

基于指针生成网络和扩展 Transformer 的多属性可控文本摘要模型^①



洗广铭, 李凡龙, 郑兆明

(华南师范大学 软件学院, 佛山 528225)

通信作者: 洗广铭, E-mail: xgm20011@163.com

摘要: 可控文本摘要模型可以生成符合用户偏好的摘要. 之前的摘要模型侧重于单独控制某个属性, 而不是多个属性的组合. 传统的 Seq2Seq 多属性可控文本摘要模型在满足多个控制属性时, 存在无法整合所有控制属性、无法准确再现文本中关键信息和无法处理单词表外单词等问题. 为此, 本文提出了一种基于扩展 Transformer 和指针生成网络 (pointer generator network, PGN) 的模型. 模型中的扩展 Transformer 将 Transformer 单编码器-单解码器的模型形式扩展成具有双重文本语义信息提取的双编码器和单个可融合指导信号特征的解码器形式. 然后利用指针生成网络模型选择从源文本中复制单词或利用词汇表生成新的摘要信息, 以解决摘要任务中常出现的 OOV (out of vocabulary) 问题. 此外, 为高效完成位置信息编码, 模型在注意力层中使用相对位置表示来引入文本的序列信息. 模型可以用于控制摘要的许多重要属性, 包括长度、主题和具体性等. 通过在公开数据集 MACSum 上的实验表明, 相较以往方法, 本文提出的模型在确保摘要质量的同时, 更加符合用户给定的属性要求.

关键词: 深度学习; 可控文本摘要; Transformer 模型; 相对位置表示; 指针生成网络

引用格式: 洗广铭, 李凡龙, 郑兆明. 基于指针生成网络和扩展 Transformer 的多属性可控文本摘要模型. 计算机系统应用, 2024, 33(4): 246-253. <http://www.c-s-a.org.cn/1003-3254/9457.html>

Multi-attribute Controllable Text Summary Model Based on Pointer Generator Network and Extended Transformer

XIAN Guang-Ming, LI Fan-Long, ZHENG Zhao-Ming

(School of Software, South China Normal University, Foshan 528225, China)

Abstract: The controllable text summary models can generate summaries that conform to user preferences. Previous summary models focus on controlling a certain attribute alone, rather than the combination of multiple attributes. When multiple control attributes are satisfied, the traditional Seq2Seq multi-attribute controllable text summary model cannot integrate all control attributes, accurately reproduce key information in the texts, and handle words outside the word lists. Therefore, this study proposes a model based on the extended Transformer and pointer generator network (PGN). The extended Transformer in the model extends the Transformer single encoder-single decoder model form into a dual encoder with dual text semantic information extraction and a single decoder form that can fuse guidance signal features. Then the PGN model is employed to select the source from the source copy words in the text or adopt vocabulary to generate new summary information to solve the OOV (out of vocabulary) problem that often occurs in summary tasks. Additionally, to efficiently complete position information encoding, the model utilizes relative position representation in the attention layer to introduce sequence information of the texts. The model can be leveraged to control many important summary attributes, including lengths, topics, and specificity. Experiments on the public dataset MACSum show that

① 收稿时间: 2023-09-28; 修改时间: 2023-11-09; 采用时间: 2023-11-24; csa 在线出版时间: 2024-01-30

CNKI 网络首发时间: 2024-02-01

compared with previous methods, the proposed model performs better at ensuring the summary quality. At the same time, it is more in line with the attribute requirements given by users.

Key words: deep learning; controlled text summary; Transformer model; relative position representation; pointer generator network (PGN)

随着近年来互联网的快速发展, 互联网产生的文本数据越来越多, 文本信息过载问题日益严重, 导致用户需要花费大量时间去获取自己想要的信息. 如何从大量数据中提取出有效的关键信息成为近年来一个研究热点. 文本摘要技术旨在将大文档的信息压缩成包含关键信息的文档, 从而帮助用户更快、更容易地消化文档中的核心信息. 然而, 不同的用户对于同一篇文章的摘要有不同喜好, 例如主题、摘要长度和具体性等. 因此, 生成个性化的摘要以满足不同用户的偏好也是摘要模型应该具备的能力.

