计算机工程与应用
Computer Engineering and Applications
ISSN 1002-8331,CN 11-2127/TP

《计算机工程与应用》网络首发论文

题目：基于深度学习的自动文本摘要研究综述
作者：其其日力格，斯琴图，王斯日古楞
网络首发日期：2025-04-28
引用格式：其其日力格，斯琴图，王斯日古楞．基于深度学习的自动文本摘要研究综述[J/OL]．计算机工程与应用．
https://link.cnki.net/urlid/11.2127.TP.20250428.1328.004

基于深度学习的自动文本摘要研究综述

其其日力格，斯琴图+，王斯日古楞

内蒙古师范大学 计算机科学技术学院，呼和浩特　010022
+ 通信作者 E-mail：sqt@imnu.edu.cn

摘　要：自动文本摘要技术是自然语言处理领域的重要研究方向，旨在实现信息的高效压缩与核心语义的保留。随着深度学习技术的快速发展，基于该技术的自动文本摘要方法逐渐成为主流。本文从抽取式与生成式两大技术路线出发，系统梳理了序列标注、图神经网络、预训练语言模型、序列到序列模型和强化学习等技术在自动文本摘要中的应用，并分析了各类模型的优缺点。同时，介绍了自动文本摘要领域常用的公开数据集、国内低资源语言数据集及评价指标。并通过多维度实验对比分析总结了现有技术面临的问题，提出了相应的改进方案。最后，探讨了自动文本摘要的未来研究方向，为后续研究提供参考。

关键词：自动文本摘要；深度学习；生成式摘要；抽取式摘要；自然语言处理
文献标志码：A　中图分类号：TP391　doi：10.3778/j.issn.1002-8331.2411-0270

A Survey of Automatic Text Summarization Based on Deep Learning

Qiqirilige, SI Qintu, WANG Siriguleng

School of Computer Science and Technology, Inner Mongolia Normal University, Hohhot 010022, China

Abstract： Automatic text summarization is a significant research direction in the field of natural language processing, aiming to achieve efficient compression of information while preserving its core semantics. With the rapid development of deep learning techniques, methods based on these technologies have gradually become mainstream. From both extractive and abstractive approaches, this paper systematically reviews the application of techniques such as sequence labeling, graph neural networks, pre-trained language models, sequence-to-sequence models, and reinforcement learning in text summarization, analyzing the strengths and weaknesses of various models. In addition, commonly used public datasets, domestic low-resource language datasets, and evaluation metrics in the field of text summarization are introduced. Multi-dimensional experimental comparisons and analyses are conducted to summarize the current challenges faced by existing technologies and to propose corresponding improvement strategies. Finally, future research directions in automatic text summarization are discussed to provide a reference for subsequent studies.

Key words： automatic text summarization; deep learning; abstractive summarization; extractive summarization; natural language processing

基金项目： 内蒙古自治区自然科学基金项目(2024LHMS06024)；内蒙古自治区自然科学基金项目(2022MS06002)；内蒙古自治区科技计划项目（2021GG0139)；内蒙古纪检监察大数据实验室开放项目（IMDBD202109)。

作者简介： 其其日力格(2000—)，女，硕士研究生，CCF 学生会员，研究方向为自然语言处理；斯琴图(1972—)，通信作者，男，硕士，副教授，研究方向为自然语言处理、计算机网络，E-mail: sqt@imnu.edu.cn；王斯日古楞(1970—)，女，博士，教授，CCF 会员，研究方向为自然语言处理、机器翻译。

随着互联网和社交媒体的迅猛发展，信息生成速度呈指数级增长，用户所面临的信息过载问题愈加突出。如何快速从海量文本中提取核心信息，帮助用户高效获取有用信息，已成为一项关键技术需求。自动文本摘要（Automatic Text Summarization,ATS）技术通过从长文本中提取关键信息生成简短摘要，显著提高了信息获取效率。1958 年，Luhn[1]首次将自然语言处理技术应用于自动文本摘要任务。早期的自动文本摘要主要使用规则和统计方法，但在处理复杂语言现象时表现受限，难以理解文本的深层语义，且生成的摘要常存在冗余和不连贯问题。随着深度学习（Deep Learning,DL）技术的快速发展，自动文本摘要取得了显著进展，特别是大语言模型（Large Language Model,LLM）的发展，极大提升了文本语义理解能力和摘要生成的质量。然而，尽管 LLM 在摘要生成上取得了显著成效，仍面临数据集匮乏和对强大算力依赖等重大挑战。因此，深入探讨基于 DL 的 ATS 技术，变得尤为重要。

目前，已有多篇综述研究[2][3][4]探讨了基于深度学习的自动文本摘要技术，但大多数综述集中于生成式摘要（Abstractive Summarization）领域，缺乏对基于深度学习的抽取式摘要（Extractive Summarization）和生成式摘要两种方法的全面分析。因此，本文对近年来基于 DL 的 ATS 技术进行系统回顾，从抽取式和生成式两个方面深入分析其核心技术、优势与不足，介绍代表性模型及其在实际应用中的表现。通过多维度实验对比与分析，探讨了 ATS 目前存在的问题及解决方案。最后，本文展望了未来的研究方向，并提出了可能的技术突破点，例如低资源语言的处理、跨语言摘要以及多文档摘要，以期为后续研究提供参考。

1 自动文本摘要任务概述

自动文本摘要根据输出类型分为抽取式摘要和生成式摘要。抽取式摘要通过选择原文中的重要句子或短语并直接拼接生成，其内容与原文一致，信息准确且不易出现语法错误。因此，在政策解读、官方文件总结及法律、医学等对信息精确性要求较高的领域具有广泛应用[3]。然而，抽取式摘要的表述常显得机械，缺乏语言连贯性与自然性，并可能导致信息冗余。

相较之下，生成式摘要通过理解文本的语义生成新的句子来表达原文核心内容，因而具备更高的灵活性和语言流畅性。尽管生成式摘要在改善文本流畅度方面表现优异，但如何确保生成内容与原文的一致性，并避免生成虚假信息，仍是一项挑战。

随着深度学习的兴起，抽取式摘要逐渐采用神经网络学习文本中句子的相对重要性。Nallapati 等[5]提出了基于递归神经网络（Recurrent Neural Network,RNN）的序列标注模型，通过学习句子特征判断其是否应被抽取为摘要。随后，Zhou 等[6]提出了结合卷积神经网络（Convolutional Neural Network,CNN）与 RNN 的混合模型，提升了模型处理长文本和语义理解能力。Liu[7]提出了 BERTSUM 模型，通过微调 BERT 进行句子编码，在句子级别打分并抽取，显著提升了抽取式摘要的质量。

生成式摘要通常依赖序列到序列（Sequence-to-Sequence,Seq2Seq）模型。Rush 等[8]在 2015 年将 Seq2Seq 模型与注意力机制[9]相结合，显著提高了摘要的流畅性和连贯性。See 等[10]提出的指针生成网络，通过结合生成和复制机制，显著增强了模型处理未登录词（Out-of-Vocabulary，OOV）的能力，从而提高了摘要的准确性。近年来，预训练语言模型（PLM）的引入进一步推动了生成式摘要的发展。Lewis 等[11]提出的 BART 模型在预训练阶段通过破坏输入文本并学习重建，大幅提升了摘要生成的流畅性和准确性。Zhang 等[12]提出的 PEGASUS 模型通过设计特定的预训练任务，增强了模型在低资源数据集上的表现。借助这些预训练技术，生成式模型在理解文本关键信息和生成符合语境的摘要方面取得了显著进展。

此外，根据文档数量，ATS 可分为单文档摘要和多文档摘要。单文档摘要从一篇文章中提炼核心信息，

而多文档摘要则需要从多个相关文档中综合信息生成摘要。多文档摘要对模型的信息整合和对比能力提出了更高要求，同时也需要解决信息冗余、冲突及跨文档语义连贯性等挑战。多文档摘要主要用于新闻聚合和综述生成等任务，是当前研究的热点之一。

2 基于深度学习的文本摘要方法

深度学习方法能够自动从大规模数据中学习文本的语义和结构特征，相比传统的规则或统计方法，其在复杂语言现象处理和模型的泛化能力上表现更为优异。

2.1 基于深度学习的抽取式文本摘要

抽取式文本摘要旨在从原文中提取关键信息，生成简洁且高效的摘要。深度学习的引入提升了抽取式摘要在处理语义和上下文关联等复杂关系方面的能力。此类方法通常利用神经网络学习句子或段落的重要性，并选取最关键的内容作为摘要输出。本文将基于深度学习的抽取式摘要方法按技术策略分为四类：序列标注、图神经网络、预训练模型和强化学习。表 1 对这四类基于 DL 的抽取式摘要方法进行了优缺点对比分析。

表 1 基于 DL 的抽取式摘要方法对比

Table 1 Comparison of DL-Based Extractive Summarization Methods

方法类别	描述	经典模型	优点	缺点
序列标注	将每个句子标记为“重要”或“非重要”，选取重要句子构成摘要	SummaRuNNer[5] NEUSUM[6] DeepSumm[13]	通过上下文学习句子重要性，适合结构化文本	无法直接建模复杂句法和语义关系，对长文本处理能力有限
图神经网络	构建句子间的图结构，捕捉句子之间的关系，识别并选取重要句子生成摘要	HeterSumGraph[14] MuchSum[15]	有效捕捉长距离依赖和多层次关系	模型复杂度高，计算资源需求大
预训练语言模型	通过自注意力机制生成句子的上下文嵌入并评分，选择高分句子作为摘要	BERTSUM[7] MatchSum[16]	深层次的语义理解，迁移学习能力强	对长文档的固定长度输入限制
强化学习	设计奖励函数来评估生成摘要的质量，逐步选择候选句子并优化决策策略	REFRESH[17] GoSum[18]	优化句子选择策略，提升摘要连贯性和覆盖度	训练效率低，奖励函数设计具有困难

2.1.1 基于序列标注的抽取式摘要方法

序列标注的抽取式摘要方法把摘要任务看作是一个分类问题。简单来说，它就是给文档里的句子打标签，标记出哪些句子应该被选为摘要。这种方法使用神经网络来提取文本的特征，然后根据句子之间的上下文和重要性来进行预测。

Cheng 等[19]首次提出基于双向 LSTM 的端到端模型，将 ATS 任务视为从文章中选取最具代表性的句子。该模型在句子层引入了注意力机制，使其能够根据上下文对句子进行打分，从而选择最重要的句子。在此基础上，Nallapati 等[5]提出 SummaRuNNer 模型，进一步优化了句子选择的过程。该模型采用了词级和句子级 RNN 建模，通过词级的平均池化生成句子表示，并在逻辑层对句子进行二分类。相比于 Cheng 等提出的模型，SummaRuNNer 更注重层次化的建模，提升了句子选择的准确性。具体结构如图 1 所示。

