

融合内容引导与多尺度注意力的摘要生成模型^①



岳 帅¹, 王 业²

¹(新疆农业大学 计算机与信息工程学院, 乌鲁木齐 830052)

²(新疆农业大学 网络与信息技术中心, 乌鲁木齐 830052)

通信作者: 王 业, E-mail: leafwy@163.com

摘 要: 长文本中的信息压缩和语义连贯性一直是摘要生成模型的难点. 为此本文提出了一种融合内容引导与多尺度注意力的摘要生成模型. 该模型通过双分支结构实现对多粒度语义的联合建模, 并利用内容引导机制聚焦于摘要相关的键信息区域. 模型在传统 BERT-Transformer 架构基础上引入双分支结构增强语义表达能力, 并通过 MSAA-SAM 融合机制设计, 进一步实现跨分支信息对齐与表达统一. 同时, 本文对指针生成网络进行了改进, 结合全局句向量引导机制提升生成控制能力, 从而增强对长文本中键信息的提取与冗余内容的抑制. 在 NLPC 2017 数据集和 LCSTS 数据集上的实验结果表明, 该模型在生成式摘要任务上均优于主流基线模型, 验证了其在语义建模、生成质量与控制能力方面的综合优势.

关键词: 深度学习; 文本摘要; 多尺度注意力; 内容引导机制; 指针生成网络; 语义建模

引用格式: 岳帅, 王业. 融合内容引导与多尺度注意力的摘要生成模型. 计算机系统应用, 2026, 35(1): 276-287. <http://www.c-s-a.org.cn/1003-3254/10044.html>

Summary Generation Model Integrating Content-guided and Multi-scale Attention

YUE Shuai¹, WANG Ye²

¹(College of Computer and Information Engineering, Xinjiang Agricultural University, Urumqi 830052, China)

²(Network and Information Technology Center, Xinjiang Agricultural University, Urumqi 830052, China)

Abstract: Information compression and semantic coherence in long texts are persistent challenges in summary generation models. To address this issue, this study proposes a summary generation model integrating content-guided and multi-scale attention. The model adopts a dual-branch architecture to jointly model multi-granularity semantics and utilizes a content-guided mechanism to focus on key information relevant to the summary. Based on the conventional BERT-Transformer framework, a dual-branch structure is introduced to enhance semantic representation, and a cross-branch fusion mechanism (MSAA-SAM) is designed to achieve semantic alignment and unified representation. In addition, the pointer-generator network is improved by incorporating a global sentence vector guidance mechanism to enhance generation control, thereby improving key information extraction and reducing redundancy in long-text summarization. Experimental results on the NLPC 2017 and LCSTS datasets demonstrate that the proposed model outperforms mainstream baseline models in generative summarization tasks, verifying its comprehensive advantages in semantic modeling, generation quality, and control capability.

Key words: deep learning; text summary; multi-scale attention; content-guided mechanism; pointer-generator network; semantic modeling

① 收稿时间: 2025-05-26; 修改时间: 2025-06-24; 采用时间: 2025-08-13; csa 在线出版时间: 2025-10-29

CNKI 网络首发时间: 2025-10-30

根据《中国互联网络发展状况统计报告》^[1]第55次发布的数据,中国网民已达11.08亿,互联网普及率为78.6%。随着互联网的快速发展,网络中的文本数据呈指数增长,进一步加剧了信息检索任务的复杂性。文本摘要技术旨在将长文档浓缩为包含关键信息的简短摘要,从而缩短用户检索时间并减少冗余信息。文本摘要方法通常分为抽取式^[2]、生成式^[3]及混合式这3类。抽取式方法通过从原文中选择关键词或关键词来生成摘要,具有语法结构完整、准确率高的优点;生成式摘要则依赖模型对文本深层语义的理解,通过语言生成能力生成更加自然、连贯的摘要;混合式方法试图结合两者优势,但如何高效整合抽取与生成内容仍是研究难点。尽管深度学习中的Seq2Seq模型在多项任务中展现出良好性能,并在摘要任务中取得一定进展,但仍面临诸如特征提取能力不足、难以处理长文本语义连贯性差及词表外(out-of-vocabulary, OOV)词处理能力弱等问题。

针对以上提到的语义理解不足、上下文建模能力弱以及生成控制能力差等问题,本文提出了一系列解决方案。

(1) 设计融合内容引导机制与多尺度注意力建模的双分支结构,实现对长文本中多粒度语义信息的全面建模,提升特征表达能力。

(2) 引入MSAA-SAM模块进行跨分支融合,对左右分支编码信息进行句向量级对齐与统一,缓解信息割裂与语义不一致问题。

(3) 改进PGN结构并结合全局句向量调控机制,增强生成阶段对关键信息的关注与对冗余信息的抑制,从而提升生成摘要的可控性和准确性。

1 相关理论

1.1 BERT-Transformer 架构

生成式文本摘要任务需要模型具备对长文本深层语义的理解能力与关键信息提取能力。近年来,基于预训练语言模型的编码器结构和高效注意力机制的引入,在摘要生成领域展现出显著优势。BERT模型^[4]基于Transformer编码器^[5]结构构建,通过双向上下文语境建模,捕捉更全面的语义信息。作为一种深层预训练语言模型,BERT采用masked language model与next sentence prediction进行训练,可在下游任务中提供鲁棒的语义表示。在摘要生成任务中,将BERT作为输入

编码器,结合Transformer的自注意力机制,有助于提升输入文本的表示能力与语义表达完整性。然而,BERT-Transformer架构在生成式摘要任务中仍存在以下3方面局限性。

(1) 缺乏目标导向建模能力。BERT本身并未针对摘要任务进行优化,难以自动聚焦于与摘要高度相关的信息区域。

(2) 全局语义感知能力不足。传统Transformer编码器虽然具备自注意力机制,但在处理超长文本时往往面临语义扩散与注意稀疏的问题,导致信息整合不充分。

(3) 表示冗余与泛化能力有限。BERT输出的编码特征较为密集,未能显式建模多粒度语义,影响后续解码器对关键信息的聚焦与生成质量。

因此,在BERT-Transformer架构的基础上,亟需引入任务特定的语义建模机制和跨粒度的信息压缩模块,以进一步提升其在摘要生成任务中的效果。

1.2 内容引导机制

内容引导机制^[6]在计算机视觉领域通过通道注意力和空间注意力机制,帮助模型聚焦于图像中的关键信息区域。具体而言,通道注意力通过加强对特定语义特征通道的关注,提升了模型对重要特征的感知能力;而空间注意力则通过定位图像中的关键区域,进一步强化了信息提取的精准度。

在自然语言处理中,文本的每个词或句子包含多重语义维度,如词义、上下文和句法结构等。通道注意力机制能够有效引导模型聚焦于这些不同的语义特征,尤其在长文本摘要任务中,能够突出关键信息区域并有效抑制冗余信息的生成。而空间注意力机制则有助于识别摘要中最具代表性的句子或段落。因此,内容引导机制在处理长文本摘要任务时,具有显著的信息提取与内容聚焦的潜力。