按照输出摘要类型可以将摘要分为抽取式摘要和生成式摘要. 抽取式摘要从最初的文本中抽取关键句和关键词组成摘要, 这种方法生成的摘要内容全部来源于原文. 该方法天生具有语法错误率低的效果. 生成式可以自由生成摘要并能够生成新颖的单词和句子. 与抽取式摘要相比, 生成式摘要更加的灵活多变, 更可能产生流畅和连贯的摘要.

伴随 AI 的快速发展, 神经网络模型在各个领域都取得了显著的成就. 在自然语言处理领域, Seq2Seq 模型结合一些其他网络或注意力机制, 在该领域表现出色. 这种模型已经在机器翻译、图像解释等领域取得了巨大的成功, 且在文本摘要领域也有不错的效果.

然而, 传统 Seq2Seq 模型在可控文本摘要任务中面临着 3 个问题: 首先, 大多数可控文本摘要方法集中于单属性控制, 对于多属性控制方法的研究较为有限; 其次, 多属性可控文本摘要方法在进行多属性控制时, 会出现控制能力下降, 无法合理兼顾所有控制属性的问题; 最后, 传统模型无法处理单词表外的单词.

为应对上述问题, 本文提出了一种基于扩展 Transformer 和指针生成网络的多要素可控文本摘要模型. 本论文的创新之处包括: 1) 使用扩展后的 Transformer 接收源文本和指导信号, 保证在输出摘要的时候通过交互注意力同时去关注源文档和指导信号, 从而使摘要更加忠诚于指导信号. 2) 引入指针生成网络, 使模型具有从源文档中复制文本的能力. 3) 在注意力层中使

用相对位置编码, 高效生成具有上下文语义信息的单词嵌入向量. 通过与其他几个模型的对比实验表明, 本文提出的模型在确保 ROUGE 指标不下降, 甚至有所提升的前提下, 可以生成更加满足用户偏好的个性化摘要.

1 相关工作

在 2014 年, Sutslevers 等人^[1]提出了一种广泛影响的序列到序列模型 (sequence-to-sequence), 成功实现了英-法互译的文本语言任务. 2015 年 Bahdanau 等人^[2]将 Seq2Seq 模型与注意力机制结合, 这一组合在翻译任务上取得了巨大的成功. 这种将 Seq2Seq 模型与注意力机制的结合的方法为后来的研究者提供了巨大的启发. 同年, Luong 等人^[3]打破了 Seq2Seq 的单任务框架, 构建了一个序列到序列的多任务的框架. 2016 年 Chopra 等人^[4]在 Bahdanau 等人的模型基础上进行改进, 将其应用于文本摘要任务. Nallapati 等人^[5]提出带有注意力的解码器编码器循环神经 (RNN) 网络, 该模型的编码器和解码器完全由 RNN 神经网络组成, 在当时的文本摘要任务上达到了 state-of-the-art 的表现. 2017 年, See 等人^[6]基于指针网络提出了指针生成网络模型, 该模型用于解决文本摘要任务中的未登录词问题. Vaswani 等人^[7]提出了一个完全摒弃 RNN 和 CNN 等网络的结构模型 Transformer, 该模型完全依赖注意力机制来实现翻译任务, 并取得了出色的效果. 这一成功使得注意力机制成为一个研究热点. 随后, 出现了一系列预训练模型, 如 BART^[8]和 GPT^[9], 它们在微调之后可以很好完成自然语言处理领域的任务. Transformer 的核心结构是 self-attention 机制, 但由于其无法对输入 token 的相对位置或绝对位置信息进行建模. 因此, 通常需要额外添加位置嵌入 (positional embedding) 来引入位置信息. 2018 年, Shaw 等人^[10]通过在自注意力机制中引入 token 之间的相对位置向量, 构建输入序列的连接图, 高效地完成了位置信息的编码, 提高了模型的性能.