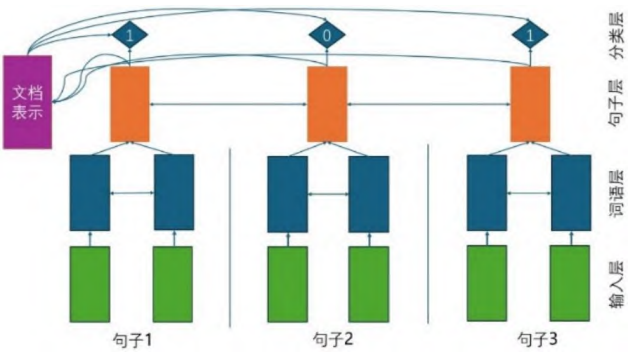

图 1 SummaRuNNer 的模型框架图

Fig.1 model framework diagram of SummaRuNNer

Zhou 等[6]提出的 NEUSUM 模型将句子打分与句子选择整合为统一的框架。该模型利用双向 GRU 提取句子表示，在生成每个句子的表示时，NEUSUM 模型采用了拼接双向 GRU 的最后隐藏状态，而不是传统的词级平均池化。通过这种方式，模型能够更好地

捕捉句子中的上下文信息，从而提高了句子表示的准确性和表达能力。NEUSUM 的一大创新在于，每次选择句子时会结合已选句子的语义信息，确保摘要的连贯性。

Joshi 等[13]提出了 DeepSumm 模型，该模型采用递归神经网络与主题模型相结合的方法，通过 Seq2Seq 框架，从内容、主题、新颖性和位置四个维度为句子生成多维评分。这种设计使 DeepSumm 能够有效捕捉全局和局部的语义信息，从而在 DUC2002 和 CNN/DailyMail 数据集上表现优于其他序列标注方法。

基于序列标注的模型通过将文档视为由句子和词构成的线性序列，逐步提取每个文本单元的特征，从而在生成摘要时能够有效识别局部重要信息。这类方法通常仅关注句子选择，而忽略了摘要的句子顺序对可读性和语义连贯性的影响。为此，Kwon 等[20]提出 OrderSum，通过引入句子顺序信息优化抽取式摘要，使摘要更具逻辑性和可读性。然而，文本中的结构不仅仅是线性序列，还存在诸如句法依赖、语义关联等复杂的拓扑关系，基于序列标注的模型无法直接建模这些复杂结构。

2.1.2 基于图神经网络的抽取式摘要方法

为克服传统序列模型在捕捉全局关系和语义依赖方面的不足，基于图神经网络（Graph Neural Networks, GNNs）的方法逐渐成为抽取式摘要的重要研究方向。GNN 通过构建句子、段落和词汇之间的图结构，并利用图中的传播机制逐层更新节点表示，从而有效建模复杂关系、捕捉长距离的语义依赖和多层次的关系。

Yasunaga 等[21]提出的模型利用图卷积网络(Graph Convolutional Network,GCN)将句子表示为图节点，并通过句子之间的余弦相似度构建节点间的边。该模型通过逐层的信息传播，让每个节点可以从相邻的节点获取信息，从而有效捕捉多文档之间的句子关系，并减少冗余信息。这样，模型能够更好地整合不同文档中的重要内容，非常适合用于新闻汇总等多文档摘要任务。Wang 等[14]提出 HeterSumGraph 模型，如图 2 所示，该模型使用二部图结构，将句子和词作为不同类型的节点，并用 TF-IDF 权重定义边的权重。通过图注意力网络(Graph Attention Network,GAT)，模型捕捉不同节点的重要性变化，并动态更新节点表示。相比 Yasunaga 等提出的模型，HeterSumGraph 加强了词与句子之间的关系表达，且该模型在单文档和多文档摘要任务都适用。

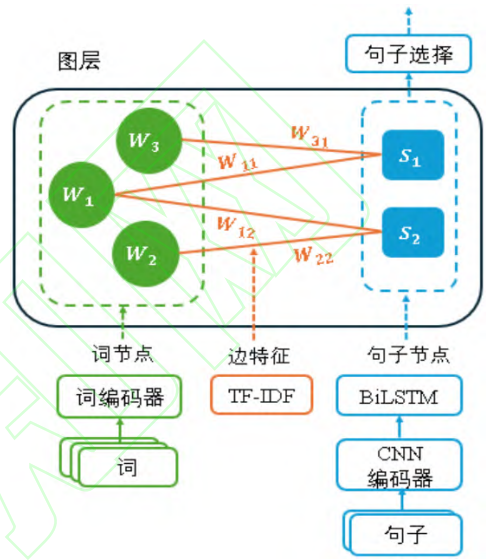

图 2 HSG 的模型框架图

Fig.2 Model framework diagram of HSG

在 HeterSumGraph 的基础上，Mao 等[15]提出 MuchSum 模型，设计了一种多通道图卷积网络。他们引入了三个图通道，分别捕捉句子的文本特征、中心度特征和位置特征，并融合各通道的句子表示进行摘要生成，提升了摘要的准确性和去冗余能力。

以上方法往往只关注单一类型的句子间关系，而忽视了句子内部词汇间的丰富信息。为了弥补这一缺陷，Jing 等[22]引入多重图卷积网络（Multi-GCN），通过构建一个包含语义相似度和自然连接等多种类型关系的多重图，更全面地捕捉句子和词汇间的复杂联系。在此基础上，他们进一步提出了 Multiplex Graph Summarization（Multi-GraS）框架，利用 Multi-GCN 的分析结果有效提取最具代表性的句子，生成高质量摘要。实验结果显示，Multi-GraS 在 CNN/DailyMail 基准数据集上取得了优于现有方法的性能。

Yasunaga 等[23]提出 ScisummNet 模型，用于科学文献摘要。该模型将摘要句和引用句作为图节点，基于余弦相似度建立句子间的连接，利用论文间的引用关系来识别相关且重要的句子。通过这种方法，Scisu

mmNet 能够生成更加全面的摘要。此外，Xu 等[24]提出话语感知神经抽取模型 DISCOBERT，该模型将文档中的话语结构和层次关系建模为图结构，并通过图神经网络 GNN 来捕捉这些关系。该方法在长文档摘要任务中表现出明显的优势，能够生成逻辑连贯、语义一致的摘要。除了上述方法外，Onan 等[25]将超图引入抽取式摘要，提出了 MCHES 模型。超图是一种允许超边同时连接多个节点的图结构，相比传统图，可更灵活地表达句子间的多元关系。MCHES 框架通过构建语义、叙事和话语三类超边的上下文超图，并利用注意力机制融合多种关系信息评估句子重要性，在 CNN/DailyMail 上显著优于多种经典模型，验证了超图神经网络在抽取式摘要中的潜力。

基于图神经网络的抽取式摘要模型通过对多层次关系的建模，克服了传统序列模型在捕捉全局依赖和句子间复杂关系上的不足。这类模型在不同的应用场景中展现了强大的扩展性和适应性，如多文档摘要、科学文献摘要及长文档摘要。然而，随着模型复杂度的提升，其计算成本和资源需求也相应增加。

2.1.3 基于预训练模型的摘要抽取

基于预训练模型的抽取式摘要方法通过在大规模语料上进行自监督或无监督预训练，学习语言的深层结构与语义知识。此类方法利用模型的迁移学习能力，通过微调适应特定任务，提升摘要质量。

Liu[7]提出了 BERTSUM 模型，首次将 BERT 应用于抽取式摘要任务。该模型在 BERT 的基础上添加 Fine-Tuning 分类层，对句子进行打分，预测其是否应纳入摘要。BERTSUM 使用句子级嵌入表示文档内容，并通过位置编码与分段编码捕捉句子间的逻辑关系。该模型在 CNN/DailyMail 数据集上成为抽取式摘要的基准模型，其 ROUGE 指标显著优于传统方法。Zhong 等[16]提出了 MatchSum 模型，针对 BERTSUM 的逐句打分机制进行了改进。该模型将抽取式摘要任务转化为文本匹配问题，提出了一个创新的摘要级框架。如图 3 所示，该框架首先构建一个候选摘要集合，然后利用 Siamese-BERT 架构计算每个候选摘要与源文档的语义相似度。从而筛选出与源文档最为匹配的摘要。相比 BERTSUM，MatchSum 不再局限于对单个句子进行独立打分，而是考虑了句子间的关联，更加关注整体语义的一致性。在 CNN/DailyMail 数据集上 Rouge-1 达到了 44.41，取得了优异的结果。但是，由于直接比较候选摘要与源文档的语义相似度，该方法容易倾向于选择包含更多句子的摘要，从而产生冗长且冗余的信息。对此，Gong 等[26]提出 SeburSum 模型，采用集合级排序策略。即将候选摘要视为一个句子集合，利用候选摘要之间的语义相似度进行排序并选择摘要，有效避免了生成冗长摘要的问题。该方法在 CNN/Daily Mail、XSum 数据集上都优于强基线方法。

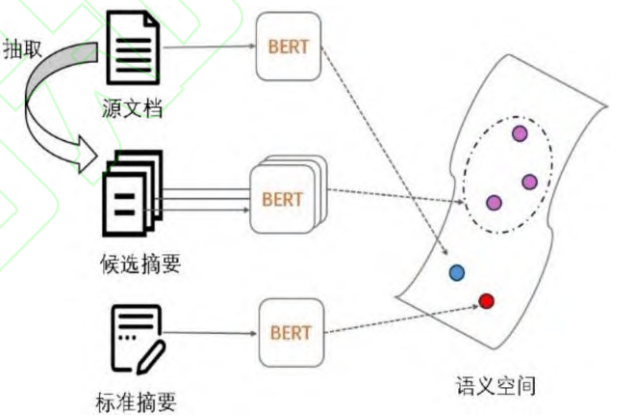

图 3 MatchSum 的框架图

Fig.3 Model framework of MatchSum

然而，以上方法在处理长文档时仍面临输入长度限制的问题。对此，Beltagy 等[27]提出 Longformer 模型。该模型通过扩展自注意力机制的感受野，使模型能够捕捉长距离依赖，从而有效处理长文档。此外，DISCOBERT 模型[24]在 BERT 的基础上通过引入语篇单元和图编码器，进一步提升了长文档摘要的效果。该模型在语篇级别选择摘要单元，克服了 BERT 对长文档的处理局限。通过结合 BERT 提取的表征和图编码器捕捉的长距离依赖及跨段指代链，DISCOBERT 模型能够生成更具逻辑性和连贯性的摘要。

在预训练模型的基础上，Cheng 等[28]提出了一种新的抽取式摘要方法—集合预测网络（SetSum）。该方法采用编码器-解码器架构，其中编码器使用 BERT，

解码器则采用非自回归解码器以提高解码效率并解决句子依赖性问题。与传统的自回归解码器相比，非自回归解码器允许模型在生成每个摘要句子时，不再依赖于前一个句子的输出，从而实现了并行化生成，显著提高了摘要的生成速度。SetSum 在单文档和多文档数据集上均取得了良好的效果。此外，有研究将预训练嵌入作为静态特征，结合传统神经网络和注意力机制进行句子评分。Gangundi 等[29]提出 RBCA-ETS 模型，该模型先用 RoBERTa 编码，再并行多通道 CNN 与双层 BiLSTM 提取特征，最终经自注意力融合进行评分，在 CNN/DailyMail 和 DUC2002 数据集上均超越多项 SOTA。

随着 LLM 的发展，Zhang 等[30]深入研究了 ChatGPT 在抽取式摘要任务中的应用。还探讨了上下文学习和链式思考推理对提升性能的有效性。他们发现，虽然 ChatGPT 在 ROUGE 评分上略逊于传统微调方法，但在基于 LLM 的评估指标 G-EVAL[31]上表现更佳。此外，通过结合抽取后再生成的框架，ChatGPT 可以显著提高摘要的忠实度，减少生成式摘要中出现的幻觉问题。这些发现为增强预训练模型在自动文本摘要领域的能力提供了新的方向。

