



计算机科学

COMPUTER SCIENCE

基于知识辅助的结构化医疗报告生成

史继筠, 张驰, 王禹桥, 罗兆经, 张美慧

引用本文

史继筠, 张驰, 王禹桥, 罗兆经, 张美慧. [基于知识辅助的结构化医疗报告生成](#)[J]. 计算机科学, 2024, 51(6): 317-324.

SHI Jiyun, ZHANG Chi, WANG Yuqiao, LUO Zhaojing, ZHANG Meihui. [Generation of Structured Medical Reports Based on Knowledge Assistance](#) [J]. Computer Science, 2024, 51(6): 317-324.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[基于跨层级多视角特征的多语言事件探测](#)

Multilingual Event Detection Based on Cross-level and Multi-view Features Fusion

计算机科学, 2024, 51(5): 208-215. <https://doi.org/10.11896/jsjcx.230200131>

[基于标签信息融合与多任务学习的中文命名实体识别](#)

Chinese Named Entity Recognition Based on Label Information Fusion and Multi-task Learning

计算机科学, 2024, 51(3): 198-204. <https://doi.org/10.11896/jsjcx.230200114>

[SemFA:基于语义特征与关联注意力的大规模多标签文本分类模型](#)

SemFA:Extreme Multi-label Text Classification Model Based on Semantic Features and Association Attention

计算机科学, 2023, 50(12): 270-278. <https://doi.org/10.11896/jsjcx.230300239>

[基于多粒度对比学习的聊天对话摘要模型](#)

Chat Dialogue Summary Model Based on Multi-granularity Contrastive Learning

计算机科学, 2023, 50(11): 192-200. <https://doi.org/10.11896/jsjcx.230300241>

[融合无监督SimCSE的短文本聚类研究](#)

Study on Short Text Clustering with Unsupervised SimCSE

计算机科学, 2023, 50(11): 71-76. <https://doi.org/10.11896/jsjcx.220900214>

基于知识辅助的结构化医疗报告生成

史继筠¹ 张驰¹ 王禹桥¹ 罗兆经² 张美慧¹

1 北京理工大学计算机学院 北京 100081

2 新加坡国立大学计算机学院 新加坡 117417

(shijiyun@bit.edu.cn)

摘要 医疗报告自动生成是文本摘要生成技术的重要应用。由于医疗问诊数据与通用领域的的数据特征存在着明显的差异,传统的文本摘要生成方法不能充分理解并利用医疗文本中高复杂性的医疗术语,因此医疗问诊中包含的关键知识并没有得到充分的利用。此外,传统的文本摘要生成方法大多是直接生成摘要,并没有针对医疗报告结构化的特点自动选择过滤关键信息并生成结构化文本的能力。针对上述问题,提出了一种知识辅助的结构化医疗报告生成方法。该方法将实体引导的先验领域知识与结构引导的任务解耦机制相结合,实现了对医疗问诊数据的关键知识与医疗报告的结构化特点的充分利用。在IMCS21数据集上的实验验证了所提方法的有效性,其生成摘要的ROUGE分数与同类方法相比提升了2%~3%,生成了更准确的医疗报告。

关键词: 医疗报告生成;预训练模型;生成式摘要;领域知识先验;任务解耦机制

中图分类号 TP301

Generation of Structured Medical Reports Based on Knowledge Assistance

SHI Jiyun¹, ZHANG Chi¹, WANG Yuqiao¹, LUO Zhaojing² and ZHANG Meihui¹

1 School of Computer Science and Technology, Beijing Institute of Technology, Beijing 100081, China

2 School of Computing, National University of Singapore, Singapore 117417, Singapore

Abstract Automatic generation of medical reports is an important application of text summarization technology. Due to the obvious difference between the medical consultation data and data of the general field, the traditional text summary generation method cannot fully understand and utilize the highly complex medical terms in the medical text, so that the key knowledge contained in the medical consultation has not been fully used. In addition, most of the traditional text summary generation methods directly generate summaries, and do not have the ability to automatically select and filter key information and generate structured text according to the structural characteristics of medical reports. In order to solve the above problems, a knowledge-assisted structured medical report generation method is proposed in this paper. The proposed method combines the entity-guided prior domain knowledge with the structure-guided task decoupling mechanism, and realizes the key knowledge of medical consultation data, taking full advantage of the structured features of medical reports. The effectiveness of the method is verified on the IMCS21 dataset. The ROUGE score of the summary generated by our method is 2% to 3% higher than that of baseline methods, and a more accurate medical report is generated.

Keywords Medical report generation, Pre-training model, Generative summarization, Domain knowledge prior, Task decoupling mechanism

医疗报告是一类结构化的医学文档,总结了医疗问诊中的关键信息,其中主要包括患者的主诉、现病史、辅助检查、既往史,以及医生作出的病情诊断和治疗建议等字段内容。医疗报告的准确性和完整性对于医生的后续诊断以及患者的治疗具有重要意义。

当前,随着医疗信息化行业的蓬勃发展,医疗问诊数据呈现爆炸式增长,撰写医疗报告给医生带来了极大的负担。近年来,随着人工智能技术的不断发展,研究者开始将自动文本摘要技术应用到医疗对话中。自动文本摘要技术旨在从用户输入的文本中提取包含原文关键信息的简短内容,并过滤

无关的文本噪声,从而为用户提供简洁且核心思想完整的信息。当前文本摘要方法的研究主要关注通用文本领域,且得益于近年来神经网络模型的快速发展,取得了显著的成果。但医疗对话文本存在术语复杂、信息繁多的特性,通用的文本摘要方法不能充分理解医疗文本中高复杂性的医疗术语,这就降低了医疗文本摘要生成的准确性。同时,现有文本摘要方法大多是直接生成完整摘要,并没有针对性地自动化选择过滤相关关键信息并生成结构化文本的能力,因此难以直接将其迁移到医疗对话领域。

针对以上问题,本文提出了一种基于知识辅助的结构化

医疗报告生成方法。该方法以生成式摘要模型为基础,结合实体识别技术充分利用医疗问诊中的知识,实现了基于实体引导的结合领域知识的先验方法。另外,为了对医疗报告中先验的结构知识加以利用,本文还借助提示词机制与 Adapter 技术,提出了结构引导的任务解耦机制,能够根据医疗报告不同字段的生成侧重点捕捉关键的结构信息,从而生成结构化的医疗报告。

