择的句子 \hat{B}_{m-1} 时的冗余性。

n -gram 匹配特征 $f_{n\text{-gram}}$ 的计算公式为

$$f_{n\text{-gram}} = \frac{|n\text{-gram}(\hat{B}_{m-1}) \cap n\text{-gram}(B_i)|}{n\text{-gram}(B_i)} \quad (4)$$

式中: $n\text{-gram}(*)$ 是任意文本片段中 n 个连续单词的集合, $n=\{1,2,3\}$ 。

语义匹配特征 f_{sem} 的计算公式为

$$f_{\text{sem}} = \max_{\hat{B}_j \in \hat{B}_{m-1}} \cos(h_{B_i}, h_{\hat{B}_j}) \quad (5)$$

由于大多数余弦值都在接近 1 的小范围内, 因此对 f_{sem} 进行放大得到 \tilde{f}_{sem} 。将 n -gram 匹配特征和语义匹配特征连接得到整体冗余特征向量 $F_r(B_i)$ 。

$$F_r(B_i) = [f'_{1\text{-gram}}; f'_{2\text{-gram}}; f'_{3\text{-gram}}; f'_{\text{sem}}] \quad (6)$$

式中: f' 表示 f 二值化后的 one-hot 向量。

(3)句子选择的排名

在句子选择步骤中, 只需要学习如何根据冗余特征 $F_r(B_i)$ 和关键性分数 $F_k(B_i)$ 对句子进行评分。使用三维矩阵 W_F 在冗余特征和关键性分数之间进行双线性匹配, 获得一个输出匹配向量, 对其应用单层 MLP 后输出最终分数。

$$f(B_i) = W_f \tanh(F_k(B_i) W_F F_r(B_i)) \quad (7)$$

式中: $f(B_i)$ 表示句子的最终得分, W_f 表示得分函数的权重矩阵。

选择得分最高的前 3 个句子作为文本摘要。句子的数量是不固定的, 可以根据实际情况和数据集文本的长度选择合适的数量。

3 实验

3.1 实验数据

本文采用的数据集来自于哈尔滨工业大学人工智能实验室 2016 年提供的 LCSTS 数据集^[14], 该数据集中的数据分布见表 1。该数据集中短文本新闻来源于新浪微博, 包含了标题和正文两个部分。每个样本都经过手工评分, 用以度量标题和正文之间的相关性, 评分区间为 1~5 分, 分数越高则表示源文本与摘要之间的相关性越强。

在数据集预处理中, 需要执行一系列关键步骤, 确保文本数据的质量和一致性。这些关键步骤包括去除 emoji 表情符号和特殊符号、拼接连续的英文单词以提高文本的连贯性、将文本中的英文标点符号转换成对应的中文符号(例如将英文逗号转换为中文逗号), 并使用中文标点符号如“,”“。”“!”“?”等进行分句操作, 从而使文本更易于处理和

分析。此外, 为了满足模型的输入要求, 需要根据其最大输入长度限制(通常为 512 个词元)进行裁剪, 同时舍弃那些少于 3 个字符的句子。在中文文本处理中, 使用 jieba 中文分词工具进行中文分词, 以便更好地理解 and 处理中文文本数据。

表 1 LCSTS 数据集中的数据分布

Tab. 1 Data distribution in the LCSTS datasets

句子分类	句子数量	评分/分
Part I	2 400 591	无
	942	1
	1 039	2
	2 039	3
	3 128	4
Part II	3 538	5
	165	1
	216	2
	227	3
	301	4
Part III	197	5

3.2 实验设置

本文模型的实现基于 PyTorch 框架和 BERTSUM 模型^[13]。使用动态遮盖进行句子级编码, 根据 BERTSUM 模型的参数设置, 将文档级 Transformer 层的数量调整为 2, 所有层的 drop-out 率为 0.1。根据 Vaswani 等^[15]的初始值 2.0×10^{-3} 和 1.0×10^4 分别安排学习率和预热步骤参数。Training_steps 为 50000, 批处理大小 Batch_size 为 3000。在文本冗余、关键度感知方面, 遵循与 BERTSUM 相同的设置, 训练一个关键度排名器。在句子选择过程中, 保持关键度排名器中的所有参数固定, 确保每个句子的关键性分数维持不变。此外, 在参数调整过程中, 扫描 {10,20,30} 等不同数值, 并最终选择最佳的离散冗余特征 bin 大小为 {5,10,20,30}。

