Large Language Models for Structuring Radiology Reports

1. General Info

Project Title: Large language models for structuring radiology reports

Contact Person: Chantal Pellegrini

Contact Email: chantal.pellegrini@tum.de

2. Project Abstract

Large language models such as ChatGPT, show great capabilities in solving tasks formulated in detailed prompts. In this project, we aim to translate free-text radiology reports into a structured finding representation, using large language models. To this end, we will investigate different language models, question answering methods [1] and prompting techniques [2]. Further we will investigate training a model for structured report generation given radiological images on the newly generated dataset.

3. Background and Motivation

Automatic report generation in radiology can reduce radiologists' workload and improve diagnostic performance [3,4]. While there is a lot of on-going research on free-text radiology report generation, the research on structured reports is very limited [5,6]. However, structured reports have a lot of advantages in comparison to free-text reports. Free-text radiology reports present findings in narrative form, which can be ambiguous and inconsistent. Structured reports use a standardized format with predefined categories, ensuring clearer and more organized information. For instance, while a free-text report may describe a lung nodule in a sentence, a structured report separates details into specific fields, like size, location, and characteristics, reducing ambiguity and enhancing report quality.

Both for clinical practice and deep learning research it would be very beneficial to have systems that can translate free-text reports to structured reports. First, such a system can generate large structured report datasets from existing public and large scale free-text report datasets. As these datasets are usually paired with radiographic images, this would allow generating a large structured report dataset for training deep learning models on this task. Furthermore, it could enable clinically accurate evaluation of generated free-text reports. Finally, as in clinical practice, radiologists often dictate their reports in free-text, such systems could be employed to allow them to keep their workflow while still ensuring completeness by filling out a structured report from free-text in real-time, and additionally asking for missing information that the radiologists did not mention.

4. Technical Prerequisites

- Good background in deep learning
- Good skills in PyTorch
- Beneficial: experience in NLP

5. Benefits:

- Scientific contribution towards a large-scale structured X-Ray reporting dataset
- Working with SOTA language models
- Possible publication of results

6. Students' Tasks Description

Students' tasks would be the following:

- WP1: Researching publicly available or accessible language models
- WP2: Researching current SOTA for knowledge retrieval from text / Question Answering (QA)
- WP3: Implementing of Text QA pipeline for population of structured reports
- WP4: Testing of different language models (e.g. Alpaca[1], gpt4all[2], PubMedGPT[3]) and prompting techniques to optimize structured report generation from free-text
- WP5: Potentially explore fine-tuning to the radiology-domain
- WP6: Apply to MIMIC-III/IV dataset [7] to generate a large structured report dataset
- WP7: Train structured report generation model on the new dataset and compare to training on smaller, manually labeled dataset
- WP8: Use translation method to evaluate clinical accuracy of SOTA report generation methods
- WP9: Documentation of the results

7. Work-packages and Time-plan:

	Description	#Students	From	To
WP1	Researching LMs	2	Start of May	End of May
WP2	Researching QA	2	Start of May	End of May
WP3	Pipeline Implementation	4	Mid of May	End of May
WP4	Testing of different LMs and prompting techniques	2	Mid of May	Mid of June
M1	Intermediate Presentation	4	Mid of June	
WP5	Potentially fine-tuning to radiology-domain	2	Mid of June	End of June
WP6	Dataset generation from MIMIC-III/IV	2	Mid of June	End of June
WP7	Train image model on new dataset	2	Mid of June	Mid of July
WP8	Clinical accuracy evaluation	2	Mid of June	Mid of July
WP9	Documentation	4	Mid of July	End of July
M2	Final Presentation	4	End of July	

[1] Zaib, M., Zhang, W. E., Sheng, Q. Z., Mahmood, A., & Zhang, Y. (2022). Conversational question answering: A survey. Knowledge and Information Systems, 64(12), 3151-3195.
[2]Zhou, D., Schärli, N., Hou, L., Wei, J., Scales, N., Wang, X., ... & Chi, E. (2022). Least-to-most prompting enables complex reasoning in large language models. arXiv preprint arXiv:2205.10625.
[3]Hou, B., Kaissis, G., Summers, R.M., Kainz, B.: Ratchet: Medical transformer for chest x-ray diagnosis and reporting. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VII 24. pp. 293–303. Springer (2021)
[4] Tanwani, A.K., Barral, J., Freedman, D.: Repsnet: Combining vision with lan- guage for automated medical reports. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2022: 25th International Conference, Singapore, September 18–22, 2022, Proceedings, Part V. pp. 714–724. Springer (2022)
[5] Keicher, M., Mullakaeva, K., Czempiel, T., Mach, K., Khakzar, A., Navab, N.: Few-shot structured radiology report generation using natural language prompts. arXiv preprint arXiv:2203.15723 (2022)
[6] Bhalodia, R., Hatamizadeh, A., Tam, L., Xu, Z., Wang, X., Turkbey, E., Xu, D.: Improving pneumonia localization via cross-attention on medical images and re- ports. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part II 24. pp. 571–581. Springer (2021)
[7] Johnson, A. E., Pollard, T. J., Shen, L., Lehman, L. W. H., Feng, M., Ghassemi, M., ... & Mark, R. G. (2016). MIMIC-III, a freely accessible critical care database. Scientific data, 3(1), 1-9.

[1] https://github.com/tloen/alpaca-lora , https://huggingface.co/spaces/tloen/alpaca-lora,
[2] https://github.com/nomic-ai/gpt4all
[3] https://crfm.stanford.edu/2022/12/15/pubmedgpt.html