1 相关工作

1.1 文本摘要

文本摘要任务是自然语言处理领域中的热门研究方向。随着深度学习技术的发展,基于神经网络的文本摘要模型已经成为该领域的主流研究方法。

文本摘要按照实现方法可分为抽取式摘要^[1]与生成式摘要^[2]两大类。抽取式摘要直接从原始文本中选取关键的单词或句子组成摘要。传统的抽取式摘要方法使用图方法、聚类等方式完成无监督摘要,后来深度神经网络也被引入文本摘要中,取得了一定的成果。近年来,预训练模型逐渐被应用于抽取式摘要,并取得了优秀的性能表现。Liu^[3]提出了 BERTSUM 模型,将其 BERT 模型应用于抽取式摘要,对 BERT 的 Embedding 层和 Fine-tune 层进行修改,使得 BERT 模型能够适用于抽取式摘要任务,并在多个数据集上取得了优秀的性能。Zhong 等^[4]考虑到传统的抽取式摘要模型在句子建模时没有考虑句子间的关系,因此提出先进行句子配对,再使用 BERTSUM 打分。

抽取式摘要的语法、句法比较符合人类的习惯,但也存在着灵活性差、连贯性差、内容选择错误等问题,而生成式摘要则不局限于原文内容,且能够保持较好的连贯性。随着近几年深度学习技术的发展,基于编码器-解码器架构的模型被广泛应用于摘要生成任务,并取得了一定的成果,但也存在重复描述事实、未登录词、生成重复等问题。为此,See 等^[5]提出了指针生成器网络 PGN(Pointer Generator Network),在基于注意力机制的编码器-解码器基础上增加了 Copy 和 Coverage 机制,能够有效缓解未登录词问题与生成内容重复的问题。随后,Paulus 等^[6]进一步提出在 PGN 上引入强化学习的反馈机制,用于解决摘要生成过程中不连贯、语句重复等问题。Li 等^[7]对原始的 Coverage 机制加以改进,基于句子级别的注意力机制,实现了句子级别的 Coverage 机制,使得模型在摘要生成时可以从不同的角度关注原文,从而缓解生成文本重复的问题。

1.2 医疗对话摘要

医学对话摘要生成任务一直是自然语言处理领域中备受关注的研究方向。由于医疗领域包含大量专业术语,且存在医疗数据集稀缺、规模和质量不一致等问题^[8-10],因此医疗对话摘要领域面临着诸多挑战。随着当前深度学习技术的不断发展,LSTM 架构已成为许多自然语言处理应用的骨干模型,并在医学对话摘要生成任务中得到了广泛的应用。Song 等^[11]在医疗领域进行了早期的探索,将结合注意力机制的 LSTM 框架应用于生成式文本摘要,取得了一定的成果。Zhang 等^[12]提出了一种药物信息提取器(MIE),使用 LSTM 框架作为编码器模块,然后采用聚合模块聚合医患

对话的交互信息。

近年来,随着 Transformer 模型出现,预训练技术在自然语言处理领域大放异彩。Enarvi 等^[13]使用序列到序列模型实现医疗对话摘要生成,并发现 Transformer 模型的生成效果优于 LSTM 模型。这是因为 Transformer 模型可以更准确地捕捉长距离依赖关系,对于医疗对话这种具有长距离依赖的文本数据有着更好的处理能力。Chintagunta 等^[14]聚焦于公开医疗对话数据有限的问题,使用经过预训练的语言生成模型 GPT-3 来生成医疗咨询数据并进行模型的训练。

为了充分利用医疗语句自身的特点,Krishna 等^[15]首次尝试使用 T5 等深度学习模型总结生成 SOAP 形式的医疗文本摘要,通过充分利用 SOAP 注释的分段结构和每个 SOAP 注释相关的医患对话信息,从而生成完整的医疗 SOAP 文本。

然而,上述方法大多只考虑对话角色与对话主题等知识,而忽略了医疗文本自身包含的丰富医疗实体信息,使得其对外部医疗知识利用不够充分。此外,医疗报告作为一种结构化的医疗文档,其不同结构内容参考的重点不同。例如,“主诉”字段更加关注医疗问诊的患者自述部分,“现病史”字段则更加关注患者的疾病症状,而“诊断”字段则更加关注对话中的疾病名称。而当前主流的医疗对话摘要技术大多是直接完整地生成摘要,并没有针对性地自动化选择过滤相关关键信息并生成结构化文本的能力。

为此,本文研究并实现了基于知识辅助的结构化医疗报告的生成方法,通过实体引导的领域知识先验和结构引导的任务解耦机制相结合的方法,来解决当前通用的文本摘要技术在医疗报告生成任务中存在的问题,从而更准确、完整地生成结构化的医疗报告。

2 知识辅助的结构化医疗报告生成模型

2.1 问题定义

定义 1(患者自述 S) 表示医疗问诊中患者自述句子序列, $S = s_1, s_2, \dots, s_m$ 。其中, s_j 表示患者自述的第 j 句话, m 表示患者自述的总句子数。

定义 2(问诊对话 D) 表示医疗问诊中问诊对话部分的句子序列, $D = d_1, d_2, \dots, d_t$ 。其中, d_j 表示问诊对话中的第 j 句话, t 表示问诊对话的总句子数。

定义 3(医疗报告 R) 医疗报告为结构化的文档, $R = f_1, f_2, \dots, f_{|R|}$ 。其中, f_i 表示医疗报告中第 i 个结构段的内容, $|R|$ 表示医疗报告的总结构段数。

定义 4(医疗报告生成) 接收患者自述 S 和医疗问诊对话 D 作为输入,输出生成医疗报告 $f_1, f_2, \dots, f_{|R|}$ 各结构段落,最终组成结构化的医疗报告 R 。