训练过程中, xent 值随训练步数变化的关系如图 5 所示。动态遮盖模型的收敛速度比传统 BERT 模型快, 整体 xent 曲线也明显低于 BERT 模型, 具有更优的收敛性质。

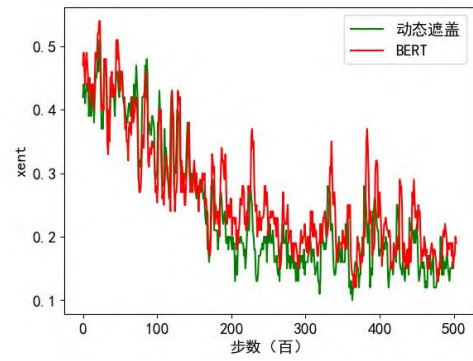


图 5 xent 值随训练步数变化的关系

Fig. 5 Relationship between xent value and the number of training steps

3.3 实验结果

模型生成的摘要见表 2。

表 2 模型生成的摘要

Tab. 2 Summary of model generation

类型	文本
原文一	昨天, 包括工农中建交五大行在内的多家银行, 不约而同地在官网发布公告称, 它们的房地产贷款政策没有变化。多家银行表示, 会支持居民购买首套住房。一名金融问题专家称, “目前房价不具备大涨大跌的基础, 特别是一二线城市狂跌的可能性小。”
参考摘要	银行集体发声: 房贷政策没变
MATCHSUM	多家银行表示, 会支持居民购买首套住房。
STARSUM	多家银行表示, 会支持居民购买首套住房。
本文模型	昨天, 包括工农中建交五大行在内的多家银行, 不约而同地在官网发布公告称, 它们的房地产贷款政策没有变化。
原文二	高考阅卷此时正在各地进行, 其他科目基本不会差到哪去, 可语文, 尤其是作文阅卷, 老师手一松或一紧, 往往结果就大不一样。若不能从专业角度去看待, 所谓的“速度影响公平”就只是一种虚妄的矫情。忽略了事实本身的监督, 也无助于真正的公平认同。
参考摘要	看待高考作文判阅需要专业角度
MATCHSUM	高考阅卷此时正在各地进行, 其他科目基本不会差到哪去, 可语文, 尤其是作文阅卷, 老师手一松或一紧, 往往结果就大不一样。
STARSUM	忽略了事实本身的监督, 也无助于真正的公平认同。
本文模型	若不能从专业角度去看待, 所谓的“速度影响公平”就只是一种虚妄的矫情。

为验证本文模型处理文本摘要任务的优越性,

选取 6 种基准模型在 LCSTS 数据集上进行对比实验。

MASS 模型^[16]: 同时利用编码器和解码器进行长序列遮盖, 提高了模型理解语义的能力。

MMR 模型^[17]: 通过计算文档中的句子与原始文档的相似度给句子打分并抽取得分最高的句子作为文档摘要。

TextRank+BERT 模型^[18]: 利用 BERT 模型将原始文档中的句子进行向量化表示, 通过算法迭代选取得分最高的句子作为文档摘要。

BERTSUM 模型: 利用 BERT 模型作为编码器, 线性分类层作为解码器, 以 ROUGE 分数为基础标注句子是否应该被抽取, 通过计算句子被选为摘要的概率, 使用 trigram blocking 方法优化抽取结果。

Matchsum 模型^[7]: 利用预训练语言模型进行文本编码, 将摘要任务转化为了文本匹配任务。通过比较文档上下文表示与真实摘要及候选摘要的上下文表示的相似度, 模型能够高效地计算匹配得分。最终, 只需选择匹配得分最高的候选结果作为摘要答案, 实现了简便而有效的摘要生成策略, 使摘要生成变得更加精准。

StarSum 模型^[19]: 一个基于星形架构的抽取式摘要模型, 创新性地运用了 BERTSUM 生成句子表示和位置嵌入相加的策略, 借助星形 Transformer 进行高效文档编码。该模型通过引入 Sigmoid 函数对最终层的句子输出进行分类, 显著提高了计算效率, 并且能够有效处理文本中的长期依赖关系。