过去的文本摘要系统生成的摘要不受用户控制,

不同的用户生成同一份摘要,这无法满足用户的个性化偏好.因此,生成可定制的个性化摘要已经成为摘要系统的必备功能.2019年,Makino等人^[11]通过一种全局优化的方法实现了摘要长度的可控性.2020年,He等人^[12]通过提示(prompt)方法实现了实体控制的生成摘要.2021年,Cao等人^[13]提出了预训练阶段的摘要生成解码器可以用于实现摘要风格控制的两种方法.2022年,Zhong等人^[14]实现了一个无监督的多粒度框架GRANUSUM,用于控制生成摘要的粒度.这些研究通常集中在单一属性的控制方面.值得注意的是,Chan等人^[15]在2021年通过受限马尔可夫决策过程的方法来对不同属性进行可控摘要生成,但是否可以实现多属性的控制仍然是一个未知数.Goyal等人^[16]通过混合多个解码器来研究多特征控制,但他们仅在解码器上进行了改进,使模型控制能力的提升比较有限.Dou等人^[17]通过混合多个编码器的方法实现了多摘要多个属性的单独控制.过去的研究大多过于专注于单个特定

属性的控制,而控制多个属性仍未得到充分探索.数据集方面,现有工作都是在改编自一般摘要数据集的伪数据集上进行评估.为了解决这个问题,2022年Zhang等人^[18]提出了第1个多属性控制的人工注释的摘要数据集,填补了多属性控制方面数据集的空缺.该论文还提出了一个自动评估系统,用于评估模型生成可控制摘要的能力.

2 模型算法设计

本文提出的模型总体架构如图1所示,主要包括两个关键模块:扩展Transformer模块和指针生成网络模块.其中扩展Transformer通过BART进行实例化.输入的文本和指导信号首先进入扩展Transformer模块的两个编码器进行特征提取,然后将输出的特征向量与解码器进行交互后作为指针生成网络的输入.指针生成网络结合生成式和抽象式摘要方法,使模型具有从源文本中复制单词的能力,同时也具有生成新单词的能力.

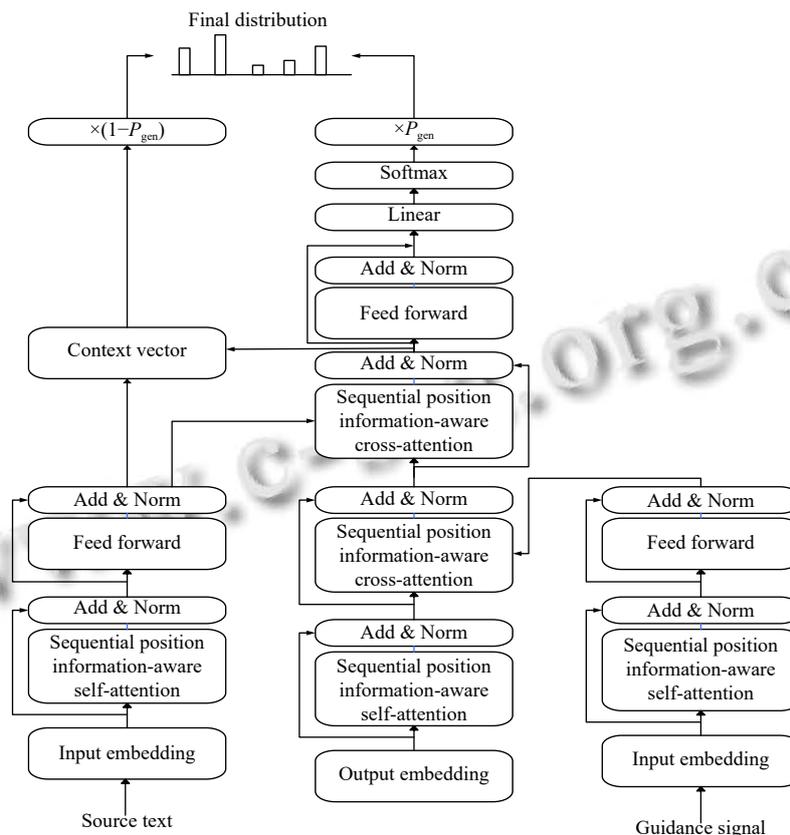


图1 模型整体架构

2.1 扩展 Transformer

在本文模型中,为了使生成的摘要更加忠诚于文

本,我们采用两个编码器,其中一个编码器用于对源文档进行编码,而另一个编码器则用于对用户输入的指

导信号进行编码. 每个编码器都包含 G_{enc} 层, 每一层由两个部分组成, 分别为序列位置感知的注意力和前馈神经网络. 编码器中每一层的过程可以通过以下公式表示:

