计算机科学
COMPUTER SCIENCE

基于多样化标签矩阵的医学影像报告生成

张俊三, 程铭, 沈秀轩, 刘玉雪, 王雷全

引用本文

张俊三, 程铭, 沈秀轩, 刘玉雪, 王雷全. 基于多样化标签矩阵的医学影像报告生成[J]. 计算机科学, 2024, 51(8): 200-208.

ZHANG Junsan, CHENG Ming, SHEN Xiuxuan, LIU Yuxue, WANG Leiquan. Diversified Label Matrix Based Medical Image Report Generation [J]. Computer Science, 2024, 51(8): 200-208.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.230600018

基于多样化标签矩阵的医学影像报告生成

张俊三　程　铭　沈秀轩　刘玉雪　王雷全
中国石油大学(华东)青岛软件学院、计算机科学与技术学院　山东 青岛 266580

摘　要　医学影像在医学诊断中具有重要作用,而准确描述的文本报告对于理解图像以及后续疾病诊断是必不可少的。目前在医学影像报告生成领域,基于模式化方法生成规范的文本报告成为近年的研究热点。但正负样本数量差距较大导致的数据偏差问题,使得生成的报告内容普遍倾向于描述正常状况,难以准确捕捉异常信息。为解决这一问题,提出了一种基于多样化标签矩阵的医学报告生成方法,可以对不同的疾病进行差异化学习,生成多样化的医疗报告;设计文本-矩阵特征损失函数,优化多样化标签矩阵;增加特征交叉模块改进 Transformer 网络,加强图像与文本的映射,提升疾病描述的准确性。在 IU-X-Ray 和 MIMIC-CXR 两个数据集上进行实验,实验结果表明,与目前的主流方法相比,所提方法在 BLEU,METEOR 等多个指标上取得了最优的效果。

关键词:深度学习;医学影像报告生成;注意力机制;图像-文本生成;多模态
中图分类号　TP391

Diversified Label Matrix Based Medical Image Report Generation

ZHANG Junsan,CHENG Ming,SHEN Xiuxuan,LIU Yuxue and WANG Leiquan
Qingdao Institute of Software,College of Computer Science and Technology,China University of Petroleum(East China),Qingdao,Shandong 266580,China

Abstract　Medical images play a vital role in medical diagnosis. Accurately described text reports are essential for understanding images and subsequent disease diagnosis. In recent years,the generation of standardized reports based on modeling methods has become a research hotspot in the field of medical imaging report generation. However,due to the data deviation problem caused by the large gap between positive and negative samples,the content of the generated report generally tends to describe the normal situation. This limitation creates challenges in accurately capturing abnormal information. To address this issue,this paper proposes a novel approach based on diversified label matrix for medical report generation. This method utilizes a diverse label matrix to perform differential learning on different diseases and generate diverse medical reports. Additionally,a text-matrix feature loss function is designed to optimize the diverse label matrix,enhancing its effectiveness. Furthermore,the Transformer network is enhanced by incorporating a feature intersection module. This module strengthens the mapping between images and text,and improves accuracy in disease description. Experimental results on the two datasets of IU-X-Ray and MIMIC-CXR show that,the proposed method achieves the best results in multiple indicators,such as BLEU and METEOR,compared with the current mainstream methods.

Keywords　Deep learning,Medical report generation,Attention mechanism,Image-Text generation,Multi-modal

1　引言

目前,通过 X 光、超声波、CT 等一系列成像技术产生医学影像后进行临床诊断,极大地提高了医学诊治的效率和准确性。然而,医学影像需要大量具有基础医学、临床医学和现代影像学的理论知识和实操能力的医学高级人才来进行疾病诊断和报告撰写,这需要大量的人工工作。近年来,随着人工智能的发展,深度学习算法在医疗领域取得了较好的效果[1-2],并为医疗影像描述提供了一个新的研究方向,即通过神经网络学习医疗影像与文字报告之间的特征关系,针对给定医疗影像自动生成准确的文本解释,实现为医生的诊断提供辅助并极大地降低医生手工撰写报告工作量的目标。

现有的医学影像报告自动生成技术一般是由一个卷积神经网络和一个自然语言处理网络构成,早期的自然语言处理网络通常为 LSTM[3],BiLSTM[4],GRU[5] 等,并通过注意力机制提取特征。近年来,随着 Transformer[6] 的发展,大部分方法使用 Transformer 及相关变形模型作为文本特征提取

到稿日期:2023-06-01　返修日期:2023-10-14
基金项目:山东省自然科学基金(ZR2020MF006,ZR2022LZH015)
This work was supported by the Natural Science Foundation of Shandong Province,China(ZR2020MF006,ZR2022LZH015).
通信作者:张俊三(zhangjunsan@upc.edu.cn)

模型与报告生成架构。这种基本结构往往更适用于传统的图像短句描述任务，而医疗报告的通常是长文本并且具有高度的模式化性质[7]。基于这样的特征，文献[8-11]从数据库中检索、提取报告模板，然后将其融入文本生成中，但是模板的构造提高了模型的复杂性。因此，文献[7,12]采用记忆网络在文本生成过程中自动学习文本信息，进一步促进了医疗报告生成的发展。尽管这些方法性能显著，但是正常样本较多、异常样本较少导致的数据不平衡性，使得生成的报告更偏向正常报告模板，导致报告文本看似正常，实则只对正常区域进行了大量描述[13]。

针对上述问题，本文提出了一种基于多样化标签矩阵的医学报告生成方法。该方法通过学习并记录不同疾病的医疗文本特征，为不同疾病的图像提供不同的语义信息；并设计了文本-矩阵特征损失函数，通过挖掘矩阵和生成文本之间的关系，进一步优化多样化标签矩阵；此外，增加特征交叉模块以改进 Transformer 网络，实现图像和文本特征的交叉，加强视觉语义信息融合，进一步提高文本对图像描述的准确性。本文在 IU-X-Ray 和 MIMIC-CXR 两个医学影像数据集上进行实验，采用 BELU-1，BLEU-2，BLEU-3，BLEU-4，METEOR 和 ROUGE-L 作为评价指标。实验结果表明，本文模型在上述指标上取得了更优的效果。

2 相关工作

2.1 图像字幕

图像字幕的任务是输入给定的图像使计算机生成相应的文字描述，生成过程通常结合计算机视觉和自然语言处理技术。基于编码器-解码器模型在自然语言处理领域的广泛应用，图像字幕任务中图像与文本的映射关系逐渐转为基于端到端的映射关系。随着深度学习的发展，图像字幕模型引入各式各样的网络与策略并取得了不错的效果，例如文献[14]将注意力机制引入编码器-解码器结构，文献[15]利用条件对抗网络实现生成描述的多样性，文献[16]提出了一种对上下文感知的 LSTM 字幕生成器和共同注意力鉴别器，实现图像和描述之间的语义对齐。除此之外，文献[17]使用强化学习的策略优化模型参数。这些方法虽然在生成短文本描述上具有不错的效果，但在长文本描述上还需要进一步的改善。另外，这些方法生成的语句仍然不足以完整地描述图像，并且模型往往不能有效提取图像中隐含的重要信息，这将使得模型泛化能力不高，无法在医疗等其他领域取得同样好的结果。

