A Data-Driven Guided Decoding Mechanism for Diagnostic Captioning

P. Kaliosis[1,2], J. Pavlopoulos[1,2], F. Charalampakos[1], G. Moschovis[1,2], I. Androutsopoulos[1,2]

[1] Department of Informatics, Athens University of Economics and Business, Greece
[2] Archimedes Research Unit, Athena Research Center, Athens, Greece

ACL 2024 Bangkok, Thailand

Abstract: Diagnostic Captioning (DC) automatically generates a diagnostic text from one or more medical images (e.g., X-rays, MRIs) of a patient. Treated as a draft, the generated text may assist clinicians, by providing an initial estimation of the patient's condition, speeding up and helping safeguard the diagnostic process. The accuracy of a diagnostic text, however, strongly depends on how well the key medical conditions depicted in the images are expressed. We propose a new *data-driven* guided decoding method that incorporates medical information, in the form of existing tags capturing key conditions of the image(s), into the beam search of the diagnostic text generation process. We evaluate the proposed method on two medical datasets using four DC systems. In most cases, the proposed mechanism improves performance with respect to all evaluation measures.

1. Background

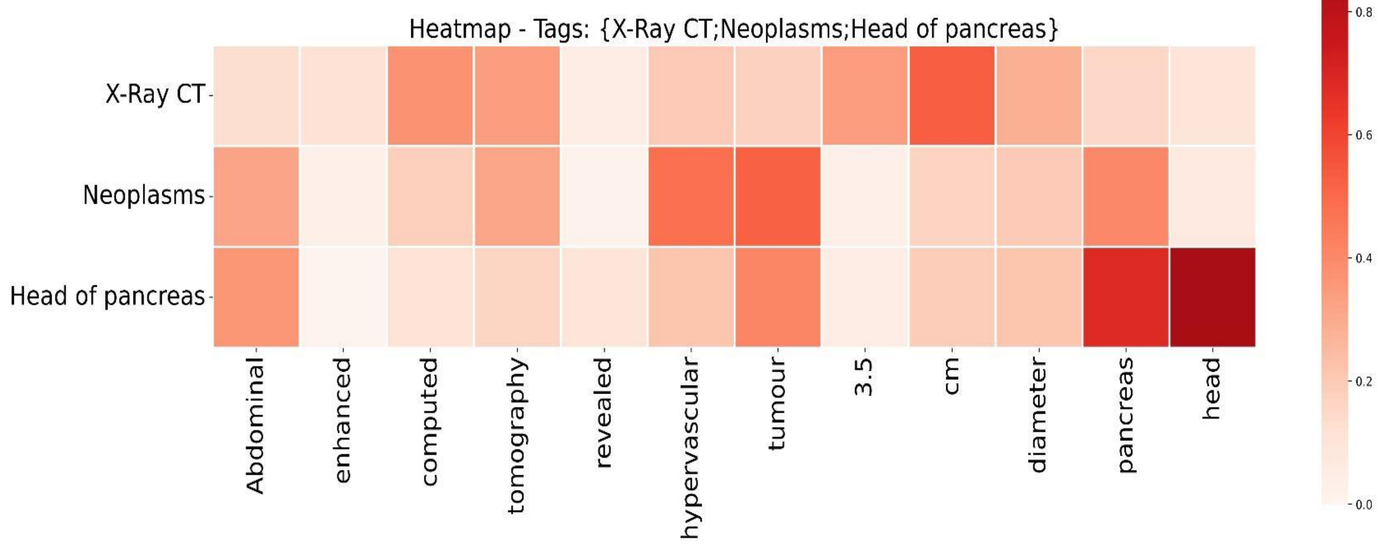
Diagnostic Captioning: Given a radiology image, generate a draft diagnostic report.

Biomedical Image Tagging: Given a radiology image, predict relevant biomedical tags.

Guided Decoding: A technique to steer the text generation of language models towards desired outcomes or characteristics.

2. Idea

The quality of a diagnostic caption is **strongly associated** with how effectively the generated report mentions the image's key biomedical tags.



3. Exploratory Step

Investigate the correlation between a tag and all the captions with which it is associated in the training set.

For a single caption s and a tag t we define:

