Computer Science and Application 计算机科学与应用, 2025, 15(11), 320-330
Published Online November 2025 in Hans. https://www.hanspub.org/journal/csa
https://doi.org/10.12677/csa.2025.1511307

Hans汉斯

基于双向GRU与Attention机制的改进Seq2Seq自动文本摘要模型研究

许美玲，裘天瑜

河北金融学院信息与人工智能学院，河北 保定

收稿日期：2025年10月22日；录用日期：2025年11月19日；发布日期：2025年11月27日

摘　要

随着大数据时代的到来，数据规模呈指数级增长，信息过载问题日益突出。自动文本摘要技术通过计算机模型提炼文本主旨，生成简洁摘要，能够有效缓解信息过载，并广泛应用于新闻标题生成、文本检索与智能问答等场景。本文在分析现有文本摘要技术的基础上，以Seq2Seq模型为研究核心，重点探讨注意力机制(Attention Mechanism)在摘要生成中的作用，并提出一种结合双向GRU与注意力机制的改进型Seq2Seq模型。该模型在编码器部分采用双向GRU (BiGRU)结构，以充分捕获上下文语义信息；在解码器部分引入注意力机制，以提升摘要生成的准确性与连贯性。本文基于CNN/Daily Mail英文数据集对所提出模型进行训练与测试，并采用ROUGE指标评估实验结果。实验表明，所提出模型在摘要质量与信息覆盖度方面均优于传统Seq2Seq模型，验证了其有效性与可行性。

关键词

自动文本摘要，Seq2Seq模型，注意力机制，双向GRU

A Bidirectional GRU and Attention-Based Improved Seq2Seq Model for Automatic Text Summarization

Meiling Xu, Tianyu Qiu

School of Information and Artificial Intelligence, Hebei Finance University, Baoding Hebei

Received: October 22, 2025; accepted: November 19, 2025; published: November 27, 2025

Abstract

With the advent of the big data era, data volumes are growing exponentially, and the problem of

文章引用: 许美玲, 裘天瑜. 基于双向 GRU 与 Attention 机制的改进 Seq2Seq 自动文本摘要模型研究[J]. 计算机科学与应用, 2025, 15(11): 320-330. DOI: 10.12677/csa.2025.1511307

information overload has become increasingly prominent. Automatic text summarization technology, which distills the main ideas of a text through computational models to generate concise summaries, can effectively alleviate information overload and is widely applied in scenarios such as news headline generation, text retrieval, and intelligent question answering. Based on an analysis of existing text summarization techniques, this paper focuses on the Seq2Seq model and explores the role of the Attention Mechanism in summary generation. A novel improved Seq2Seq model combining bidirectional GRU (BiGRU) and attention mechanism is proposed. In this model, the encoder employs a BiGRU structure to fully capture contextual semantic information, while the decoder incorporates an attention mechanism to enhance the accuracy and coherence of generated summaries. The model is trained and tested on the CNN/Daily Mail English dataset, and evaluation is conducted using the ROUGE metric. Experimental results demonstrate that the proposed model outperforms the traditional Seq2Seq model in both summary quality and information coverage, validating its effectiveness and feasibility.

Keywords

Automatic Text Summarization, Seq2Seq Model, Attention Mechanism, Bidirectional GRU (BiGRU)

1. 引言

随着互联网与信息技术的快速发展，新闻、社交媒体以及信息分享平台的普及，网络文本数据量呈指数级增长。伴随海量信息的涌现，信息过载问题日益突出，人们在信息获取过程中需要耗费大量时间和精力，从冗余文本中提取有价值信息[1]。在这种背景下，如何高效地从海量文本中提取核心内容，成为自然语言处理(NLP)领域的重要研究方向。自动文本摘要技术正是在此需求下应运而生。该技术通过计算机模型自动分析文本的语义结构与关键信息，从而生成简洁的摘要，帮助用户快速理解文本主旨[2]。与人工摘要相比，自动文本摘要不仅能显著提高信息获取效率，还能降低人力成本，广泛应用于新闻标题生成、文本检索、知识问答与舆情分析等领域[3]。因此，自动摘要技术的研究具有重要的现实意义与应用价值。