2.2 医学影像报告生成

医学影像报告生成任务旨在通过计算机技术读取患者的放射影像，识别影像中的异常或疾病，并生成相应的文本报告。基于图像字幕任务各深度学习模型在多个领域的成功延展，目前的医学影像报告生成方法也采用了类似的方法，并取得了较大进展[18]。基于模式化生成方法是该领域目前的一个研究热点。为了生成连贯的描述医学图像的长文本，文献[8]提出了通过手动方式提取模板的基于检索的主题生成方法，并在文本生成过程中利用强化学习进行模型训练。文献[9]通过将图像特征转换为图，并进行模板检索和扩充。为了提高报告对异常的准确描述，文献[19]通过生成疾病主题，构建知识图谱促进报告生成。文献[13]提取先验知识和后验知识，并将其融合到报告生成中。为了提高文本多样性并降低模型的复杂度，文献[7]提出了一种记忆网络，用于记录文本生成过程中的关键信息，并将其融合到 Transformer 中的解码器中。文献[12]提出了一种跨模态记忆网络增强的编码器解码器框架，用于存储图像和文本之间的对齐信息，促进跨模态的交互和生成。虽然上述研究已经取得不错的效果，但这些方法不能同时关注报告的模板特征与异常的准确描述。受此启发，本文学习并记录不同疾病的模板，为不同的图像提供不同的增强信息，从而生成更加准确、规范的医疗报告。

3 本文方法

本文模型的整体架构是基于编码器-解码器的网络结构。如图 1 所示。

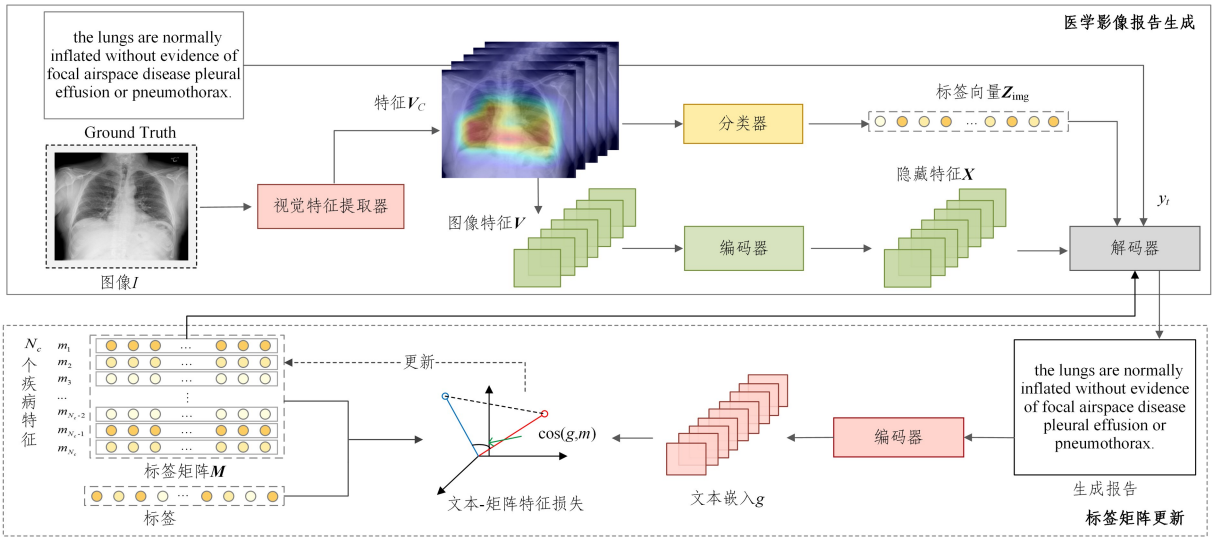


图 1 模型结构图

Fig. 1 Model structure diagram

模型主要包括 4 个部分，分别是视觉特征提取器、分类器、编码器和解码器。其中，视觉特征提取器对输入的医疗影像 I 提取图像特征 $\boldsymbol{V}$，并分别输入编码器和分类器中获得隐藏向量 $\boldsymbol{X}=\{\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n\}$ 和影像的疾病类别向量 $\boldsymbol{Z}_{\text{img}}$，然后与标签矩阵 $\boldsymbol{M}$、t 时刻前的文本序列 $\{y_1,y_2,\cdots,y_{t-1}\}$ 共同输入解码器中解码生成 t 时刻的目标文本 y_t。t 时刻文本生成过程可以表述为：

$$p(y_t)=p(y_t\mid y_1,y_2,\cdots,y_{t-1},\boldsymbol{M},I) \tag{1}$$

完整的医学影像报告生成过程可以概括为：

$$p(Y)=\prod_{t=1}^{T}p(y_t\mid y_1,y_2,\cdots,y_{t-1},\boldsymbol{M},I) \tag{2}$$

其中，T 为生成报告的长度。

3.1 视觉特征提取器与分类器

视觉特征提取模块旨在提取图像特征。首先将图片送入通过预训练的卷积神经网络(CNN)中，本文使用的是预训练的 ResNet101 网络，在平均池化操作之前，提取网络最后一个卷积层输出的特征 $\boldsymbol{V}_c\in\mathbb{R}^{h\times w\times d}$，其中 h，w 和 d 分别是图像的高度、宽度和通道数，即图像分解成大小相等的区域，然后直接将每行的所有特征连接，最后扩展成一个序列，并将其作为后续模块的源输入 $\boldsymbol{V}=\{v_1,v_2,\cdots,v_{n_v}\}$，$\boldsymbol{V}\in\mathbb{R}^{n_v\times d}$，其中 $n_v=h\times w$，该过程如式(3)所示：

$$\{v_1,v_2,\cdots,v_{n_v}\}=f_v(I) \tag{3}$$

其中，$f_v(\cdot)$表示视觉特征提取器，I 为输入的医疗影像。

将 CNN 网络最后一个卷积层输出的特征 $\boldsymbol{V}_c\in\mathbb{R}^{h\times w\times d}$ 输入分类器中对医疗影像进行分类。如图 2 所示，首先使用池化层对特征进行压缩，去除冗余信息，然后使用全连接层实现特征的高度整合和抽象，并利用 ReLU 激活函数和 Dropout 避免梯度消失和过拟合，最后使用全连接层和 Sigmoid 函数激活，将特征值归一化到 0 和 1 之间，获得图像的标签向量 $\boldsymbol{Z}_{\text{img}}$，过程如式(4)所示：

$$\boldsymbol{Z}_{\text{img}}=\text{Sigmoid}(ReLU(AvePool(\boldsymbol{V})\boldsymbol{M}_{v_1})\boldsymbol{M}_{v_2}) \tag{4}$$