$$x = LN(x + Self-Attention(x)) \quad (1)$$

$$x = LN(x + FFN(x)) \quad (2)$$

$$FFN(x) = \max(0, xW_1 + b_1)W_2 + b_2 \quad (3)$$

其中, LN 表示层归一化. FFN 表示前馈神经网络, 与 Transformer 中的相同, 由两个线性变化组成. 第 1 个全连接层的激活函数为 $ReLU$ 激活函数. 其中 W_1 、 W_2 、 b_1 、 b_2 都为可学习参数. 需要注意的是在编码过程中, 两个编码器不会相互影响. 除了最高层外, 剩下的 $G_{enc} - 1$ 层和词嵌入层参数都是共享的. 这样设计的好处是: 1) 这样可以减少计算和内存需求. 2) 我们认为源文档和指导信号之间的差异主要集中在模型的高层表示中, 这些差异可以在解码器的最顶层被充分捕获.

通常的 Transformer 只需处理一个编码器的输出, 但是本文模型的解码器必须同时处理源文档和引导信号两个输入. 解码器是由 G_{dec} 个相同的层组成, 每一层包含有 4 个关键部分. 输入首先经过序列位置信息感知自注意力部分, 然后解码器将通过序列位置信息感知交互注意力部分关注用户提供的指导信号并生成相应的表示, 这一方法使得指导信号可以指导解码器应该关注源文档的哪一部分. 接下来, 解码器将根据指导感知表示关注整个源文档. 最后, 输出表示将被馈送到前馈神经网络中进行进一步处理, 具体过程可以通过以下公式表示:

$$y = LN(y + Self-Attention(y)) \quad (4)$$

$$y = LN(y + Cross-Attention(y, g)) \quad (5)$$

$$y = LN(y + Cross-Attention(y, x)) \quad (6)$$

$$y = LN(y + FFN(y)) \quad (7)$$

$$FFN(y) = \max(0, yW_1 + b_1)W_2 + b_2 \quad (8)$$

其中, LN 表示层归一化. FFN 是前馈神经网络, 由两个线性变化即两个全连接层构成. 第 1 个全连接层的激活函数为 $ReLU$ 激活函数. 公式中 W_1 、 W_2 、 b_1 、 b_2 都为可学习参数, g 为指导信号的表征, x 为输入文本的表征.

2.2 指针生成网络

编码器逐词读取源文本之后, 生成一个编码器隐

藏状态的序列 h_i . 在每一个时间步 t 上, 解码器将摘要中之前的单词作为输入, 并用它们来更新时间步 t 解码器的隐藏状态 s_t . 然后解码器隐藏状态会被用来计算注意力分布, 通过式 (9) 和式 (10) 计算得到.

$$e_i^t = v^t \tanh(W_h h_i + W_s s_t + b_{attn}) \quad (9)$$

$$a_t = Softmax(e^t) \quad (10)$$

其中, v^t 、 W_h 、 W_s 、 b_{attn} 都为可学习参数. 然后将得到的注意力分布与编码器的隐藏状态进行加权求和, 可以得到上下文向量 (context vector) h_t^* :

$$h_t^* = \sum_i a_i^t h_i \quad (11)$$

将上下文向量和解码器隐藏状态拼接后输入两个线性层可以得到词汇分布 P_{vocab} :

$$P_{vocab} = Softmax(V'(V[s_t, h_t^*] + b) + b') \quad (12)$$

其中, V' 、 V 、 b 、 b' 为可学习参数. P_{vocab} 为词汇表中所有单词的概率分布.

指针生成网络为了解决 OOV 问题, 模型可以直接复制源文本中的单词也可以从词汇表中选择单词. 对于每一个时间步 t , 生成摘要的概率 P_{gen} 通过上下文向量 h_t^* 、解码器状态 s_t 和解码器输入 x_t 三者得到:

$$P_{gen} = \sigma(w_h^T \times h_t^* + w_s^T s_t + w_x^T x_t + b_{ptr}) \quad (13)$$