基于预训练模型的抽取式摘要方法增强了对语言和上下文关系的建模能力，但仍面临长文档处理的效率问题。未来研究可探索模型轻量化和与 GNN 结合等方向，以提升摘要质量和计算效率。

2.1.4 基于强化学习的抽取式摘要方法

在抽取式摘要中，强化学习（Reinforcement Learning, RL）将句子选择转化为序列决策任务。这一方法通过逐步选择候选句子，以最大化特定奖励函数（如 ROUGE 分数），不仅考虑句子的局部重要性，还能够在整体选择上进行优化，从而提升摘要的连贯性、信息覆盖度并减少冗余。Narayan 等[17]提出了一种基于强化学习的训练算法 REFRESH。该算法通过全局优化 ROUGE 评估指标来优化抽取式文本摘要并首次将强化学习中的策略梯度方法应用于句子排序问题。如图 4 所示，该模型架构含句子编码器、文档编码器和句子抽取器。句子编码器用 CNN 将句子转为连续表示，文档编码器用带 LSTM 单元的 RNN 处理文档序列，句子抽取器利用强化学习框架训练，依据文档表示和已标记句子做二元预测，选取最高分句子组成摘要，提升了句子判别与选择能力。在 CNN/DailyMail 数据集上的实验表明，REFRESH 优于现有的抽取式和生成式系统。

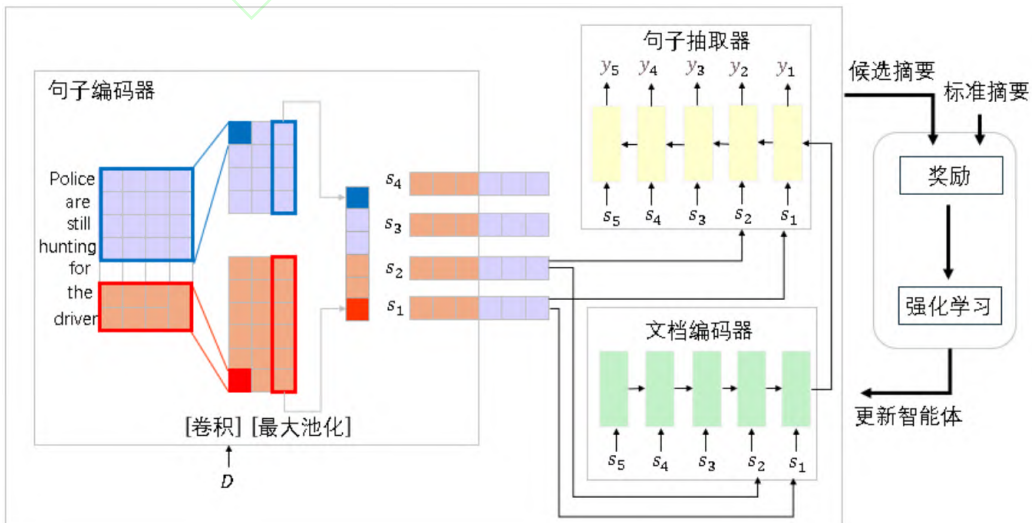

图 4 REFRESH 的模型框架图

Fig.4 Model framework diagram of REFRESH

Chen 等[32]提出了一种结合抽取式和生成式的摘要生成模型，该模型首先通过监督学习从候选句子中

提取显著性高的句子，然后利用策略梯度方法对句子选择进行微调，以最大化 ROUGE 分数作为奖励。最后，生成模块对选择的句子进行重写。这种方法在提升摘要的流畅性及可读性方面取得了显著进展。然而，这些方法面临着一个共同的问题，即 ROUGE 的局限性。ROUGE 主要测量文本之间的表面相似性，无法充分捕捉语义的深度。基于此，研究者们引入 BERTScore[33]、GPTscore[34]、G-EVAL[31]等新的评价指标，旨在提升摘要评价的效果。

近年来，GoSum 模型[18]通过将强化学习与 GNN，为长文档摘要任务提供了一种新颖的解决思路。虽然 GoSum 依然使用 ROUGE 分数作为奖励函数，但它通过引入 GNN 来构建异构图，从而增强了对长文档中句子间复杂关系的建模能力。GoSum 采用策略梯度算法优化句子选择策略，并设计了一个停止机制，以确保在适当时刻终止句子抽取，从而生成简洁且信息丰富的摘要。实验结果表明，GoSum 在科学论文数据集 PubMed 和 arXiv 上达到了最先进的性能，进一步验证了强化学习在抽取式摘要中的应用潜力。然而，这一复杂模型的计算成本和训练效率问题仍然是需要解决的重要挑战。

尽管基于 DL 的抽取式摘要在新闻、法律等结构清晰的文本中取得了显著进展，但其局限性依然明显。由于只能从原文中提取现有句子，抽取式方法在语言自然性与灵活性方面不如生成式摘要。同时，长文档摘要仍然是挑战。尽管图神经网络和预训练模型缓解了部分问题，但仍需进一步探索新方法。

2.2 基于深度学习的生成式文本摘要

生成式摘要模型通过生成新句子概括原文信息，具有较高的表达灵活性和信息重组能力。然而，生成式摘要在实际应用中面临诸多挑战，包括长距离依赖处理不足、OOV 识别错误、信息冗余与遗漏，以及生成事实性错误等问题。因此，研究者逐步引入了多种深度学习框架，以提升摘要生成质量。如序列到序列模型、注意力机制、复制机制、强化学习以及预训练语言模型，这些框架在一定程度上解决了生成式摘要中的关键问题，并推动了 ATS 的发展。生成式摘要的研究始于 Rush 等[8]提出的编码器-解码器架构。该模型通过编码器将输入文本转换为隐藏向量，再由解码器生成摘要。然而，由于该模型无法有效处理长文本中的语义依赖，其在处理复杂文本生成任务时表现不佳。这一局限性推动了后续模型对注意力机制和长序列处理方法的探索，为生成式摘要领域奠定了基础。表 2 列出了目前基于 DL 的生成式摘要方法的对比分析。

表 2 基于 DL 的生成式摘要方法对比

Table 2 Comparison of DL-Based Abstractive Summarization Methods

方法类别	描述	经典模型	优点	缺点
序列到序列模型	将输入文本编码为固定长度的上下文向量，再解码为简洁的摘要	BRIO[35] E2S2[36]	生成结果灵活多样，能够适应不同长度和结构的输入文本	对长距离依赖处理不足，容易忽略关键信息，生成冗余内容
注意力机制	通过动态聚焦输入文本的不同部分，逐步生成摘要	Transformer[37]	提高生成文本的一致性和连贯性，能够捕捉全局与局部关系	计算复杂度高，需优化生成一致性
复制机制	从源文本中直接选择并复制关键字或短语，以增强摘要的准确性	CopyNet[38] PGN[39] SAGCopy[40]	解决 OOV 问题， 减少冗余信息	RNN 架构在处理长文本时效率较低，复杂依赖关系处理困难
强化学习	通过奖励信号评价摘要质量，逐步优化生成策略以提高摘要效果	MDO[41] IRL[42]	改善摘要连贯性和语义一致性，活优化多项性能指标	计算成本高，奖励函数设计复杂
预训练语言模型	微调学习文本的语义和结构，从而生成连贯且相关的摘要	SumBART[43] PEGASUS[12]	深入掌握语言结构和语义信息，长距离依赖捕捉能力强	需要大量微调数据支持且计算成本较高

2.2.1 基于序列到序列模型的方法

序列到序列模型最早是用于机器翻译领域，随后广泛应用于 ATS 任务。它主要依赖编码器-解码器结构，将输入文本编码为隐藏处理，再逐步生成摘要内容。目前，大部分研究者都致力于优化编码和解码过

程，以提升生成摘要性能。

Cohan 等[44]提出了一个层次化的 Seq2Seq 模型，该模型对文档的论述章节分层编码，使模型能够同时捕获全局信息和局部细节，从而在摘要生成上取得了显著成效。在此基础上，为了解决传统序列到序列模型存在的训练和下游实际任务不一致的曝光偏差问题，Liu 等[35]提出了 BRIO 方法。该方法结合最大似然估计（Maximum Likelihood Estimation,MLE）与对比学习，通过引入对比损失来微调预训练的摘要模型，鼓励模型为质量更高的候选摘要分配更高的概率。如图 5 所示，在传统 Seq2Seq 模型训练中，MLE 损失函数因假设只有唯一正确输出序列而限制了模型学习输出多样性的能力。相比之下，无参考对比损失函数不依赖参考输出，能更好地处理多种合理答案输出任务，捕捉输出多样性。在 CNN/DailyMail、XSum 等数据集上的实验结果表明，BRIO 框架都取得了 SOTA 结果，显著提升了自动文本摘要的性能。

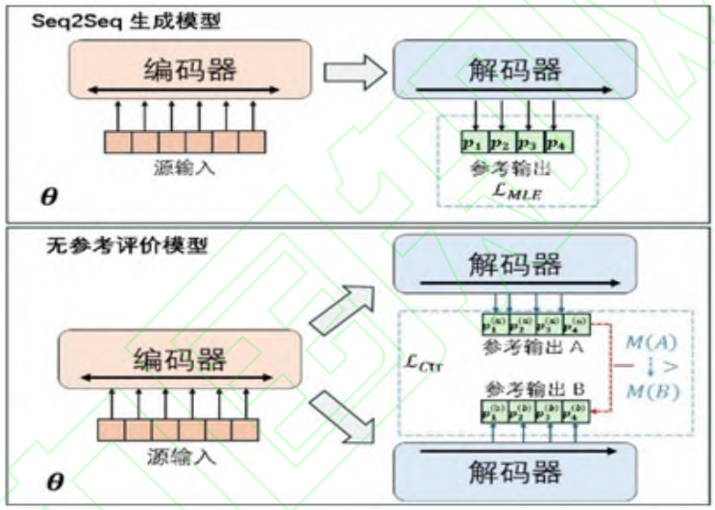

图 5 BRIO 的模型框架图

Fig.5 Model framework diagram of BRIO

然而，BRIO 的固定边距对比损失难以区分不同候选摘要的质量差异，且未优化训练数据筛选策略。为此，Sun 等[45]提出数据选择课程（DSC）方法，通过动态调整边距的对比损失函数，结合高斯分布筛选中等难度样本及课程学习策略，显著提升模型效率。这些改进有效地应对了信息冗余和长文档处理的挑战。但传统的 Seq2Seq 模型在处理长距离依赖和长文档摘要时仍存在不足，常常忽略关键信息或产生冗余内容。为此，Zhong 等[36]提出了一种编码增强的 Seq2Seq 预训练策略 E2S2，该策略通过整合更多自监督信息，增强了模型的条件生成能力。实验结果表明，E2S2 在机器翻译、自动文本摘要和语法错误修正等多个自然语言处理任务中，显著提升了 BART 等模型的性能。这一研究为优化 Seq2Seq 模型的编码过程提供了新思路，助力解决长文档摘要中的相关挑战。

