Token Imbalance Adaptation for Radiology Report Generation

Yuexin Wu, I-Chan Huang, Xiaolei Huang *

{xhuang7, ywu10}@memphis.edu

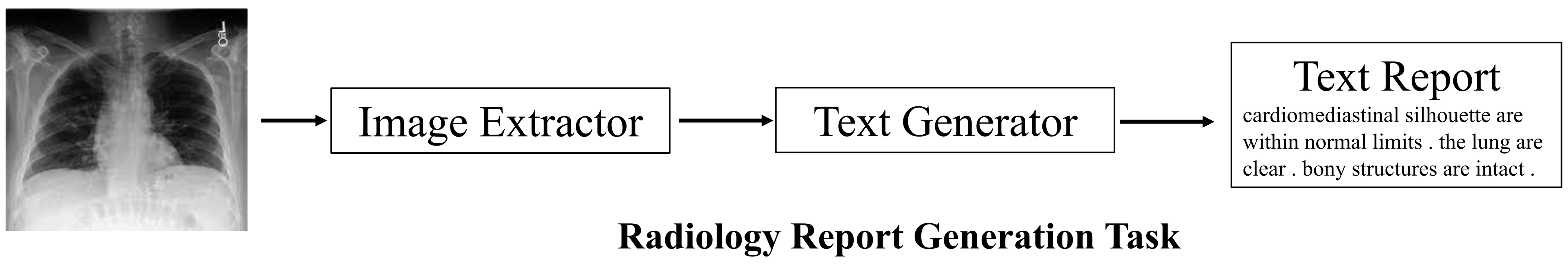
St. Jude Children's Research Hospital — THE UNIVERSITY OF MEMPHIS

❖Imbalance: Infrequent tokens have much higher medical- & disease-related words.

❖Robustness: Machine-generated reports severely **underfit** infrequent tokens.

❖Solution: We design the unlikelihood loss and dynamic adaptation mechanisms to augment model robustness on the token imbalance issue.