其中, w_h^T 、 w_s^T 、 w_x^T 、 b_{ptr} 为可学习参数, σ 为激活函数 Sigmoid 函数. 接下来, 生成单词 w 的概率等于从词汇集生成 w 的概率再加上指向原文本任何位置出现该词的概率, 我们用式 (14) 表示:

$$P(w) = P_{gen} P_{vocab}(w) + (1 - P_{gen}) \sum_{i:w_i=w} a_i^t \quad (14)$$

值得注意的是, 在式 (14) 中, 如果 w 是一个词汇外的单词, P_{vocab} 将等于 0, 同样, 如果 w 没有出现在源文本中, 则后面部分将等于 0. 在训练过程中, 时间步 t 的损失值为目标单词 w_t^* 的负对数似然值:

$$loss_t = -\log P(w_t^*) \quad (15)$$

整体损失值为:

$$loss = \frac{1}{T} \sum_{t=0}^T loss_t \quad (16)$$

2.3 序列位置信息感知的注意力

在 Transformer 模型中, 输入进编码器的嵌入由两

个向量相加,分别为单词嵌入 (word embedding) 和位置嵌入 (position embedding). 本文提出的模型中,单词嵌入采用了随机初始化的方法,但同时将其设置为可训练的,这意味着在训练过程中会不断优化嵌入表示. 这一处理方式与 Transformer 模型相似. 然而,与传统 Transformer 不同的是,本文模型不再使用位置嵌入来打破自注意力机制的置换不变性,而是采用了更加高效的方法来输入序列的位置信息,以达到生成更优质摘要的效果.

序列位置感知的注意力扩展普通注意力机制以考虑输入元素之间的成对关系. 直观来说,它将输入建模为有标记的、有向的、完全连接的图. 对于两个输入的元素 t_i 和 t_j 之间的边通过两个向量来表示 $q_{ij}^V, q_{ij}^K \in R^{d_a}$, 在多个 head 之间这些向量是共享的, $d_a = d_b$. 通过引入边的表征,序列位置信息感知注意力机制计算方法如下:

$$z_i = \sum_{j=1}^n q_{ij} (t_j W^V + q_{ij}^V) \quad (17)$$

$$q_{ij} = \frac{\exp e_{ij}}{\sum_{k=1}^n \exp e_{ik}} \quad (18)$$

$$e_{ij} = \frac{t_j W^Q (t_i W^K + q_{ij}^K)^T}{\sqrt{d_z}} \quad (19)$$

即对于各个 Value 和 Key 来说,都会引入一个相互的位置关系表示.

考虑到计算量、内存消耗以及远距离的精确位置信息效用不是很足等因素,本文对最远的相对位置距离限制为 k .

对于公式中的 q_{ij}^V, q_{ij}^K 的定义如下:

$$q_{ij}^K = m_{clip(j-i,k)}^K \quad (20)$$

$$q_{ij}^V = m_{clip(j-i,k)}^V \quad (21)$$

$$clip(x,k) = \max(-k, \min(k, x)) \quad (22)$$

模型在训练过程去学习相对位置表示 $m^K = (m_{-k}^K, \dots, m_k^K)$ 和 $m^V = (m_{-k}^V, \dots, m_k^V)$ 其中 $m_i^K, m_i^V \in R^{d_a}$, 具体如图 2 所示.

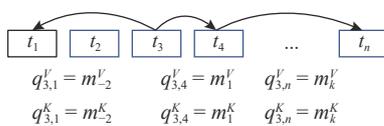


图 2 相对位置向量实例

图 2 表示了相对位置或元素之间距离的实例边. 本文提出的模型会学习距离为 k 以内的每个相对位置的表示.

3 实验分析

3.1 数据集

实验所采用的数据集是 MACSum^[18], 这是第 1 个专门用于多属性文本摘要控制的人工注释数据集. 该数据集包含了新闻文章和对话两个领域的源文本. 具体来说, MACSum 源文本来自 CNNDM 的部分被称为 MAC-Doc, 来自 QMSum 的部分被称为 Mac-Dial.

MACSum 涵盖了 5 个不同的控制属性, 分别是长度、提取性、特异性、主题和演讲者.

长度: 表示摘要的字数, 服务于用户阅读该摘要的时间预算.

提取性: 表示从源文本中提取摘要的比例, 当用户希望生成的摘要更多地从源文本中提取内容时, 这个控制属性可以起到作用.

特异性: 表示我们需要在摘要中包含细节和描述性内容的比例.

主题: 用户特别感兴趣的源内容.

演讲者: 对话中用户更喜欢的某些发言者.