从研究方法上看，自动文本摘要主要分为抽取式摘要与生成式摘要两大类。抽取式摘要通过对文本中的关键词、句子进行评分与筛选生成摘要，典型方法包括 Luhn 的词频统计法[4]、Kupiec 等人的朴素贝叶斯方法[5]，以及 Mihalcea 提出的 TextRank 算法[6]等。生成式摘要则借助神经网络与自然语言生成技术，根据语义重构生成全新的摘要内容，具有更强的语言生成与语义理解能力。自 Sutskever 等人提出 Seq2Seq [7]模型以来，生成式摘要成为研究热点。Rush 等人[8]将注意力机制引入 Seq2Seq 模型，提升了摘要生成质量；See 等人[9]提出指针生成网络(Pointer-Generator Network)，有效缓解了摘要重复与信息遗漏问题。近年来，基于 Transformer 结构的模型(如 BERT、T5、PEGASUS、BART)及大规模预训练语言模型(如 GPT-4、LLaMA) [10] [11]，进一步推动了生成式摘要向高质量、可解释性和多语言方向发展。

基于上述研究进展，本文提出一种基于双向门控循环单元(BiGRU)与注意力机制的改进型 Seq2Seq 文本摘要模型。在编码器部分采用 BiGRU 结构，以充分捕捉上下文语义信息；在解码器部分引入注意力机制[12]，提高摘要生成的连贯性与准确性。本文在 CNN/Daily Mail [13]英文数据集上对模型进行训练与测试，并采用 ROUGE [14]指标对生成结果进行客观评估，以验证模型的有效性与优越性。通过改进生成式摘要模型并进行实验验证，本文旨在为自动文本摘要的智能化与高质量生成提供新的思路与技术支持。

2. 数据获取和预处理

2.1. 数据获取

在自动文本摘要任务中，数据集的选择对模型训练效果具有重要影响。由于不同数据集在文本长度、主题分布及摘要形式等方面存在差异，模型的性能表现也可能出现一定程度的差异。目前，常用的英文摘要数据集主要包括 Gigaword、CNN/Daily Mail 和 DUC2004 等(表 1)。其中，Gigaword 与 DUC2004 数据集主要用于单句摘要任务，适合对简短文本进行抽象生成；而 CNN/Daily Mail 数据集属于多句摘要数据集，更贴近新闻类长文本的摘要生成需求，能够有效提升模型对长篇语义的理解与表达能力。

Table 1. Dataset introduction
表 1. 数据集介绍

摘要	类型	数据集	训练集	验证集	测试集
DUC2004	news	500	-	-	-
Gigaword	news	420 万	380 万	20 万	20 万
CNN/Daily Mail	news	311,632	287,226	13,368	14,490

基于上述考虑，本文选用 CNN/Daily Mail 数据集作为模型训练与测试的数据来源。该数据集由美国有线电视新闻网(CNN)和《每日邮报》(Daily Mail)的新闻报道组成，包含新闻正文及对应的参考摘要。数据量大且结构清晰，因而被广泛应用于生成式摘要研究任务中，能够有效支撑模型在长文本摘要生成上的训练与评估。

2.2. 数据预处理

原始数据文件采用 Unicode 编码而非 ASCII 格式，因此需首先进行编码转换以确保数据的一致性。随后，对数据进行归一化处理，包括将字符统一为小写以及去除文本首尾的冗余空格。由于所使用的数据集为英文语料，在分词阶段以空格作为词语分隔符，处理后的示例结果如表 2 所示。

Table 2. Data preprocessing
表 2. 数据预处理

原始句子	处理后句子
> As they make a final push to approve presidential nominations before Republicans take control of the Senate, Democrats said Tuesday the confirmation of a record number of federal judges was evidence they were right to make controversial changes to filibuster rules, despite objections from Republicans. “Yes,” Senate Majority Leader Harry Reid responded loudly when asked if still believes he was right to employ the so-called “nuclear option” a year ago in order to clear a backlog of nominees. The No. 2 Senate Democrat explained that at the time there was a “breakdown in the relationship between the executive and legislative branch.” “If you just look at where we were, with all of the nominations stacked on the calendar, most of which had been reported from committees with overwhelming bipartisan votes,” Sen. Dick Durbin said.	< as they make a final push to approve presidential nominations before republicans take control of the senate democrats said tuesday the confirmation of a record number of federal judges was evidence they were right to make controversial changes to filibuster rules despite objections from republicans. yes senate majority leader harry reid responded loudly when asked if still believes he was right to employ the so called nuclear option a year ago in order to clear a backlog of nominees .the no . senate democrat explained that at the time there was a breakdown in the relationship between the executive and legislative branch . if you just look at where we were with all of the nominations stacked on the calendar most of which had been reported from committees with overwhelming bipartisan votes sen. dick durbin said.