1.3 多尺度注意力机制

多尺度建模思想源于计算机视觉领域^[7],在自然语言处理中则可视作对不同语义粒度(如短语、句子、段落)建模能力的补充。通过多尺度卷积核(如 3×3 、 5×5 、 7×7),MSAA机制能够并行提取文本中不同层次的语义信息,从而增强模型对文本多层次语义结构的感知能力。空间注意力则能帮助模型在文本序列中识别出最具有摘要价值的部分,进一步优化信息提取效果。这一机制特别适用于处理长文本,因为它有助于捕捉文本中的关键信息并减少冗余信息。因此,MSAA机

制在长文本摘要任务中具有提升生成摘要的语义完整性和逻辑连贯性的潜力。

2 模型设计

2.1 总体架构

为了解决处理长文本摘要时仍面临语义扩散与冗余信息的问题, 本文提出了一种融合内容引导与多尺度注意力机制的摘要生成模型, 其模型框架如图 1 所示. 该模型的主要思想是通过双分支编码结构分别捕捉文本中的全局依赖关系与多尺度语义特征, 并在融

合机制中对齐信息分布, 从而提升摘要的完整性与连贯性.

具体来说, 本文通过引入 CGA 和 MSAA 模块, 以增强模型的语义建模能力, 帮助其更好地聚焦摘要任务中的关键信息, 减少冗余内容, 并提升摘要的语义连贯性和信息密度. 特别是, CGA 模块中的通道注意力机制通过引导模型关注文本中具有高语义密度的区域, 有效增强了模型对关键信息的聚焦能力. 该机制能够适应文本数据中的多语义维度 (如词义、上下文等), 特别是在长文本摘要任务中表现得尤为有效.

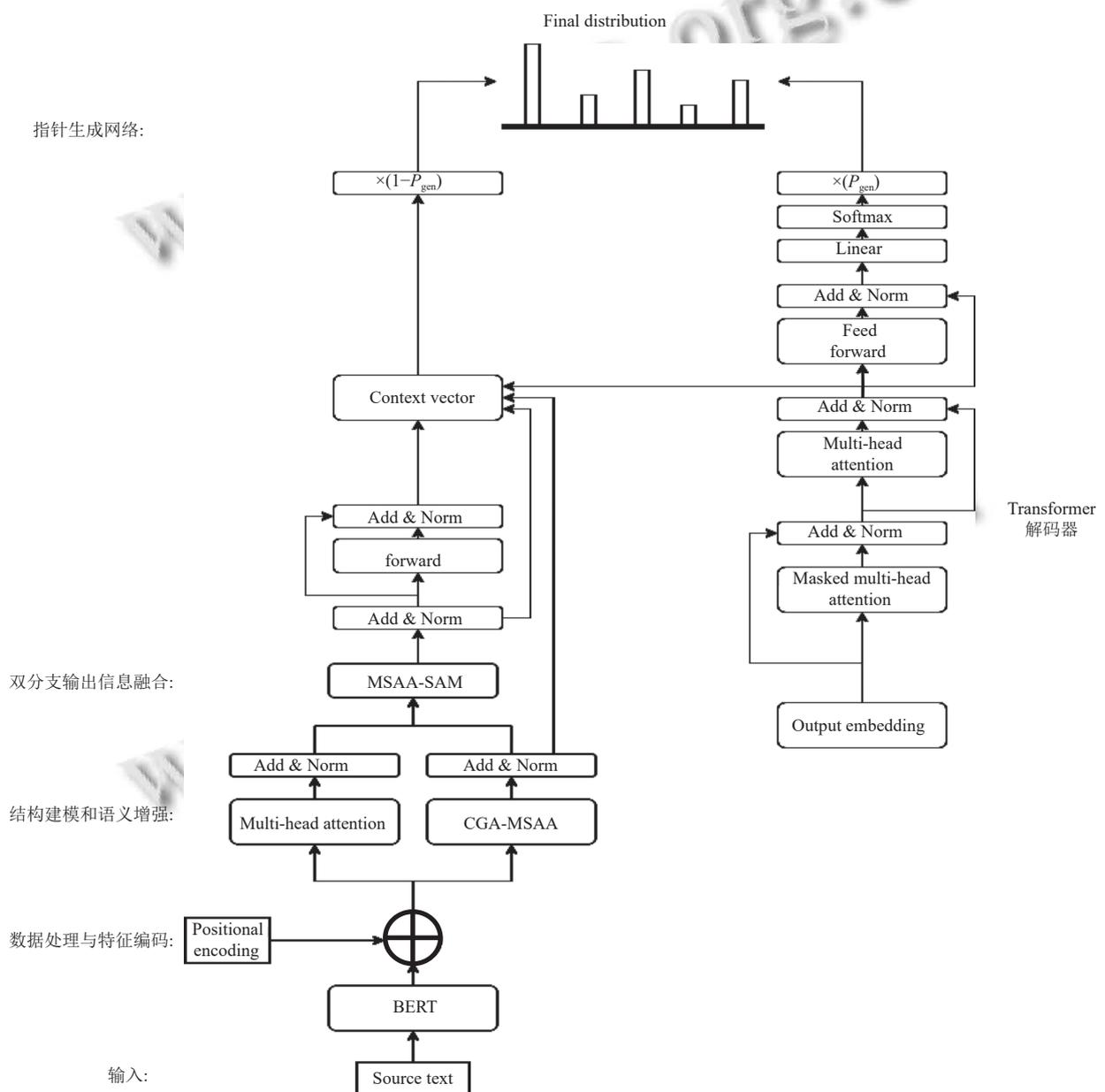


图 1 模型总体架构

在输入层中,模型首先通过 BERT 对原始源文本进行特征编码,随后结合位置编码,将编码表示输入至改进的 Transformer 编码器中以增强语义建模能力.在双分支输出阶段,模型设计了一个基于多尺度注意力聚合与句向量对齐机制 (MSAA-SAM) 的融合模块,该模块通过残差结构、注意力加权与线性变换实现特征融合,生成统一上下文表示.

在解码器部分,模型采用标准 Transformer 解码结构,在输出层中,模型引入了改进的指针生成网络 (PGN),通过融合右分支多尺度语义表示与全局句向量引导机制,实现了解码阶段在词表生成与源文复制之间的动态平衡.同时,模型结合多任务学习策略,在优化词汇选择的同时,引入句向量控制与句级覆盖机制,共同增强摘要生成的语义完整性与句法连贯性.

通过上述结构设计,本文模型在保持原始语义完整性的基础上,实现了多粒度语义建模、任务导向注意引导与全局句向量对齐的有机融合.使得模型在长文本摘要场景下的语义理解能力、信息压缩效率与摘要生成质量,为构建高性能、可控性的自动文本摘要系统提供了更加稳健的结构支撑.

2.2 改进的 Transformer 编码器

2.2.1 基于结构建模与语义增强的双分支编码器设计

为了提升模型对长文本中全局结构关系建模能力和多尺度语义特征提取能力,本文设计了一种基于结构建模与语义增强互补机制的双分支编码器结构.这种双分支结构既能充分捕捉句子内部的长距离依赖与句法结构特征,又能融合不同尺度下的语义表示与局部关键信息,增强模型对摘要内容的理解与压缩能力.