Introduction



Radiology Report Generation Task

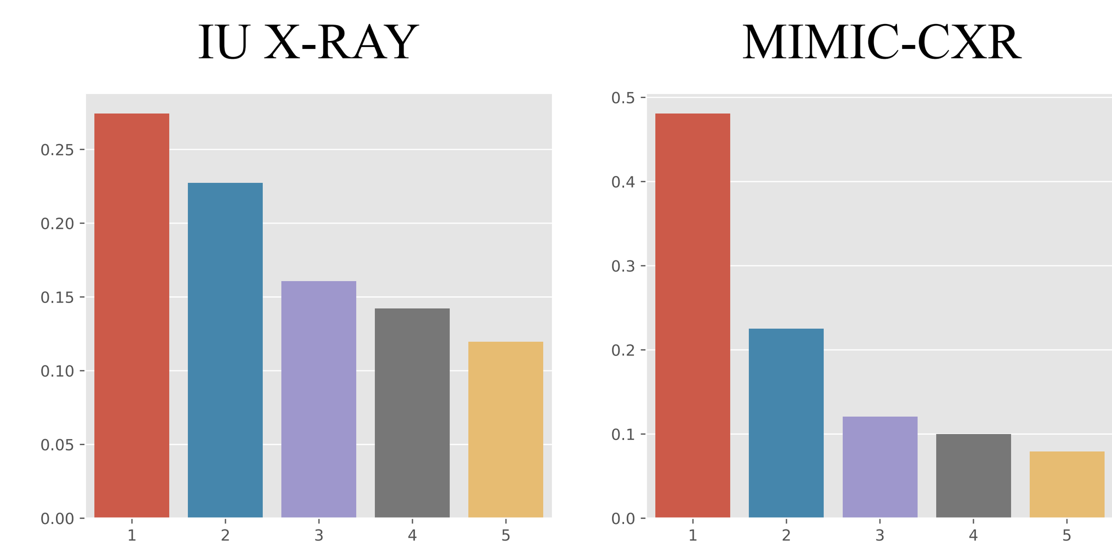


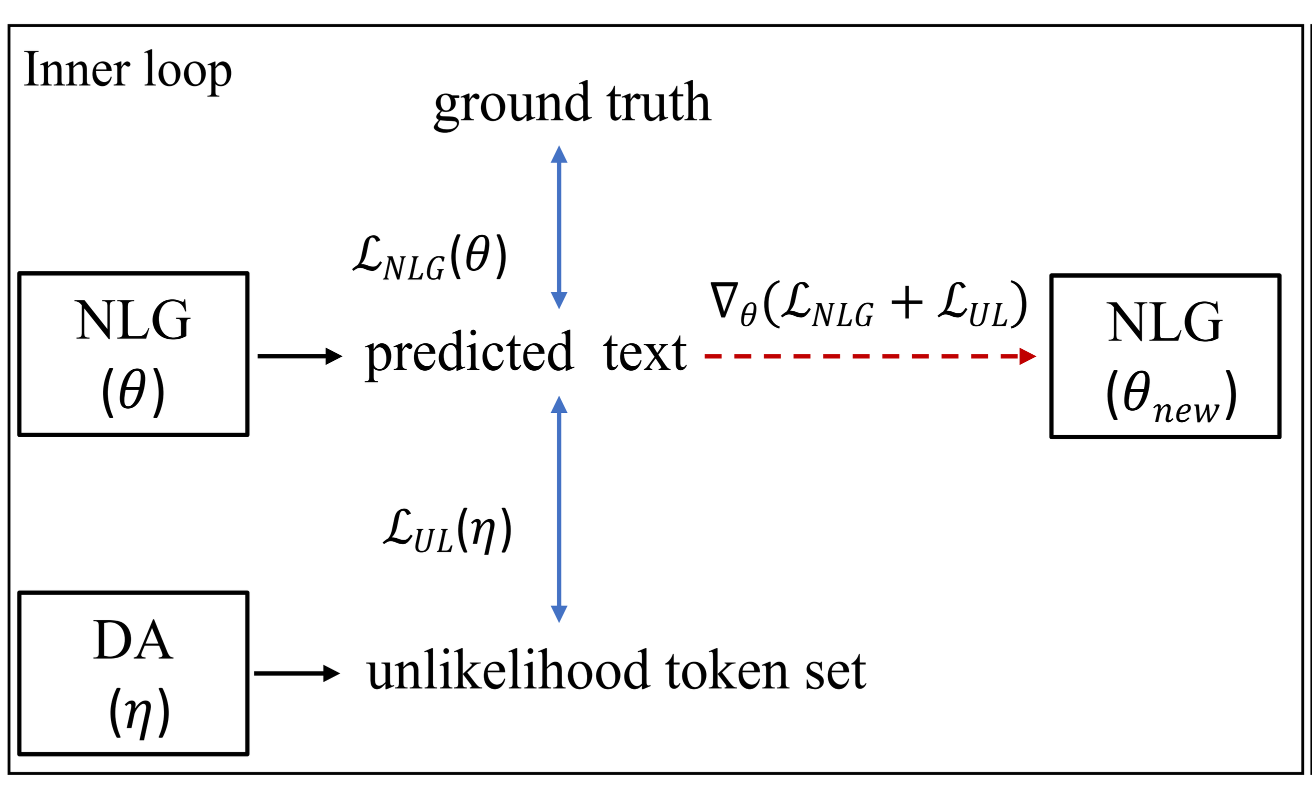
Figure 1: Ratios of medical terms across five equal splits of vocabulary.