2.3. 文本向量化

分词完成后，需要对文本进行向量化处理，将每个词表示为数值向量。本研究采用 Word2Vec [15]方法训练词向量，包含 CBOW 和 Skip-Gram 两种模式。CBOW 模型通过上下文词预测目标词，梯度均匀分布于上下文词，每次预测需遍历整个词典大小 V；而 Skip-Gram 模型以目标词预测上下文词，梯度作用于目标词，每次输入一个词需预测多个上下文词，因此训练次数更多、速度较慢，但对新词及低频词的语义表示能力更强。在本文中，通过 Word2Vec 对语料进行向量化处理，为 Seq2Seq 自动摘要模型提供高质量的词向量输入，从而提升模型在语义理解和摘要生成上的性能。

3. 模型训练和参数选取

在自动文本摘要任务中，Encoder-Decoder (编码器–解码器)架构具有良好的序列建模能力，但传统的 Seq2Seq 模型在实际应用中仍存在一定不足。由于该模型将输入序列压缩为固定长度的向量，容易造成部分语义信息丢失，从而影响摘要生成质量。为解决这一问题，本文在模型中引入了注意力(Attention)机制，通过对输入序列中不同位置分配动态权重，使解码器在生成每个词时能够关注与其最相关的上下文信息，从而有效提升模型的表达能力和摘要的语义完整性，如图 1 所示。

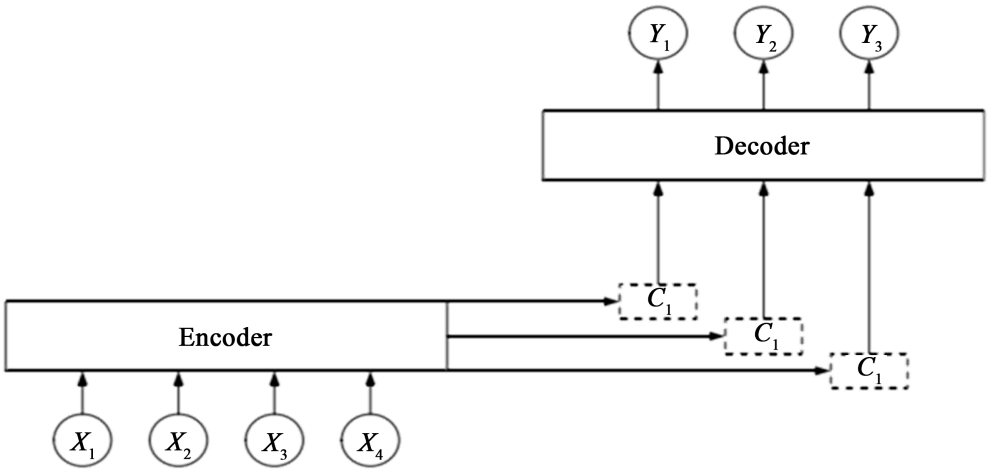

Figure 1. Sequence-to-sequence with attention model diagram
图 1. 注意力机制的序列到序列模型示意图

在编码器和解码器的具体实现中，循环神经网络(RNN)被广泛应用，但在处理长序列时容易出现梯度消失和长期依赖问题。为此，本文在传统 Seq2Seq 结构的基础上进行了改进：编码器部分采用双向门控循环单元，以充分捕捉上下文的前向和后向语义信息；解码器部分采用单向 GRU，并结合注意力机制对上下文信息进行加权融合，从而提升摘要生成的准确性与连贯性。

3.1. 编码器

编码器接收变长输入序列，通过 GRU 单元计算每个时间步的隐藏状态，并将隐藏层输出汇总生成上下文表示向量 $\boldsymbol{C}$ 。其计算过程如公式(1)所示：

$$\boldsymbol{C} = q\left(h_1, \cdots, h_T\right) \tag{1}$$

传统的单向 GRU 只能捕获序列的前向依赖信息，可能导致语义理解不完整。为充分地利用上下文信息，本文在编码器部分引入双向 GRU 结构。BiGRU 由正向 GRU 与反向 GRU 两个网络组成：正向网络