左侧分支侧重于全局依赖建模,其结构与标准 Transformer 编码器保持一致,主要包括:一个多头自注意力机制,使用 8 个注意力头,每个注意力头的键/值映射维度为 16 维,输出维度为 128 维;一个前馈全连接网络,其中间维度为 512,输出维度映射回原始的 128 维.

右侧分支则侧重于语义增强与多尺度感知建模,其结构由两个核心模块组成:CGA 模块和 MSAA 模块.其中 CGA 模块通过通道注意力与空间注意力机制,引导模型聚焦于摘要任务相关的重要区域与特征通道,通过加性融合、通道交叉与卷积操作生成内容引导下的特征注意力图.MSAA 模块通过多尺度卷积核 (3×3、5×5、7×7) 融合不同感受野的上下文语义,结合

空间注意力与通道注意力机制提取语义粒度多样的特征表达.

最后,两个模块的输出在通道维度上进行拼接,随后通过 1×1 卷积进行映射融合,形成最终的右分支编码输出表示.

2.2.2 左分支多头注意力机制

首先,该分支的输入为 BERT 输出的带有位置信息的上下文特征表示 X ,在此基础上,模型引用 h 头多头自注意力机制,用于在不同语义子空间中并行捕捉词与词之间的关联关系,提升表示表达能力,对于查询 Q 、键 K 和值 V ,第 h 个注意力头的计算过程为:

$$head_h = Attention(QW_h^q, KW_h^k, VW_h^v) \quad (1)$$

$$Attention(Q, K, V) = Softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (2)$$

其中, QW_h^q 、 KW_h^k 和 VW_h^v 分别为查询、键、值的线性投影矩阵, d_k 是键向量的维度, $\sqrt{d_k}$ 用于缩放点积,防止梯度消失.

最终,多头注意力的输出由所有子头拼接后经线性变换得到:

$$MultiHead(Q, K, V) = Concat(head_1, \dots, head_h)W^o \quad (3)$$

注意力输出后接前馈神经网络,用于非线性特征变换,其形式为两层全连接结构.

$$FFN(x) = \max(0, xW_1 + b_1)W_2 + b_2 \quad (4)$$

为提升训练稳定性并保留原始信息流,在每个子层(多头注意力、前馈网络)后均引入残差连接与层归一化.

$$X = LayerNorm(x + MultiHead(x)) \quad (5)$$

$$\tilde{X} = LayerNorm(x + FFN(x)) \quad (6)$$

其中, $LayerNorm$ 表示层归一化操作, $Dropout$ 率统一设置为 0.2.

2.2.3 CGA-MSAA 右分支内容引导与多尺度语义建模模块

在长文本摘要任务中,CGA 模块通过引导注意力机制帮助模型聚焦于文本中最相关的语义部分,减少冗余信息.同时,MSAA 模块通过多尺度卷积增强了对文本不同语义层次(如短语、句子、段落)的感知能力,使得模型能够准确提取关键信息并压缩冗余内容,从而显著提升摘要生成质量.具体来说,CGA 模块中的通道注意力机制能够适应 NLP 任务中词汇和句法结构的多重语义特征,通过为不同语义维度(如上下文

信息、语法结构等) 分配不同的注意力权重, 使得模型能够更好地识别和提取文本中的关键信息。

为此本文在编码器右分支中设计了一个融合内容引导注意力机制与多尺度语义注意机制的复合模块。

该模块由两个子结构组成, 如图 2 所示. 左侧为 CGA 模块, 右侧为 MSAA 模块, 二者分别作用于输入上下文表示 X , 最终在特征维度上融合, 形成语义增强后的输出表示 $Y_{CGA-MSAA}$.

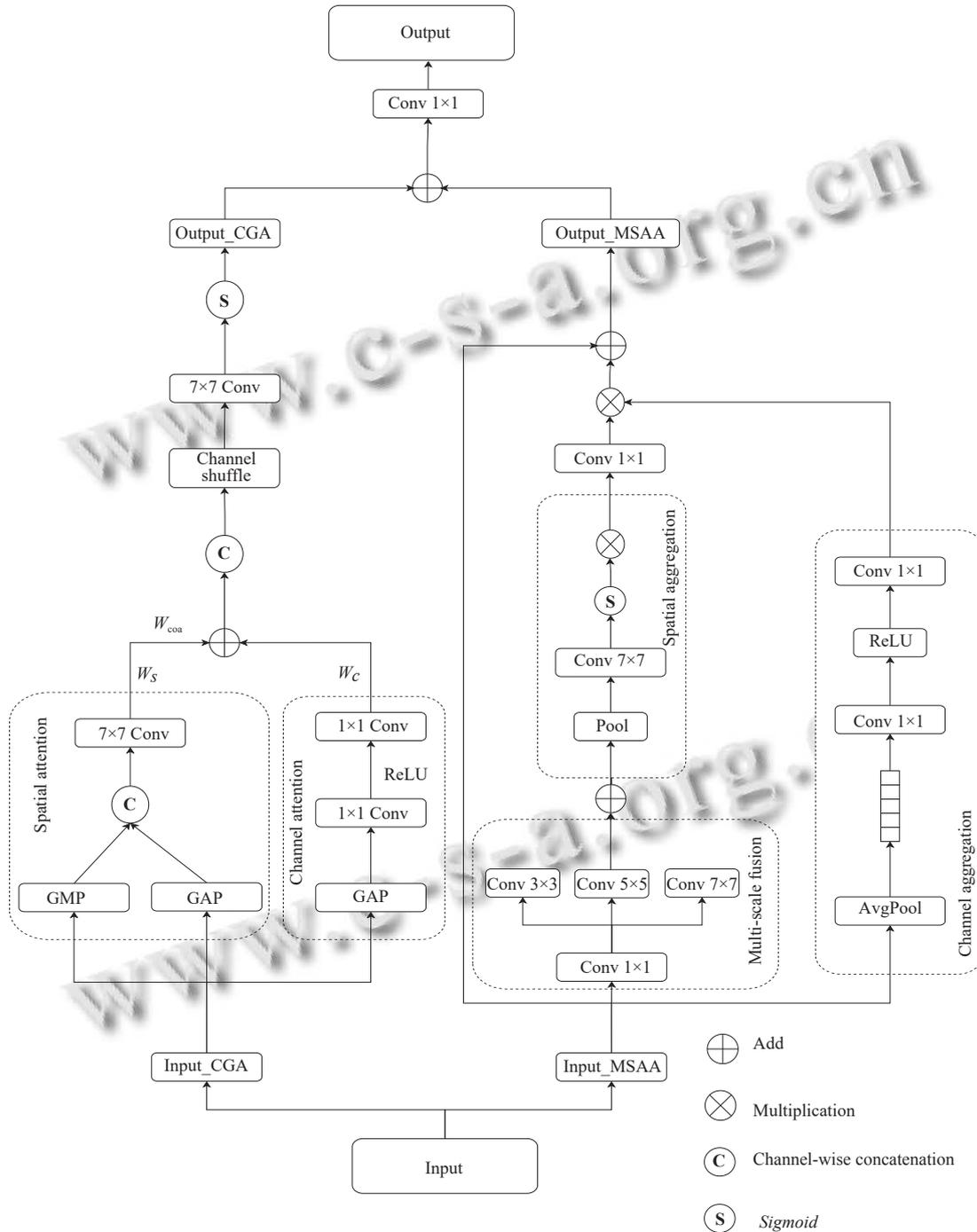


图 2 CGA-MSAA 结构图

具体实现过程如下。

CGA 模块的目标是生成特定于通道的空间重要

性特征, 首先, 对输入表示 X 通过在空间方向上执行全局平均池化, 然后通过一个两层的全连接建模通道上

的相关性,得到通道注意力权重.