其中，标签向量 $\boldsymbol{Z}_{\text{img}}\in\mathbb{R}^{N_c}$，$N_c$ 为医学影像类别数，$\boldsymbol{W}_{v_1}$ 和 $\boldsymbol{W}_{v_2}$ 是分类器中全连接层的可训练权重。

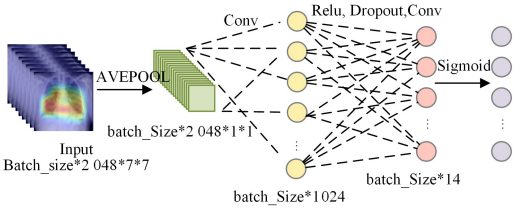


图 2 分类器结构图

Fig. 2 Classifier structure diagram

3.2 编码器-解码器

本文方法使用 Transformer 中的编码器为视觉特征提取模块中获得的图像特征进行编码，输出隐藏向量 $\boldsymbol{X}$，该过程可以表示为：

$$\{x_1,x_2,\cdots,x_n\}=f_e(v_1,v_2,\cdots,v_{n_v}) \tag{5}$$

其中，$f_e(\cdot)$表示编码器。

然后将该隐藏向量输入解码器中与文本向量进行融合对齐，生成预测文本 y。为了加强不同模态特征之间的语义关联，生成更准确、多样性的文本，建立一个标签矩阵来训练并记忆不同标签的文本特征，并利用图像分类产生的标签向量提取图像相关的特征。此外，对 Transformer 网络进行改进，捕捉不同模态之间的相关性和交互信息，产生更丰富、更全面的特征表示，提高模型对图像和文本的理解和表达能力。解码器的结构如图 3 所示。

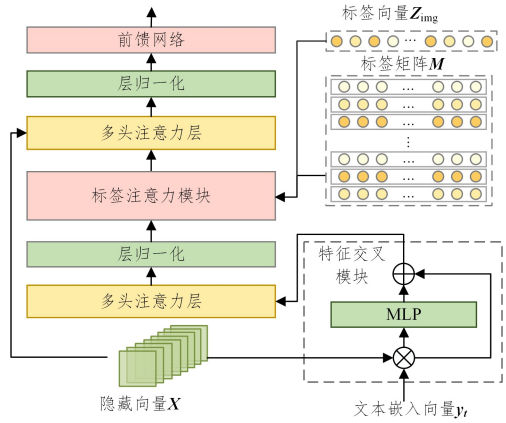


图 3 解码器结构图

Fig. 3 Decoder structure diagram

3.2.1 多样化标签矩阵

目前的研究主要关注生成文本与目标文本的相似性，忽略了数据集的不平衡性，导致模型在样本数量较少的疾病类别上生成的文本效果较差，从而影响模型的整体性能。为了解决这一问题，本文提出标签矩阵为文本特征增加同类别的模板向量。具体来说，在训练中利用本文定义的可训练的标签矩阵 $\boldsymbol{M}=\{m_1,m_2,\cdots,m_{N_c}\}$ 学习不同疾病的文本信息，来自上一层的向量通过注意力模块对标签矩阵中的文本特征进行查询与响应，并进行后续的解码操作。标签注意力模块的结构如图 4 所示，使用标签矩阵 $\boldsymbol{M}$、来自上一层的向量 $\boldsymbol{Q}$ 和标签向量 $\boldsymbol{Z}_{\text{img}}$ 共同计算注意力分数，并与矩阵 $\boldsymbol{M}$ 进行加权计算，从而实现为相同疾病类别的文本特征建立特征映射，抑制其他类别的文本特征。标签注意力模块具体的计算过程如下。

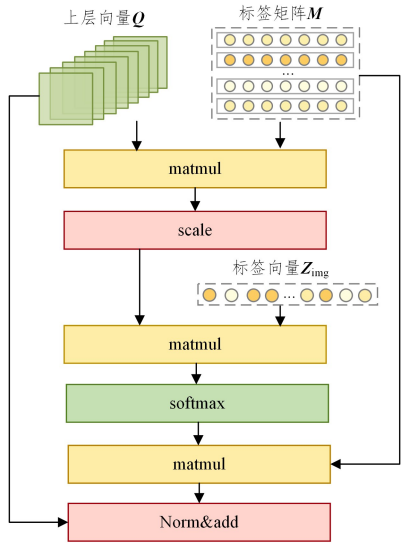


图 4 标签注意力模块结构图

Fig. 4 Structure of label-attention module

将标签矩阵 $\boldsymbol{M}$ 作为值和键嵌入，上一层输出的向量 $\boldsymbol{Q}$ 作为注意力机制中的查询向量，然后通过式(6)计算标签矩阵和向量 $\boldsymbol{Q}$ 的相似度 $\boldsymbol{D}$。

$$\boldsymbol{D}=\frac{\boldsymbol{M}\boldsymbol{Q}^{\mathrm{T}}}{\sqrt{d}} \tag{6}$$

其中，标签矩阵 $\boldsymbol{M}=\{m_1,m_2,\cdots,m_{N_c}\}$，$m_i\in\mathbb{R}^d$，$N_c$ 为数据集的疾病类别数目，矩阵的每行向量 $\boldsymbol{m}_i$ 表示疾病 i 的文本特征向量，d 表示其维度。

为了增强相同标签的全局文本特征注意力，减小不同疾病的特征权重，本文将分类器生成的类别向量 $\boldsymbol{Z}_{\mathrm{img}}$ 作为标签权重，并将其与相似度 $\boldsymbol{D}$ 相乘作为最终的相似度，然后通过对所有相似度进行归一化，获得每个类别的文本特征对于源序列的注意力分数 $\boldsymbol{S}$。该过程可被概括为：

$$\boldsymbol{S}=\mathrm{softmax}(\boldsymbol{D}\cdot\boldsymbol{Z}_{\mathrm{img}}) \tag{7}$$

其中，Softmax 为归一化指数函数。

最后利用注意力分数将不同类别的特征进行融合，得到最终与源文本序列相关的特征，同时为了防止梯度消失和爆炸，引入归一化和残差连接。该过程被表述为：

$$\boldsymbol{Q}'=\boldsymbol{Q}+f_{\mathrm{norm}}(\boldsymbol{S}\cdot\boldsymbol{M}) \tag{8}$$

$$f_{\mathrm{norm}}(x)=\frac{x-\mu}{\sigma} \tag{9}$$

其中，$\boldsymbol{Q}$ 为上层输出向量，μ 表示均值，σ 表示标准差。

3.2.2 特征交叉模块

现有的医学影像报告生成模型在训练过程中对 Ground-Truth 的依赖性强，导致生成的报告对图像的描述不够准确，为了实现不同特征之间的信息互补并获取更丰富的特征表示，本文在 Transformer 的解码器中增加了特征交叉模块。