2.2.2 基于注意力机制的方法

注意力机制最早在 2015 年应用于机器翻译任务，以解决传统 RNN 在处理长文档生成中的局限性[37]。注意力机制作为 Seq2Seq 框架的关键扩展技术，在生成式摘要中，该机制让模型在每一步生成时聚焦于最相关的输入内容，从而极大地改善了生成文本的一致性和连贯性。Transformer 的引入使得大部分研究都围绕自注意力来生成摘要。Kumar 等[46]结合自注意力机制 Transformer 架构，解决了文本中的共指问题，这使得模型能更好地理解文本信息。为了更好地处理长文档中的信息，研究者们还提出了结构化注意力机制。Chowdhury 等[47]在多文档摘要任务中采用了层次化注意力机制，将注意力焦点限制在文档的不同结构单元

上，从而更精细地捕捉全局与局部的信息，解决了生成过程中上下文不连贯的问题。在中文文本生成任务中，Zhao 等[48]提出了 CNsum 模型，采用 BERT 编码器与多层 GPT-2 解码器构建完整的 Transformer 架构，充分利用多层自注意力机制进行上下文建模与摘要生成。该模型在 NLPCC2017、LCSTS 中文数据集上取得了优异的 ROUGE 成绩，验证了注意力机制在中文生成式摘要中的广泛适用性与有效性。此外，Dilawari 等[49]提出了一个神经注意力模型，该模型不仅利用了注意力机制来捕捉输入文本中的重要信息，还引入了句子位置、词性标签、命名实体标签和词权重等语言学特征，解决了生成摘要事实不一致和 OOV 的问题，进一步提升了摘要生成的质量。

除了文本生成，注意力机制逐渐被应用于结合不同模态的信息生成任务中，许多研究进一步探索了它在多模态生成任务中的潜力。Argade 等[50]提出了一种基于 BERT 与注意力机制结合的多模态摘要模型。该模型通过使用文本和视觉特征来生成综合摘要，采用自注意力机制有效地融合来自文本和图像的信息，使生成的摘要更加全面。Minaidi 等[51]进一步将自注意力机制与生成对抗网络结合，提出了一种无监督的视频摘要生成方法。这种无监督方法特别适合无标注数据的视频处理任务。

在跨语言生成摘要领域，注意力机制也显示出了强大的适应性和灵活性。Yang 等[52]提出了一种基于注意力机制的生成模型，旨在通过注意力机制在源语言与目标语言之间建立紧密联系，有效减少了翻译过程中信息的丢失，还保证了不同语言之间生成内容的无缝转换。尤其在多语言语料有限的环境下，这种方法表现很出色。

尽管上述这些技术在处理长距离依赖、复杂多样的数据以及跨语言生成方面取得了显著进展，但在计算效率和生成一致性上仍存在改进空间。

2.2.3 基于复制机制的方法

复制机制（Copy Mechanism）通过允许模型在生成过程中从原文直接复制关键内容，有效解决了生成式摘要任务中的 OOV 和生成重复问题。同时提高了摘要的准确性和连贯性。这一机制发展迅速，并在多个框架中得到了优化。

Gu 等[38]提出的 CopyNet 模型通过在生成过程中实现生成模式与复制模式的动态切换，有效解决了 OOV 的问题。该模型会判断当前生成的词是否在输入文本中出现，如果是，它会直接复制原文中的词，反之，则切换回生成模式，确保了生成摘要的准确性和连贯性。然而，由于其依赖于 RNN 架构，在处理长文本和复杂依赖关系时仍面临效率问题。See 等[39]进一步提出了 Pointer-Generator 模型，该模型将复制机制与指针机制结合，使模型能够直接从输入文档中复制词汇或短语。此外，模型还引入了覆盖机制（Coverage Mechanism），通过跟踪每个词的注意力分布累积情况来避免重复生成。这一方法在 CNN/DailyMail 等数据集上表现出色，但在处理长文档时仍存在一定的冗余控制挑战。Xu 等[40]提出了 SAGCopy 摘要模型，他们将复制机制与 Transformer 架构结合，通过自注意力层构建的有向图指导复制过程。该模型通过中心性分析选取重要词汇，显著提升了摘要的准确性和信息覆盖度且解决了传统 RNN 的局限。这一方法在 CNN/DailyMail 和 Gigaword 数据集上表现优越，展示了复制机制在复杂文本生成中的潜力。此外，Li 等[53]提出了一种改进的 BIO 复制机制，针对传统复制方法在生成中文摘要时存在的片段不连续和主题偏离问题，通过为生成的每个词预测 BIO 标签（B：开始，I：内部，O：其他），强化模型对连续性片段的学习。该方法在 LCSTS 中文数据集上显著提升了摘要的连贯性和主题一致性，验证了精细化复制策略在实际应用中的有效性。

基于复制机制的方法通过在生成与复制间灵活切换，有效提升了摘要的准确性与信息覆盖度，并减

少了 OOV 问题的影响。在保证生成流畅性的同时，这些方法减少了冗余信息，提高了摘要的整体质量。

2.2.4 基于强化学习的方法

强化学习在生成式文本摘要中的应用大致分为两类，一类专注于策略优化，通过不断尝试找到最优解。另一类结合语义分析与迁移学习以使模型更好地理解上下文并适应不同任务需求。

Paulus[54]提出的模型通过 MLE 与策略梯度结合，形成了一个混合训练目标。该模型先使用 MLE 进行预训练，帮助生成器掌握基础的生成能力。随后，通过策略梯度直接优化 ROUGE 分数，从而减轻了训练和测试之间的“曝光偏差”。然而，ROUGE 主要衡量词汇重叠，难以捕捉深层的语义一致性。

为了克服 ROUGE 分数的局限性，研究者们开始探索多维度的优化方法。Ryu 等[41]提出了 MDOmin 和 MDOpro 两种多维优化策略，通过强化学习同时优化摘要的一致性、连贯性、相关性和流畅性。MDOmin 侧重于提升当前最低的维度分数，而 MDOpro 则旨在同时优化多个维度，实现类似多任务学习的效果。这种方法不仅提升了摘要的整体质量，还为后续研究提供了新的思路和方向。此外，Fu 等[42]引入了逆强化学习（Inverse Reinforcement Learning, IRL），将其作为一种新的策略优化方法应用于 ATS 任务。IRL 通过观察专家示范（人类生成的摘要）来学习奖励函数，从而能够灵活地优化模型的多项性能指标，如覆盖度、忠实度等。

除了上述方法，Fikri 等[55]针对 ROUGE 分数在捕捉语义一致性方面的不足，进一步提出了基于语义相似度奖励的强化学习框架。他们利用 BERT 等预训练模型，计算生成摘要与原文的语义相似度，将其作为新的奖励信号。这种方法确保生成的摘要不仅内容一致，而且逻辑上符合人类的阅读习惯。Keneshloo 等[56]将迁移学习与强化学习结合，通过共享参数实现了不同任务间的知识迁移。这种方法让模型能够适应不同领域的摘要任务，尤其在小数据集的情况下，避免了过拟合。

基于强化学习的方法通过策略优化和语义相似度奖励等技术，显著提升了生成摘要的连贯性、信息覆盖度和语义一致性。然而，高计算成本和奖励函数设计的复杂性依然是其广泛应用的主要挑战。

2.2.5 基于预训练语言模型的方法

预训练语言模型，如 BERT[57]、GPT[58][59]、BART[60]、T5[61]在大规模语料上的预训练能掌握语言的深层结构和语义信息，并通过微调适应不同的生成任务。相比传统的 CNN 或 RNN，BERT 和其他基于 Transformer 的模型解决了梯度消失和爆炸的问题，使其在捕捉长距离依赖方面更具优势。GPT 系列模型在语言生成方面表现突出，采用单向生成模式捕捉上下文关系。BART 在此基础上进一步优化，通过降噪自编码器预训练，结合了 BERT 和 GPT 展现了更优越的性能。

在生成式摘要领域，一个常见的问题是生成的摘要与原文事实存在不一致，为了解决这个问题，研究者们积极探索并提出了多种方法。Chen 等[62]通过整合知识图谱数据和结构化语义，提升了生成模型的事实准确性，克服了 LLM 在理解复杂语义时的不足。Xu 等[63]提出了一种创新的方法，该方法通过任务分类来引导 LLM 进行推理，从而显著提高了识别摘要中事实不一致性的能力。这种方法不仅适用于零样本推理场景，而且在监督范式下也取得了显著成效，为生成式摘要的事实准确性提供了新的解决方案。此外，Vivek 等[43]提出了 SumBART，对 BART 进行了改进，以减少 BART 模型在生成摘要时可能出现的事实错误和不一致问题。Balachandran 等[64]提出 FACTEDIT 的方法，通过译后编辑和语言模型填充来纠正事实错误。该方法首先利用语言模型填充技术生成包含事实性错误的合成摘要数据集，然后使用一个序列到序列的模型对这些错误摘要进行修正。FACTEDIT 在

CNN/DailyMail 和 XSum 数据集上取得了显著的效果，提高了摘要的事实准确性。

在多语言及小语种任务领域，T5 模型通过其统一文本输入输出设计，展现了卓越的多任务处理能力。其多语言扩展版本 mT5 更是覆盖了多达 101 种语言，在零样本场景效果非常显著。但数据稀缺对 mT5 这样的强大模型同样构成挑战。PEGASUS 模型[12]采用自监督学习策略，通过从原始文本中抽取句子生成训练数据，成功克服了摘要数据集匮乏的挑战。这一方法的思路也可用于低资源语言摘要任务，通过自监督学习提升模型在低资源语言上的表现。尽管如此，处理小语种任务时，这些模型仍面临数据稀缺导致的生成不稳定等挑战。为此，Reda 等[65]专为阿拉伯文设计了一种混合摘要系统，首先利用 AraBERT 提取关键句子，然后将这些句子作为输入，利用 mT5 生成高质量的摘要。这种方法显著提高了摘要的质量，展现了在小语种任务上的有效改进。

在此基础上，Wang 等[66]将多语言摘要（MLS）和跨语言摘要（CLS）统一为一个多对多摘要（M2MS）框架，并引入了 PISCES 模型。PISCES 模型通过三阶段的预训练学习语言建模、跨语言能力和摘要能力，提高了模型在处理不同语言对摘要任务时的性能。实验结果表明，PISCES 模型在跨语言摘要任务上取得了显著性能提升，同时在单语言摘要任务上也保持了竞争力，为自动文本摘要技术的发展提供了新的思路和方法。

预训练语言模型在生成式文本摘要任务中展现了强大的能力，但其复杂性和高昂的计算成本限制了实际应用。为此，Jiang 等[67]提出 TriSum 框架，从 GPT-3.5 生成的结构化摘要示例中提取知识，并基于课程学习策略训练轻量化模型，实现对 LLM 摘要能力的有效迁移。该框架在多个摘要数据集上均取得优于基线的效果。未来研究可进一步探索降低计算成本的策略，并提升模型在低资源环境中的适应性。

3 自动文本摘要数据集

3.1 常用自动文本摘要数据集

在 ATS 任务中，数据集的选择对于模型的训练和评估至关重要。为了全面了解和选择适合的数据集，本文总结了目前常用的自动文本摘要数据集，并从数据的规模、语言、摘要方式、文档形式以及适用范围等多个角度进行了分类和汇总。表 3 为目前公开的主流中英文文本摘要数据集。