2.2 模型架构

为了充分利用医疗问诊对话中的语义信息,本文将医疗报告生成问题建模为一个由基于报告结构的自适应生成式文本摘要模型。首先,模型需要充分理解医疗问诊对话中的相关实体,并将实体、角色、句子类型等领域知识加入学习过程中,从而更完整地理解医疗问诊对话中的语义信息。此外,模型在生成摘要时,需要根据医疗报告的规定结构,针对每个医疗报告的不同字段进行相应的调整,生成最合适的文本,从而

达到任务解耦的目的。因此,本文提出的生成式摘要模型主要通过结合领域知识先验和任务解耦机制来优化生成效果。

本文提出的知识辅助的结构化医疗报告生成模型的整体架构如图1所示,该模型主要包含3个部分,即实体识别模块、数据抽取模块与摘要生成模块。其中,实体识别模块使用基于BERT-CRF的实体识别模型来进行实体识别,得到输入文本的 BIO 标签;数据抽取模块主要基于实体识别模块得到的 BIO 标签以及预先设置的抽取规则来进行数据抽取,从而根据医疗报告的结构来构建摘要生成模型的输入;而摘要生成模块则主要基于 BART^[16]摘要生成模型来进行最终的医疗报告生成。

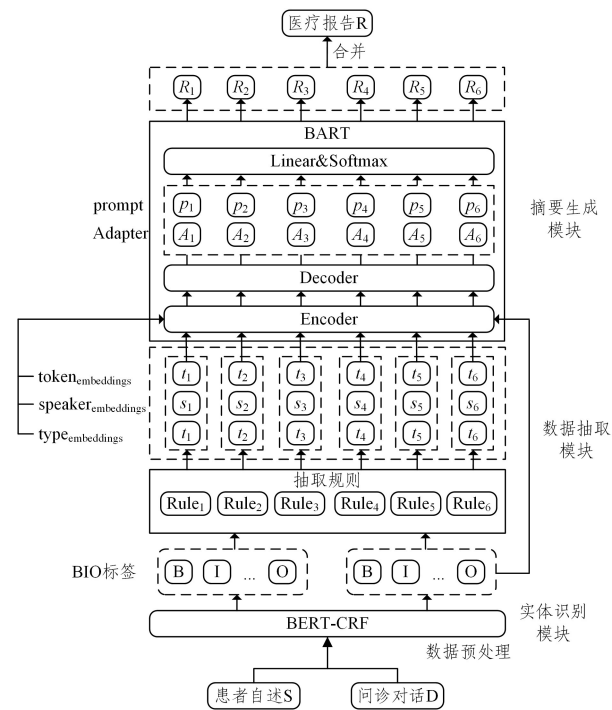


图1 知识辅助的结构化医疗报告生成模型的整体架构

Fig. 1 Overall architecture of knowledge-assisted structured medical report generation model

框架的输入为患者自述与问诊对话,文本被输入实体识别模块后,实体识别模块首先对其进行实体识别,得到每个字的 BIO 标签,然后进入数据抽取模块,根据 BIO 标签和设置的抽取规则进行数据抽取,从而对先验知识进行表征。数据抽取模块输出的结果为根据医疗报告每个字段的结构抽取后的文本以及文本中每一个字的 BIO 标签,同时作为摘要生成模块的输入。摘要生成模块在接收到输入后在编码器中将相关的先验知识信息添加到输入的句子中,并在解码器部分根据医疗报告的结构引导分别生成报告的每个字段,从而实现任务解耦的目的。

2.3 实体引导的领域知识先验

2.3.1 基于 BERT-CRF 的实体识别

在领域知识先验的方法中,首先需要输入语料进行实体识别,为后续摘要模型的知识嵌入提供知识辅助,并在针对医疗报告不同字段的数据抽取中以实体类型与字段之间的关系为先验知识,为各个子任务的生成提供指导。

BERT-CRF^[17]模型是实体识别任务的常用模型,在各个领域的实体识别任务中都取得了优秀的效果。BERT-CRF

模型主要分为两层结构,如图2所示。首先 BERT 层使用预训练 BERT 模型对标注数据进行训练和编码,从而获取准确的字符语义表示,然后 CRF 层对 BERT 层的输出结果进行状态转移约束。

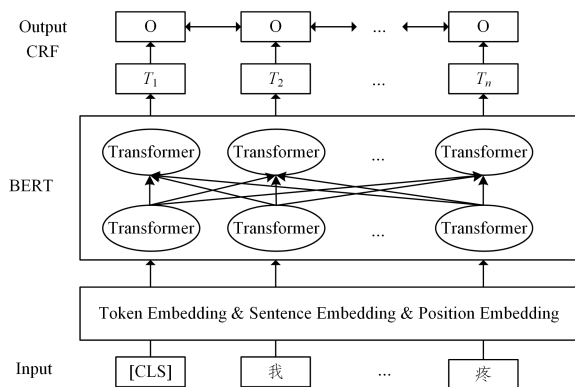


图2 基于 BERT-CRF 的实体识别模型的结构

Fig. 2 Structure of entity recognition model based on BERT-CRF

在本文提出的医疗报告生成模型中,患者自述和问诊对话被输入到基于 BERT 与 CRF 的实体识别模型中进行实体识别,得到输入文本每个字对应的 BIO 标签,从而为后续摘要模型的知识嵌入提供重要的参考知识。同时,在针对报告中不同字段的数据抽取时,数据抽取模块根据实体类型和字段之间的关系设置了抽取规则,这将作为一种先验的结构知识为摘要生成模型提供指导。本文共设置了 13 种 BIO 标签,分别表示药品名称、医学检查、疾病症状等实体类型。基于实体类型的数据抽取模块则利用医疗报告先验的结构知识,在实体识别模型得到输入文本每个字的 BIO 标签后,基于实体类型,根据设置的规则对医疗报告中每一个字段的输入进行抽取。

2.3.2 基于外部知识嵌入的知识增强

本文提出的摘要生成框架在完成数据抽取后,需要进行文本摘要生成任务。如何构建合适的摘要生成模型将直接影响到摘要生成的效果。目前,基于编码器-解码器架构的生成模型在文本摘要生成等领域获得了广泛的应用,已经成为文本生成领域的基本框架。但由于传统的基于编码器-解码器架构的生成模型通常需要大量的数据来进行训练,而目前关于医疗报告生成的中文公开数据集较少,无法满足训练的需要,因此本文使用了基于预训练的生成式模型来完成医疗报告生成的任务,其架构如图3所示。