3.2 实验评价指标

我们采用 ROUGE^[19] 评估中的 ROUGE-1、ROUGE-2、ROUGE-L 的 $F1$ 值来度量文本摘要的质量. 同时, 还采用一个自动指标系统, 用于评估模型生成可控摘要的能力. 对于每个控制属性, 该系统定义了属性度量函数来表示控制程度. 在属性度量函数基础上提出了控制错误率 (CER) 和控制相关性 (CC). CER 使用度量函数来衡量生成摘要与标准摘要之间在控制程度上的距离. 一个优秀的模型应该具有较小的 CER , 表明其生成的摘要与用户期望的控制属性更加一致. CC 则是测量不同属性值的生成摘要之间度量函数的分布, 代表模型的控制属性定义相关的能力. 一个出色的模型应该具有与标准摘要相似的 CC 分布.

对于一个控制属性 r 和它的属性度量函数 f_r , 给定一个预测出来的摘要 \hat{y} , 标准摘要 y , 控制错误率的定义如下:

$$CER(\hat{y}, y) = \frac{|f_r(\hat{y}) - f_r(y)|}{f_r(y) + \varepsilon} \quad (23)$$

其中, ε 是一个小值, 用来避免 $f_r(y)$ 等于 0. 对于控制属性 r 与控制值对 $[v_1, v_2]$, 这两个值的生成摘要预测为

$[\hat{y}_1, \hat{y}_2]$, 控制相关性可以被定义为:

$$CC(\hat{y}_1, \hat{y}_2) = \frac{f_r(\hat{y}_1) - f_r(\hat{y}_2)}{Distance(v_1, v_2)} \quad (24)$$

其中, $Distance(v_1, v_2)$ 计算从控制值 v_1 到 v_2 的距离, 可以是负数. 当 CC 高于 0 的时候, 表明被评价的模型与控制属性有正相关性, 当 CC 低于 0 时, 表明两者之间有负相关性. 值得注意的是, 在实验中, 多个样本的 CER 和 CC 是其算数平均值.

综合使用这些指标, 我们能够全面评估模型的文本摘要质量和多属性可控性能

3.3 模型说明

我们基于 BART 模型构建了本文的扩展 Transformer 模型, 并按照其超参数来训练总结器. 整个模型总共包含 24 个编码层, 其中顶层用预训练的参数初始化, 但两个编码器之间的参数是单独训练的. 解码器同样由 24 层组成, 其中解码器的第 1 个交互注意力块采用了随机初始化, 而第 2 个交叉注意力块则使用了预训练参数初始化. 在实验过程中, 我们将最远的相对位置距离 k 设置为 64. 需要特别值得注意的是, 对话文本被扁平化处理, 通过使用“</s>”来分割对话转折, 这种处理方式可以提升模型效果. PGN 模型的词向量嵌入维度

为 512, 隐藏层的维度为 256.

3.4 对比模型

为了验证本文模型的有效性, 我们选择了 5 种不同的对比模型与本文提出的模型进行比较.

1) 扩展 Transformer (Trans-1): 融合普通注意力与双编码器-单解码器的基线模型.

2) PGN: 指针生成网络通过使用指针网络来使模型具备从源文本中复制单词的能力. 编码器和解码器都采用单层双向 LSTM.

3) BART: 通过去噪自动编码进行了预训练, 是目前较为优秀的抽象摘要模型.

4) MatchSum^[20]: 一种抽取模型, 可对 BERTExt 生成的候选摘要进行重新排序.

5) 硬提示 (HP)^[21]: 使用控制属性的描述作为硬提示, 每个属性的构成都是“属性: 值”. 同时使用“;”连接 5 个控制属性并且其添加到输入源. 利用 BART^[9] 作为骨架模型. 将该模型作为多属性可控文本摘要的基线模型与本文模型进行对比.

3.5 实验结果

上述模型与本文提出的模型在数据集上的对比实验结果如表 1、表 2 所示.