从序列起始位置到终止位置依次处理输入，反向网络则从序列末尾反向处理输入。两者在每个时间步的输出进行拼接或融合，从而获得更全面的上下文特征表示，其计算过程如公式(2)所示。

$$h_t = f\left(h_{t-1}, x_t\right) \tag{2}$$

BiGRU 模型能够同时捕获输入序列的双向语义信息，有效减少信息丢失，增强模型的表达能力。其结构示意如图 2 所示。

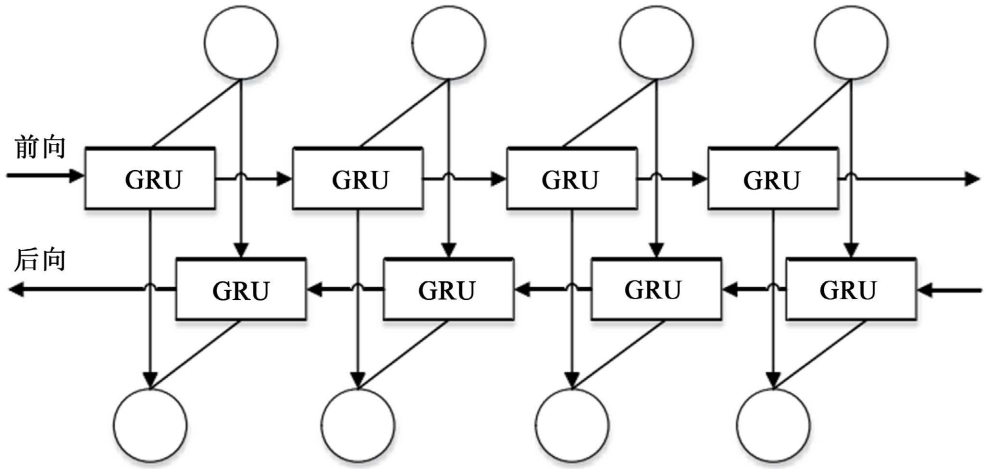

Figure 2. Architecture of the BiGRU network
图 2. BiGRU 网络结构图

在信息计算过程中，经过嵌入层(Embedding Layer)转换得到的文本向量序列作为 BiGRU 层的输入，BiGRU 依次提取其中的语义特征,并输出包含历史信息的向量表示，其计算过程如公式(3)所示：

$$h = g\left(h_{t-1}, e_t; \theta_g\right) \tag{3}$$

其中，$g(\cdot)$ 表示非线性激活函数，θ_g 表示循环神经网络的参数集合。

本文实现了一个双向 GRU (BiGRU)编码器，用于 Seq2Seq 模型的文本编码。编码器首先通过 nn.Embedding 将输入词索引映射为向量，再由 BiGRU 捕捉序列的前向与后向上下文信息，生成隐藏表示。随后，GRU 输出通过全连接层映射回隐藏层维度，得到最终的编码向量。在前向传播过程中，输入的词向量与前一时间步的隐藏状态被 GRU 处理，并返回更新后的隐藏状态与编码结果。该结构能够有效整合上下文信息，为解码器提供语义丰富的输入。其主要代码如下所示：

```
class EncoderBiGRU(nn.Module):
    def __init__(self, input_size, hidden_size):
        super(EncoderBiGRU, self).__init__()
        self.hidden_size = hidden_size
        # 词嵌入层
        self.embedding = nn.Embedding(input_size, hidden_size)
        # 双向 GRU
        self.gru = nn.GRU(
            input_size=hidden_size,
            hidden_size=hidden_size,
```

```
        bidirectional=True
    )
    # 输出映射层
    self.fc = nn.Linear(2 * hidden_size, hidden
```

3.2. 解码器

解码器的主要任务是根据编码器输出的上下文向量以及先前生成的词语，逐步预测目标摘要序列中的下一个词。本文在解码器部分采用单向 GRU (Gated Recurrent Unit)结构，并引入注意力机制(Attention Mechanism)，使模型在生成每个词语时能够动态关注输入序列中与当前预测最相关的部分，从而提高摘要的语义一致性与信息完整性。

解码器通过编码阶段得到的上下文向量 $\boldsymbol{C}$ 与前一时刻的输出序列 y_{t-1} 共同预测当前词的生成结果。隐藏状态的计算如公式(4)所示，输出的条件概率如公式(5)所示：

$$h_t = f\left(h_{(t-1)}, y_{t-1}, \boldsymbol{C}\right) \tag{4}$$

$$p\left(y_t \mid y_1, y_2, \cdots, y_{t-1}, \boldsymbol{C}\right) = g\left(h_{(t-1)}, y_{t-1}, \boldsymbol{C}\right) \tag{5}$$