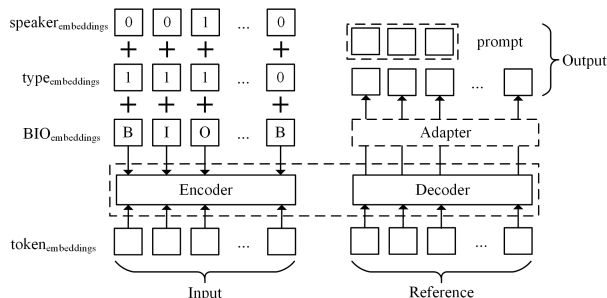


图3 预训练生成式模型的架构

Fig. 3 Architecture of pre-training generative model

本文在选取的主干模型的基础上,对主干模型等编码器进行微调,嵌入实体类型、角色、句子类型等知识,以达到知识

增强的效果。形式化来看,设输入模型的问诊语料序列为 $W = \{d_1, d_2, \dots, d_n\}$, 其中 d_i 表示序列中的第 i 句话, n 表示序列长度。 d_i 经过编码, 得到词嵌入序列 $token_{embeddings} = \{x_{i,1}, x_{i,2}, \dots, x_{i,t}\}$, 其中 $x_{i,j}$ 表示句子 d_i 中的第 j 个词, t 表示 d_i 的总词数。根据句子 d_i 的角色和句子类型, 有序列 $speaker$ 与 $type$:

$$speaker = \begin{cases} \{1, 1, \dots, 1\}, & \text{角色为医生时} \\ \{0, 0, \dots, 0\}, & \text{角色为患者时} \end{cases} \quad (1)$$

$$type = \begin{cases} \{1, 1, \dots, 1\}, & d_i \text{ 为患者自述时} \\ \{0, 0, \dots, 0\}, & d_i \text{ 为问诊对话时} \end{cases} \quad (2)$$

这两个序列经过编码后可得到序列 $speaker_{embeddings}$ 和 $type_{embeddings}$ 。同时, 根据模型的输入, 有句子 d_i 对应的实体 BIO 标签序列 bio , 其经过编码后得到序列 $bio_{embeddings}$ 。然后将实体信息、角色、句子类型的知识添加到词嵌入序列中。

$$token_{embeddings} = token_{embeddings} + bio_{embeddings} + speaker_{embeddings} + type_{embeddings} \quad (3)$$

通过外部知识的输入, 模型能够学习的知识进一步增加, 模型能够自动学习添加的外部知识对生成结果的影响, 调整注意力分布, 并在生成结果中倾向于拷贝完整的实体, 从而提升生成报告的效果。

2.4 结构引导的任务解耦机制

2.4.1 基于提示词机制的摘要生成

本文的摘要生成模型在接收到数据抽取模块的结果并将其作为输入后, 需要对医疗报告中不同字段进行子任务生成。为了利用医疗报告自身的结构信息, 有效区分不同的报告字段, 本文在摘要生成模型中加入了提示词 (Prompt) 机制^[18], 其可以根据设置的提示词、关键词等信息, 生成更准确和可控的文本摘要, 从而提高模型的生成效果和可控性。

对于训练阶段输入的医疗报告, 需要将原始的医疗报告数据按照指定的 6 个字段进行分割, 并在生成摘要时为每个字段的生成内容加上相应的提示词。在主干模型的解码器中, 模型在解码器的每个位置计算与所有提示词的注意力得分, 得到提示词向量, 指导生成下一个词语。在生成医疗报告时, 将生成的第一个词语指定为“主诉:”“现病史:”“辅助检查:”“既往史:”“诊断:”以及“建议:”, 作为每个字段的起始点。具体地, 每个提示词向量的计算式如下:

$$p_i = \sum_{j=1}^n \beta_j h_j \quad (4)$$

其中, β_j 表示当前位置的隐状态 s_i 和编码器的输出向量 h_j 与第 i 个提示词的注意力得分。

在模型训练时, 需要使用带有提示词的训练数据来训练模型。具体地, 需要在不同字段的生成内容前加上相应的提示词, 并使用类似于最大似然估计的方法来优化模型参数。在生成医疗报告时, 需要将不同字段的输入序列分别输入模型中, 并在生成摘要时, 将提示词作为条件加入到解码器中, 生成该字段的内容。在使用提示词机制生成结构化的医疗报告时, 主干模型的数学表达式如下:

$$\hat{p}(y_i | y_{i-1}, x_1, x_2, \dots, x_n, p_i) = \text{softmax}(f(y_{i-1}, s_i, p_i)) \quad (5)$$

其中, y_{i-1} 表示上一个输出词语, s_i 表示当前解码器的隐状态, p_i 表示提示词向量。

解码器需要在上一个词语和提示词向量之间进行注意力计算, 得到一个加权向量, 表示模型应该关注的内容。具体地, 可以用注意力函数来计算上下文词向量 c_i 和提示词向量 p_i 的加权平均值。

$$c_i = \sum_{j=1}^n \alpha_{ij} h_j \quad (6)$$

$$\alpha_{ij} = \frac{\exp(a(s_{i-1}, h_j, p_i))}{\sum_{k=1}^n \exp(a(s_{i-1}, h_k, p_i))} \quad (7)$$

其中, h_j 表示编码器的输出向量, $a(\cdot)$ 表示计算注意力得分的函数。提示词向量 p_i 可以直接用解码器的输出向量来计算, 或者通过一层全连接网络进行计算。最后, 使用所生成的序列作为摘要输出。

提示词机制的加入, 使得输入的医疗语料能够按照医疗报告的结构进行针对性的生成, 从而将医疗报告生成任务解耦为 6 个字段对应的子任务, 提升生成报告的效果。

2.4.2 基于 Adapter 的子任务自适应调整

在基于结构引导的任务解耦机制中, 本文还在摘要生成模型中加入了 Adapter^[19] 技术, 使模型能够对每个字段生成的子任务进行自适应的调整。Adapter 是一种在预训练模型中实现参数共享的技术。传统的预训练模型在处理结构化医疗报告的生成这种包含多个子任务的任务时, 需要对所有子任务一起训练, 而共享的参数可能会出现冲突, 影响模型的效果。为了解决该问题, Adapter 技术中使用一组额外的参数来记录每个任务的特征, 而不影响模型的共享参数。在训练中, 模型的共享参数会被冻结, 只对不同于子任务的 Adapter 参数进行调整, 实现子任务自适应的效果。