选择摘要生成的基准评价指标 ROUGE^[20]评分评估本文方法的表现。ROUGE 分数用于计算生成的摘要和参考之间的重叠程度。在本文中选取 ROUGE-1 (R-1)、ROUGE-2 (R-2)和 ROUGE-L(R-L)的 F1 分数作为评估指标。LCSTS 数据集的实验结果见表 3。

表 3 LCSTS 数据集的实验结果

Tab. 3 Experimental results of the LCSTS datasets

模型	F ₁ 分数/%		
	ROUGE-1	ROUGE-2	ROUGE-L
MMR	21.71	11.77	20.41
MASS	22.31	9.54	19.29
TextRank+BERT	23.63	12.63	21.98
BERTSUM	29.41	15.57	25.21
Matchsum	30.78	15.54	27.94
StarSum	31.26	15.61	29.03
本文模型	33.99	15.64	31.24
最大提升比	+12.28	+6.10	+11.95

LCSTS 是典型的短文本摘要数据集, 不能充分说明模型的性能和泛化能力, 因此本文引入了中文长文本数据集 CSL。CSL 的数据来自计算机领域的论文摘要和标题, 其正文部分平均字数为 200, 几乎是 LCSTS 的 2 倍。CSL 数据集的实验结果见表 4。

表 4 CSL 数据集的实验结果

Tab. 4 Experimental results of the CSL datasets

模型	F ₁ 分数/%		
	ROUGE-1	ROUGE-2	ROUGE-L
MMR	23.69	13.24	22.67
MASS	24.54	11.56	21.45
TextRank+BERT	25.34	15.38	23.79
BERTSUM	30.03	17.64	26.63
Matchsum	31.19	17.79	28.92
StarSum	33.43	18.54	29.69
本文模型	35.04	19.23	31.02
最大提升比	+11.35	+7.67	+9.57

由表 3、表 4 可知, 本文模型在所有对比实验中获得了最好的性能。MMR 和 TextRank+BERT 两种模型对初始权重的定义较为单一, 仅仅考虑了句子相似度, 且基于词汇重合度的句子相似度计算方法对文本的语义不能进行充分挖掘, 造成在 ROUGE 指标上的巨大差异, 说明以句子信息关键性与冗余性联合评分的可行性。MASS 模型使用自注意力机制处理输入序列中的每个位置, 但它可能无法充分捕捉到文档中的全局信息。这说明全局信息对确定文档的重要内容和主题至关重要, 而本文的模型使用双层 Transformer 编码文档级特征弥补了这个缺陷。最后与 BERTSUM、Matchsum、StarSum 模型依次相比 ROUGE 指标均有一定提升, 说明通过动态遮盖模型进行句子嵌入能够取得更好的抽取效果。

3.4 消融实验

为了研究应用在句子嵌入和句子选择过程中的改进对模型性能的具体影响, 进行消融实验, 结果见表 5。

表 5 消融实验结果

Tab. 5 Results of ablation experiment

处理类型	方法	F ₁ 分数/%		
		ROUGE-1	ROUGE-2	ROUGE-L
句子	动态遮盖	24.54	13.24	22.67
嵌入	BERT-base-chinese	23.69	11.56	21.45
句子	三角重叠	25.34	15.38	23.79
选择	冗余、关键性感知	30.03	17.64	26.63

从复杂度来看,动态遮盖版本的参数量比 BERT-base-chinese 版本提高了约 3 倍,因此可初步预测动态遮盖版本应该比 BERT-base-chinese 版本更强大。实验比较了以上两个版本模型在 LCSTS 短文摘要任务中的效果,结果表明动态遮盖版本确实优于 BERT-base-chinese 版本。

默认模型采取的句子选择策略是三角重叠 (TRIBLK),考虑到模型在选择过程在陷入关键性最优解而不是综合冗余全局最优解,实验分别测试了模型在三角重叠和冗余、关键性感知下的性能。由表 5 可知,冗余、关键性感知优于三角重叠,但差距很小,提升程度有限。

4 结 语

本文提出一种基于句子平衡策略的动态遮盖文本摘要方法,使用动态遮盖策略和冗余、关键性联合评分有效地指导模型生成更具可读性和更高 ROUGE 得分的摘要,减少了模型的累计误差。在 LCSTS 和 CSL 数据集上的实验结果和摘要实例都证明了本文模型的优势。但是本文模型仍存在剪枝高、训练时间长等局限,在今后的研究中,将进一步对模型进行合理剪枝以减少模型的参数量和计算量。

