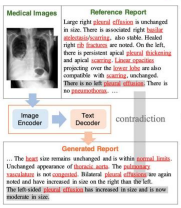
Summarizing Biomedical Texts with Encoder Decoder Models

Raul Salles de Padua, Imran Qureshi
Department of Computer Science, Stanford University
Research Mentors: Kathy Yu

Problem



- Doctors need to summarize complex findings efficiently
- Make quick decisions on urgent patients cases.

Background

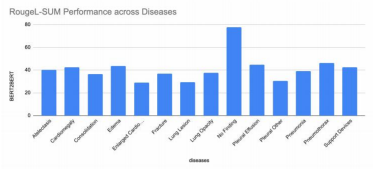
Radiologists write reports to report observation on chest X-rays and synthesize those findings into impressions to hand off to other specialists. We introduce a Biomedical-BERT2BERT model to help radiologists generate concise impressions from reports, with **state of the art performance**.

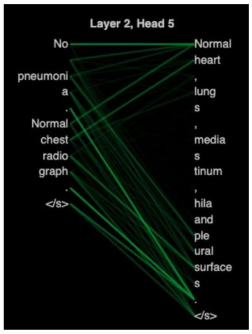
Methods

- Model centric methods
  - Text-to-Text: T5 (small / large)
  - Encoder-Decoder: BERT-to-BERT
  - Linear Attention: BigBird
- Data centric methods:
  - Augmentation techniques with inputs fields randomization (example below)

EXAMINATION: Ap chest x-ray
INDICATION: ___ year old woman with Altered ... // Evidence of pneumonia
TECHNIQUE: AP chest radiograph.
COMPARISON: Chest radiograph from ___.
FINDINGS: No evidence of consolidation, ... //lower thoracic spine, unchanged from prior.
IMPRESSION: No radiographic evidence of pneumonia.

Analysis

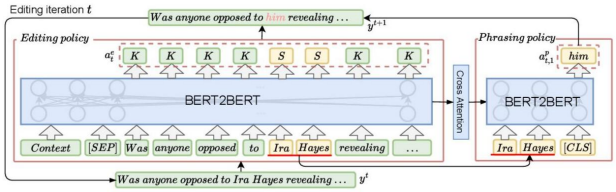




- Model Robustness: All models learned to detect No Findings very well
- Learned disease level fairly well
- Learned several radiology facts such as pneumonia corresponds with “pleural surfaces”

Results

Model	Epochs	time/ epoch	ROUGE-1	ROUGE-2	ROUGE-L	baseline
Finetuned T5-Small	3	1.5 h	48.71	37.98	47.42	
DistillBART	3	2.5 h	30.81	19.51	26.88	
Finetuned T5-base-long	12	**0.63 h**	55.68	45.52	54.74	
Finetuned BERT2BERT	6	0.65 h	**59.61**	**48.99**	**58.75**	
Finetuned PubMed BigBird	7	1.55 h	57.83	47.12	56.66	



Conclusions

- New **state-of-art ROUGE-L score of 58.75**
  - Previous knowledge on similar tasks and vocabulary showed advantageous
  - Very efficient model, compared to 2nd best performing model
- **Data-centric approach** enabled new opportunities for improvement in performance
  - Data augmentation techniques with input fields shuffling: **"paraphrasing"**
- Analysis on diseases shows that may be still room for improvement if class imbalance is addressed

References

Zhuk Chen, Gong. Predicting doctor’s impression for radiology reports with abstractive text summarization.
https://web.stanford.edu/class/cs224n/reports/final_reports/report005.pdf

Kenton Lee Kristina Toutanova Jacob Devlin, Ming-Wei Chang. Bert: Pre-training of deep bidirectional transformers for language understanding.
https://arxiv.org/abs/1810.04805

Acknowledgments

- Kathy Yu (project Mentor)