如图 3 所示，该模块输入文本嵌入向量 $\boldsymbol{y}$ 和来自编码器的隐藏向量 $\boldsymbol{X}$，通过乘积和多层感知机(MLP)网络学习不同特征之间的相关性，实现特征交叉，获得融合向量 $\boldsymbol{r}$。为了避免网络退化，通过增加残差层将融合向量与源嵌入向量相加，最终将其输入到源 Transformer 的解码器中，整个过程如式(10)、式(11)所示：

$$\boldsymbol{r}=f_{\mathrm{mlp}}(\boldsymbol{y}\cdot\boldsymbol{X}^{\mathrm{T}}) \tag{10}$$

$$\tilde{\boldsymbol{y}}=\boldsymbol{y}+\boldsymbol{r} \tag{11}$$

其中，$\boldsymbol{y}\in\mathbb{R}^{n_t\times d}$ 为 t 时刻前的所有文本特征，隐藏向量 $\boldsymbol{X}\in\mathbb{R}^{n_w\times d}$，融合向量 $\boldsymbol{r}\in\mathbb{R}^{n_t\times d}$，其中 n_t 为输入到解码器中词向量的个数，在训练过程中，该值即为报告的长度。

3.3 文本-矩阵标签损失

定义标签矩阵使用正态分布进行初始化，往往会导致训练时不能很好地学习到不同疾病的文本特征，在查询响应模块产生较大误差。为了减少文本特征误差的产生，本文提出文本-矩阵特征损失 L_{matrix} 来监督标签矩阵的学习。

本文利用 Transformer 的编码器作为文本特征提取器，将模型输出的文本 $\boldsymbol{Y}'=\{y_1,'y_2',\cdots,y'_{N_{tw}}\}$ 进行位置编码后输入 Transformer 的编码器模块提取训练文本特征，使用编码器输出的第一个向量作为全局向量，该过程如式(12)所示：

$$\boldsymbol{G}_{\mathrm{text}}=f_{\mathrm{encoder}}(\boldsymbol{Y}') \tag{12}$$

其中，N_{tw} 为模型生成文本的长度，$G_{\mathrm{text}}=\{\boldsymbol{g}_0,\boldsymbol{g}_1,\cdots,\boldsymbol{g}_{N_{tw}}\}$，$\boldsymbol{G}_{\mathrm{text}}\in\mathbb{R}^{N_{tw}\times d}$ 为编码器输出的文本向量表示，取 $\boldsymbol{g}_0$ 作为全局特征向量。

然后，使用报告真实标签和标签矩阵得到标签响应的矩阵特征向量，如式(13)所示：

$$\boldsymbol{m}_{\mathrm{text}}=\sum_{i=0}^{N_c}z_i'\cdot\boldsymbol{m}_i \tag{13}$$

其中，z_i' 为真实疾病 i 的标签，$\boldsymbol{m}_i$ 为疾病 i 对应的标签矩阵中第 i 个标签向量，$\boldsymbol{m}_{\mathrm{text}}\in\mathbb{R}^d$ 为生成的矩阵特征向量。

通过计算矩阵文本特征和全局特征向量之间的余弦相似度来定义文本-矩阵特征损失 L_{matrix}。计算过程如式(14)所示：

$$L_{\mathrm{matrix}}=1-\cos(\boldsymbol{g}_0,\boldsymbol{m}_{\mathrm{text}}) \tag{14}$$

其中，$\boldsymbol{g}_0$ 为生成文本的全局特征向量。

医学影像报告自动生成任务是典型的自然语言生成任务，模型通过最小化交叉熵损失来最大化生成的报告与目标报告的相似性，其损失 L_{text} 如式(15)、式(16)所示：

$$L_{\mathrm{text}}=-\frac{1}{N_W}\sum_{i=0}^{N_W}\log(p(y_i)) \tag{15}$$

$$p(y_i)=p(y_i\mid y_1,y_2,\cdots,y_{i-1},\boldsymbol{M},I) \tag{16}$$

其中，N_W 为生成报告的单词数量，y_t 为 t 时刻生成的文本。

为了训练视觉特征提取模块生成准确的视觉标签，模型增加多标签分类损失，即将视觉特征提取模块中获得的标签向量与真实向量输入二元交叉熵函数中计算相似度，旨在产生更准确的图像标签。该过程可概括为：

$$L_{\mathrm{label}}=-\frac{1}{N_c}\sum_{i=0}^{N_c}z_i'\log z_i \tag{17}$$

其中，z_i 表示医疗图像预测的 i 疾病的标签，z_i' 为真实疾病 i 的标签，N_c 为数据集的疾病类别数目。

文献[20-21]已经表明，可以通过结合多种损失来更新模型。本文通过共同最小化文本损失、标签损失和文本-矩阵特征损失来优化所提模型，最终损失如式(18)所示：

$$L=L_{\mathrm{text}}+\lambda_1L_{\mathrm{label}}+\lambda_2L_{\mathrm{matrix}} \tag{18}$$

其中，λ_1 和 λ_2 是平衡 3 个损失的系数。

4 实验结果与分析

4.1 数据集与评价指标

4.1.1 数据集

本文使用了两个公开的医疗图像数据集，即 IU-X-Ray 和 MIMIC-CXR。IU-X-Ray 数据集包括 7 470 张包括正视图和侧视图的胸部 X 光图像和 3 933 份相应报告，是目前医学影像报告生成任务中应用广泛的基准数据集。MIMIC-CXR 数据集包含 371 920 张带标签的医疗图像，是近年来公布的最大的医学影像报告生成任务的数据集之一。每个图像都有人工标注的报告和疾病类别。实验遵循文献[12]的做法，按照 7∶1∶2 的数据拆分比例将数据集拆分成训练集、验证集和测试集。

4.1.2 评价指标

为了客观且全面地评价模型的文本生成性能，并与同任务的其他模型进行对比，本文使用针对生成文本与源文本之间相似度的评价指标 BLEU，ROUGE，METEOR。BLEU 是一种主流的用于衡量目标文本与源文本相似度的机器翻译评价指标，根据比对词的连续个数存在多个变种，本实验使用常见的

BLEU-1,BLEU-2,BLEU-3,BLEU-4 这 4 种。不同于 BLEU 计算文本准确率,ROUGE 是一种基于召回率的指标的统称,包含 ROUGE-N,ROUGE-L,ROUGE-W,ROUGE-S 这 4 个指标,本文使用更适用于文本生成任务的 ROUGE-L 指标。METEOR 关注生成文本与源文本中的相似词的替换。