其中，f 和 g 均为非线性神经网络函数。通过最大化输出序列的联合概率公式(6)来优化模型参数，并以交叉熵损失函数公式(7)作为训练目标：

$$P\left(y_t \mid y_{t-1}, y_{t-2}, \cdots, y_1\right) = \prod_{t=1}^{T} p\left(y_t \mid y_1, y_2, \cdots, y_{t-1}, \boldsymbol{C}\right) \tag{6}$$

$$L = -\log P\left(y_t, y_{t-1}, y_{t-2}, \cdots, y_1\right) \tag{7}$$

3.3. 注意力机制

在传统 Seq2Seq 模型中，编码器输出的上下文向量在整个解码过程中保持不变，无法体现输入序列中不同词语的重要性。注意力机制(Attention Mechanism)通过为输入序列的各部分分配权重，使模型在生成摘要时能够聚焦于信息量较大的词或短语，从而提升摘要生成的准确性与可解释性。其计算过程如公式(8)~(10)所示：

$$C_i = \sum_{j=1}^{n} a_{ij} h_j \tag{8}$$

$$h_j = g\left(x_j\right) \tag{9}$$

$$s_i = g\left(y_{i-1}, C_i, s_{i-1}\right) \tag{10}$$

其中，$g(\cdot)$ 表示编码器函数，h_j 为编码器在 j 时刻的隐藏状态，a_{ij} 为输入序列中 x_j 对解码器的注意力权重，C_i 为加权后的上下文向量，s_i 为解码器在第 i 时刻的隐藏状态。

变量 e_{ij} 为编码器隐藏状态 h_j 和解码器上一步隐藏状态 s_{i-1} 的相关性评分，如公式(11)所示。

$$e_{ij} = \delta\left(s_{i-1}, h_j\right) \tag{11}$$

注意力机制能够一定程度上缓解传统 RNN 模型的梯度消失问题，并增强模型的可解释性。通过可视化注意力分布，可以直观地观察模型在生成摘要时关注的文本区域，为错误分析和模型优化提供依据。

本文实现了一个带注意力机制的双向 GRU 编码器，用于将输入序列映射为上下文丰富的隐藏表示，并计算每个时间步的注意力权重。其主要代码如下所示：

```
def forward(self, input, hidden, encoder_outputs):
    # 嵌入并添加 dropout
    embedded = self.embedding(input).view(1, 1, -1)
    embedded = self.dropout(embedded)
    # 计算注意力权重
    attn_weights = F.softmax(
        self.attn(torch.cat((embedded[0], hidden[0]), 1)), dim=1
    )
    # 应用注意力权重
    attn_applied = torch.bmm(
        attn_weights.unsqueeze(0), encoder_outputs.unsqueeze(0)
    )
    # 结合嵌入与注意力
    output = torch.cat((embedded[0], attn_applied[0]), 1)
    output = self.attn_combine(output).unsqueeze(0)
    output = F.relu(output)
    # 输入 GRU
    output, hidden = self.gru(output, hidden)
    # 输出概率
    output = F.log_softmax(self.out(output[0]), dim=1)
    return output, hidden, attn_weights
```

3.4. 模型训练整体流程

本文所提出的改进型 Seq2Seq 文本摘要模型训练流程如下：

(1) 词向量训练：使用 Word2Vec 对源文本序列进行训练，生成词嵌入向量；

(2) 编码器处理：编码器端采用 BiGRU 处理输入词向量，输出包含上下文信息的隐藏状态序列；

(3) 目标序列向量化：解码器端采用与编码器相同的词嵌入方式，将目标序列映射为词向量；

(4) 序列生成预测：解码器通过单向 GRU 结构，以编码器最后的隐藏状态作为初始状态，根据上下文信息和前一步输出预测下一个词；

(5) 输出生成：解码器输出经全连接层映射后，通过 Softmax 函数计算词汇表概率分布，最终利用解码算法生成目标摘要序列。具体流程如图 3 所示。

3.5. 实验评价方法

在文本摘要研究中，常用的评价方法可分为人工评价和自动评价两类。人工评价是指由评审者阅读模型生成的摘要，并根据准确性(Accuracy)、完整性(Completeness)和可读性(Readability)等主观指标进行打分的方法。该方法能够从人类理解角度评估摘要质量，但由于评价者的教育背景、语言水平及主观偏好存在差异，评分结果往往具有较大波动。此外，人工评价成本高、效率低，不适用于大规模文本摘要任务。因此，本研究主要采用 ROUGE (Recall-Oriented Understudy for Gisting Evaluation)指标进行自动化评价。ROUGE 通过计算模型生成摘要与人工参考摘要之间的重叠单元数，衡量模型在信息覆盖和语义保留方面的性能。

