The Role of Historical Reports in Radiology Report Summarisation

Mong Yuan Sim[1] **(Presenter)** **Wei Emma Zhang**[1] **Xiang Dai**[2] **Cecile Paris**[2]
[1]The University of Adelaide [2]CSIRO Data61
mongyuan.sim@student.adelaide.edu.au
wei.e.zhang@adelaide.edu.au
{dai.dai;cecile.paris}@csiro.au

1 Introduction

Radiology report acts as a bridge of communication between radiologists and physicians (Kahn et al., 2009; Gershanik et al., 2011). Radiology report usually contains "Reason for the exam", "Comparison" (with any available previous exams), "Findings", and "Impression" sections (Naik et al., 2001; Wallis and McCoubrie, 2011). "Findings" section describes what the radiologist saw in the medical image(s) while "impression" section summarises crucial radiology findings and possible causes (differential diagnosis) (Wallis and McCoubrie, 2011).

Radiology report summarisation is a task that generates "impression" section given "findings" section as an input (Zhang et al., 2018; Ben Abacha et al., 2021) as physicians often only focus on the "impression" section (Lafortune et al., 1988; Gershanik et al., 2011). Existing work on radiology report summarisation solely rely on the "findings" section directly related to current medical image (Zhang et al., 2018; Miura et al., 2021; Ben Abacha et al., 2021), without taking historical reports into consideration when generating "impression" section. This could pose a problem because radiologists may not consistently report the same findings, especially when the examination is intended for comparison with previous exams.In this work, we investigate the role of historical reports in radiology report summarization. The main research question we focus on is, "Can historical radiology reports be used together with the "findings" to generate a more accurate "impression?".

2 Dataset

The dataset used in this work is MIMIC-IV Notes (Johnson et al., 2023), it contains 2,321,355 de-identified radiology reports from 237,427 patients. As we are focusing on patient-specific report, we first group all reports by patient. Then, to avoid the problem of wide time-span between each radiology report, we group reports by hospitalisation to ensure their relevance. Different from the framework in Zhang et al. (2018) and Ben Abacha et al. (2021) where they generate impression section based on current findings section only, our aim is to generate impression section of radiology report given current and previous findings, and we take the current impression section as our gold summary. We argue that impression section of the latest radiology report can be used as gold summary as radiologists have access to previous medical images when writing the report (Johnson et al., 2019).

3 Methodology

Our proposed method leverages pre-trained BART-large (Lewis et al., 2020), a seq2seq model for natural lanaguage generation. We fine-tune BART-large under two different settings, using i) latest findings only and ii) latest and all previous findings.

4 Analysis

We then evaluate our models using four different setups:

i) Given the latest "findings" section only, generate latest "impression" section

ii) Given all "findings" from the same hospitalisation, generate the impression section

iii) Given findings section from all previous reports except the latest one, generate the latest impression

iv) Given randomly sampled reports from other patients and the latest report finding from the patient, generate the latest impression

On one hand, comparing results from i) and ii), we can answer the question "Is additional information from the same patient and hospitalisation necessary and effective in improving the quality of generated summary?". On the other hand, comparing results from ii) and iv) answers the question "Is the improvement made by related information about the patient or any arbitrary information helps