基于 Adapter 的子任务自适应调整是一种使用 Adapter 技术来实现任务解耦的方法, 通过在预训练模型中添加 Adapter 来为每个子任务提供单独的参数, 从而使模型可以同时处理多个任务, 而不会受到其他任务的影响。调整模型时, 可以选择使某些 Adapter 模块处于开启状态, 而其他模块处于关闭状态, 从而仅仅使用与当前子任务相关的参数, 这样可以有效地减少不相关参数的干扰, 提高模型的效果。

设医疗报告生成任务为 T , 分解后得到 N 个子任务, 分别为 T_1, T_2, \dots, T_N 。Adapter 会针对每个子任务 T_i , 添加一个独立的 Adapter 模块 $Adapter_i$, $Adapter_i$ 中存储了对应于子任务 T_i 所需要的知识。该模块与共享的预训练主干模型 M 共同构成一个新的模型 M_i' , 即 $M_i' = M \oplus A_i$, 其中 \oplus 表示两个模块的拼接操作。并且, 在 M_i' 中, Adapter 模块 A_i 只被连接到任务 T_i 所对应的部分, 而不会影响到其他部分。

对于一个预训练模型 M , 该模型包含 k 个层, 每个层的输出为 $h^{(i)}$, 其中 $i=1, 2, \dots, k$ 。Adapter 为每个子任务 T_i 添加的 Adapter 模块, $Adapter_i$ 则由两个线性变换矩阵组成, 分别为 $W_i^{(i)}$ 和 $W_i^{(i)}$, 其参数形状分别为 (d_i, d_h) 和 (d_h, d_i) , 其中 d_i 和 d_h 分别表示输入和输出的维数。设当前要处理的

任务为 T_i , 对应的 Adapter 模块为 A_i , 则对于任意输入 x , 模型输出的计算式如式(8)和式(9)所示:

$$\mathbf{h}^{(1)} = M^{(1)}(x) \quad (8)$$

$$\mathbf{h}^{(i)} = \text{Adapter}_i(\mathbf{h}^{(i-1)}) \odot \mathbf{h}^{(i-1)} + \mathbf{h}^{(i-1)} \quad (9)$$

其中, \odot 表示元素乘法。而每个 Adapter 模块的计算式如式(10)所示:

$$\text{Adapter}_i(\mathbf{h}) = \mathbf{W}_2^{(i)} \text{relu}(\mathbf{W}_1^{(i)} \mathbf{h}) \quad (10)$$

其中, $\text{relu}(x) = \max(0, x)$ 表示一个 ReLU 函数, 用于非线性激活。通过上式可知, Adapter_i 把输入 x 转换为 \mathbf{h} , 并且和前一层的输出 $\hat{\mathbf{h}}$ 拼接起来, 然后通过两个线性变换矩阵进行计算, 得到最终的结果 h_i 。这个结果与原来的输出 $\mathbf{h}^{(i-1)}$ 加以融合, 得到 $\mathbf{h}^{(i)}$ 。

为了获得更好的效果, 可以根据不同的子任务动态地开启或关闭相应的 Adapter 模块。设当前需要完成的任务为 T_i , 则其对应的 Adapter 模块 A_i 将被开启, 而其他所有的 Adapter 模块将被关闭, 以消除它们对输出结果的影响。Adapter 模块的控制是通过二值掩码(Mask)来实现的。定义一个长度为 N 的二元数组 $mask$ 来控制 Adapter 模块的开关状态, 其中 $mask_i = 1$ 表示第 i 个 Adapter 模块处于开启状态, $mask_i = 0$ 表示处于关闭状态。在模型推理时, 可以通过一个采样比率(Dropout Rate)参数 $p_i \in [0, 1]$ 自适应地生成掩码 $mask_i$, 如式(11)所示:

$$mask_i = \begin{cases} 1, & \text{概率为 } 1 - p_i \\ 0, & \text{概率为 } p_i \end{cases} \quad (11)$$

由此, 可通过掩码 Mask 实现 Adapter 模块的开启和关闭。当需要执行任务 T_i 时, 可以将 M_i 设置为 1, 而其他 $M_{j \neq i}$ 则保持为 0, 这样就可以实现仅仅使用与当前子任务相关的参数, 从而提高模型的效果。模型的输出如式(12)、式(13)所示:

$$x_{\text{Adapted}}^{(i)} = mask_i \odot \text{Adapter}_i(\mathbf{h}^{(i-1)}) \quad (12)$$

$$\mathbf{h}^{(i)} = x_{\text{Adapted}}^{(i)} + \mathbf{h}^{(i-1)} \quad (13)$$

可以看出, 通过调整采样比率 p_i , 可以控制每个 Adapter 模块参与计算的概率, 使得计算结果更符合当前子任务对应的特征。这样可以保证模型参数的精简, 减少不必要的计算, 提高整个任务的处理效率。

需要注意的是, 掩码 Mask 的生成和应用只在推理阶段完成, 而在训练阶段则不需要考虑掩码的影响。在训练过程中, 每个 Adapter 模块都会接收完整的输入, 并且参与同步更新。只有在推理时, 才需要根据当前任务动态地生成调整掩码, 并将掩码应用到 Adapter 模块中。这样, 就可以实现 Adapter 的自适应调整, 从而提高多个子任务学习的效果。

在结构化医疗报告的生成任务中, 本文针对医疗报告的 6 个字段预先设置了 6 个独立于数据集的 Adapter, 在每个字段的生成中模型会自动使用对应 Adapter 的参数, 而不会使整个模型的参数共享, 影响生成效果。因此, 基于 Adapter 的子任务自适应调整技术可以通过动态调整 Adapter 模块的参数来适应不同的子任务, 为任务的处理提供更好的支持, 提高任务处理的效率和准确度。