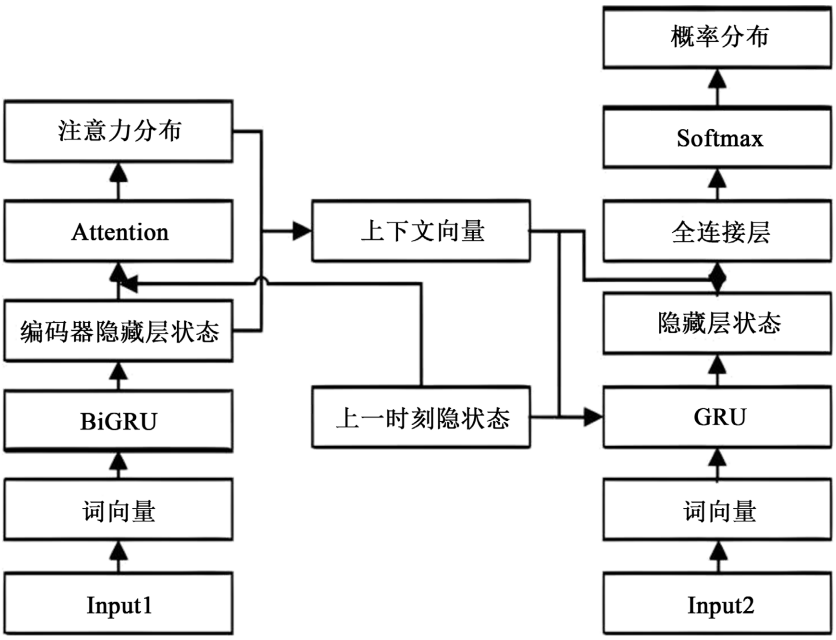

Figure 3. Seq2Seq + Attention model flowchart
图 3. Seq2Seq + Attention 模型算法流程图

ROUGE 指标主要包括 ROUGE-N、ROUGE-L、ROUGE-S 和 ROUGE-W 四种类型，其中最常用的是 ROUGE-N 和 ROUGE-L。ROUGE-N 基于 N-gram 匹配原理，计算参考摘要与生成摘要中共有的 N 元词序列比例；ROUGE-L 基于最长公共子序列(Longest Common Subsequence, LCS)，能够反映摘要在句法结构与整体语义上的相似度。

ROUGE-N 评价方法的计算如下所示：

$$\text{ROUGE-N}=\frac{\sum_{S\in\{\text{RefSum}\}}\sum_{\text{N-gram}\in S}\text{Count}_{\text{match}}\left(\text{N-gram}\right)}{\sum_{S\in\{\text{RefSum}\}}\sum_{\text{N}_{\text{N-gram}}\in S}\text{Count}\left(\text{N-gram}\right)} \tag{12}$$

其中，分母表示参考摘要中 N-gram 的总数，分子表示参考摘要与模型摘要中共有的 N-gram 数。对于 ROUGE-L 指标，L 表示参考摘要与生成摘要之间的最长公共子序列(LCS)长度，其计算公式如下：

$$\text{ROUGE-L}_{\text{RecaLL}}=\frac{\text{LCS}\left(\text{C},\text{S}\right)}{\text{len}\left(\text{S}\right)} \tag{13}$$

$$\text{ROUGE-L}_{\text{Precision}}=\frac{\text{LCS}\left(\text{C},\text{S}\right)}{\text{len}\left(\text{C}\right)} \tag{14}$$

$$\text{ROUGE-L}=\frac{\left(1+\beta^2\right)\text{Precision}\times\text{Recall}}{\text{Recall}+\beta^2\text{Precision}} \tag{15}$$

其中，β 为调节系数，当 $\beta>1$ 时，更加关注召回率(Recall)，反之更侧重精确率(Precision)。通过 ROUGE 指标，本研究能够客观评价模型在信息覆盖、语义保留及摘要质量方面的表现，为模型优化提供量化依据。

4. 结果分析

表 3 展示了基于传统 Seq2Seq 模型生成的摘要示例。可以看出，该模型能够捕捉文本中的部分关键信息，但在长文本摘要生成中仍存在一定局限性，如信息遗漏、语义不连贯以及句式不完整等问题。

