MIMIC-CXR Structured Reporting Dataset

Project Management and Software Development
for Medical Applications

General Info

Contact Person: Matthias Keicher

Contact Email: matthias.keicher@tum.de

Project Abstract

The goal of this project is to create a radiology reporting template for the popular chest radiograph dataset MIMIC-CXR and create a new dataset for the task of structured radiology report generation The MIMIC-CXR RadGraph dataset consists of more than 220000 patients with pairs of chest X-ray images and graphs describing the content of radiology reports. We want to map these graphs to a radiology reporting template and create a new benchmark for structured report generation that will be made public.

Background and Motivation

The generation of structured reports is trending within the medical machine-learning community. The detailed and machine-readable nature of structured reports makes them an excellent data source to explore and analyze. In contrast to radiological free text reports used in current NLP models, structured reports do not suffer from ambiguity in clinical findings.

This project will be conducted in collaboration with Smart Reporting GmbH.

Student's Tasks Description

The student will perform a data exploration of the RadGraph dataset and analyze how the information provided by the RadGraph graphs can be mapped to the reporting template provided by Smart Reporting. Next, an interface needs to be implemented to populate the structured reports automatically, and the free-text reports generated with the template need to be extracted. Then, the add new annotation for MIMIC-CXR dataset is created, consisting of structured reports and generated free-text reports. Finally, the generated reports are compared to the ground truth reports with NLP and RadGraph metrics

You will learn:

- Semantics, ontology, and logic of structured radiology reports
- Modeling structured data in a graph
- NLP metrics

Technical Prerequisites

- Python

References

Pino et al., Clinically Correct Report Generation from Chest X-Rays Using Templates. https://www.springerprofessional.de/en/clinically-correct-report-generation-from-chest-x-rays-using-tem/19698608.

Keicher, et al. "Few-shot Structured Radiology Report Generation Using Natural Language Prompts." https://arxiv.org/abs/2203.15723

Johnson et al. "MIMIC-CXR Database (version 1.0.0)." PhysioNet. https://physionet.org/content/mimic-cxr/1.0.0/

Jain et al. "RadGraph: Extracting Clinical Entities and Relations from Radiology Reports." https://physionet.org/content/radgraph/1.0.0/

Smart Reporting GmbH
www.smart-reporting.com

Please send the completed proposal to ardit.ramadani@tum.de, zl.jiang@tum.de, lennart.bastian@tum.de and tianyu.song@tum.de. Please note that this proposal will be evaluated by the BMC coordinators and will be assigned to a student only in case of acceptance.