$$W_c = \text{Conv}(\text{AvgPool}(x)) \quad (7)$$

然后在通道方向上分别执行全局平均池化和全局最大池化,然后拼接,并通过卷积降维,得到空间注意力权重.

$$X_A = \text{AvgPool}(x) \quad (8)$$

$$X_M = \text{MaxPool}(x) \quad (9)$$

$$W_s = \text{Conv}(\text{Concat}(X_A, X_M)) \quad (10)$$

接着,将通道注意力权重与空间注意力权重相加,得到粗糙的空间重要性特征.

$$W_{\text{coa}} = W_c + W_s \quad (11)$$

为了得到最终精准的空间重要性特征,我们将 W_{coa} 与 X 输入进行拼接并通道重排后,并使用 7×7 卷积与 Sigmoid 激活生成最终注意力权重.

$$\tilde{X} = \text{ChannelShuffle}(\text{Concat}(X, W_{\text{coa}})) \quad (12)$$

$$W = \text{Sigmoid}(\text{Conv}_{7 \times 7}(\tilde{X})) \quad (13)$$

CGA 为每个通道分配唯一的 SIM,引导模型关注每个通道对应的重要区域,因此,可以强调特征中的更多有用信息.最后,通过权重 W 对特征 X_1 和特征 X_2 进行加权求和:

$$Y_{\text{CGA}} = W \odot X_1 + (1 - W) \odot X_2 + X \quad (14)$$

其中, \odot 表示逐元素乘法, Y_{CGA} 为该模块添加残差连接之后的输出表示.

在 CGA 模块处理输入表示 X 的同时,多尺度语义注意模块 (MSAA) 同时处理输入表示 X .多尺度语义注意模块通过不同尺度的感受野建模句级、段级等多层次语义结构,增强模型对复杂上下文关系的建模能力,具体实现过程如下.

首先,MSAA 模块以编码张量 X 为输入,先通过一个 1×1 卷积层对通道维度进行压缩变换,得到中间表示:

$$\tilde{x} = \text{Conv}_{1 \times 1}(X) \quad (15)$$

然后,为了捕捉不同尺度的上下文语义信息,分别使用 3×3 、 5×5 和 7×7 的卷积核对 \tilde{x} 进行并行处理,得到 3 组不同感受野:

$$X_1 = \text{Conv}_{3 \times 3}(\tilde{x}) \quad (16)$$

$$X_2 = \text{Conv}_{5 \times 5}(\tilde{x}) \quad (17)$$

$$X_3 = \text{Conv}_{7 \times 7}(\tilde{x}) \quad (18)$$

$$U = X_1 + X_2 + X_3 \quad (19)$$

接着,MSAA 模块分别计算空间注意力权重和通道注意力权重.空间注意力分支先对融合特征 U 进行平均池化,通过卷积与 Sigmoid 激活获得位置敏感的注意力分布:

$$W_s = \text{Sigmoid}(\text{Conv}(\text{Pool}(U))) \quad (20)$$

$$U_s = \text{Conv}(W_s \odot U) \quad (21)$$

通道注意力分支对原始输入 X 执行全局池化,并通过两个串联的卷积生成通道权重:

$$W_c = \text{Sigmoid}(\text{Conv}(\text{Conv}(\text{Pool}(X)))) \quad (22)$$

然后,生成的通道权重 W_c 用于调整输入特征:

$$U_c = W_c \odot X \quad (23)$$

其中, \odot 表示逐元素乘法操作.最终,将空间注意力的输入和通道注意力的输入通过逐元素点积进行交互,并通过一个 Conv 层恢复与输入相同的通道,然后添加一个残差连接得到最终的输出:

$$Y_{\text{MSAA}} = \text{Conv}(U_s + U_c) + X \quad (24)$$

最后将左右两个子模块的输出表示 Y_{CGA} 和 Y_{MSAA} 在特征维度上进行拼接操作,形成联合表示并映射回原始维度:

$$Y_{\text{concat}} = \text{Conv}_{1 \times 1}(\text{Concat}(Y_{\text{CGA}}, Y_{\text{MSAA}})) \quad (25)$$

2.2.4 MSAA-SAM 双分支融合模块

为实现结构建模分支与语义增强分支在全局语义层面的有效融合,本文在双分支之后设计了多尺度注意力聚合和句向量对齐机制 (MSAA-SAM)^[8,9],旨在提升融合特征在长文本语义组织中的一致性和连贯性.

需要特别指出的是,MSAA-SAM 中的 MSAA 机制被用于对结构特征与语义特征进行多尺度语义层次上的融合聚合,目的是提升双分支之间的特征协调性与整体语义表示的统一性.具体实现如下.

首先,MSAA 模块以编码张量 Y_{concat} 为输入,此处设为 Y ,其卷积核大小仍为 3×3 、 5×5 、 7×7 ,但卷积的目标是模拟结构语义对与内容语义之间的多层交互关系.

$$\tilde{Y} = \text{Conv}_{1 \times 1}(Y) \quad (26)$$

$$Y_1 = \text{Conv}_{3 \times 3}(\tilde{Y}) \quad (27)$$

$$Y_2 = \text{Conv}_{5 \times 5}(\tilde{Y}) \quad (28)$$

$$Y_3 = \text{Conv}_{7 \times 7}(\tilde{Y}) \quad (29)$$

$$U = Y_1 + Y_2 + Y_3 \quad (30)$$

随后引入空间注意力机制,引导模型关注句子中对于融合任务最为关键的位置区域:

$$W_c = \text{Sigmoid}(\text{Conv}(\text{Conv}(\text{Pool}(U)))) \quad (31)$$

$$U_s = \text{Conv}(W_s \odot U) \quad (32)$$

通过这种结构语义-内容语义的重构方式,实现双分支信息在空间位置上的动态聚焦。

此外,为缓解长文本中不同子句或段落之间语义不一致的问题,MSAA-SAM模块同时引入句向量对齐机制.该机制首先将每个时间步表示映射为固定长度的句子级语义向量,并构造全局上下文句向量:

$$S_{\text{global}} = \text{MeanPooling}(Y) \quad (33)$$

再将 S_{global} 与当前每一位置 Y_i 进行对齐建模:

$$A_i = \text{MLP}([Y_i; S_{\text{global}}; Y_i - S_{\text{global}}; X_i \odot S_{\text{global}}]) \quad (34)$$