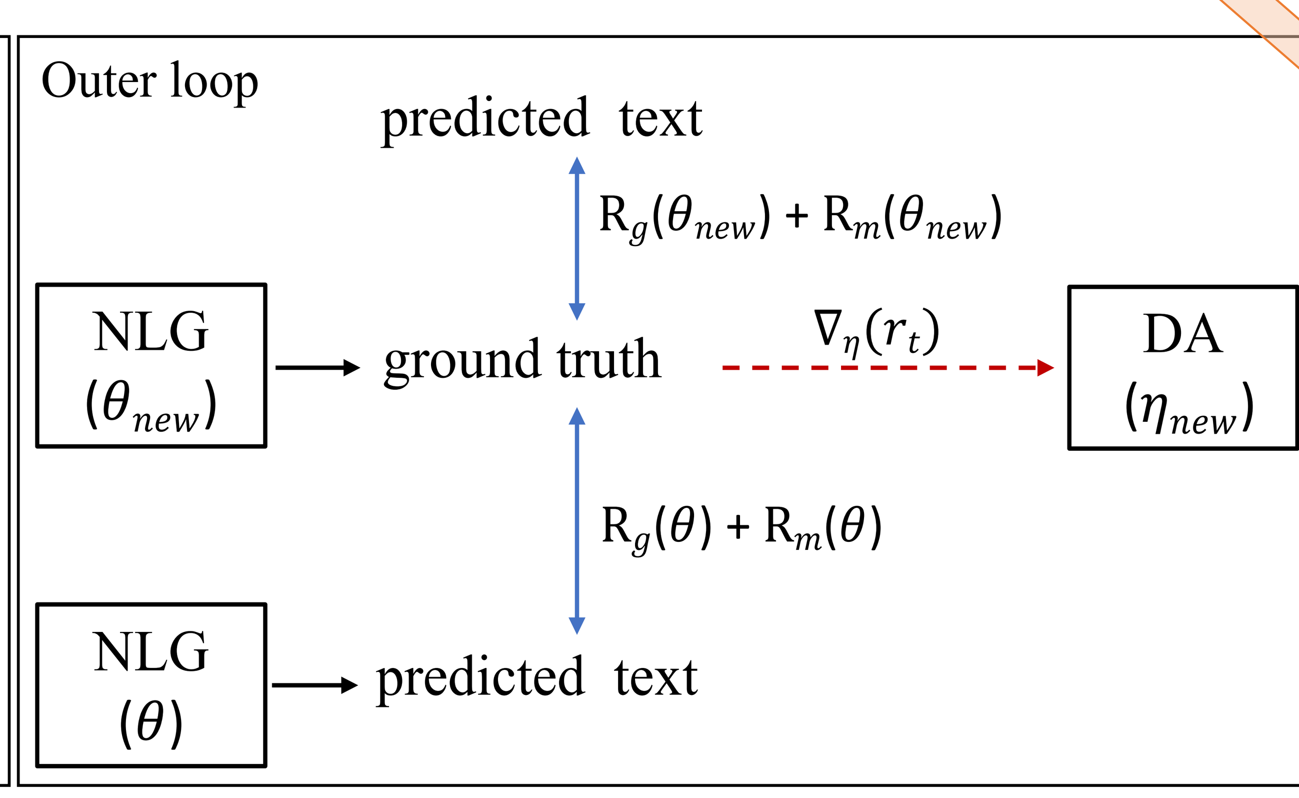
- The token imbalance may dampen the robustness of radiology report generators, while over 80% medical terms are infrequent tokens.
- Token Imbalance Adapter (TIMER) penalizes incorrect prediction of high-frequent tokens by the unlikelihood loss with dynamic unlikelihood tokens.