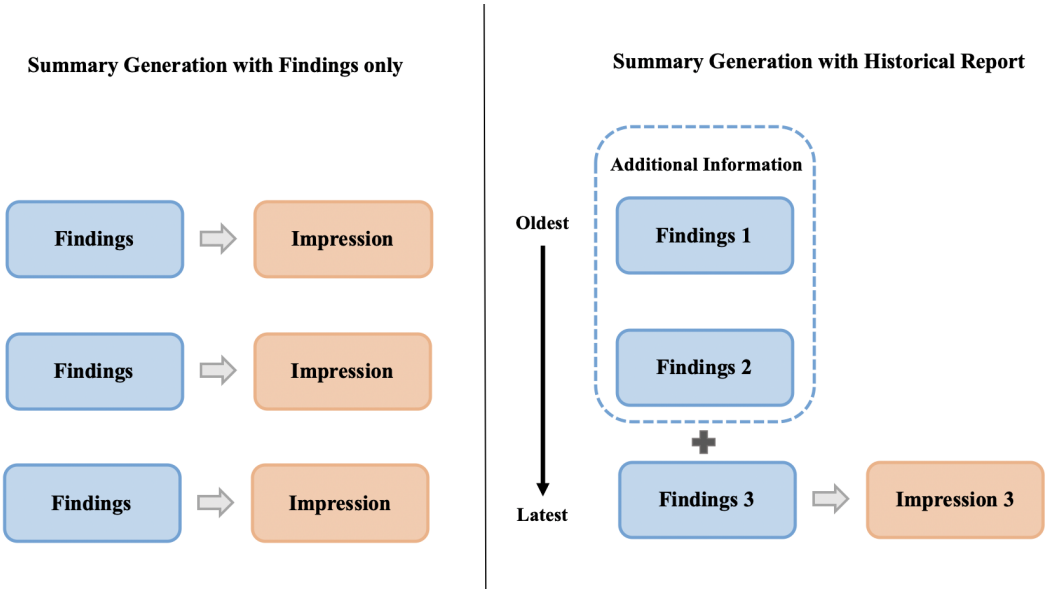

Figure 1: This figure shows the framework we proposed and the grouping method we used. The left part shows the original radiology report summarisation framework as in Zhang et al. (2018) and MEDIQA 2021 shared task (Ben Abacha et al., 2021), where each findings section is used to generate its corresponding impression section. The right part shows our proposed framework where our aim is to generate the latest impressions section given current and previous findings as additional information.

improve the summary quality?". Question i), ii), and iii) helps to analyse to what extend the historical information can improve the quality of the summary.

References

Asma Ben Abacha, Yassine Mrabet, Yuhao Zhang, Chaitanya Shivade, Curtis Langlotz, and Dina Demner-Fushman. 2021. Overview of the MEDIQA 2021 Shared Task on Summarization in the Medical Domain. In *BioNLP@NAACL*.

E. F. Gershanik, R. Lacson, and R. Khorasani. 2011. Critical finding capture in the impression section of radiology reports. *AMIA Annu Symp Proc*, 2011:465–469.

A. E. W. Johnson, L. Bulgarelli, L. Shen, A. Gayles, A. Shammout, S. Horng, T. J. Pollard, S. Hao, B. Moody, B. Gow, L. H. Lehman, L. A. Celi, and R. G. Mark. 2023. MIMIC-IV, a freely accessible electronic health record dataset. *Sci Data*, 10(1):219.

A. E. W. Johnson, T. J. Pollard, S. J. Berkowitz, N. R. Greenbaum, M. P. Lungren, C. Y. Deng, R. G. Mark, and S. Horng. 2019. MIMIC-CXR, a de-identified publicly available database of chest radiographs with free-text reports. *Sci Data*, 6(1):317.

C. E. Kahn, C. P. Langlotz, E. S. Burnside, J. A. Carrino, D. S. Channin, D. M. Hovsepian, and D. L. Rubin. 2009. Toward best practices in radiology reporting. *Radiology*, 252(3):852–856.

M. Lafortune, G. Breton, and J. L. Baudouin. 1988. The radiological report: what is useful for the referring physician? *Can Assoc Radiol J*, 39(2):140–143.

Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. 2020. BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension. In *ACL*.

Yasuhide Miura, Yuhao Zhang, Emily Tsai, Curtis Langlotz, and Dan Jurafsky. 2021. Improving Factual Completeness and Consistency of Image-to-Text Radiology Report Generation. In *NAACL*.

Sandeep S. Naik, Anthony Hanbidge, and Stephanie R. Wilson. 2001. Radiology reports. *American Journal of Roentgenology*, 176(3):591–598. PMID: 11222186.

A. Wallis and P. McCoubrie. 2011. The radiology report–are we getting the message across? *Clin Radiol*, 66(11):1015–1022.

Yuhao Zhang, Daisy Yi Ding, Tianpei Qian, Christopher D. Manning, and Curtis P. Langlotz. 2018. Learning to summarize radiology findings. In *Proceedings of the Ninth International Workshop on Health Text Mining and Information Analysis*, pages 204–213, Brussels, Belgium. Association for Computational Linguistics.