Table 3. Example of model-generated summaries
表 3. 模型生成摘要实例

源文本 1	london cnn the new james bond movie will be called skyfall the producers announced in london thursday. oscar winner sam mendes will direct daniel craig as in the rd bond movie. we ll start shooting today mendes said. spanish actor javier bardem will play the villain and judi dench will reprise her role as bond s boss m. the cast also includes french actress berenice marlohe. the announcement of the title of the movie came on the th anniversary of the date sean connery was revealed as the first actor to play ian fleming s spy
参考摘要	the film is to be called skyfall daniel craig will reprise his role as in the rd bond filmjavier bardem will play the villain the announcement comes on the th anniversary of sean connery s casting new
生成摘要	french officials bella was part from a walk for tibet from st to west palm beachhe <EOS>
源文本 2	cnn student news september download pdf maps related to today s show egypt libya democratic republic of congoyen ennew york city chicago illinoisclick here to access the transcript of today s cnn student news program. please note that thene may be a delay between the time when the video is available and when the transcript is published
参考摘要	the daily transcript is a wnitten version of each day s cmn student news programuse this transcript to help students with reading comprehension andvocabularyuse the weekly newsquiz to test your knowledge of stories you saM on cnn student new
生成摘要	the daily transcript is a written version of each days cmn student news programuse this transcript to help students with reading comprehension and vocabulary use the end of year news quiz to test your knowledge of stories you saw on cmn student new

由表 3 可见，传统 Seq2Seq 模型虽然能够在句法层面生成合理的语言结构，但生成的摘要中常混入无关信息或遗漏关键内容，表明其在全局语义建模方面仍存在不足。

表 4 展示了基于 BiGRU + Attention 机制的改进模型生成的摘要示例。与传统模型相比，该改进模型生成的摘要内容更加完整且连贯，尤其在多句摘要任务中表现出明显优势，能够更有效地捕捉长文本中的关键信息，实现语义的准确表达。

Table 4. Examples of summaries generated using the GRU + Attention model
表 4. GRU + Attention 模型生成摘要实例

源文本 1	cmn the national flag of the united states of america can be found hanging off homes across the country flapping atop mount everest and sitting on the mon s surface. here are some of the most unique places ireporters have spotted americasstars and stripes. have you seen the american flag in an unexpected place ? share your photos with cnn ireport for a chance to be featured
参考摘要	ireporters share unusual spots they ve seen the stars and stripesthe flag of the united states was first adopted on june
生成摘要	new a truck is on the to death in in total prize
源文本 2	cnn student news september download pdf maps related to today s show egypt libya democratic republic of congoyen ennew york city chicago illinoisclick here to access the transcript of today s cnn student news program. please note that thene may be a delay between the time when the video is available and when the transcript is published
参考摘要	the daily transcript is a wnitten version of each day s cmn student news programuse this transcript to help students with reading comprehension andvocabularyuse the weekly newsquiz to test your knowledge of stories you saM on cnn student new
生成摘要	the daily transcript is a written version of each day s cmn student news programuse this transcript to help students with reading comprehension and vocabularyuse the end of year newsquiz to test your knowledge of stories you saw on cmn student new

由表 4 可见，尽管改进模型生成的部分摘要仍存在少量语法错误或词序问题，但整体语义相关性明显提升，模型在信息保留和句间连贯性方面表现出明显改善。表 5 展示了不同模型在 ROUGE 指标上的评分对比。

Table 5. Comparison of evaluation metrics across models
表 5. 模型评分对比

模型	ROUGE-1	ROUGE-2	ROUGE-L
Seq2Seq 模型	30.69	23.08	28.87
基于 BiGRU + Attention 的 Seq2Seq 模型	32.12	23.27	30.05

从整体指标来看，传统 Seq2Seq 模型在 ROUGE-1、ROUGE-2 和 ROUGE-L 指标上分别达到 30.69、23.08 和 28.87，而改进后的 BiGRU + Attention 模型分别为 32.12、23.27 和 30.05，均有所提升。这表明，引入双向 GRU 和 Attention 机制能够有效捕捉输入文本的上下文信息与关键特征，从而提升摘要的完整性和语义一致性。

进一步分析可知，ROUGE-1 和 ROUGE-L 指标提升幅度较大，分别提高了 1.43 和 1.18，而 ROUGE-2 提升幅度相对较小(仅 0.19)。这可能与 ROUGE-2 主要衡量二元词序列匹配有关。由于摘要生成中句式较灵活、词序多变，二元匹配的提升空间有限。然而，总体提升仍表明改进模型在语义捕捉、信息覆盖及摘要连贯性方面优于传统 Seq2Seq 模型。