表 1 不同模型在 MAC-Doc 上的评估结果

模型	长度		具体性		抽取性		主题	质量		
	CC	CER	CC	CER	CC	CER	CER	R1	R2	R3
标准摘要	32.444	0.000	0.103	0.000	0.141	0.000	0.000	1.000	1.000	1.000
Trans-1	29.785	0.350	0.163	0.371	0.121	0.921	0.496	0.290	0.094	0.251
PGN	28.472	0.361	0.179	0.377	0.253	0.949	0.515	0.287	0.093	0.249
MatchSum	29.997	0.352	0.172	0.374	0.137	0.896	0.507	0.289	0.097	0.250
BART	0.000	0.348	0.000	0.366	0.000	0.915	0.495	0.297	0.101	0.251
HP	30.563	0.355	0.122	0.365	0.244	0.913	0.493	0.298	0.103	0.254
Trans-1+PGN	30.112	0.349	0.157	0.357	0.130	0.900	0.487	0.301	0.098	0.252
Ours	30.293	0.345	0.154	0.355	0.138	0.893	0.478	0.308	0.100	0.259

表 2 不同模型在 MAC-Dial 上的评估结果

模型	长度		具体性		抽取性		演讲者	主题	质量		
	CC	CER	CC	CER	CC	CER	CER	CER	R1	R2	R3
标准摘要	42.045	0.000	1.610	0.000	0.088	0.000	0.000	0.000	1.000	1.000	1.000
Trans-1	10.457	0.596	1.493	0.597	0.055	0.513	0.260	0.486	0.314	0.104	0.261
PGN	10.441	0.597	1.491	0.655	0.054	0.544	0.267	0.488	0.316	0.101	0.262
MatchSum	10.613	0.592	1.498	0.595	0.058	0.501	0.259	0.483	0.317	0.103	0.267
BART	0.000	0.603	0.000	0.593	0.000	0.511	0.261	0.480	0.324	0.109	0.281
HP	11.539	0.601	1.523	0.591	0.053	0.505	0.258	0.493	0.323	0.108	0.280
Trans-1+PGN	11.501	0.590	1.517	0.584	0.056	0.502	0.258	0.481	0.321	0.110	0.273
Ours	11.512	0.589	1.519	0.553	0.061	0.499	0.257	0.473	0.322	0.111	0.284

为了评估控制质量, 我们采用了两个关键指标: 控制错误率 (CER) 和控制相关性 (CC). 此外, 我们还使

用 ROUGE 分数来评估生成的摘要质量. 对于一个好的模型来说, 我们期望它生成的摘要具有较低的 CER

值,越高的 ROUGE 值,并且 CC 值越接近标准摘要。表 1 展示了在 MAC-Doc 数据集上的实验结果。本文提出的模型在 ROUGE 指标上与其他对比模型持平,甚至在 R1 和 R3 上有提升。其中 R1 达到了 0.308, R3 达到了 0.259。在 4 个控制属性上,CC 和 CER 几乎都达到了最佳的性能水平。这表明本文提出的模型在保证摘要质量的同时更加符合用户需求。PGN 模型由于结合了生成式策略和提取式策略,使得摘要生成的效果有了大幅的提升。通过与 Trans-1 的对比实验,我们可以发现,对于许多低频词,生成式方法生成并不可靠,而指针生成网络的复制机制能够有效解决这个问题。与 PGN 的对比实验结果显示扩展后的 Transformer 可以使生成的摘要更加忠诚于指导信号,从而解决摘要不符合控制信号的问题。此外模型引入序列位置感知注意力后相较于普通注意力可以有效提升生成摘要质量。值得注意的是,与 HP 模型相比,本文提出的模型在抽取性和主题的控制能力上表现得更出色,但在长度和具体性控制属性上略显不足。这可能是因为不同的控制属性之间的存在一定的矛盾关系,例如像长摘要需要包括更多的单词,这不可避免地就会包含更多的信息。相反,简短的摘要单词数较少,因此具体性自然会偏低。为了实现更好的整体控制,模型倾向于生成一种折中的摘要。表 2 呈现了在 MAC-Dial 数据集上的结果。与 MAC-Doc 数据集类似,本文提出的模型在大多数指标上都获得了最高分。但是长度控制方面性能略有下降。这可能是因为 CNNDM 预训练为模型学习控制新闻文章的属性提供了某些有用的信息。值得注意的是,CER 不应跨数据集进行比较,因为不同的数据集规模不同。