3.1.1 英文数据集

CNN/DailyMail[68]数据集包含来自 CNN 和 Daily Mail 的新闻文章及其对应的人类撰写摘要。该数据集覆盖多个主题，如政治、经济和娱乐，被广泛用于新闻摘要任务，且支持单文档和多文档摘要。由于其规模大，含有数万篇文章，因此非常适合深度学习模型的训练和评估。

Gigaword 主要用于短文本摘要任务。数据规模超过 400 万，可满足深度神经网络训练的需求。由于其简短、丰富的句子结构，Gigaword 在标题生成和简短摘要生成领域中扮演了重要角色。

DUC/TAC 是由美国国家标准技术研究院(NIST)提供的 DUC 和 TAC 数据集，这些数据集多用于跨文档信息的提取和整合，广泛涉及新闻、政策、科技等领域的研究。DUC 数据集自 2001 年至 2007 年在 DUC 会议期间发布，包含三种类型的摘要，手动创建的摘要、自动生成的基准摘要，以及来自挑战赛参与者生成的摘要。TAC 数据集则自 2008 年起接管了 DUC 摘要任务。这些数据集在抽取式摘要系统领域被广泛使用。

Xsum[69]是一个专为生成极简摘要设计的数据集。该数据集包含 BBC 新闻中的单句摘要，特别强调信息的浓缩性，因此在极端压缩摘要任务中表现出色。

3.1.2 中文数据集

LCSTS[70](Large-scale Chinese Short Text Summarization Dataset)由哈尔滨工业大学团队创建，是专为中文生成式摘要任务设计的数据集。该数据集的主要内容为从新浪微博等社交媒体平台收集的短文本及其对应的摘要。尽管 LCSTS 数据规模大，但其文本相对较小，因此，非常适合训练短文本摘要的神经网络模型。可应用于新闻标题生成、社交媒体分析等任务。

NLPCC2017 数据集是中文 ATS 任务中的重要资源，专为支持生成式摘要研究而设计。该数据集包含了数万篇新闻文本及其对应的摘要，数据来源广泛，适用于各类中文文本处理任务。

表 3 主流 ATS 数据集

Table 3 Mainstream ATS Datasets

数据集名称	语言	数据规模	适用摘要方式	单文档/多文档	适用性	获取网址
CNN/DailyMail	英文	大	抽取式摘要/生成式摘要	单文档	深度神经网络方法	https://github.com/abisee/cnn-dailymail
Gigaword	英文	大	生成式摘要	单文档	深度神经网络方法	https://catalog.ldc.upenn.edu/LDC2012T21
DUC/TAC	英文	中	抽取式摘要/生成式摘要	单文档/多文档	深度神经网络方法	http://www-nlpir.nist.gov/projects/duc/data.html http://tac.nist.gov/data/
Xsum	英文	大	生成式摘要	单文档	深度神经网络方法	https://github.com/EdinburghNLP/XSum
LCSTS	中文	大	生成式摘要	单文档	深度神经网络方法	http://icrc.hitsz.edu.cn/Article/show/139.html
NLPCC 2017	中文	小	生成式摘要	单文档	深度神经网络方法	http://tcci.ccf.org.cn/conference/2017/

3.2 国内低资源语言摘要数据集

尽管在高资源语言中，ATS 领域已有许多公开的数据集，但针对国内低资源语言的自动文本摘要数据集仍然相对匮乏。为了解决这一问题，研究者们开始致力于构建低资源语言的摘要数据集。这些数据集在国内低资源语言和跨语言摘要研究中，发挥了至关重要的作用。

表 4 为一些国内低资源语言自动文本摘要数据集，涵盖了数据集的规模大小、语言、适用范围以及获取方式。

表 4 国内低资源语言 ATS 数据集

Table 4 Domestic Low-resource Language ATS Datasets

数据集名称	语言	数据规模	适用范围	获取地址
Ti-SUM	藏文	小	单文档	http://www.doi.org/10.11922/sciencedb.j00001.00352
TiCLS	中-藏	小	单文档	https://doi.org/10.57760/sciencedb.15452
MMDS	中-蒙-藏-维	小	多文档	https://doi.org/10.57760/sciencedb.j00001.01084

闫晓东等[71]自建了小型藏文多文本摘要数据集 Ti-SUM，由 1000 篇真实藏文新闻组成。翁彧等[72]构建了面向多文档摘要生成的数据集 MMDS，该数据集包含中文、蒙古文、藏文和维文多文档摘要版本。每个版本包含 1044 个新闻簇（6234 篇新闻文章），每个新闻簇均配有精准的人工摘要。此外，欧阳新鹏等[73]针对跨语言需求，构建了藏汉跨语言摘要数据集，该数据集包含了 2000 条高质量样本，每条数据由藏文原文和中文目标语言摘要组成。这些数据集对于推动低资源语言的信息处理研究具有重要价值。

基于 DL 的 ATS 技术需要大量高质量的数据集作为支撑，这使得国内低资源语言 ATS 研究面临诸多

挑战。无论是中文还是其他低资源语言，在构建自动文本摘要数据集方面仍存在较大困难。尤其是在低资源语言的研究中，由于缺乏足够的高质量数据集，低资源语言的研究始终难以取得实质性进展，目前仍处于起步阶段。

4 自动文本摘要的评价指标

在 ATS 任务中，评价指标是衡量生成摘要质量的关键工具。这些指标不仅帮助我们判断自动生成摘要的质量，还确保其与人工撰写的理想摘要之间的一致性和准确性。通过各种评价方法，我们可以量化和分析摘要的准确性、覆盖度和流畅度。自动文本摘要评价指标分为两种，自动评价和人工评价。

4.1 自动评价

常用自动评价方法有 ROUGE、BLEU 和 METEOR 等。与人工评价相比，自动评价通过计算指标和算法来衡量摘要的质量，无需人类干预。这些自动评价指标利用语言特征、统计分析或机器学习算法，能够快速且高效地处理大量数据。然而，自动评价方法无法完全捕捉到优秀摘要的细微差别，且其结果有时与人类的判断不一致。通常，这些指标侧重于表面特征，而缺乏对深层语义理解的考量。在表 5 中，我们梳理了一些常用自动评价指标的应用场景以及优缺点。

表 5 常用自动评价指标对比

Table 5 Comparison of Common Automatic Evaluation Metrics

评价指标	定义	应用场景	优势	缺点
ROUGE[74]	通过比较系统生成摘要与参考摘要之间的 n-gram 重叠来评估摘要质量	广泛应用于抽取式文本摘要评价	计算方式相对简单，易于实现	忽视文本语义层面
BLEU[75]	通过比较系统生成文本与参考文本的 n-gram 重叠来评估结果，注重“精确度”	最初广泛用于机器翻译评测，也常用于自动文本摘要	基于精确度衡量，计算方法成熟，使用广泛	缺乏对释义或同义词的灵活处理
BERTScore[33]	基于 BERT 词向量，从语义层面衡量系统生成摘要与参考摘要之间的相似度	适用于抽取式和生成式文本摘要评价，可更好地捕捉上下文语义关联	能更好地理解和对比上下文语义，对词序和同义替换等语义变化更具鲁棒性	对硬件和预训练模型依赖较高，计算成本较大
G-Eval[31]	模拟人工评价方式，对摘要的流畅度、信息量、连贯性等多维度指标进行综合评估	适用于需要综合考量语言质量、信息完整度和可读性的场景	人性化、综合性强	评价成本高、结果带有主观性

4.1.1 ROUGE

ATS 通常使用 ROUGE[74]（Recall-Oriented Understudy for Gisting Evaluation）作为评价手段，它主要用于衡量生成摘要与参考摘要之间的重叠情况。ROUGE 有多个子标准，如 ROUGE-N、ROUGE-L 和 ROUGE-W。ROUGE-N 指标通过比较生成摘要和参考摘要中相同的 n-gram 片段的数量来评价摘要的覆盖度。常见的 ROUGE-N 变体包括 ROUGE-1 和 ROUGE-2，分别用来衡量单词级和二元组级的重叠情况。计算公式如下：

$$\mathrm{ROUGE-N}=\frac{\sum_{S\in R}\sum_{gram_n\in S}C_{match}(gram_n)}{\sum_{S\in R}\sum_{gram_n\in S}C(gram_n)} \tag{1}$$

其中，R 表示参考摘要 Reference Summaries，n 代表 n-gram 的长度，$C_{match}(gram_n)$表示系统摘要与参考摘要中共同出现的 n-gram 的个数。尽管 ROUGE-N 简单而有效，但它没有考虑到词序或语义相似，只考虑了表面特征，忽略了深层语义理解的考量。

ROUGE-L 和 ROUGE-W 分别计算重叠的最长公共子序列和加权最长公共子序列。ROUGE-W 采用 ROUGE-L 的方法，但根据单词在摘要中的位置为每

个单词赋予权重。摘要开头或结尾的单词权重较高，而中间的单词权重较低。尽管 ROUGE 系列指标在实际应用中表现出色，但其对文本深层语义的捕捉仍然存在局限性。

4.1.2 BLEU 和 METEOR

BLEU[75]（Bilingual Evaluation Understudy）也被应用于摘要评价，尽管它主要用于机器翻译，其 n-gram 精确度和流畅度指标在摘要任务中也具有一定的参考价值。近年来，BERTScore[33]方法逐渐受到关注，这种方法不仅看重词汇的相似性，还考虑了上下文和位置的差异。通过对候选文本和参考文本进行比较，BERTScore 会给出一个从 0 到 1 的评分，得分为 1 表示候选文本与参考文本完全一致。

此外，METEOR[76]（Metric for Evaluation of Translation with Explicit ORdering）也是基于 n-gram 的评价指标。METEOR 弥补了 BLEU 的部分局限性，加入了词义相似度、词干匹配和同义词匹配等机制，能够更好地评价文本的语义一致性。相较于 BLEU，METEOR 能够捕捉到更细粒度的语义相似度，尤其在多样化的摘要任务中表现出色。随着 LLM 在 ATS 领域的广泛应用，传统评价方法逐渐无法有效评价生成摘要的语义深度和内容一致性。为解决这一问题，Fu 等提出了 GPTscore[34]，利用预训练 LLM 对生成摘要进行评分，克服了传统方法的局限性，提供了更精准的语义匹配和质量评价。然而，近期 Shen 等[77]的研究表明，虽然 ChatGPT 和 GPT-4 等 LLM 在摘要评价方面展现出了一定的潜力，其评价能力仍未达到人类水平。实验结果表明，LLM 评价器在区分性能相近的候选摘要以及处理高质量候选摘要时仍存在不足。未来有望通过进一步优化 LLM 技术，提升其在自动文本摘要质量评价中的精度与鲁棒性。

4.2 人工评价

尽管自动化评价指标为摘要任务提供了快速的定量评价手段，但它们在评价文本的可读性、流畅性和语义一致性方面仍存在局限。因此，人工评价仍是 ATS 研究中的关键环节。人工评价一般通过专家或非专家打分的方式，评价以下几个方面：

4.2.1 信息覆盖度

摘要应涵盖原文的核心信息，评审者根据摘要保留关键信息的程度评分，高覆盖度能准确传达原文主旨。

4.2.2 可读性与语言流畅性

生成文本的流畅性和语法正确性是人工评价的重点，确保摘要易于理解，自动化指标通常无法评价这一点。除可读性外，还需考察文本是否自然连贯，风格一致，确保符合人类表达习惯。