从机制上分析，BiGRU 编码器可同时获取输入序列的正向与反向上下文信息，减少信息丢失；Attention 机制通过为输入各部分分配权重，使模型能够动态关注重要的词或短语，从而生成更符合原文语义的摘要。该结构在处理长文本与多句摘要任务时尤其有效。

5. 结论

综上所述，本文提出的 BiGRU + Attention 改进模型在摘要的准确性、信息覆盖率与连贯性方面均显著优于传统 Seq2Seq 模型，验证了该结构在自动文本摘要任务中的可行性与有效性。未来研究可进一步结合多头注意力机制(Multi-Head Attention)、预训练语言模型(如 BERT、T5)或集成学习方法，以进一步提升摘要生成质量。

基金项目

2025 年度河北省教育厅科学研究项目资助及项目立项编号：基于多模态信息的虚假新闻检测研究(课题编号：QN2025018)。

参考文献

[1] 王文静, 张宏宇, 李明. 自动文本摘要技术研究综述[J]. 计算机研究与发展, 2023, 60(8): 1572-1588.

[2] 邢淼. 国内外大语言模型生成中文论文摘要对比研究[J]. 知识管理论坛, 2024, 9(2): 45-52.

[3] 裴炳森. 基于大语言模型的司法文本摘要生成与评价技术研究[J]. 情报科学, 2024, 42(6): 88-97.

[4] Luhn, H.P. (1958) The Automatic Creation of Literature Abstracts. *IBM Journal of Research and Development*, **2**, 159-165. https://doi.org/10.1147/rd.22.0159

[5] Kupiec, J., Pedersen, J. and Chen, F. (1995) A trainable Document Summarizer. *Proceedings of the* 18*th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval-SIGIR* '95, Washington, 9-13 July 1995, 68-73. https://doi.org/10.1145/215206.215333

[6] Mihalcea, R. and Tarau, P. (2004) TextRank: Bringing Order into Texts. *Proceedings of the* 2004 *Conference on*

Empirical Methods in Natural Language Processing, Barcelona, 25-27 July 2004, 404-411.

[7] Sutskever, I., Vinyals, O. and Le, Q.V. (2014) Sequence to Sequence Learning with Neural Networks. *Advances in Neural Information Processing Systems*, **27**, 3104-3112.

[8] Rush, A.M., Chopra, S. and Weston, J. (2015) A Neural Attention Model for Abstractive Sentence Summarization. *Proceedings of the* 2015 *Conference on Empirical Methods in Natural Language Processing*, Lisbon, 17-21 September 2015, 379-389. https://doi.org/10.18653/v1/d15-1044

[9] See, A., Liu, P.J. and Manning, C.D. (2017) Get to the Point: Summarization with Pointer-Generator Networks. *Proceedings of the* 55*th Annual Meeting of the Association for Computational Linguistics*, Vancouver, 30 July-4 August 2017, 1073-1083. https://doi.org/10.18653/v1/p17-1099

[10] 张扬, 金涵蕾, 孟丹, 王骏, 谭晶华. 基于大语言模型的自动文本摘要研究综述[J]. 数据分析与知识发现, 2025, 9(1): 1-14.

[11] 祁天, 杨建安, 赵铁军, 杨沐昀. 基于思维链的跨语言多文档摘要生成技术研究[C]. 中国计算语言学年会论文集(CCL 2024). 北京: 中国中文信息学会, 2024: 98-108.

[12] Bahdanau, D., Cho, K. and Bengio, Y. (2015) Neural Machine Translation by Jointly Learning to Align and Translate. arXiv:1409.0473

[13] Hermann, K.M., Kocisky, T., Grefenstette, E., Espeholt, L., Kay, W., Suleyman, M. and Blunsom, P. (2015) Teaching Machines to Read and Comprehend. *Advances in Neural Information Processing Systems*, Montreal, 1693-1701.

[14] Lin, C.-Y. (2004) Rouge: A Package for Automatic Evaluation of Summaries. *Proceedings of the Workshop on Text Summarization Branches Out*, Barcelona, 25-26 July 2004, 74-81.

[15] Mikolov, T., Chen, K., Corrado, G. and Dean, J. (2013) Efficient Estimation of Word Representations in Vector Space. arXiv:1301.3781. https://arxiv.org/abs/1301.3781