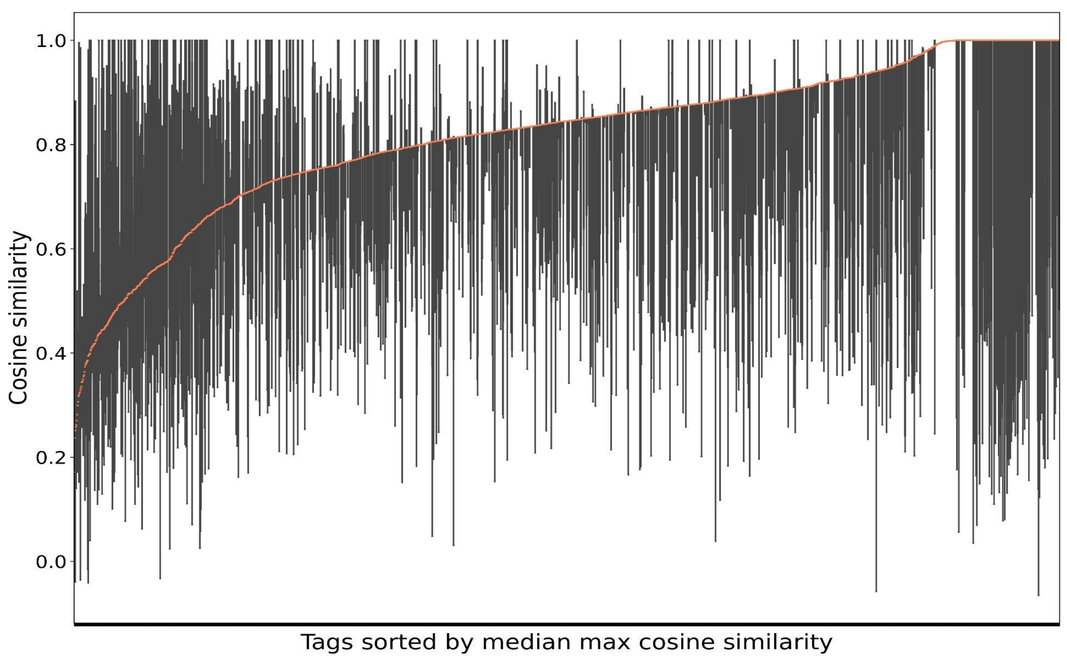
$$MCS(t,s) = \max_{1 \le j \le |s|} sim(h(t), h(s_j))$$

shows how strongly the tag t is mentioned in the caption s

$$MMCS(t) = median(\{MCS(t,s) | s \in S\})$$

shows how strongly the tag t is mentioned on average among its associated training captions S

Question: Are all the biomedical tags always strongly mentioned in the captions with which they are associated?



4. Proposed Mechanism (DMMCS)

Goal: By imposing a penalty at each decoding step, we incentivize the decoder to choose captions that encapsulate the associated biomedical tags more or less explicitly.Given a set of tags T and a caption s:

$$DMMCS_p(T,s) = \frac{1}{|T|} \cdot \sum_{t \in T} (MCS(t,s) - MMCS(t))^2$$

We want the sum to be as close to zero as possible!

The overall form looks like:

$$DMMCS(s) = \alpha \cdot DMMCS_p(T,s) + (1-\alpha) \cdot (1 - D_{score})$$

D*score*: Typical LM score.

5. Computational Overhead

Temporal Overhead: ~~25-27%~~ 17-19% **(Update!)**

Memory Overhead: 4-6%

6. Qualitative Results

Gold Caption	Vanilla Beam Search Decoding	DMMCS Decoding	Predicted Tags
Echocardiogram showing trivial **pericardial effusion.**	**Parasternal long axis view** of the **left ventricular outflow tract**. The arrow indicates the left ventricular tract.	Transthoracic **echocardiography** parasternal long axis view showing **pericardial effusion**.	*Transthoracic Echocardiography* *Pericardial Effusion*
Ultrasound of the **testes** showing a left testicular **mass** (encompassed within the yellow plus symbols).	**Ultrasound** scan of the right **breast** showing a well defined **lesion** with irregular margins.	**Ultrasonography** of the *right* **testis** showing a heterogeneous hypoechoic **mass** in the lower pole of the right **testis**.	*Ultrasonography* *Ultrasonography - Doppler - Duplex* *Structure of right testis*

7. Quantitative Results

Evaluation metrics: BLEU, BLEURT, Clinical Accuracy (CA)

	Show and Tell	ViT-GPT2	InstructBLIP	OpenFlamingo
ImageCLEFmedical2023	BS	ConBS ∃	DMMCS	BS
MIMIC-CXR	ConBS ∀	ConBS ∀	DMMCS	ConBS ∀

Model	ImageCLEFmedical 2023								MIMIC-CXR							
	BLEU				BLEURT				BLEU				BLEURT			
	BS	ConBS ∀	ConBS ∃	DMMCS	BS	ConBS ∀	ConBS ∃	DMMCS	BS	ConBS ∀	ConBS ∃	DMMCS	BS	ConBS ∀	ConBS ∃	DMMCS
Show and Tell	20.61	20.52	21.21	**21.27**	29.99	30.03	30.39	**30.47**	20.61	20.52	21.21	**21.27**	29.99	30.03	30.39	**30.47**
ViT-GPT2	15.34	15.75	16.29	**16.31**	26.50	26.31	26.92	**27.01**	15.34	15.75	16.29	**16.31**	26.50	26.31	26.92	**27.01**
InstructBLIP	11.81	15.89	**16.14**	15.93	29.68	29.71	30.08	**30.10**	11.81	15.89	**16.14**	15.93	29.68	29.71	30.08	**30.10**
OpenFlamingo	15.34	15.81	**15.92**	15.47	28.49	30.11	30.67	**31.34**	15.34	15.81	**15.92**	15.47	28.49	30.11	30.67	**31.34**