3 实验分析

3.1 数据集与评价指标

本文实验使用的问诊对话数据集 IMCS21^[20] 来源于中文医疗信息处理评测标准 CBLUE(Chinese Bi-omedical Language Understanding Evaluation)^[21] 的医疗报告生成任务 IMCS-MRG, 共包含 4 116 组细粒度标注的医患对话数据案例, 覆盖 10 种儿科疾病, 详细统计数据如表 1 所列。IMCS21 数据集的一组数据样例如表 2 所列。

表 1 IMCS21 数据集的详细统计数据

统计指标	数据
总疾病数	10
总对话数	4 116
总句子数	164 731
平均每个对话包含的句子数	40
平均每个对话包含的字符数	523
平均每个对话包含的字符数(包含患者自述)	580

表 2 IMCS21 数据集样例

Table 2 Example of IMCS21 dataset

医疗问诊数据	参考报告
患者自述: 6个月宝宝,5月21日确诊为支气管炎,住院治疗1个星期,治愈出院。6月5日因母亲咳嗽感染又一次被诊断为支气管炎,是病毒性的。医生开了复方鱼腥草糖浆和小儿止咳糖浆吃,我能否搭配射干利咽口服液和阿奇霉素服用?	(1)主诉:咳嗽一天。 (2)现病史:患儿一月前因支气管炎,住院治疗1个星期。一天前出现咳嗽症状,诊断为病毒性支气管炎,伴腹泻,口服复方鱼腥草糖浆和小儿止咳糖浆,目前症状无明显改善。
问诊对话(只给出片段): …… 医生:大夫验血诊断的病毒性支气管炎吗? 患者:是的。 医生:验血确实支持病毒感染。 医生:孩子咳嗽厉害吗? 患者:还好,不厉害。 ……	(3)辅助检查:血常规提示病毒性感染。 (4)既往史:支气管炎。 (5)诊断:小儿支气管炎。 (6)建议:继续服用原药物,加金双歧。

为了评估本文方法的性能, 本文使用了机器翻译及自动文本摘要等文本生成领域的经典评测指标 ROUGE^[22] 作为评测指标。ROUGE 指标通过比较参考文本和模型生成文本, 计算两者中同时出现的文本单元在参考文本中所占的比例, 从而衡量模型生成文本的质量。常用的 ROUGE 指标主要有 ROUGE-1, ROUGE-2 以及 ROUGE-L 这 3 类, 本文选择这 3 类常见的 ROUGE 指标进行评测。

3.2 实验环境

本文实验使用的 IMCS21 数据集共有 4 116 条样本, 其中训练集共 2 472 条样本, 验证集共 833 条样本, 测试集共 811 条样本。本文的实验环境如表 3 所列。

表 3 实验环境

Table 3 Experimental environment

配置项	配置
服务器 CPU 型号	Intel(R)Xeon(R)CPU W-2275 @ 3.30 GHz
服务器内存大小	512 GB
服务器 GPU 型号	NVIDIA RTX A5000 × 2
服务器操作系统	Ubuntu 22.04.1 LTS
CUDA 版本	11.3
Python 版本	3.8
深度学习框架版本	PyTorch 1.12.1

3.3 与基线模型对比分析

为了验证本文方法在医疗报告生成任务上的有效性,本文选择了基于 BART 的模型架构为主干模型,该模型在多种摘要生成任务中均取得了最好的效果。同时,本文与医疗文本摘要领域中 Zhang 等^[12]所采用的 LSTM 模型、Enarvi 等^[13]采用的 Transformer 模型,以及 Krishna 等^[15]采用的 T5 模型进行了对比,相关模型的介绍如下。

1) LSTM: 基于单向长短期记忆网络(Long Short-Term Memory)的编码器-解码器架构,即使用单向 LSTM 作为编码器-解码器架构中的编码器和解码器。

2) Transformer: 基于 Attention 机制的编码器-解码器架构。

3) T5: 基于 Transformer 的编码器-解码器结构,将多种文本到文本的任务统一作为通用的文本转换任务处理。

此外,为了进一步验证本文方法的有效性,我们还选取了 6 种在文本摘要领域广泛应用的算法模型作为基线模型进行性能对比实验,基线模型的介绍如下。

1) CNN: 基于卷积神经网络(CNN)的编码器-解码器架构,即使用 CNN 作为编码器-解码器架构中的编码器以及解码器。

2) Bi-LSTM: 基于双向 LSTM 的编码器-解码器架构,即使用双向 LSTM 作为编码器-解码器架构中的编码器,解码器则依旧使用单向 LSTM。

3) PGN: 指针生成器网络(Pointer Generator Network),既能够从原文中复制词语,也具有生成原文中没有的词语的能力。

4) BART: 双向自回归变压器(Bidirectional and Auto-Regressive Transformers)模型使用标准的基于 Transformer 的序列到序列结构。

5) ProphetNet^[23]: 基于 Transformer 的编码器-解码器架构,能够高效地在训练过程中完成每个时刻同时预测未来的 n 个字符。

6) gpt-3.5-turbo^[24]: 人工智能驱动的自然语言处理工具 ChatGPT 所使用的生成式语言模型,在各种自然语言处理任务中都取得了极其出色的表现。

与基线模型的对比实验结果如表 4 所列。从表中可以看出,本文方法在验证集上所有的 ROUGE 指标均优于参与对比的 9 种基线方法。其中,本文方法相比 CNN, LSTM 以及 Bi-LSTM 这 3 个传统的基于编码器-解码器架构的方法以及 PGN 和 Transformer 这两类基于 Attention 机制的编码器-解码器架构的方法在 3 个指标上均取得了较大的优势,原因主要在于这几类方法在训练时不使用经过预训练的词向量,且可用的数据集规模较小,因此其取得的性能较差。而相比 ProphetNet, BART, T5 这 3 类使用预训练模型的方法,本文方法在使用预训练模型的基础上,通过领域知识先验和任务解耦机制,使模型更加关注实体信息等关键知识,以及其与医疗报告自身结构之间的关系,确保医疗报告不同字段文本的准确性,因此在预训练模型之上取得了进一步的性能提升。其中,本文所采用的主干模型 BART 在作为基线模型时,其性能指标比基线模型如 ProphetNet 和 T5 要差,而结合本文工作后其性能指标优于所有基线模型,由此证明了本文工作