最后,将对齐结果作为引导权重调节特征表示,得到最终输出:

$$Y_{\text{MSAA-SAM}} = \text{Conv}(U_s \oplus A) + Y \quad (35)$$

2.2.5 双分支模型的协同作用

在生成式摘要任务中,面向长文本的语义建模、上下文压缩和摘要导向表达的实际挑战提出的结构优化方案,具体作用体现如下。

(1) 理论创新性:设计了具备理论创新性的双分支结构,联合构建了多粒度、多维度、多层次的语义建模路径,显著增强了模型对长文本中主题主线、关键信息和结构层级的表达能力。

(2) 结构设计合理性:多尺度语义融合机制融合通道注意力与空间注意力,其中通道注意力负责筛选具有摘要意义的语义维度,空间注意力突出文本序列中具有关键性内容的位置信息,两者融合后的注意图具备更强的信息选择性和摘要任务适应性。

(3) 提升语义一致性与连贯性:提出MSAA-SAM融合模块,通过引入全局句向量对齐机制,进一步压缩与统一多尺度语义信息,提升了编码表示的一致性与

摘要导向性。

综上所述,本文提出的改进Transformer编码器从语义建模、上下文压缩、信息聚焦等关键维度系统地回应了生成式摘要中长文本处理的核心难点。

2.2.6 双分支结构与残差连接的差异分析

在长文本摘要任务中,双分支结构和残差连接在信息流传递和建模能力上的差异,直接影响模型在处理复杂上下文关系时的效果.具体差异体现在以下几个方面。

(1) 结构差异:残差连接通过单一路径传递信息,重点保留重要特征,而双分支结构采用并行路径分别处理全局和局部信息,再进行融合.因此双分支结构能有效解决信息割裂问题,增强语义建模能力。

(2) 建模能力差异:残差连接在处理复杂、多粒度的文本时能力有限,而双分支结构通过并行路径处理不同层次的语义特征,使模型在长文本摘要中具有更好的适应性。

(3) 信息融合差异:残差连接通过简单加法融合信息,双分支结构则通过特征拼接和卷积等方式更精细地融合信息.这使得双分支结构在语义表达和信息压缩方面表现更强。

综上所述,双分支结构相比残差连接,能更好地处理长文本摘要任务,提供更强的语义建模能力和信息整合能力。

2.3 面向长文本摘要的PGN模块

首先,指针生成网络会计算每个时间步 t 的注意力得分 e_i^t ,用于衡量解码器隐藏状态 S_t 和编码器隐藏状态 h_i 之间的相关性。

$$e_i^t = v^T \tanh(W_h h_i + W_s S_t + b_{\text{attn}}) \quad (36)$$

其中, W_h 和 W_s 是可学习的权重矩阵, b_{attn} 是偏置项, v 为投影向量.该得分经Softmax归一化后得到注意力分布:

$$\text{attn}^t = \text{Softmax}(e_i^t) \quad (37)$$

为增强上下文建模能力,本文将编码器右分支CGA-MSAA模块的输出表示 $Y_{\text{CGA-MSAA}}$ 与当前注意力上下文进行融合,获得融合上下文向量:

$$\text{Context}_t = \alpha \cdot \text{Context}_t + (1 - \alpha) \cdot Y_{\text{CGA-MSAA}} \quad (38)$$

融合后的上下文向量与当前解码器状态 S_t 联合计算词汇表上的生成概率分布:

$$P_{\text{vocab}} = \text{Softmax}(V^T(V[S_t], \text{Context}_t) + b) \quad (39)$$

其中, V 是权重矩阵, Context_t 是计算后的上下文向量, b 是偏置项。

为提升生成控制能力, 本文引入 MSAA-SAM 模块输出的全局句向量 S_{global} , 增强生成概率 P_{gen} 的判别能力:

$$P_{\text{gen}} = \sigma(w_{\text{Context}}^T \text{Context}_t + w_S^T S_t + w_{S_{\text{global}}}^T S_{\text{global}} + w_X^T X_{\text{input}} + b_{\text{ptr}}) \quad (40)$$

其中, σ 是 Sigmoid 激活函数, X_{input} 是上一个生成的单词。最终, 每个词的生成概率由词表生成概率与源文复制概率加权获得:

$$p(\omega) = P_{\text{gen}} P_{\text{vocab}}(\omega) + (1 - P_{\text{gen}}) \sum_i \text{attn}_i \quad (41)$$

训练过程中, 模型优化目标为最小化每个时间步预测词与真实目标词之间的交叉熵损失:

$$\text{loss} = -\frac{1}{T} \sum_{t=0}^T \log P(w_t) \quad (42)$$

该设计能够融合多粒度语义表示与全局句向量信息, 增强模型在解码阶段的语义感知与生成控制能力。采用改进后的 PGN 模块具有以下优势。

(1) 提升了对关键信息的聚焦能力, 有效缓解了长文本生成中的重复与冗余问题

(2) 增强了语义一致性, 实现更精确的词汇生成与复制策略调控。

实验结果表明, 增强的 PGN 模块在摘要准确性、信息压缩率和生成连贯性等方面均有提升。具体表现为: 相比传统 PGN 模型^[10], 改进后的模型在长文本摘要任务中对关键信息的捕捉更加准确, 语义连贯性更强, 验证了引入句向量引导、语义融合与上下文控制机制的有效性。该设计为复杂长文本场景下的摘要生成任务提供了新的建模思路与优化方向。

2.4 模型复杂度

为了全面评估模型的计算效率与部署能力, 本文从编码阶段与解码阶段分别分析了模型的时间复杂度与空间复杂度。

在训练阶段, 模型整体时间复杂度包括左分支标准的多头自注意力部分 $O(N \cdot L \cdot h \cdot d^2)$ 、右分支 CGA-MSAA 部分 $O(S \cdot L \cdot d)$, 其中 S 表示尺度数 (本文为 3)、融合阶段的 MSAA-SAM 部分 $O(L \cdot d)$, 解码阶段的自

注意力机制、交叉注意力机制和增强的 PGN 部分总共为 $O(L_t^2 \cdot d + L_t \cdot L \cdot d + d^2)$, 因此, 模型训练阶段总时间复杂度为 $O(N \cdot L \cdot h \cdot d^2 + S \cdot L \cdot d + d^2 + L_t \cdot L \cdot d + d^2)$ 。

在空间复杂度方面, 左分支标准的多头自注意力部分 $O(L^2 \cdot L \cdot d)$ 、右分支 CGA-MSAA 部分 $O(L \cdot d)$ 、此外为引入全局句向量控制, MSAA-SAM 部分增加了少量句向量-状态融合的矩阵投影操作产生额外的空间消耗为 $O(d^2)$, 但不影响整体主阶复杂度。解码阶段为 $O(L_t \cdot d + L \cdot d)$, 总空间复杂度为 $O(L^2 + L \cdot d + d^2 + L_t \cdot d)$ 。

综上, 本文模型在保证表达能力的同时, 通过并行多尺度建模、轻量级注意机制与句向量控制策略, 有效控制了计算复杂度与内存开销。尤其在解码阶段, 引入句向量调控的 PGN 机制仅带来常数级的额外消耗, 相较于深层堆叠或多阶段摘要模型, 具备更高的计算效率与部署适应性, 适用于长文本摘要等资源受限场景下的高效生成任务。