4.2 模型参数设置

实验使用预训练过的 ResNet101 网络提取医疗影像的视觉特征,然后将生成的 512 维的图像特征输入全连接层获取 14 维的标签向量。设置 Transformer 网络的图像特征输入层数为 3、多头注意力 head 为 8、隐藏维度为 512,输入的标签矩阵大小为 14×512,设置平衡参数 λ_1 和 λ_2 为 1 和 0.5。在训练 IU-X-Ray 数据集的实验中使用 Adam 优化器、1×10^{-4} 的模型学习率、5×10^{-5} 的视觉提取学习率进行模型优化,并且设置 IU-X-Ray 数据集训练轮数为 70,模型在训练过程中的损失变化如图 6 所示,经过 50 轮迭代后损失曲线趋于平稳。训练 MIMIC-CXR 数据集的实验中使用 Adam 优化器、5×10^{-4} 的模型学习率、5×10^{-5} 的视觉提取学习率进行模型优化,训练轮数为 50 轮,模型在训练过程中的损失变化如图 5、图 6 所示,经过 25 轮迭代后损失曲线趋于平稳,模型达到最好效果。

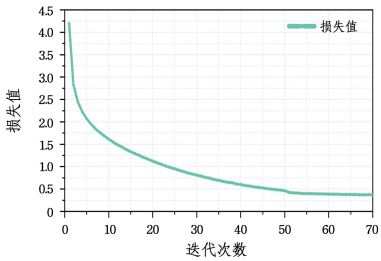


图 5 在 IU-X-Ray 上的损失值变化曲线

Fig. 5 Loss curve on IU-X-Ray

同时,本文通过图 7 中的箱线图展示了模型训练过程中评价指标的分布情况。明显可见,各个训练指标的分布呈现出稳定而平衡的趋势。基于以上分析可以得出结论,本文模型在迭代过程中呈现稳定的推进态势。

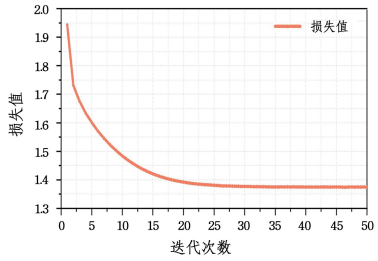


图 6 在 MIMIC-CXR 上的损失值变化曲线

Fig. 6 Loss curve on MIMIC-CXR

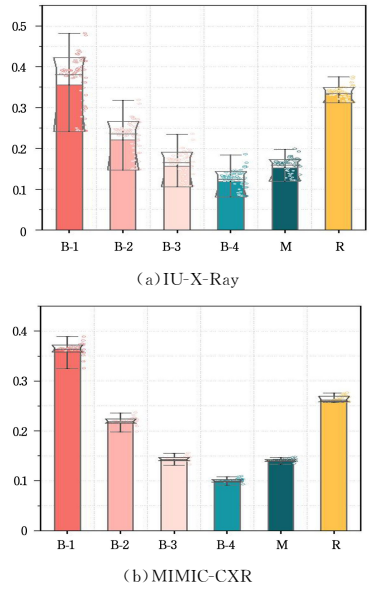


图 7 IU-X-Ray 和 MIMIC-CXR 数据集上的箱线图

Fig. 7 Box plot on IU-X-Ray and MIMIC-CXR datasets

4.3 实验结果分析

4.3.1 消融实验结果分析

为了进一步证明提出的特征交叉模块、多样化标签矩阵以及文本-矩阵特征损失函数的有效性,分别在 IU-X-Ray 和 MIMIC-CXR 两个数据集上进行消融实验。实验结果如表 1、表 2 所列。

表 1 IU-X-Ray 数据集上的消融实验

Table 1 Ablation study on IU-X-Ray

模型结构	特征交叉	标签矩阵	文本-矩阵损失	B-1	B-2	B-3	B-4	METEOR	ROUGE-L
Base				0.440	0.284	0.206	0.158	0.179	**0.377**
(a)	√			0.434	0.283	0.210	0.165	0.179	0.360
(b)		√		0.462	0.294	0.215	0.167	0.182	0.359
(c)		√	√	0.471	0.305	0.224	0.174	0.187	0.365
(d)	√	√	√	**0.482**	**0.317**	**0.234**	**0.184**	**0.197**	0.370

表 2 MIMIC-CXR 数据集上的消融实验

Table 2 Ablation study on MIMIC-CXR

模型结构	特征交叉	标签矩阵	文本-矩阵损失	B-1	B-2	B-3	B-4	METEOR	ROUGE-L
Base				0.320	0.190	0.124	0.088	0.120	0.255
(a)	√			0.351	0.213	0.142	0.101	0.138	0.273
(b)		√		0.361	0.220	0.146	0.102	0.139	0.274
(c)		√	√	0.364	0.224	0.150	0.105	0.143	**0.280**
(d)	√	√	√	**0.388**	**0.235**	**0.154**	**0.107**	**0.146**	0.275

表中,base 为基于 Transformer 的编码器-解码器模型,实验(a)为在 base 模型的基础上增加特征交叉模块的结果,实验(b)增加了多样化标签矩阵,实验(c)通过文本-矩阵特征损失对标签矩阵进行优化,实验(d)则为采用本文模型的实验。

通过比较两个数据集实验的 base 和实验(a)的结果可以发现,增加特征交叉的 Transformer 网络在大部分指标上

提升显著。由于传统的 Transformer 模型仅仅在多头注意力模块中融合图像特征，往往不足以生成准确描述医疗疾病的文本，因此增加特征交叉模块将视觉和文本特征进行整合，使模块关注重要特征，并对不相关的信息进行抑制，进而为模型提供更加丰富的特征表示。在实验(b)中，首先使用正态分布对矩阵进行初始化，然后在后续训练中学习不同标签报告的特征，与 base 的结果相比，各个评价指标有明显的提升，证明通过实验训练的标签矩阵学习不同标签的文本特征能有效提高模型性能。

实验(c)通过增加文本-矩阵特征损失进一步优化标签矩阵，实验指标相比实验(b)有明显提升。可见设计的文本-矩阵特征损失函数能够进一步优化标签矩阵，明显减小矩阵学习中产生的误差。

模型使用 λ_1 和 λ_2 两个参数用于平衡文本损失、标签损失和文本-矩阵特征损失。为了不影响模型生成的文本报告的性能，实验设置标签损失和文本-矩阵特征损失参数值小于文本损失参数。响应参数 λ_1 的敏感度如图 8 所示。

