摘要

融合计算机视觉和自然语言处理的多模态学习能够处理和理解来自多种模态的信息，已成为多媒体内容分析与理解的主要手段。比如输入一张图片到多模态网络模型中则会输出一个句子来描述图片中的内容，从而实现了由图像到文本的跨模态转换。与之类似，医疗报告生成是从给定的医学图像来生成相应的诊断报告，多模态学习使医疗报告自动生成具备了可能性。

X 射线图像作为典型且最常见的医学图像，广泛的应用于胸部疾病的诊断，但是对其撰写 X 射线影像报告则会占据医生的大量时间，这种情况促进了对 X 射线影像医疗报告自动生成的研究。尽管 X 射线影像医疗报告生成可以使用图像描述的方法建模，但是两者的任务要求并不相同，因此将图像描述方法直接应用在 X 射线影像报告生成任务上，并不能生成最准确的诊断报告。因此本文基于图像描述，就如何生成准确且可靠的 X 射线影像医疗报告进行研究。本文的主要贡献点如下：

1）提出了基于分层双解码器的 X 射线影像医疗报告生成方法。该方法首先考虑了 X 射线图像正面图像和侧面图像的互补关系，引入正面视觉注意力机制和侧面视觉注意力机制来捕获语言与正面图像和侧面图像的上下文信息；其次，X 射线影像报告生成的是一个长段落，为了克服普通循环神经网络及其变体在生成长段落任务中面临的梯度消失问题，本文提出了分层解码架构，即将 X 射线影像报告的生成任务分解为几个诊断句子的生成任务，然后将多个生成的句子整合成最终的报告；最后本文提出了双解码器来分别生成对正常情况描述和异常情况描述的句子，有效地缓解了数据分布偏差的问题。

2）提出了基于自适应图像-标签融合的 X 射线影像医疗报告生成方法。该方法首先对输入的 X 射线图像进行疾病预测，产生一组疾病标签，包含了 X 射线图像中所有异常的明确信息，而 X 射线图像中的视觉信息包含了异常的具体细节，基于此提出了疾病标签注意力机制和图像视觉注意力机制；针对医疗报告过长问题，设计基于多重注意力的文本生成解码器，可以有效生成长段落。然后通过引进一个自适应注意力融合门，将疾病标签信息和图像视觉信息进行自适应融合，以产生以疾病为导向的视觉特征，可以更好的表征 X 射线图像的异常区域。这些以疾病为导向的视觉特征结合基于多重注意力的解码器可以生成全面且详细的医疗报告。一系列的实验证明了该方法取得了较好的性能。

关键词：X 射线影像医疗报告生成，图像描述，注意力机制，融合门