3 实验与分析

3.1 实验设置

3.1.1 数据集

本文采用 NLPCC 2017 中文新闻摘要数据集^[11]: 该数据集由 2017 年 CCF 国际自然语言处理与中文计算会议 (NLPCC 2017) 发布, 属于中长文本单文档摘要任务, 文本内容主要来源于新闻领域, 共计包含 49944 条新闻样本, 每条样本均包含新闻标题 (摘要) 与正文内容。本文按照 8:2 的比例划分训练集与测试集, 用于实验的性能评估。

LCSTS 微博摘要数据集^[12]: 该数据集由哈尔滨工业大学发布, 内容主要来源于新浪微博, 共收录 2108915 条微博样本。每条样本包含一条微博正文与其对应的标题式摘要, 文本风格偏口语化、句式较短, 具有一定的摘要挑战性。本文采用其中清洗后的数据集, 并以 8:2 的比例进行训练集与测试集划分。

数据集的统计信息如表 1。

表 1 数据集的统计信息

数据集	来源	规模 (条)	训练集 (条)	测试集 (条)	平均长度 (字)
NLPCC 2017	新闻领域	49944	39955	9988	25
LCSTS	新浪微博	2108915	1687132	421783	20

3.1.2 评估指标

为了客观评估模型摘要结果与参考摘要之间的匹

配程度, ROUGE-1、ROUGE-2 和 ROUGE-L 作为评价标准。

3.2 实验结果分析

3.2.1 整体性能比较

将提出的模型与以下基线方法比较。

(1) BART^[13]: 结合 BERT 的编码器与 GPT 的解码器结构, 支持生成式任务 (如摘要、翻译)。

(2) BERTSUM^[14]: 一种基于 BERT 的抽取式摘要模型, 通过句子级分类器选择重要句子生成摘要。

(3) T5^[15]: 将所有 NLP 任务统一为文本到文本形式的预训练模型, 支持高质量抽象式摘要生成。

(4) PEGASUS^[16]: 引入了 gap sentence generation 的预训练目标, 通过删除输入文档中的重要句子并让模型预测这些句子, 增强了模型的摘要生成能力。

(5) BRIO^[17]: 通过对候选摘要进行排序优化, 提升了生成摘要的质量和一致性。

(6) UniSumm^[18]: 一个统一的多任务摘要框架, 支持抽取摘要, 适用于不同风格的文本。

(7) HiStruct+^[19]: 引入了层次结构建模机制, 显式编码文本的层次结构信息, 特别适用于结构化长文本 (如新闻、论文) 的摘要任务。

(8) LongLLMLingua^[20]: 针对长文本场景优化的大型语言模型压缩方法, 通过引入问题感知的粗细粒度压缩、文档重排序和子序列恢复策略, 提升了模型对关键信息的感知能力, 在长文本摘要任务中表现出色。

在相同实验条件下, 各实验结果如图 3 和图 4 所示。

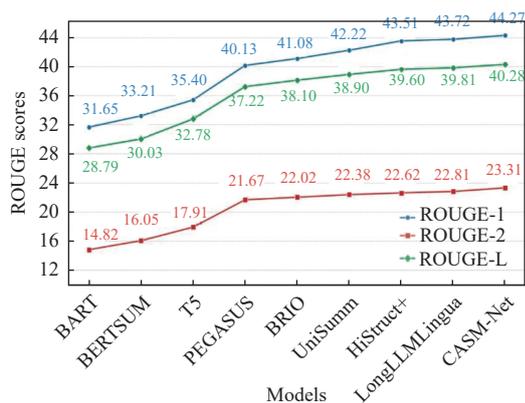


图 3 本模型与基线模型在 NLPCC 2017 数据集上的实验结果图

从实验结果的整体对比看, 该模型 (CASM-Net) 在所有评估指标上都展示出优势: 在 NLPCC 2017 数据集上, 相较当前最优的 LongLLMLingua 模型, ROUGE-1

提升了 1.26% (从 43.72 提升至 44.27), ROUGE-2 提升了 2.19% (从 22.81 提升至 23.31), ROUGE-L 提升 1.18% (从 39.81 提升至 40.28)。

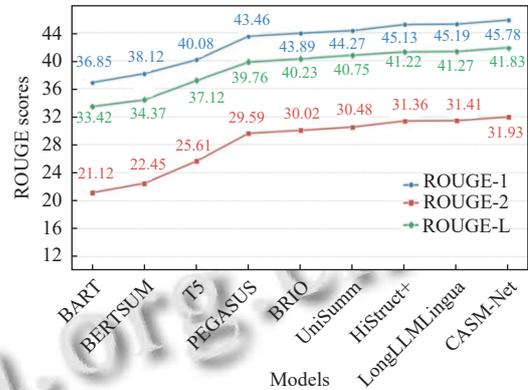


图 4 本模型与基线模型在 LCSTS 数据集上的实验结果图

在 LCSTS 数据集上, CASM-Net 同样取得领先性能。相比于 LongLLMLingua 模型, ROUGE-1 提升了 1.31% (从 45.19 提升至 45.78), ROUGE-2 提升 1.66% (从 31.41 提升至 31.93), ROUGE-L 提升了 1.36% (从 41.27 提升至 41.83)。

从方法演进的角度分析, 从传统的 BERT 到本研究提出的 CASM-Net 的性能提升, 清晰地展现了多尺度语义建模与内容引导机制在生成式摘要中的重要性。尤其是双分支结构设计、CGA 引导注意力机制、MSAA 多尺度感知模块与增强型 PGN 解码器的表现, 验证了各子模块对提升摘要质量的有效性。

本模型的优势主要体现在以下几个方面。

(1) 结构设计合理: 采用双分支语义建模策略, 将自注意力与内容感知建模解耦, 有效提升了文本的语义理解深度。

(2) 信息压缩能力^[21]: 通过 MSAA 和 SAM 融合机制, 引导模型聚焦多粒度语义区域, 提升了对长文本主干信息的压缩表达能力。

(3) 生成质量提升: 增强 PGN 解码器融合全局句向量, 有效缓解了传统生成式模型中出现的重复内容和语义漂移问题。

从各项指标的深入分析来看, 相对高的 ROUGE-2 值表明模型整体已优于多数对比模型; 而较高的 ROUGE-1 和 ROUGE-L 值, 则说明模型在词汇覆盖与长句结构保持方面具有显著优势。这些实验结果充分证明了模型的有效性与实用性, 其创新的内容引导机制和多尺度融合结构能够更好地捕捉摘要任务所需的