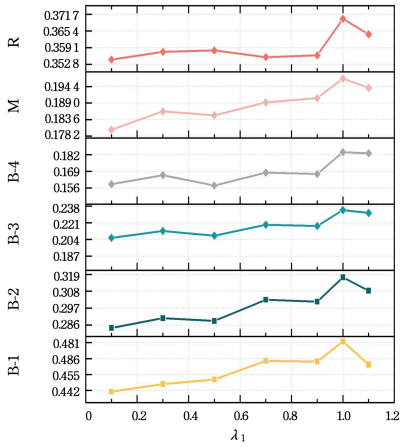


图 8 响应参数 λ_1 的敏感度

Fig. 8 Sensitivity analysis of λ_1

实验结果显示，随着 λ_1 的增加，各个指标的大小存在较大的波动，直到 λ_1 达到 1，大部分指标的值达到最大，并趋于稳定。响应参数 λ_2 的敏感度如图 9 所示。当 λ_2 由 0.1 升至 0.5 时，大部分评价指标的值逐渐升至最高，当 λ_2 的大小再次增加到 1 时，评价指标的值逐渐减小。参数 λ_2 是调节优化矩阵损失的参数，参数值过小，导致矩阵不能精准反映类别向量，参数值过大，则模型引入过多的噪声信息，导致性能下降。实验结果显示，在 λ_1 为 1、λ_2 为 0.5 时实验取得最优的结果。

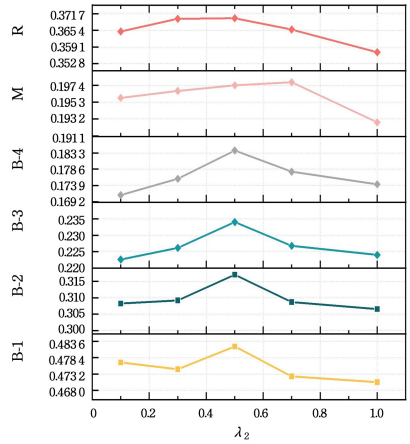


图 9 响应参数 λ_2 的敏感度

Fig. 9 Sensitivity analysis of λ_2

4.3.2 与其他算法的对比结果

如表 3 所列，为了表明模型的有效性，本文在 IU-X-Ray 数据集上与 9 种现有方法进行对比，本文方法在 BLEU-1，BLEU-3，BLEU-4 和 METEOR 指标上取得了最优结果。在 BLEU-2 指标上取得了次优结果。在 MIMIC-CXR 数据集上与 6 种现有方法进行对比，本文模型在 BLEU-1，BLEU-2，BLEU-3，BLEU-4 和 METEOR 指标上取得最优结果。由于本文对不同疾病标签的特征进行加权学习，生成的疾病描述往往同义但采取不同表达，因此模型在指标 ROUGE-L 上效果不佳。在这些对比方法中，传统的图像字幕方法 ATT2IN，ADAATT 直接应用于医学影像报告生成领域未能取得较优的结果。而使用记忆网络的 R2Gen，R2GenCMN 模型只考虑使用内存记录生成信息，虽然取得了较好的结果，但没有解决数据偏差问题。与引入强化学习、课程学习的 CMAS-RL、CMCL 模型相比，本文模型的结构更加简单，并实现了更好或者相同的结果。总体来说，本文方法在各个指标上的表现优于大多数对比方法。

表 3 不同模型在两个数据集上的对比实验

Table 3 Comparative experiments of different models on IU-X-Ray and MIMIC-CXR

数据集	模型	B-1	B-2	B-3	B-4	METEOR	ROUGE-L
IU-X-Ray	ATT2IN[22]	0.224	0.129	0.089	0.068	—	0.308
	ADAATT[23]	0.220	0.127	0.089	0.068	—	0.308
	HRGR[8]	0.436	0.278	0.197	0.150	—	0.341
	CMAS-RL[24]	0.464	0.301	0.210	0.154	—	0.362
	R2Gen[7]	0.470	0.304	0.219	0.165	0.187	0.371
	R2GenCMN[12]	0.475	0.309	0.222	0.170	0.191	0.375
	CMCL[25]	0.473	0.305	0.217	0.162	0.186	0.378
	JPG[26]	0.479	**0.319**	0.222	0.174	0.193	0.377
	Prior Guided Trans[27]	0.482	0.313	0.232	0.181	**0.203**	**0.381**
	Ours	**0.482**	0.317	**0.234**	**0.184**	0.197	0.370
MIMIC-CXR	ADAATT[23]	0.299	0.185	0.124	0.088	0.118	0.266
	ATT2IN[22]	0.325	0.203	0.136	0.096	0.134	0.276
	R2Gen[7]	0.353	0.218	0.145	0.103	0.142	0.277
	R2GenCMN[12]	0.353	0.218	0.148	0.106	0.142	0.278
	CMCL[25]	0.344	0.217	0.140	0.097	0.133	**0.281**
	Prior Guided Trans[27]	0.356	0.222	0.151	**0.111**	0.140	0.280
	Ours	**0.388**	**0.235**	**0.154**	0.107	**0.146**	0.275

4.3.3 可视化结果分析

为了进一步分析模型的文本生成情况，图 10 给出了对 IU-X-Ray 数据集中两组正视图和侧视图的影像案例进行定性分析的结果，图中下划线部分为生成报告中与目标报告的相似文本，其中红色字体为生成报告的异常描述，蓝色字体为生成报告的正常描述。可以发现，与传统的 Transformer 网络以及 R2Gen，R2GenCMN 模型相比，本文模型生成的文本更接近放射科医生所写的报告。

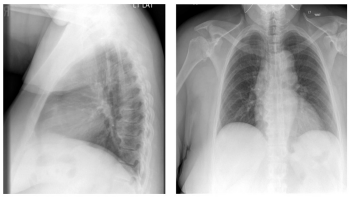
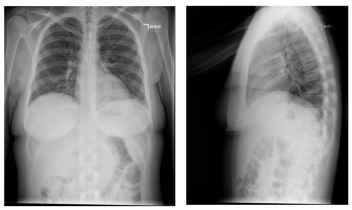
Images	Ground-truth	Base	R2Gen	R2GenCMN	Ours
	heart size within normal limits . tortuous aorta . low lung volumes with no focal consolidations . no pneumothorax or effusion . moderate degenerative disc disease in the midthoracic spine .	the heart is normal in size . the mediastinum is stable . the aorta is atherosclerotic . the lungs are clear .	heart size within normal limits . low lung volumes . no focal airspace consolidations . no pneumothorax or pleural effusion .	low lung volumes . heart size normal . the trachea is midline . lungs are clear . no pneumothorax . no pleural effusion .	low lung volumes are present . the heart size and pulmonary vascularity appear within normal limits . the lungs are free of focal airspace disease . no pleural effusion or pneumothorax is seen . mild degenerative changes are present in the spine .
	Cardiomediastinal contour and pulmonary vascularity stable and within normal limits. Lung volumes are slightly low. There are streaky left basal opacities. No pleural effusion or pneumothorax. No acute osseous findings. No free air is demonstrated.	heart size is within normal limits . low lung volumes . no focal airspace consolidations . no pneumothorax or pleural effusion .	low lung volumes are present . the heart size and pulmonary vascularity appear within normal limits . the lungs are free of focal airspace disease . no pleural effusion or pneumothorax is seen .	stable cardiomegaly stable mediastinal contour . low lung volumes . no pneumothorax or pleural effusion .	low lung volumes . cardiomediastinal silhouette and pulmonary vasculature are within normal limits . streaky bibasilar opacities . no pneumothorax or pleural effusion . no acute osseous findings .