Non-Medical (high frequence)	Medical (low frequency)
is, the, are, of, or, and, no, normal	cardiomediastinal, sihouette, vascularity, pleural, effusion, pulmonary

Method

Module	Loss	Goal	Loss and Reward	Symbol	Training
Natural Language Generation (NLG, θ)	MLE	train a generation model	$\mathcal{L}_{NLG}(\theta) = -\sum_{l=1}^{L} \log p(y_l \mid y_1, \ldots, y_{l-1}, \boldsymbol{x}; \theta)$	x: image; L: the length of a report; **y**: a report;	Inner loop: Minimize the sum of MLE and Unlikelihood loss by gradient descent
	Unlikelihood loss (UL)	penalize predicted probabilities for frequent tokens	$\mathcal{L}_{UL}(U_h) = -\sum_{u \in U_h} \log(1 - p(u))$	U_h : the unlikelihood token set;	
Dynamic Adaptation (DA, η)	Imbalance reward	train a dynamic unlikelihood set	$R_m = (\lvert F(U_h) - F(U_l)\rvert + \lvert F(U_m) - F(U_l)\rvert + F(U_h) - F(U_m)\rvert)/3$	F(U_l), F(U_m), F(U_h): the F1 score of the low, medium and high-frequency token sets;	Outer loop: maximize the reward by reinforcement learning
	Imbalance reward with baselines	adapt to model's learning ability	$r_t = R_g(\theta_{new}) + R_m(\theta_{new}) - R_g(\theta) - R_m(\theta)$		