4 结论与展望

本文提出了一种基于扩展 Transformer 和指针生成网络的多属性可控文本摘要模型。模型包含两个编码器,分别对源文档和指导信号进行编码。在解码器中引入了两个交互注意力机制,以指导模型关注源文档的特定部分,从而生成满足用户个性化需求的摘要。此外,通过采用指针生成网络,我们的模型能够在自动生成摘要的同时还具备从源文本中复制单词的能力。同时,引入相对位置表示于注意力机制中,以高效地编码位置信息,从而生成更高质量的摘要。实验在 MACSum 数据集上进行了训练和测试,实验结果表明,该模型在

保证摘要质量的前提下更加符合用户的个性化偏好。尽管取得了显著进展,但多属性可控文本摘要仍然是一个具有挑战性的研究领域,值得我们继续深入探索和改进。

参考文献

- 1 Sutskever I, Vinyals O, Le QV. Sequence to sequence learning with neural networks. Proceedings of the 27th International Conference on Neural Information Processing Systems. Montreal: MIT Press, 2014. 3104–3112.
- 2 Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate. Proceedings of the 3rd International Conference on Learning Representations. San Diego: ICLR, 2015.
- 3 Luong MT, Le QV, Sutskever I, et al. Multi-task sequence to sequence learning. Proceedings of the 4th International Conference on Learning Representations. San Juan: ICLR, 2016.
- 4 Chopra S, Auli M, Rush AM. Abstractive sentence summarization with attentive recurrent neural networks. Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. San Diego: Association for Computational Linguistics, 2016. 93–98.
- 5 Nallapati R, Zhou BW, Dos Santos C, et al. Abstractive text summarization using sequence-to-sequence RNNs and beyond. Proceedings of the 20th SIGNLL Conference on Computational Natural Language Learning. Berlin: Association for Computational Linguistics, 2016. 280–290.
- 6 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: Association for Computational Linguistics, 2017. 1073–1083.
- 7 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.
- 8 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. Association for Computational Linguistics, 2019. 7871–7880.
- 9 Radford A, Narasimhan K, Salimans T, et al. Improving

- language understanding by generative pre-training. https://s3-us-west-2.amazonaws.com/openai-assets/research-covers/language-unsupervised/language_understanding_paper.pdf. [2023-08-10].
- 10 Shaw P, Uszkoreit J, Vaswani A. Self-attention with relative position representations. Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers). New Orleans: Association for Computational Linguistics, 2018. 464–468.
- 11 Makino T, Iwakura T, Takamura H, *et al.* Global optimization under length constraint for neural text summarization. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence: Association for Computational Linguistics, 2019. 1039–1048.
- 12 He J, Kryściński W, McCann B, *et al.* CTRLsum: Towards generic controllable text summarization. Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing. Abu Dhabi: Association for Computational Linguistics, 2020. 5879–5915.
- 13 Cao SY, Wang L. Inference time style control for summarization. Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Association for Computational Linguistics, 2021. 5942–5953.
- 14 Zhong M, Liu Y, Ge SY, *et al.* Unsupervised multi-granularity summarization. Proceedings of the 2022 Findings of the Association for Computational Linguistics: EMNLP 2022. Abu Dhabi: Association for Computational Linguistics, 2022. 4980–4995.
- 15 Chan HP, Wang L, King I. Controllable summarization with constrained Markov decision process. Transactions of the Association for Computational Linguistics, 2021, 9: 1213–1232. [doi: 10.1162/tacl_a_00423]
- 16 Goyal T, Rajani NF, Liu WH, *et al.* Hydrasum: Disentangling stylistic features in text summarization using multi-decoder models. arXiv:2110.04400, 2021.
- 17 Dou ZY, Liu PF, Hayashi H, *et al.* GSum: A general framework for guided neural abstractive summarization. Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Association for Computational Linguistics, 2020. 4830–4842.
- 18 Zhang YS, Liu Y, Yang ZY, *et al.* MACSum: Controllable summarization with mixed attributes. Transactions of the Association for Computational Linguistics, 2022, 11: 787–803.
- 19 Lin CY. ROUGE: A package for automatic evaluation of summaries. Proceedings of the 2004 Workshop on Text Summarization Branches Out. Barcelona: Association for Computational Linguistics, 2004. 74–81.
- 20 Zhong M, Liu PF, Chen YR, *et al.* Extractive summarization as text matching. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics, 2020. 6197–6208.
- 21 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.

(校对责编: 孙君艳)