图 10 实验结果可视化(电子版为彩图)

Fig. 10 Visualization of experiment results

针对正常描述，在第一个例子中，本文模型生成的“the heart size and pulmonary vascularity appear within normal limits，the lungs are free of focal airspace disease”不仅描述了 Ground Truth 中的正常部分，并进行了大量的补充。第二个例子中基于多样化标签矩阵的模型生成的“cardiomediastinal silhouette and pulmonary vasculature are within normal limits”也使用了相似的单词进行相应的正常描述。对于更具有临床意义的异常描述的诊断，相比不能准确生成异常描述的 Base 方法，本文模型生成的“low lung volumes are present”和“mild degenerative changes are present in the spine”对第一个例子中的疾病进行了正确的描述。第二个例子中，本文模型也通过生成“low lung volumes”和“streaky bibasilar opacities”，使用不同的语句对 Ground Truth 中的疾病进行了准确的判断。

通过分析可以发现，相比传统 Transformer 模型生成的文本，本文型生成的文本与器官和异常之间的映射关系更紧密，同时与基于记忆网络的 R2Gen 以及基于多模态特征映射的 R2GenCMN 相比，针对不同的疾病提取不同的报告特征的方法对疾病描述也更加详尽和准确。此外，该模型通过学习相同标签的文本特征，生成的文本与目标文本不完全相同，使用了相同含义的替代性词汇与语句，使得生成的报告更具有灵活性。

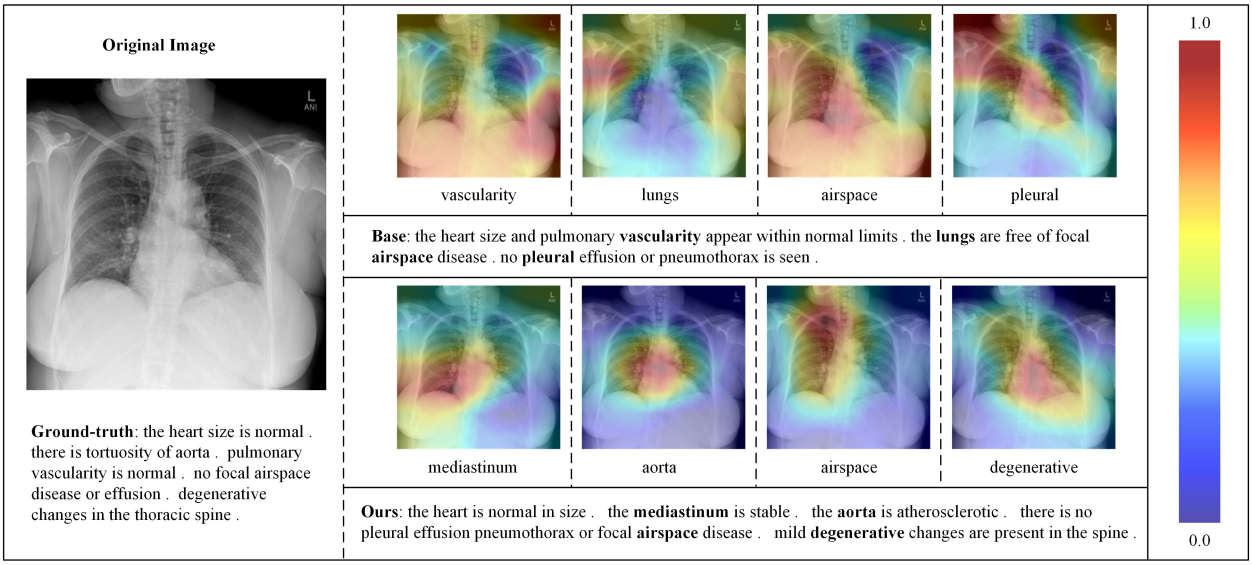


图 11 图像文本映射可视化

Fig. 11 Visualization of image-text mappings

为了进一步观察模型在异常和文本上的映射情况，本文从 IU-X-Ray 中选择一张胸部 X 光图像对本文模型和 Base 模型生成的文本进行了注意力可视化。图 11 给出了该图像及其真实报告以及不同模型生成的部分文本在图像上的注意力映射，图像上的映射区域用不同的颜色表示从 0 到 1 的注意力权重。从图中可以发现，base 模型生成报告不能对疾病进行正常描述，并且文本与图像之间的映射也不够准确，而本文模型实现了疾病的精准判断，并且图像上的相关区域与文本的位置一一对应，例如图像对“mediastinum”的关注区域即为其在胸片中的大致位置。由此可见本文模型不仅增强了放射学报告生成的能力，而且增强了多模态特征的表达能力。

结束语 针对医学影像报告生成任务中由数据偏差导致的基于单一模板生成的文本难以精确捕捉异常信息的问题，本文提出了一种为不同疾病提供不同模板的新的医学影像生成方法。该方法通过训练标签矩阵记录不同标签的报告特征，并用于指导文本生成。本文方法首先在 Transformer 网络中增加特征交叉模块，旨在使模型强化图像特征与文本特征的映射，然后利用多样化矩阵学习不同标签的文本特征，并将其记录在标签矩阵中，最后利用标签注意力模块将相同标签的文本特征融入文本生成序列中，生成能准确描述异常的医疗报告。本文模型在 IU-X-Ray 和 MIMIC-CXR 两个数据集的大多数指标上都取得了最优的效果。

参 考 文 献

[1] ZHANG M M，QIN P L，CHAI R，et al. CT-Generated MRI Algorithm for Acute Ischemic Stroke[J]. Computer Engineering，2024，50(2)：317-326.

[2] JIA H Y，XIA R，LYU A Q，et al. Panoramic mosaic approach of ultrasound medical images based on template fusion[J]. Journal of Jilin University(Engineering and Technology Edition)，2022，52(4)：916-924.

[3] HOCHREITER S，SCHMIDHUBER J. Long short-term memory[J]. Neural Computation，1997，9(8)：1735-1780.

[4] GRAVES A，SCHMIDHUBER J. Framewise phoneme classification with bidirectional LSTM and other neural network architectures[J]. Neural Networks，2005，18(5/6)：602-610.

[5] CHUNG J，GULCEHRE C，CHO K，et al. Empirical evaluation of gated recurrent neural networks on sequence modeling[C]// NIPS 2014 Workshop on Deep Learning. 2014.