深层语义信息,从而生成更加紧凑、连贯、信息密度高的自动摘要结果。

3.2.2 消融实验通道注意力机制在 NLP 中的适配性验证

为了验证通道注意力机制在 NLP 任务中的适配性,我们将其作为额外的机制加入标准的 Transformer 架构中.具体实验设置如下。

(1) BTP: 该模型使用 BERT+Transformer+PGN 框架。

(2) BTP+Channel: 该模型在标准的 BERT+Transformer+PGN 框架中,额外加入了通道注意力机制。

对比实验结果如图 5 和图 6 所示。

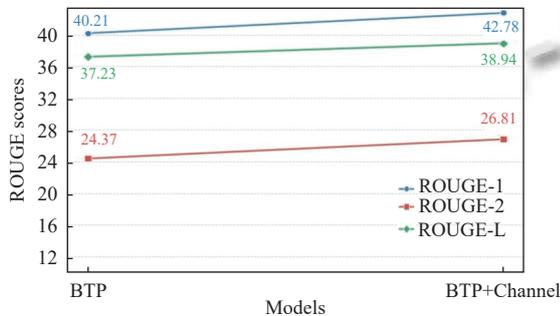


图 5 通道注意力机制在 NLPCC 2017 数据集上的验证结果图

根据图 5 和图 6 的实验结果,可得出如下分析结论:在 NLPCC 2017 数据集上,BTP+Channel 模型相较于 BTP 模型,在 ROUGE-1 上提升了 6.39%,ROUGE-2 提升了 10%,ROUGE-L 提升了 4.6%。在 LCSTS 数据集上,ROUGE-1 提升了 9.04%,ROUGE-2 提升了 15.42%,ROUGE-L 提升了 5.39%。这些结果表明,加入通道注意力机制显著改善了模型在关键信息提取和冗余信息压缩方面的能力,从而提高了文本摘要的质量。

表 2 n -gram 覆盖率统计结果 (%)

Model	NLPCC 2017			LCSTS		
	1-gram	2-gram	3-gram	1-gram	2-gram	3-gram
BTP	66.80	48.92	33.07	59.60	43.72	29.81
BTP+MSAA	74.51	57.77	41.81	72.30	56.77	42.54

3.2.4 统计显著性检验

为了验证 CASM-Net 模型在 ROUGE-1 指标上的性能提升是否具有统计显著性,本文采用配对样本 t 检验,对 CASM-Net 与 LongLLMLingua 模型在 NLPCC 2017 和 LCSTS 两个数据集上的表现差异进行评估.实验在相同条件下重复运行 10 次,每次结果如表 3 所示。

对上述数据进行配对 t 检验后,在 NLPCC 2017

通道注意力机制有效地引导模型聚焦于最具信息密度的部分,优化了摘要生成的效果,并提高了模型对长文本处理的适应性。

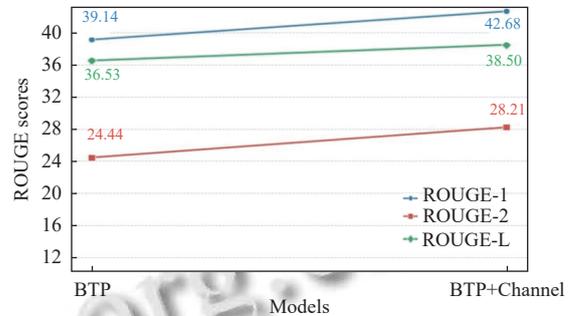


图 6 通道注意力机制在 LCSTS 数据集上的验证结果图

3.2.3 基于 n -gram 覆盖率的多尺度注意力机制必要性验证

为验证多尺度注意力机制在 NLP 任务中的必要性,我们将其作为额外的机制加入标准的 Transformer 架构中.具体实验设置如下。

(1) BTP: 该模型使用 BERT+Transformer+PGN 框架。

(2) BTP+MSAA: 该模型在标准的 BERT+Transformer+PGN 框架中额外加入了多尺度注意力机制。

n -gram 覆盖率统计结果如表 2 所示。

从表 2 中的 n -gram 覆盖率统计结果可以看出,加入多尺度注意力机制 (BTP+MSAA) 显著提高了模型在各 n -gram 维度上的覆盖率.这一结果表明,多尺度注意力机制能够有效提升模型对不同粒度语义信息的感知能力,尤其在长文本处理中的 n -gram 特征捕捉上,进一步提升了模型的性能.因此,加入多尺度注意力机制对提升模型在 NLP 任务中的性能是必要的。

和 LCSTS 两个数据集上均获得 p -value<0.05 的结果.这表明,CASM-Net 在 ROUGE-1 指标上的性能提升在统计上是显著的,性能差异并非由偶然波动引起,而是具有稳定性和统计可信度。

3.2.5 消融实验

为了验证模型各个模块的有效性,本节进行了详细的消融实验,通过逐步引入模型的核心创新组件,分

析各部分对模型性能的贡献. 具体实验设置如下.

(1) Base: 采用标准的 BERT 模型+标准的 Transformer+PGN 结构.

(2) Base+CGA-MSAA: 加入内容引导与多尺度建模右分支结构, 以增强模型对语义密度和重要内容的建模能力.

(3) Base+CGA-MSAA+MSAA-SAM: 在左分支和右分支之间引入多尺度注意力融合模块, 实现跨分支语义对齐与信息整合.

(4) 完整 (CSAM-Net): 包含所有组件.

表 3 CASM-Net 模型运行结果统计表

序号	NLPCC 2017		LCSTS	
	LongLLMLingua	CASM-Net	LongLLMLingua	CASM-Net
1	43.82	44.18	45.10	45.73
2	43.69	44.18	45.02	45.80
3	43.85	44.32	45.05	45.83
4	44.02	44.04	45.15	45.76
5	43.67	44.04	45.17	45.82
6	43.67	44.16	45.09	45.73
7	44.04	44.07	45.13	45.77
8	43.87	44.33	45.18	45.85
9	43.63	44.09	45.14	45.70
10	43.83	44.04	45.11	45.81

消融实验结果如图 7 和图 8 所示.

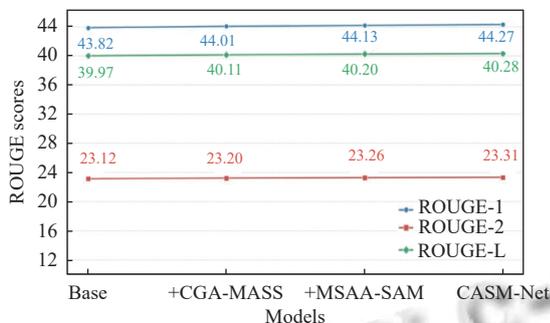


图 7 本模型在 NLPCC 2017 数据集上的消融实验结果图

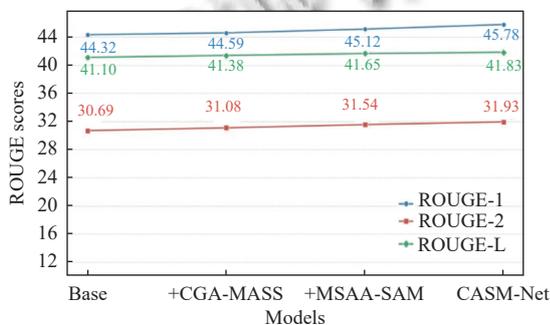


图 8 本模型在 LCSTS 数据集上的消融实验结果图

根据图 7 和图 8 的消融实验结果, 可得出如下分析结论.