4.2.3 简洁性与语义一致性

摘要在传达足够信息的同时应避免冗余和重复，评分标准基于摘要的简明扼要程度。确保生成摘要与原文在语义上高度一致，不产生歧义或误导。

由于人工评价成本高且耗时，通常在小规模数据集上进行，但它能有效弥补自动化评价的不足，为质量评价提供更细致的视角。在高质量摘要任务中，常结合自动化评价指标与人工评价的优势，以确保生成结果的全面性与可靠性。

4.3 主流方法实验对比与分析

4.3.1 实验结果对比

为明确不同自动文本摘要技术的实际效果与潜在问题，我们对第二章所述的主流方法进行了多维度实验对比。目前，主流的 ATS 模型通常采用 CNN/DailyMail 数据集作为主要评价基准。评价指标方面，绝大多数工作采用 ROUGE 指标体系，其中 R-1、R-2 和 R-L 分别衡量单词、二元组和最长公共子序列的重合程度，作为衡量摘要质量的重要依据。针对 ROUGE 指标无法考虑语义层面信息的问题，我们还采用基于 GPT-4 的 G-Eval 评分综合衡量摘要质量。表 6 展示了各方法在 CNN/DailyMail 数据集上的综合评分结果。为了全面考察各模型在不同文本风格和长度上的泛化

能力，在表 7 中，我们进一步对比了部分模型在 XSum、DUC2002 和 LCSTS 等数据集上的表现。通过多数据集、多指标的对比分析，可以更全面地揭示各模型的优缺点，为后续深入研究提供参考。

表 6 ATS 主流方法实验结果对比

Table 6 Comparison of Experimental Results of Mainstream ATS Methods

摘要方式	技术策略	主流方法	年份	CNN/DailyMail 实验结果			
				R-1（%）	R-2（%）	R-L（%）	G-Eval（1-5）
抽取式	序列标注	SummaRuNNer[5]	2017	39.6	16.2	35.3	-
		NEUSUM[6]	2018	41.59	19.01	37.98	-
		DeepSumm[13]	2023	43.3	19.0	38.9	-
		OrderSum[20]	2025	44.44	21.31	30.5	-
	图神经网络	HSG+Tri-Blocking[14]	2020	42.95	19.76	39.23	-
		Multi-GraS[22]	2021	43.16	20.14	39.49	-
		MuchSum[15]	2022	43.85	20.93	40.72	-
		MCHES[25]	2024	44.76	24.97	42.48	-
	预训练语言模型	BERTSUM[7]	2019	43.25	20.24	39.63	-
		MatchSum[16]	2020	44.41	20.86	40.55	3.28
		SeburSum[26]	2023	**45.49**	**22.36**	**41.67**	-
		ChatGPT[30]	2023	39.25	17.09	25.64	3.24
		RBCA-ETS[29]	2024	43.45	19.95	39.45	-
	强化学习	REFRESH[17]	2018	40.00	18.20	36.60	-
		rnn-ext +RL[32]	2018	41.47	18.72	37.76	-
生成式	序列到序列模型	BRIO[35]	2022	47.78	23.55	44.57	3.25
		DSC[45]	2023	**48.62**	**24.14**	**45.31**	-
		E2S2[36]	2023	44.18	21.32	40.91	-
	注意力机制	Two-stage network[52]	2023	37.76	17.81	33.83	-
	复制机制	PGN+coverage[39]	2017	39.53	17.28	36.38	-
		SAGCopy-Outdegree[40]	2020	42.53	19.92	39.44	-
		BIOCopy[53]	2024	43.97	37.22	41.55	-
	强化学习	BART-large+IRL[42]	2022	46.12	21.98	43.15	-
	预训练语言模型	PEGASUS[12]	2020	44.17	21.47	41.11	-
		SumBART[43]	2022	37.31	-	34.98	-
		FACTEDIT+FactCCFilter (FF)[64]	2022	42.53	20.48	39.74	-
		ChatGPT[30]	2023	38.48	14.46	28.39	**3.46**
		TriSum[67]	2024	46.7	23.5	40.7	-

表 7 其他数据集上的实验结果对比

Table 7 Experimental Result Comparisons Across Multiple Datasets

方法	Xsum（%）			DUC2002（%）			LCSTS（%）		
	R-1	R-2	R-L	R-1	R-2	R-L	R-1	R-2	R-L
BRIO[35]	**49.07**	**25.59**	**40.40**	-	-	-	-	-	-
SumBART[43]	33.83	-	30.31	-	-	-	-	-	-
E2S2[36]	44.96	21.96	36.95	-	-	-	-	-	-
PEGASUS[12]	47.21	24.56	39.25	-	-	-	-	-	-
ChatGPT[30][78]	26.30	7.53	20.21	-	-	-	31.30	-	-
SummaRuNNer[5]	-	-	-	47.4	24.0	14.7	-	-	-
DeepSumm[13]	-	-	-	53.2	28.7	49.2	-	-	-
rnn-ext + RL[32]	-	-	-	39.46	17.34	36.72	-	-	-
RBCA-ETS[29]	-	-	-	**54.15**	**27.23**	**49.32**	-	-	-
PGN+coverage[39]	-	-	-	-	-	-	37.15	24.00	34.05
BERTSum[79]	-	-	-	-	-	-	**55.00**	**42.33**	**46.85**
BIOCopy[53]	-	-	-	-	-	-	44.21	33.55	43.85
CNsum[48]	-	-	-	-	-	-	36.49	17.26	33.24

从表 6 的实验结果来看，不同摘要生成方法在 CNN/DailyMail 数据集上存在显著差异，抽取式方法中，基于预训练语言模型的方法表现更为优异。但抽取式方法无法对文本进行灵活的重组与压缩，影响了摘要的灵活性。而生成式方法中，BRIO 和 DSC 凭借对候选摘要的精细对比及优化，在 ROUGE 指标上取得了

明显提升。表 7 的多数据集对比结果显示，BRIO、PEGASUS 等模型在多个数据集上均能保持较高性能，反映出其较强的泛化能力和稳定性，但部分方法尚缺乏跨数据集验证。此外，RL、GNN 等技术在 ATS 任务中凭借其各自的优势也获得了较优的结果。基于上述实验表现，我们筛选出在 CNN/DailyMail、XSum、DUC2002 和 LCSTS 等数据集中表现突出的模型，结合其技术策略和实验结果，总结其核心优缺点，如表 8 所示。

表 8 模型优缺点分析

Table 8 Analysis of Model Advantages and Disadvantages

模型	技术策略	优点	缺点
SeburSum[26]	抽取式（PLM）	通过比较候选摘要之间的最大语义相似度而非与原文的相似度，避免了对长摘要的偏好	候选摘要数量随文档长度指数增长，排名开销大，对长文效果下降
BRIO[35]	抽取式（Seq2Seq）	通过对比学习引入多摘要质量信号，缓解传统 MLE 暴露偏差问题，跨数据集泛化能力强	依赖预先生成的候选摘要进行对比训练，候选摘要生成计算成本高
DSC[45]	生成式（Seq2Seq）	仅使用 CNN/Daily Mail 中约 20%的训练实例即可超越强基线	仅基于英文新闻数据，尚未验证跨语言或跨领域的泛化能力
TriSum[67]	生成式（PLM）	利用 LLM 生成的结构化信息指导小模型训练，实现了在资源受限环境下的轻量化部署	模型性能高度依赖上游 LLM 质量，若 LLM 输出存在偏差或遗漏，可能导致本地模型性能受限或信息缺失
BART-large+IRL[42]	生成式 (RL)	通过逆向学习专家摘要的奖励权重，无需人工调参，即可提升摘要质量	需交替更新奖励模型和策略模型，训练流程复杂且收敛速度较慢
BERTSum[7]	抽取式（PLM）	能够捕捉文本的深层语义信息，擅长处理中文的复杂语义	长序列依赖处理不足，计算资源消耗高，生成灵活性差

4.3.2 ATS 面临挑战与解决方案

结合表 6、7、8 的信息和当前 ATS 领域的发展，我们在表 9 中从四个方面梳理了 ATS 面临的挑战及解决方案：

1) 跨领域与低资源场景的泛化瓶颈

主流预训练模型在单文档任务上表现均衡，显示了较好的通用性。但在多文档、跨领域和低资源场景下，其泛化能力明显受限。尽管在通用数据集上抽取式和生成式方法各有优势，但在专业领域（如医学、法律、金融）和低资源语言任务中，由于缺乏针对性训练，模型在专业术语和领域表达上表现不够稳定，易出现信息遗漏或事实错误。未来可通过领域自适应微调[80][81]和跨语言迁移学习[82]等策略，提升模型在各应用场景下的鲁棒性和适应性。

2) 评估体系的语义盲区与改进方向

现有的 ROUGE 等指标主要关注词汇匹配，难以全面反映摘要的语义连贯性和事实准确性。尽管这些指标在一定程度上提供了量化评估，但在实际应用中仍有提升空间。对此。可融合多维指标或采用强化学习策略[55]，以强化对事实一致性、信息覆盖和生成流畅度的综合评估，并针对抽取式与生成式摘要各自特点优化评价标准。

3) 事实一致性问题

生成式摘要技术虽然能生成流畅的文本，但仍存在事实错误问题。主要原因在于大多数 Seq2Seq 生成模型采用 MLE 训练，优化目标侧重语言流畅度而非事实准确性。另外，模型在生成过程中往往过度依赖语言预测，忽视了原文中明确的事实信息，导致在长文本或信息密集场景下易产生虚构内容。此外，摘要数据集可能存在标注不严谨或评价指标不足以客观衡量真实性的问题，从而加剧事实错误。对此，研究者探索使用知识图谱数据和结构化语义信息等提升了生成模型的事实准确性[62]。还有，当前检索增强生成

（RAG）技术在多数 NLP 任务中表现优异[83]，未来有望通过该技术提升摘要的事实性。总之，生成式摘要的事实一致性问题不仅涉及模型架构和训练方法，也反映了评价体系和数据标注的不足。

4) LLM 在自动文本摘要任务中的优势与局限

LLM 在 ATS 任务中凭借其强大的语言理解能力和生成能力，可以生成流畅且包含全局关键信息的摘要，为自动文本摘要任务带来了突破性进展[84]。从表 6 和表 7 中可知，以 ChatGPT 为代表的大语言模型在传统 Rouge 指标上虽然不及专门调优的生成或抽取模型，但在 G-Eval 上表现优异，说明其生成摘要更贴近人类语言习惯。然而，LLM 仍存在“幻觉”现象，即模型生成与事实不符的内容、输出高度依赖提示设计和计算需求高等问题，限制了其实时和大规模部署。为应对这些挑战，可采用以下措施。首先，可利用 RAG 技术降低 LLM 幻觉风险。其次，优化提示工程和调优策略，如采用思维链与提示链方法[85]，以便模型更准确地理解任务要求，减少提示设计不当引起的偏差。最后，通过模型压缩和蒸馏技术[86]，在降低计算资源消耗的同时保持高效推理，从而提升摘要质量和事实一致性。