[6] VASWANI A，SHAZEER N，PARMAR N，et al. Attention is all you need[C]// Advances in Neural Information Processing Systems. 2017：5998-6008.

[7] CHEN Z，SONG Y，CHANG T H，et al. Generating Radiology Reports via Memory-driven Transformer[C]// Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing(EMNLP). 2020：1439-1449.

[8] LI Y，LIANG X，HU Z，et al. Hybrid Retrieval-Generation Reinforced Agent for Medical Image Report Generation[C]// NeurIPS. 2018.

[9] LI C Y，LIANG X，HU Z，et al. Knowledge-driven encode，retrieve，paraphrase for medical image report generation[C]// Proceedings of the AAAI Conference on Artificial Intelligence. 2019：6666-6673.

[10] HARZIG P，EINFALT M，LIENHART R. Automatic disease detection and report generation for gastrointestinal tract examination[C]// Proceedings of the 27th ACM International Conference on Multimedia. 2019：2573-2577.

[11] HAN Z，WEI B，LEUNG S，et al. Towards automatic report generation in spine radiology using weakly supervised framework[C]// International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham：Springer，2018：185-193.

[12] CHEN Z，SHEN Y，SONG Y，et al. Cross-modal Memory Networks for Radiology Report Generation[C]// Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1：Long Papers). 2021：5904-5914.

[13] LIU F，WU X，GE S，et al. Exploring and distilling posterior and prior knowledge for radiology report generation[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021：13753-13762.

[14] XU K，BA J，KIROS R，et al. Show，Attend and Tell：Neural Image Caption Generation with Visual Attention[J]. Computer Science，2015：2048-2057.

[15] BO D，FIDLER S，URTASUN R，et al. Towards diverse and natural image descriptions via a conditional gan[C]// Proceedings of the IEEE International Conference on Computer Vision. 2017：2970-2979.

[16] AMIRIAN S，RASHEED K，TAHA T R，et al. Image Captioning with Generative Adversarial Network[C]// 2019 International Conference on Computational Science and Computational Intelligence(CSCI). 2019.

[17] LIU S，ZHU Z，NING Y，et al. Improved Image Captioning via Policy Gradient optimization of SPIDEr[C]// 2017 IEEE International Conference on Computer Vision(ICCV). IEEE，2017.

[18] JING B，XIE P，XING E. On the Automatic Generation of Medical Imaging Reports[C]// Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1：Long Papers). 2018：2577-2586.

[19] ZHANG Y，WANG X，XU Z，et al. When radiology report generation meets knowledge graph[C]// Proceedings of the AAAI Conference on Artificial Intelligence. 2020：12910-12917.

[20] PAULUS R，XIONG C，SOCHER R. A Deep Reinforced Model for Abstractive Summarization[C]// International Conference on Learning Representations. 2018.

[21] ZHANG Y，MERCK D，TSAI E B，et al. Optimizing the Factual Correctness of a Summary：A Study of Summarizing Radiology Reports[C]// ACL. 2020.

[22] RENNIE S J，MARCHERET E，MROUEH Y，et al. Self-critical

sequence training for image captioning[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017:7008-7024.
[23] LU J, XIONG C, PARIKH D, et al. Knowing when to look: Adaptive attention via a visual sentinel for image captioning [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017:375-383.
[24] JING B, WANG Z, XING E. Show, Describe and Conclude: On Exploiting the Structure Information of Chest X-ray Reports [C]// Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. 2019:6570-6580.
[25] LIU F, GE S, WU X. Competence-based Multimodal Curriculum Learning for Medical Report Generation[C]// ACL/IJCNLP. 2021.
[26] YOU J, LI D, OKUMURA M, et al. JPG-Jointly Learn to Align: Automated Disease Prediction and Radiology Report Generation[C]// Proceedings of the 29th International Conference on Computational Linguistics. 2022:5989-6001.
[27] YAN B, PEI M, ZHAO M, et al. Prior Guided Transformer for Accurate Radiology Reports Generation[J]. IEEE Journal of Biomedical and Health Informatics, 2022, 26(11):5631-5640.

ZHANG Junsan, born in 1978, Ph.D, associate professor, is a member of CCF (No. 74487M). His main research interests include information retrieval and recommender systems.

(责任编辑:喻藜)

CCF 苏州换届,新一届执委会产生!

2024 年 7 月 11 日,CCF 苏州会员活动中心(简称“CCF 苏州”)换届选举会议在 CCF 业务总部召开。来自苏州各高校、企事业单位的 CCF 苏州委员、会员代表近 50 人现场参会。换届活动分为三个环节,分别是换届筹备情况介绍、CCF 苏州工作介绍和选举会议。会议由换届筹备组组长、苏州大学应用技术学院陈志峰教授主持。

CCF 苏州主席、常熟理工学院龚声蓉教授介绍了 CCF 苏州的发展历程以及两年来取得的成绩。他表示,CCF 苏州围绕服务企业数字转型、院校人才培养、产业技术赋能、地方人才需求五个方面开展了走进大院大所活动、金鸡报晓论坛、产教融合论坛和院长论坛四个系列品牌活动。目前,CCF 苏州拥有专业会员 1644 人,企业会员单位 23 家,获得 2022 年 CCF 优秀会员活动中心和 2023 年 CCF 总部特别贡献奖。

随后,主持人介绍了 CCF 苏州换届选举规则和流程,宣读了换届筹备组名单和委员名单。选举过程中,候选人们上台进行竞选演说并回答会员提问,参会委员进行无记名投票。产生了新一届执委会和监委会成员,苏州科技大学胡伏原教授当选 CCF 苏州新一届主席。

最后,CCF 秘书长唐卫清对此次换届大会作总结发言,肯定了 CCF 苏州近年来取得的成绩,希望 CCF 苏州在胡伏原主席的带领下,继续保持高质量发展势头,再创辉煌。

附:CCF 苏州新一届执行委员会、监督委员会成员名单

执行委员会主席

胡伏原　苏州科技大学电子与信息工程学院院长、教授

副主席

王　进　苏州大学未来科学与工程学院副院长、教授
翁志勇　苏州朗捷通智能科技有限公司董事长

秘书长

王　喜　苏州工业职业技术学院人工智能学院副院长、副教授

执行委员

邢晓双　常熟理工学院计算机科学与工程学院院长、教授
马洁明　西交利物浦大学西浦人工智能产业学院副院长、副教授
刘　正　苏州工业园区服务外包职业学院计算机科学与工程学院院长、教授
曹　敏　苏州大学计算机科学与技术学院副教授
支洪平　科大讯飞(苏州)科技有限公司总经理

监督委员会主席

韩月娟　苏州微兔信息科技有限公司总经理

监督委员

罗　颖　苏州工业职业技术学院人工智能学院院长、副教授
王　涛　星科智汇(苏州)科技有限公司董事长

据 CCF 微信公众号