Multi-modal Radiology Report Generation Using Image Classification and LLMs

Applicant's name: He Zhicheng

Program: Ph.D.

Abstract

Radiology reports are critical in patient diagnosis, yet producing accurate and comprehensive reports poses challenges due to the complexity of interpreting medical images and the detailed narrative required. By integrating image classification with Large Language Models (LLMs) like GPT-4, there is an opportunity to enhance the quality and clinical relevance of these reports. To address this, I propose developing a multi-modal framework that processes medical images to identify key clinical findings, generates structured descriptions, and uses these to produce coherent and actionable radiology reports. This approach aims to streamline report generation while ensuring the accuracy necessary for clinical use.

Introduction

Radiology imaging serves as a fundamental tool in modern healthcare, providing detailed insights that are essential for diagnosing and managing patient conditions. The process of translating these images into structured, accurate radiology reports, however, is complex and time-intensive, often subject to variability based on the radiologist's expertise and workload. Traditionally, this task has been largely manual, requiring careful review of images and the crafting of detailed narratives to communicate the findings effectively.

Advances in machine learning, particularly in natural language processing (NLP), have opened new avenues for automating parts of this workflow. Large Language Models (LLMs) such as GPT-4 have demonstrated a remarkable ability to generate human-like text from contextual cues. Despite their capabilities, these models have not yet been fully leveraged in radiology due to the challenge of effectively integrating visual data from medical images.

A growing body of research underscores the need for a multi-modal approach that bridges the gap between image analysis and text generation. By combining these modalities, it becomes possible to create a system that not only identifies key clinical findings but also generates a coherent narrative that aligns with clinical best practices. This integrated approach could significantly reduce the time required for report generation, improve consistency, and ensure that subtle findings are accurately captured.

In response to this need, I plan to develop a multi-modal framework that combines advanced image classification techniques with LLMs to generate comprehensive radiology reports. This framework is designed to enhance the precision and utility of the reports, ensuring they meet the stringent demands of clinical environments.

Research Plan

1. Data Acquisition and Preprocessing

To ensure the robustness and accuracy of the model, I will utilize well-established datasets such as MIMIC-CXR and CheXpert, which provide paired radiology images and reports. Preprocessing these datasets will be a crucial step to enhance model performance. The images will be subjected to normalization and augmentation to support robust feature extraction. Simultaneously, the textual data will be standardized, ensuring consistent terminology and report structure, which is vital for effective training of the LLM.

Given the importance of precise data preparation in multi-modal learning, I aim to preprocess the images and reports meticulously. This will ensure that the multi-modal framework can seamlessly integrate visual and textual information, laying a strong foundation for the subsequent stages of the research.

2. Development of the multi-modal Framework

The core of this research lies in the integration of image classification and natural language generation within a single framework. This integration is intended to leverage the strengths of both approaches to generate high-quality radiology reports.

Initially, radiology images will be analyzed using a convolutional neural network (CNN) or Vision Transformer. These models are chosen for their efficacy in extracting relevant features from complex medical images and will be trained to classify conditions such as atelectasis, cardiomegaly, and pneumothorax.

Following classification, the findings will be converted into structured descriptions that are clear, accurate, and clinically relevant. These descriptions will serve as inputs for GPT-4, a leading LLM, which will generate the full radiology report. To ensure that the generated reports are both coherent and clinically sound, I plan to apply In-Context Instruction Learning (ICIL) and Chain of Thought (CoT) reasoning. These techniques will guide the LLM in producing narratives that closely mimic the reasoning process of human radiologists.

3. Training and Evaluation

The effectiveness of the multi-modal model will hinge on a carefully balanced training approach that optimizes both image classification accuracy and report quality. I will train the model on the preprocessed datasets, focusing on refining its ability to generate both accurate classifications and high-quality text.

For evaluation, a combination of traditional NLP metrics (such as BLEU and METEOR) and clinical efficacy metrics (including precision, recall, and F1 score) will be employed. In addition to these quantitative measures, qualitative feedback from expert radiologists will be sought. Their insights will be instrumental in refining the model to ensure that it meets the rigorous standards expected in clinical practice.

By incorporating this dual approach to evaluation, I aim to ensure that the model not only performs well according to standard metrics but also delivers outputs that are practically useful in real-world medical settings.

4. Optimization and Fine-Tuning

To achieve the best possible performance, I plan to implement a range of optimization strategies. These will include learning rate scheduling to adapt the training process dynamically, dropout to mitigate overfitting, and model ensembling to leverage the strengths of multiple model architectures. Fine-tuning will be conducted with domain-specific datasets to ensure the model's generalizability across various clinical scenarios.

The goal of these optimization efforts is to develop a model that performs consistently and reliably across a wide range of radiology tasks, making it suitable for deployment in diverse healthcare environments.

5. Real-World Application and Deployment

The final phase of this research will focus on deploying the developed model in a clinical setting, where its performance can be evaluated in real-world conditions. The integration of this model into existing radiology workflows will be critical for assessing its impact on diagnostic accuracy, report generation time, and overall workflow efficiency.

By embedding the model within the clinical environment, I aim to gather valuable feedback and data that can be used to further refine the system. The ultimate objective is to demonstrate that the model not only meets technical benchmarks but also provides tangible benefits in clinical practice, improving both efficiency and patient outcomes.