表 9 ATS 面临挑战及解决方案

Table 9 Challenges Facing ATS and Their Solutions

存在问题	详细描述	改进策略	参考文献
跨领域与低资源适应性	模型在多文档、跨领域和低资源场景下泛化能力差	领域自适应微调、跨语言迁移技术	文献[66]、文献[80] 文献[81]、文献[82]
评估体系不足	现有评价指标侧重词汇匹配，难以全面反映语义连贯性与事实准确性	结合等多项评价指标，采用 RL 策略，构建更全面的综合评价体系	文献[55]
事实一致性问题	生成模型易产生虚假信息	引入 KG 技术，RAG 技术、优化训练目标和评价标准	文献[62]、文献[83]
大语言模型应用局限	对提示依赖强、计算资源消耗高	优化提示工程、采用模型压缩与蒸馏技术	文献[85]、文献[86]

5 自动文本摘要未来研究方向

虽然目前基于深度学习的自动文本摘要方法取得了一定进展，但该领域仍存在一些亟需解决的问题和挑战。需要进一步深入研究和优化现有方法。本文将从四个方面对自动文本摘要的未来研究工作进行展望。

1) 构建高质量数据集

开发高质量的基准数据集对于训练、评价和比较 ATS 模型至关重要。目前，大多数 ATS 模型主要依赖于新闻和对话数据集，针对金融、法律、医疗等专业领域，由于语料稀缺和数据分布特殊，现有数据集难以满足领域应用的需求。而低资源语言的数据集则更加稀缺，进一步限制了模型在这些语言下的泛化能力。未来，应整合多源信息，既关注低资源语言数据集的构建，也应针对特定领域开发定制化的数据集，以全面提升 ATS 在多样化应用场景下的表现。

2) 研究更全面的评价指标

在 ATS 任务中，传统的 ROUGE 等指标主要关注词汇级别的匹配，难以全面反映摘要的深层语义、逻辑连贯性以及事实准确性。现有工作虽然揭示了这些评价方法的局限性，但仍未形成一个能够同时衡量语义深度、事实一致性和用户体验的多维评价体系。未来研究应探索基于 LLM 的无参考评价方法[87][88]，利用其强大的语义理解能力直接评估生成摘要的质量。同时，可设计融合自动评价与人工评价的混合指标体系，从多个维度对摘要进行综合评价。此外，还可以考虑将评价指标作为反馈信号，融入到模型训练过程

中，实现生成目标的端到端优化。

3) 研究跨语言和低资源语言摘要技术

跨语言摘要旨在为一种语言的文档生成另一种语言的摘要，这一任务对低资源语言尤为具有挑战性。由于低资源语言缺乏大规模平行数据，传统基于端到端训练的方法在跨语言摘要任务中往往难以取得理想效果。近年来，研究者们从多个角度出发，尝试通过少样本学习和多任务预训练等策略来弥补这一瓶颈。Park 等[89]利用少量跨语言摘要示例对 LLM 进行微调，从而在低资源环境下显著提升跨语言摘要性能，这为低资源语言的跨语言摘要提供了可行的微调策略。Bai 等[90]提出了多任务学习框架，通过将单语摘要任务作为跨语言摘要任务的前置步骤，在统一解码器中共享知识，从而实现了高资源语言到低资源语言的知识迁移，有效缓解了平行数据稀缺的问题。未来研究应进一步结合少样本学习和多任务预训练策略，优化大语言模型的跨语言迁移能力，并针对低资源语言设计专门的数据构建和训练方法，以期在低资源环境下生成高质量的跨语言摘要。

4) 研究多文档摘要技术

在自动文本摘要领域，多文档摘要技术旨在从多个相关文档中提炼出一个全面且连贯的摘要，但由于文档之间信息量大且关联复杂，容易忽视跨文档的关系，且仍存在生成连贯性、信息覆盖和冗余控制等诸多挑战。近年来，基于图的模型，特别是 GNN 和 Transformer 模型，在这方面表现出很大的潜力[91][92]。未来研究应进一步探讨如何利用图结构和语言学知识，更高效地捕捉文档间共享的语义信息，并针对不同领域文档的特点设计专门的摘要策略，从而在保证摘要连贯性的同时提高信息精炼度和准确性。

6 结束语

本文对基于 DL 的 ATS 方法进行了全面调查与分析。首先，将当前 ATS 技术分为抽取式摘要和生成式摘要两大类，详细阐述了各自的核心技术，包括序列标注、图神经网络、预训练模型和强化学习等。其次，梳理了常用的数据集和评价指标，比较分析了主流方法在多个数据集和评价指标上的实验结果，探讨了当前 ATS 领域面临的挑战和解决方案。最后，展望了自动文本摘要未来的发展方向，强调了在构建高质量数据集、研究更全面的评价指标、跨语言和低资源语言摘要技术、多文档摘要技术等方面的探索和发展，以期为后续研究提供参考。

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] 李金鹏,张闯,陈小军,等.自动文本摘要研究综述[J].计算机研究与发展,2021,58(01):1-21.
LI J P,ZHANG C,CHEN X J, et al. Survey on automatic text summarization[J]. Journal of Computer Research and Development, 2021,58(01):1-21.

[3] 田萱,李嘉梁,孟晓欢.基于深度学习的抽取式摘要研究综述[J].计算机科学与探索,2024,18(11):2823-2847.
TIAN X,LI J L,MENG X H. A Survey of Deep Learning-Based Extractive Summarization[J].Journal of Frontiers of Computer Science and Technology, 2024, 18(11):2823-2847.

[4] ZHANG M, ZHOU G, YU W, et al. A comprehensive survey of abstractive text summarization based on deep learning[J]. Computational intelligence and neuroscience, 2022, 2022(1): 7132226.

[5] NALLAPATI R,ZHAI F, ZHOU B.Summarunner: A recurrent neural network based sequence model for extractive summarization of documents[C]//Proceedings of the AAAI conference on artificial intelligence. 2017, 31(1).

[6] ZHOU Q,YANG N, WEI F, et al. Neural document summarization by jointly learning to score and select sentences[J]. arXiv preprint arXiv:1807.02305, 2018.

[7] LIU Y. Fine-tune BERT for extractive summarization[J]. arxiv preprint arxiv:1903.10318, 2019.

[8] RUSH A M,CHOPRA S, WESTON J. A neural attention model for abstractive sentence summarization[J]. arxiv preprint arxiv:1509.00685, 2015.

[9] BAHDANAU D.Neural machine translation by jointly learning to align and translate[J]. arXiv preprint arXiv: 1409.0473, 2014.

[10] SEE A,LIU P J, MANNING C D. Get to the point: Summarization with pointer-generator networks[J]. arXiv preprint arXiv:1704.04368, 2017.

[11] LEWIS M.Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension[J]. arxiv preprint arxiv:1910.13461, 2019.

[12] ZHANG J,ZHAO Y, SALEH M, et al. Pegasus: Pre-train-

ing with extracted gap-sentences for abstractive summarization[C]//International conference on machine learning. PMLR, 2020: 11328-11339.

[13] JOSHI A, FIDALGO E, ALEGRE E, et al. DeepSumm: Exploiting topic models and sequence to sequence networks for extractive text summarization[J]. Expert Systems with Applications, 2023, 211: 118442.

[14] WANG D, LIU P, ZHENG Y, et al. Heterogeneous graph neural networks for extractive document summari- zation[J]. arXiv preprint arXiv:2004.12393, 2020.

[15] MAO Q,ZHU H, LIU J, et al. Muchsum: Multi-channel graph neural network for extractive summarization[C]//Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval. 2022: 2617-2622.

[16] ZHONG M,LIU P,CHEN Y, et al. Extractive summari- zation as text matching[J]. arxiv preprint arxiv:2004. 08795, 2020.

[17] NARAYAN S, COHEN S B, LAPATA M. Ranking sentences for extractive summarization with rein- forcement learning[J]. arxiv preprint arxiv:1802.08636, 2018.

[18] BIAN J, HUANG X, ZHOU H, et al. GoSum: extractive summarization of long documents by reinforcement learning and graph-organized discourse state[J]. Know- ledge and Information Systems, 2024: 1-24.

[19] CHENG J, LAPATA M. Neural summarization by extracting sentences and words[J]. arXiv preprint arXiv:1603.07252, 2016.

[20] KWON T, LEE S. OrderSum: Semantic Sentence Ordering for Extractive Summarization[J]. arxiv preprint arxiv:2502.16180, 2025.

[21] YASUNAGA M,ZHANG R,MEELU K, et al. Graph-based neural multi-document summarization[J]. arXiv preprint arXiv:1706.06681, 2017.

[22] JING B,YOU Z, YANG T, et al. Multiplex graph neural network for extractive text summarization[J]. arxiv preprint arxiv:2108.12870, 2021.

[23] YASUNAGA M,KASAI J, ZHANG R, et al. Scisummnet: A large annotated corpus and content-impact models for scientific paper summarization with citation networks[C]// Proceedings of the AAAI conference on artificial intelligence. 2019, 33(01): 7386-7393.

[24] XU J,GAN Z, CHENG Y, et al. Discourse-aware neural extractive text summarization[J]. arXiv preprint arXiv: 1910.14142, 2019.

[25] ONAN A, ALHUMYANIl H. Contextual hypergraph networks for enhanced extractive summarization: Introducing multi-element contextual hypergraph extractive summarizer (mches)[J]. Applied Sciences, 2024, 14(11): 4671.

[26] GONG S, ZHU Z, QI J, et al. SeburSum: a novel set-based summary ranking strategy for summary-level extractive summarization[J]. The Journal of Supercom- puting, 2023, 79(12): 12949-12977.

[27] BELTAGY I,PETERS M E,COHAN A. Longformer: The long-document transformer[J]. arXiv preprint arXiv:2004. 05150, 2020.

[28] CHENG X,SHEN Y,LU W. A set prediction network for extractive summarization[C]//Findings of the Association for Computational Linguistics: ACL 2023. 2023: 4766-4777.

[29] GANGUNDI R, SRIDHAR R. RBCA-ETS: enhancing extractive text summarization with contextual embedding and word-level attention[J]. International Journal of Information Technology, 2024: 1-9.

[30] ZHANG H,LIU X, ZHANG J. Extractive summarization via chatgpt for faithful summary generation[J]. arXiv preprint arXiv:2304.04193, 2023.

[31] LIU Y, ITER D, XU Y, et al. G-eval: Nlg evaluation using gpt-4 with better human alignment[J]. arXiv preprint arXiv:2303.16634, 2023.

[32] CHEN Y C, BANSAL M. Fast abstractive summarization with reinforce-selected sentence rewriting[J]. arXiv preprint arXiv:1805.11080, 2018.

[33] ZHANG T, KISHORE V, WU F, et al. Bertscore: Evaluating text generation with bert[J]. arxiv preprint arxiv:1904.09675, 2019.

[34] FU J, NG S K, JIANG Z, et al. Gptscore: Evaluate as you desire[J]. arXiv preprint arXiv:2302.04166, 2023.

[35] LIU Y, LIU P, RADEV D, et al. BRIO: Bringing order to abstractive summarization[J]. arXiv preprint arXiv:2203. 16804, 2022.

[36] ZHONG Q, DING L, LIU J, et al. E2S2: Encoding- enhanced sequence-to-sequence pretraining for language understanding and generation[J].IEEE Transactions on Knowledge and Data Engineering, 2023.

[37] VASWANI A.Attention is all you need[J]. Advances in Neural Information Processing Systems, 2017.