的有效性。同时,在多种文本生成任务中表现出色的 gpt-3.5-turbo 在本实验中表现较差的主要原因是,其不能将本文的训练集数据放入其中对模型进行微调,只能直接生成。此外,为了验证本文方法的有效性,我们与各基线算法的复杂度进行比较。如表 5 所列,本文方法的计算复杂度与 Transformer 架构的方法基本相同,在设置序列长度 $l=512$ 、隐藏层表征维度 $d=768$ 的条件下,优于非 Transformer 架构的模型,可见本文算法的有效性。本文还针对所提方法在医疗报告不同字段上的生成性能进行了实验,并将其与基线方法中表现较好的 ProphetNet 和 T5 进行对比,结果如表 6 所列。

表 4 本文方法与基线方法的性能对比

Table 4 Performance comparison between our method and baseline methods

方法	ROUGE-1	ROUGE-2	ROUGE-L
CNN	57.92	43.91	55.95
LSTM	57.88	43.72	56.06
Bi-LSTM	58.89	44.07	56.39
PGN	57.27	43.41	53.64
Transformer	53.99	39.38	49.78
BART	59.38	42.30	55.27
ProphetNet	60.48	45.73	56.41
T5	60.97	44.18	57.63
gpt-3.5-turbo	46.99	28.26	39.83
本文方法	62.59	49.99	58.55

表 5 本文方法与基线方法的算法复杂度对比

Table 5 Algorithm complexity comparison between our method and baseline methods

序号	方法	时间复杂度
1	CNN	$O(k * l * d^2)$
2	LSTM	$O(l * d^2)$
3	PGN	$O(l * d^2)$
4	Transformer	$O(l^2 * d)$
5	BART	$O(l^2 * d)$
6	ProphetNet	$O(l^2 * d)$
7	T5	$O(l^2 * d)$
8	本文方法	$O(l^2 * d)$

注:序列长度为 l , d 为隐藏层表征维度, k 为卷积核的大小。

表 6 本文方法在医疗报告不同字段的生成中与基线方法的性能对比

Table 6 Performance comparison between our method and baseline methods in the generation of different fields of medical reports

方法	字段	ROUGE-1	ROUGE-2	ROUGE-L
ProphetNet	主诉	71.72	62.26	69.79
T5	主诉	72.29	59.35	71.11
本文方法	主诉	77.57	67.21	77.51
ProphetNet	现病史	43.63	29.69	42.46
T5	现病史	45.02	28.11	41.16
本文方法	现病史	52.04	36.40	50.73
ProphetNet	辅助检查	80.74	73.92	78.22
T5	辅助检查	79.95	71.85	79.35
本文方法	辅助检查	84.59	77.86	84.59
ProphetNet	既往史	79.25	64.50	71.64
T5	既往史	78.69	61.38	74.33
本文方法	既往史	82.06	72.45	82.07
ProphetNet	诊断	68.82	56.46	67.18
T5	诊断	67.49	58.71	66.21
本文方法	诊断	74.23	63.95	74.23
Prophe-Net	建议	45.99	25.52	41.68
T5	建议	42.54	25.94	39.12
本文方法	建议	52.07	34.27	49.80

从表6中可以看出,本文方法在医疗报告每个字段的生成任务中均比基线方法 ProphetNet 和 T5 有着更加出色的表现。从字段整体表现规律来看,本文方法与预训练模型 ProphetNet 以及 T5 在不同字段上的性能表现规律基本一致,都在“主诉”“辅助检查”“既往史”“诊断”这4个字段的 rouge 值较高,都在“现病史”与“建议”字段的 rouge 值较低。主要原因在于,“主诉”“辅助检查”“既往史”“诊断”这4个字段内容长度较短且结构相似,例如“诊断”字段的内容基本都为疾病名称;而“现病史”与“建议”字段内容则较长,其更需要对医疗问诊对话的语义进行充分的理解,并生成合适的内容,这对于模型的学习有着更高的要求。

综上所述,通过本文方法与基线方法分别在医疗报告生成任务与不同字段生成的任务中的对比可以看出,相比基线方法,本文方法在 ROUGE-1, ROUGE-2 以及 ROUGE-L 这3个指标上均有所提升。本节给出的对比实验结果充分展示了本文方法的有效性。

3.4 消融实验分析

为了验证本文提出的医疗报告生成方法的各个组件对整个模型的贡献程度,本节对模型进行消融实验分析,在本文模型上分别去掉领域知识先验和任务解耦机制。任务解耦机制主要借助提示词机制与 Adapter 验证。消融实验的性能如表7所列。

表7 消融实验分析

Table 7 Ablation experiment analysis

方法	ROUGE-1	ROUGE-2	ROUGE-L
-领域知识先验	62.06	48.34	58.45
-提示词验证	62.26	47.92	58.44
-Adapter 验证	62.14	47.85	58.34
本文方法	62.59	49.99	58.55

从表中可以看出,去掉领域知识先验之后,模型的性能在 ROUGE 指标上均表现出了一定程度的下降,原因主要在于,领域知识先验是为了对输入语料文本进行实体等先验知识的提取,并将相关知识嵌入模型,使模型有意识地学习知识,从而使模型的注意力分布发生改变,帮助模型更充分地理解输入语料文本的语义信息。实验结果说明,使用实体引导的领域知识先验能够通过知识辅助使得生成式模型更加专注于输入语料文本的关键信息中,并为医疗报告的生成提供指导。

此外,在分别去掉提示词机制和 Adapter 技术之后,模型性能有一定下降,这也说明了以报告结构引导医疗报告生成的重要性。其中,提示词机制使得输入的医疗语料能够按照医疗报告的结构进行针对性的生成,Adapter 机制使模型能够对每个字段生成的子任务进行自适应的调整,从而能够有效地去除输入语料文本中与医疗报告无关的噪声数据,并根据报告结构筛选出不同字段生成子任务需要学习的重点内容,使得模型能够根据医疗报告不同字段的侧重点生成更加符合字段特点的内容。任务解耦机制不仅从数值上表现出了比较直观的性能提升,更确保了任务解耦后生成医疗报告的事实正确性。