参考文献:

- [1] LUHN H P. The automatic creation of literature abstracts[J]. IBM Journal of research and development, 1958, 2 (2): 159-165.
- [2] EDMUNDSON H P, WYLLYS R E. Automatic abstracting and indexing:survey and recommendations[J]. Communications of the ACM, 1961, 4 (5): 226-234.
- [3] PADMALAHARI E, KUMAR D V N S, PRASAD S. Automatic text summarization with statistical and linguistic features using successive thresholds[C]// IEEE. Proceedings of 2014 IEEE International Conference on Advanced Communications, Control and Computing Technologies. Ramanathapuram:IEEE, 2014: 1519-1524.
- [4] FATTAH M A. A hybrid machine learning model for multi-document summarization[J]. Applied intelligence, 2014, 40: 592-600.
- [5] NALLAPATI R, ZHAI F F, ZHOU B W. Summarunner: a recurrent neural network based sequence model for extractive summarization of documents[C]// AAAI. Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence. Palo Alto:AAAI Press, 2017.
- [6] DEVLIN J, CHANG M W, LEE K, et al. BERT: pre-training of deep bidirectional transformers for language understanding [EB/OL]. (2019-05-24)[2023-10-10]. <https://doi.org/10.48550/arXiv.1810.04805>.
- [7] ZHONG M, LIU P F, CHEN Y R, et al. Extractive summarization as text matching [EB/OL]. (2020-04-19)[2023-10-10]. <https://doi.org/10.48550/arXiv.2004.08795>.
- [8] ROIT P, FERRET J, SHANI L, et al. Factually consistent summarization via reinforcement learning with textual entailment feedback [EB/OL]. (2023-05-31)[2023-10-10]. <https://doi.org/10.48550/arXiv.2306.00186>.
- [9] MCDONALD R. A study of global inference algorithms in multi-document summarization[C]// AMATI G, CARPINETO C, ROMANO G. Proceedings of 2007 European Conference on Information Retrieval. Berlin:Springer, 2007: 557-564.
- [10] LIN H, BILMES J. A class of submodular functions for document summarization[C]// Association for Computational Linguistics. Proceedings of the 49th annual meeting of the association for computational linguistics: human language technologies. Portland:Association for Computational Linguistics, 2011: 510-520.
- [11] REN P J, WEI F R, CHEN Z M, et al. A redundancy-aware sentence regression framework for extractive summarization[C]//MATSUMOTO Y, PRASAD R. Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics. Osaka:The COLING 2016 Organizing Committee, 2016: 33-43.
- [12] ZHOU Q Y, YANG N, WEI F R, et al. Neural document summarization by jointly learning to score and select sentences [EB/OL]. (2018-07-06)[2023-10-10]. <https://doi.org/10.48550/arXiv.1807.02305>.
- [13] LIU Y, LAPATA M. Text summarization with pretrained encoders[EB/OL]. (2019-09-05)[2023-10-10]. <https://doi.org/10.48550/arXiv.1908.08345>.
- [14] HU B T, CHEN Q C, ZHU F Z. LCSTS: a large scale chinese short text summarization dataset[EB/OL]. (2016-02-19)[2023-10-10]. <https://doi.org/10.48550/arXiv.1506.05865>.
- [15] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. Advances in neural information processing systems, 2017, 30:6000-6010.
- [16] SONG K T, TAN X, QIN T, et al. MASS: masked sequence

- to sequence pre-training for language generation[EB/OL]. (2019-06-21)[2023-10-10]. <https://doi.org/10.48550/arXiv.1905.02450>.
- [17] CARBONELL J, GOLDSTEIN J. The use of MMR, diversity-based reranking for reordering documents and producing summaries[C]//ACM. Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. New York:ACM, 1998: 335-336.
- [18] 肖莉才.文本摘要生成技术研究与应用[D].成都:电子科技大学,2021.
- [19] SHI K, CAI X, YANG L, et al. StarSum: a star architecture based model for extractive summarization[J]. IEEE/ACM Transactions on audio, speech, and language processing, 2022, 30: 3020-3031.
- [20] ROUGE L C Y. A package for automatic evaluation of summaries[C]//ACL. Proceedings of Workshop on Text Summarization of ACL.Barcelona: ACL, 2004: 74-81.