[38] GU J, LU Z, LI H, et al. Incorporating copying mechanism in sequence-to-sequence learning[J]. arxiv preprint arxiv:1603.06393, 2016.

[39] SEE A, LIU P J, MANNING C D. Get to the point: Summarization with pointer-generator networks[J]. arXiv preprint arXiv:1704.04368, 2017.

[40] XU S,LI H,YUAN P, et al. Self-attention guided copy mechanism for abstractive summarization[C]//Pro- ceedings of the 58th annual meeting of the association for computational linguistics. 2020: 1355-1362.

[41] RYU S,DO H, KIM Y, et al. Multi-Dimensional Optimization for Text Summarization via Reinforcement Learning[J]. arXiv preprint arXiv:2406.00303, 2024.

[42] FU Y,XIONG D, DONG Y. Inverse Reinforcement Learning for Text Summarization[J]. arXiv preprint arXiv:2212.09917, 2022.

[43] VIVEK A, DEVI V S. SumBART-An Improved BART Model for Abstractive Text Summarization[C]// International Conference on Neural Information Pro- cessing. Singapore: Springer Nature Singapore, 2022: 313-323.

[44] COHAN A, DERNONCOURT F, KIM D S, et al. A discourse-aware attention model for abstractive summarization of long documents[J]. arXiv preprint arXiv:1804.05685, 2018.

[45] SUN S, YUAN R, HE J, et al. Data selection curriculum for abstractive text summarization[C]//Findings of the Association for Computational Linguistics: EMNLP 2023. 2023: 7990-7995.

[46] KUMAR S, SOLANKI A. An abstractive text summarization technique using transformer model with self-attention mechanism[J]. Neural Computing and Applications, 2023, 35(25): 18603-18622.

[47] CHOWDHURY T, KUMAR S, CHAKRABORTY T. Neural abstractive summarization with structural attention[J]. arXiv preprint arXiv:2004.09739, 2020.

[48] ZHAO Y, HUANG S, ZHOU D, et al. CNsum: Automatic Summarization for Chinese News Text[C]//International Conference on Wireless Algorithms, Systems, and Applications. Cham: Springer Nature Switzerland, 2022: 539-547.

[49] DILAWARI A, KHAN M U G, SALEEM S, et al. Neural attention model for abstractive text summarization using linguistic feature space[J]. IEEE Access, 2023, 11: 23557-23564.

[50] ARGADE D,KHAIRNAR V, VORA D, et al. Multimodal Abstractive Summarization using bidirectional encoder representations from transformers with attention mechanism[J]. Heliyon, 2024, 10(4).

[51] MINAIDI M N,PAPAIOANNOU C, POTAMIANOS A. Self-attention based generative adversarial networks for unsupervised video summarization[C]//2023 31st European Signal Processing Conference (EUSIPCO). IEEE, 2023: 571-575.

[52] YANG F,CUI R,YI Z, et al. Cross-language generative automatic summarization based on attention mechanism[C]//Web Information Systems and Applications:17th International Conference, WISA 2020, Guangzhou, China, September 23–25, 2020, Proceedings 17. Springer International Publishing, 2020: 236-247.

[53] LI Q, WAN W, ZHAO Y, et al. Improved BIO-Based Chinese Automatic Abstract-Generation Model[J]. ACM Transactions on Asian and Low-Resource Language Information Processing, 2024, 23(3): 1-16.

[54] PAULUS R.A deep reinforced model for abstractive summarization[J]. arXiv preprint arXiv:1705.04304, 2017.

[55] FIKRI F B,OFLAZER K, YANIKOĞLU B. Abstractive summarization with deep reinforcement learning using semantic similarity rewards[J]. Natural Language Engineering, 2024, 30(3): 554-576.

[56] KENESHLOO Y, RAMAKRISHNAN N, REDDY C K. Deep transfer reinforcement learning for text summarization[C]//Proceedings of the 2019 SIAM International Conference on Data Mining. Society for Industrial and Applied Mathematics, 2019: 675-683.

[57] DEVLIN J. Bert: Pre-training of deep bidirectional transformers for language understanding[J]. arxiv preprint arxiv:1810.04805, 2018.

[58] RADFORD A.Improving language understanding by generative pre-training[J]. 2018.

[59] RADFORD A,Wu J, Child R, et al. Language models are unsupervised multitask learners[J].OpenAI blog, 2019, 1(8): 9.

[60] LEWIS M. Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension[J]. arxiv preprint arxiv:1910.13461, 2019.

[61] RAFFEL C, SHAZEER N, ROBERTS A, et al. Exploring the limits of transfer learning with a unified text-to-text transformer[J]. Journal of machine learning research, 2020, 21(140): 1-67.

[62] CHEN T,WANG X, YUE T, et al. Enhancing abstractive summarization with extracted knowledge graphs and multi-source transformers[J]. Applied Sciences, 2023, 13(13): 7753.

[63] XU L, SU Z, YU M, et al. Identifying Factual Inconsistencies in Summaries: Grounding LLM Inference via Task Taxonomy[C]//Findings of the Association for Computational Linguistics: EMNLP 2024. 2024: 14626- 14641.

[64] BALACHANDRAN V, HAJISHIRZI H, COHEN W W, et al.Correcting diverse factual errors in abstractive summarization via post-editing and language model infilling[J]. arXiv preprint arXiv:2210.12378, 2022.

[65] REDA A, SALAH N, ADEL J, et al. A hybrid arabic text summarization approach based on transformers[C]//2022 2nd International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC). IEEE, 2022: 56-62.

[66] WANG J, MENG F, ZHENG D, et al. Towards unifying multi-lingual and cross-lingual summarization[J]. arxiv preprint arxiv:2305.09220, 2023.

[67] JIANG P, XIAO C, WANG Z, et al. Trisum: Learning summarization ability from large language models with structured rationale[J]. arxiv preprint arxiv:2403.10351, 2024.

[68] NALLAPATI R, ZHOU B, GULCEHRE C, et al. Abstractive text summarization using sequence-to- sequence rnns and beyond[J]. arxiv preprint arxiv:1602. 06023, 2016.

[69] 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.

[70] HU B,CHEN Q, ZHU F. Lcsts: A large scale chinese short text summarization dataset[J]. arXiv preprint arXiv: 1506.05865, 2015.

[71] 闫晓东,王羿钦,黄硕,等.藏文多文本摘要数据集[J/OL]. 中国科学数据,2022,7(2).(2022-06-27). DOI:10.11922/11-6035.csd.2021.0098.zh.
YAN X D,WANG Y Q,HUANG S, et al. A dataset of Tibetan text summarization[J/OL].Science Data Bank, 2022,7(2).(2022-06-27).DOI:10.11922/11-6035.csd.2021.0098.zh.

[72] 翁彧,邢天娇,叶旭明,等.中-蒙-藏-维文多文档摘要数据集[J/OL].中国科学数据,2024,9(4).(2024-12-26).DOI:10.11922/11-6035.csd.2024.0038.zh.
WENG Y,XING T J,YE X M, et al.A dataset of Chinese-Mongolian-Tibetan-Uyghur multi-document summaries[J/OL]. Science Data Bank, 2024,9(4).(2024-12-26). DOI:10.11922/11-6035.csd.2024.0038.zh.

[73] 欧阳新鹏,闫晓东.藏汉跨语言摘要数据集 TiCLS[J/OL]. 中国科学数据,2024,9(4). (2024-12-25). DOI:10.11922/11-6035.csd.2024.0024.zh.
OUYANG X P, Yan X D. Tibetan Chinese cross language summary dataset TiCLS[J/OL].Science Data Bank, 2024,9(4).(2024-12-25).DOI:10.11922/11-6035.csd.2024.0024.zh.

[74] LIN C Y. Rouge: A package for automatic evaluation of summaries[C]//Text summarization branches out. 2004: 74-81.

[75] PAPINENI K, ROUKOS S, WARD T, et al.Bleu: a method for automatic evaluation of machine transla- tion[C]//Proceedings of the 40th annual meeting of the Association for Computational Linguistics. 2002: 311- 318.

[76] BANERJEE S, LAVIE A. METEOR: An automatic metric for MT evaluation with improved correlation with human judgments[C]//Proceedings of the acl workshop on intrinsic and extrinsic evaluation measures for machine translation and/or summarization. 2005: 65-72.

[77] SHEN C, CHENG L, NGUYEN X P, et al. Large language models are not yet human-level evaluators for abstractive summarization[J]. arxiv preprint arxiv:2305. 13091, 2023.

[78] LI L, ZHANG H, LI C, et al. Evaluation on ChatGPT for Chinese language understanding[J]. Data Intelligence, 2023, 5(4): 885-903.

[79] CHEN Y, CHEN H, LIU S, et al. Research on Information Extraction of LCSTS Dataset Based on an Improved BERTSum-LSTM Model[C]//2024 2nd International Conference on Mechatronics, IoT and Industrial Informatics (ICMIII). IEEE, 2024: 226-231.

[80] YU T, LIU Z, FUNG P. AdaptSum: Towards low-resource domain adaptation for abstractive summarization[J]. arxiv preprint arxiv:2103.11332, 2021.

[81] LI Y, MIAO S, HUANG H, et al. Word Matters: What Influences Domain Adaptation in Summarization?[J]. arxiv preprint arxiv:2406.14828, 2024.

[82] ŽAGAR A, ROBNIK-ŠIKONJA M. Cross-lingual transfer of abstractive summarizer to less-resource language[J]. Journal of Intelligent Information Systems, 2022, 58(1): 153-173.

[83] ARSLAN M, GHANEM H, MUNAWAR S, et al.A Survey on RAG with LLMs[J]. Procedia Computer Science, 2024, 246: 3781-3790.

[84] PU X, GAO M, WAN X. Summarization is (almost) dead[J]. arxiv preprint arxiv:2309.09558, 2023.

[85] SUN S, YUAN R, CAO Z, et al. Prompt chaining or stepwise prompt? refinement in text summariza- tion[C]//Findings of the Association for Computational Linguistics ACL 2024. 2024: 7551-7558.

[86] LAMAAKAL I, MALEH Y, EL MAKKAOUI K, et al. Tiny Language Models for Automation and Control: Overview, Potential Applications, and Future Research Directions[J]. Sensors, 2025, 25(5): 1318.

[87] BADSHAH S, SAJJAD H. Reference-Guided Verdict: LLMs-as-Judges in Automatic Evaluation of Free-Form Text[J]. arxiv preprint arxiv:2408.09235, 2024.

[88] CHEN S, GAO S, HE J. Evaluating factual consistency of summaries with large language models[J]. arxiv preprint arxiv:2305.14069, 2023.

[89] PARK G, HWANG S, LEE H. Low-Resource Cross- Lingual Summarization through Few-Shot Learning with Large Language Models[J]. arxiv preprint arxiv:2406. 04630, 2024.

[90] BAI Y, GAO Y, HUANG H. Cross-lingual abstractive summarization with limited parallel resources[J]. arxiv preprint arxiv:2105.13648, 2021.

[91] LIU Y, LAPATA M. Hierarchical transformers for multi-document summarization[J]. arxiv preprint arxiv: 1905.13164, 2019.

[92] LI M, QI J, LAU J H. Compressed heterogeneous graph for abstractive multi-document summarization[C]// Proceedings of the AAAI Conference on Artificial Intelligence. 2023, 37(11): 13085-13093.