综上所述,本节消融实验的结果充分展示了领域知识先验和任务解耦机制对于模型整体性能提升的有效性与必要性。

结束语 本文提出了一种基于知识辅助的结构化医疗报告生成方法,分别从医疗报告生成任务输入与输出的知识学习利用两个角度,提出了实体引导的领域知识先验与结构引导的任务解耦机制两种具体方法,通过这两种方法的结合,使用先验的领域专业知识与医疗报告的结构化特点来引导医疗报告的生成,从而提高生成结果的准确性和合理性。

本文方法在中文医疗信息处理评测标准 CBLUE 的医疗报告生成任务 IMCS-MRG 的公开数据集 IMCS21 上进行测评,实验结果表明,相比传统的基于编码器-解码器架构的模型以及常用的预训练模型,基于领域知识先验和任务解耦机制的医疗报告生成方法能够取得更优的性能。同时,本文还针对该方法进行了消融实验分析,从而验证了方法中各部分的有效性。

本文在医疗报告生成中使用了预训练模型,但由于其计算负载过重,使得使用预训练模型进行推理需要的时间过长。因此,之后的工作可以考虑使用模型蒸馏、模型压缩等技术,在不影响模型性能或者影响很小的情况下,降低模型大小,缩短推理时间。

参考文献

- [1] ZHOU Q, YANG N, WEI F, et al. Neural document summarization by jointly learning to β score and select sentences[C]// ACL 2018—56th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference (Long Papers), Melbourne, VIC, Australia; 2018: 654-663.
- [2] RUSH A M, CHOPRA S, WESTON J. A neural attention model for abstractive sentence summarization[C]// Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2015: 379-389.
- [3] LIU Y. Fine-tune BERT for extractive summarization [J]. arXiv:1903.10318, 2019.
- [4] ZHONG M, LIU P, CHEN Y, et al. Extractive summarization as text matching[C]// Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Seattle, Washington, USA; ACL, 2020: 6197-6208.
- [5] SEE A, LIU P J, MANNING C D. Get to the point; Summarization with pointer-generator networks[J]. arXiv:1704.04368, 2017.
- [6] PAULUS R, XIONG C, SOCHER R. A deep reinforced model for abstractive summarization[J]. arXiv:1705.04304, 2017.
- [7] LI W, XIAO X, LYU Y, et al. Improving neural abstractive document summarization with structural regularization[C]// Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. 2018: 4078-4087.
- [8] LUO Z, YEUNG S H, ZHANG M, et al. MLCask: Efficient management of component evolution in collaborative data analytics pipelines[C]// 2021 IEEE International Conference on Data Engineering (ICDE). IEEE, 2021: 1655-1666.
- [9] LUO Z, CAI S, GAO J, et al. Adaptive lightweight regularization tool for complex analytics[C]// 2018 IEEE International Conference on Data Engineering (ICDE). IEEE, 2018: 485-496.

- [10] LUO Z, CAI S, WANG Y, et al. Regularized Pairwise Relationship based Analytics for Structured Data [C] // Proceedings of the ACM on Management of Data, 2023:1-27.
- [11] SONG Y, TIAN Y, WANG N, et al. Summarizing medical conversations via identifying important utterances [C] // Proceedings of the 28th International Conference on Computational Linguistics, 2020:717-729.
- [12] ZHANG Y Z, JIANG Z T, ZHANG T, et al. MIE: A medical information extractor towards medical dialogues [C] // Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, 2020:6460-6469.
- [13] ENARVI S, AMOIA M, TEBA M A, et al. Generating medical reports from patient-doctor conversations using sequence-to-sequence models [C] // Proceedings of the First Workshop on Natural Language Processing for Medical Conversations, 2020:22-30.
- [14] CHINTAGUNTA B, KATARIYA N, AMATRIAIN X, et al. Medically aware gpt-3 as a data generator for medical dialogue summarization [C] // Machine Learning for Healthcare Conference, PMLR, 2021:354-372.
- [15] KRISHNA K, KHOSLA S, BIGHAM J, et al. Generating soap notes from doctor-patient conversations using modular summarization techniques [C] // ACL-IJCNLP 2021 - 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, Proceedings of the Conference, Virtual, Online: 2021:4958-4972.
- [16] LEWIS M, LIU Y, GOYAL N, et al. Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension [J]. arXiv:1910.13461, 2019.
- [17] SOUZA F, NOGUEIRA R, LOTUFO R. Portuguese named entity recognition using BERT-CRF [J]. arXiv:1909.10649, 2019.
- [18] LIU P, YUAN W, FU J, et al. Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing [J]. ACM Computing Surveys, 2023, 55(9):1-35.
- [19] REBUFFI S A, BILEN H, VEDALDI A. Learning multiple visual domains with residual adapters [J]. arXiv:1705.08045, 2017.
- [20] CHEN W, LI Z, FANG H, et al. A benchmark for automatic medical consultation system: frameworks, tasks and datasets [J]. Bioinformatics, 2023, 39(1):817.
- [21] ZHANG N, CHEN M, BI Z, et al. Cblue: A chinese biomedical language understanding evaluation benchmark [J]. arXiv:2106.08087, 2021.
- [22] LIN C Y. Rouge: A package for automatic evaluation of summaries [C] // Text Summarization Branches out, 2004:74-81.
- [23] QI W, GONG Y, YAN Y, et al. Prophetnet-x: Large-scale pre-training models for english, chinese, multi-lingual, dialog, and code generation [J]. arXiv:2104.08006, 2021.
- [24] CHEN X, YE J, ZU C, et al. How Robust is GPT-3.5 to Predecessors? A Comprehensive Study on Language Understanding Tasks [J]. arXiv:2303.00293, 2023.



SHI Jiyun, born in 1991, Ph. D, is a member of CCF (No. J2410M). Her main research interests include big data and artificial intelligence.



ZHANG Meihui, born in 1985, professor, Ph. D supervisor, is a member of CCF (No. 92466M). Her main research interests include big data, blockchain and artificial intelligence.

(责任编辑:何杨)