(1) 基础结构 Base (即 BERT+Transformer+PGN) 提供了基本的语义建模与生成能力, 但各项指标表现仍有提升空间, 尤其在 ROUGE-2 上偏低, 说明仅依赖基础结构难以充分捕捉长文本中的深层语义关系与关键信息.

(2) 引入 CGA-MSAA 模块后, 模块通过内容引导注意力与多尺度语义建模机制, 增强了模型对摘要相关内容的识别能力, 有效缓解了冗余生成与信息遗漏问题.

(3) 引入 MSAA-SAM 模块后, 通过双分支语义融合与全局句向量对齐机制, 有效提升了左右分支的信息协同能力, 使模型能够在结构上形成对关键信息的统一建模视角, 从而提升摘要的一致性与完整性.

(4) 最终集成改进的 PGN 模块后, 该模块在调控生成模式、控制词汇复制与选择方面起到了关键作用, 增强了模型的解码精度和控制能力.

CASM-Net 模型在 NLPCC 2017 数据集上相较基础 Base 模型, 其各项评价指标: ROUGE-1、ROUGE-2 和 ROUGE-L 分别提升了 1.03%、0.82% 和 0.78%. 在 LCSTS 数据集上, 相较于 Base 模型, 其各项评价指标: ROUGE-1、ROUGE-2 和 ROUGE-L 分别提升 3.29%、4.04% 和 1.77%.

每个模块的引入都为解决特定问题提供了有效手段, 在语义理解和内容生成两个层面上协同优化: CGA-MSAA 模块增强了模型对关键信息和多粒度语义的捕捉能力, MSAA-SAM 模块实现了跨分支特征的融合与句级语义对齐, 改进 PGN 进一步提升了生成控制能力与句子连贯性. 通过这些模块的协同工作, 模型能够更全面地建模长文本中的关键信息与语义结构, 提供高质量、准确性和连贯性的自动摘要结果.

4 结论

针对长文本摘要存在的语义信息分布稀疏、冗余重复生成、上下文连贯性不足等问题, 本文从语义建模增强、跨尺度信息融合、生成控制优化等角度出发, 提出一种融合内容引导与多尺度注意力机制的摘要生成模型 CASM-Net. 首先, 在 BERT-Transformer-PGN 编码结构的基础上引入双分支架构, 其中左分支用于建模全局上下文语义, 右分支则结合 CGA-MSAA 模块以便强化对多粒度关键信息的感知. 其次, 通过 MSAA-SAM 模块进行双分支语义对齐与融合, 以此来

缓解信息割裂、建模片面的问题。然后,在对传统的PGN 结构进行生成机制扩展的基础上,引入了全局句向量引导与上下文融合机制,有效提升了生成内容的聚焦性与控制力。通过与多个基线模型的性能比较以及消融实验验证结果表明,CASM-Net 模型在 NLPC 2017 数据集和 LCSTS 数据集上 ROUGE-1、ROUGE-2 和 ROUGE-L 等指标均有提升,以此验证了所提出模型结构及各模块设计的有效性。未来的研究将继续深入探索摘要生成中的语义强化机制与生成可控性设计,进一步提升模型在多场景长文本摘要任务中的泛化能力与实际应用价值。

参考文献

- 1 CNNIC. 第 55 次《中国互联网络发展状况统计报告》. <https://www1.cnnic.cn/n4/2025/0117/c88-11229.html>. [2025-01-17].
- 2 田萱, 李嘉梁, 孟晓欢. 基于深度学习的抽取式摘要研究综述. 计算机科学与探索, 2024, 18(11): 2823–2847. [doi: 10.3778/j.issn.1673-9418.2308100]
- 3 陈明轩, 肖诗斌, 王洪俊. 基于深度学习的生成式文本摘要综述. 软件导刊, 2024, 23(5): 212–220.
- 4 Devlin J, Chang MW, Lee K, *et al.* BERT: Pre-training of deep bidirectional Transformers for language understanding. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). Minneapolis: ACL, 2019. 4171–4186.
- 5 Vaswani A, Shazeer N, Parmar N, *et al.* Attention is all you need. Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach: Curran Associates Inc., 2017. 6000–6010.
- 6 Chen ZX, He ZW, Lu ZM. DEA-Net: Single image dehazing based on detail-enhanced convolution and content-guided attention. IEEE Transactions on Image Processing, 2024, 33: 1002–1015. [doi: 10.1109/TIP.2024.3354108]
- 7 Liu MS, Dan J, Lu ZQ, *et al.* CM-UNet: Hybrid CNN-mamba UNet for remote sensing image semantic segmentation. arXiv:2405.10530, 2024.
- 8 Wang H, Xiong WH, Yu M, *et al.* Sentence embedding alignment for lifelong relation extraction. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). Minneapolis: ACL, 2019. 796–806.
- 9 Cui P, Hu L, Liu YC. Inducing alignment structure with gated graph attention networks for sentence matching. arXiv:2010.07668v2, 2020.
- 10 See A, Liu PJ, Manning CD. Get to the point: Summarization with pointer-generator networks. Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Vancouver: ACL, 2017. 1073–1083.
- 11 Hua LF, Wan XJ, Li L. Overview of the NLPC 2017 shared task: Single document summarization. Proceedings of the 6th CCF International Conference on Natural Language Processing and Chinese Computing. Dalian: Springer, 2017. 942–947.
- 12 Hu BT, Chen QC, Zhu FZ. LCSTS: A large scale Chinese short text summarization dataset. Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. Lisbon: ACL, 2015. 1967–1972.
- 13 Lewis M, Liu YH, Goyal N, *et al.* BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. ACL, 2020. 7871–7880.
- 14 Liu Y. Fine-tune BERT for extractive summarization. arXiv:1903.10318v2, 2019.
- 15 Raffel C, Shazeer N, Roberts A, *et al.* Exploring the limits of transfer learning with a unified text-to-text Transformer. The Journal of Machine Learning Research, 2020, 21(1): 140.
- 16 Zhang JQ, Zhao Y, Saleh M, *et al.* PEGASUS: pre-training with extracted gap-sentences for abstractive summarization. Proceedings of the 37th International Conference on Machine Learning. JMLR.org, 2020. 11328–11339.
- 17 Liu YX, Liu PF, Radev D R, *et al.* BRIO: Bringing order to abstractive summarization. Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Dublin: ACL, 2022. 2890–2903.
- 18 Chen YL, Liu Y, Xu RC, *et al.* UniSumm and SummZoo: Unified model and diverse benchmark for few-shot summarization. Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Toronto: ACL, 2023. 12833–12855.
- 19 Ruan Q, Ostendorff M, Rehm G. HiStruct+: Improving extractive text summarization with hierarchical structure information. Proceedings of the 2022 Findings of the Association for Computational Linguistics. Dublin: ACL, 2022. 1292–1308.
- 20 Jiang HQ, Wu QH, Luo XF, *et al.* LongLLMLingua: Accelerating and enhancing LLMs in long context scenarios via prompt compression. Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Bangkok: ACL, 2024. 1658–1677.
- 21 Park J, Pilault J, Pal C. Learning to summarize long texts with memory compression and transfer. arXiv:2010.11322, 2020.

(校对责编: 李慧鑫)