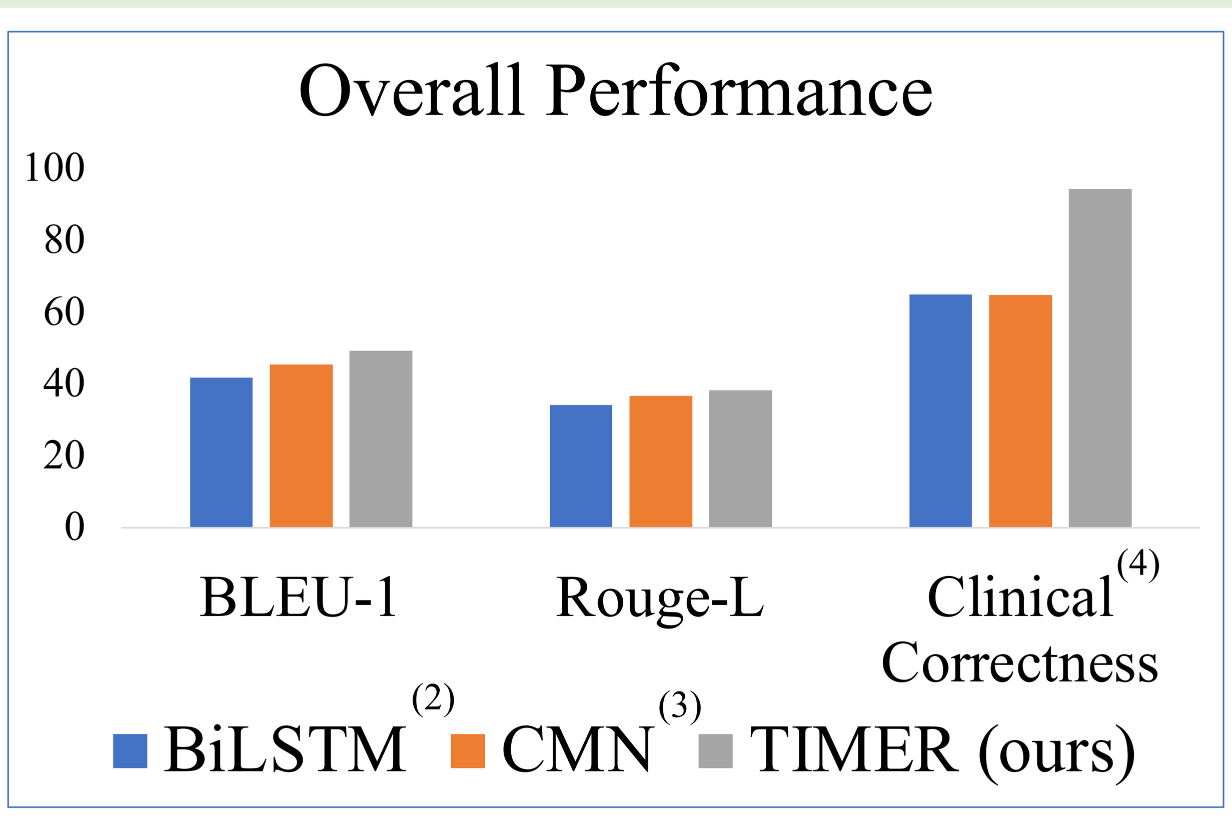


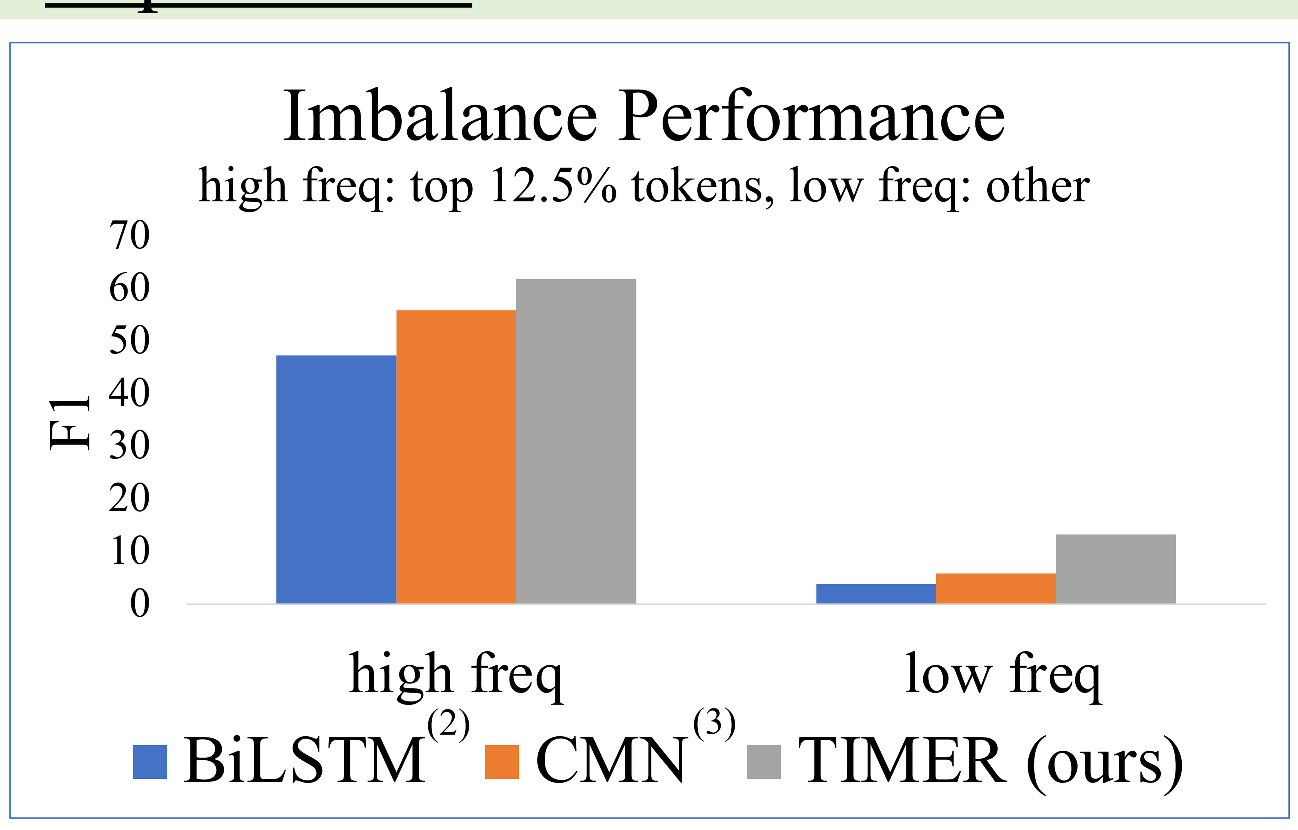


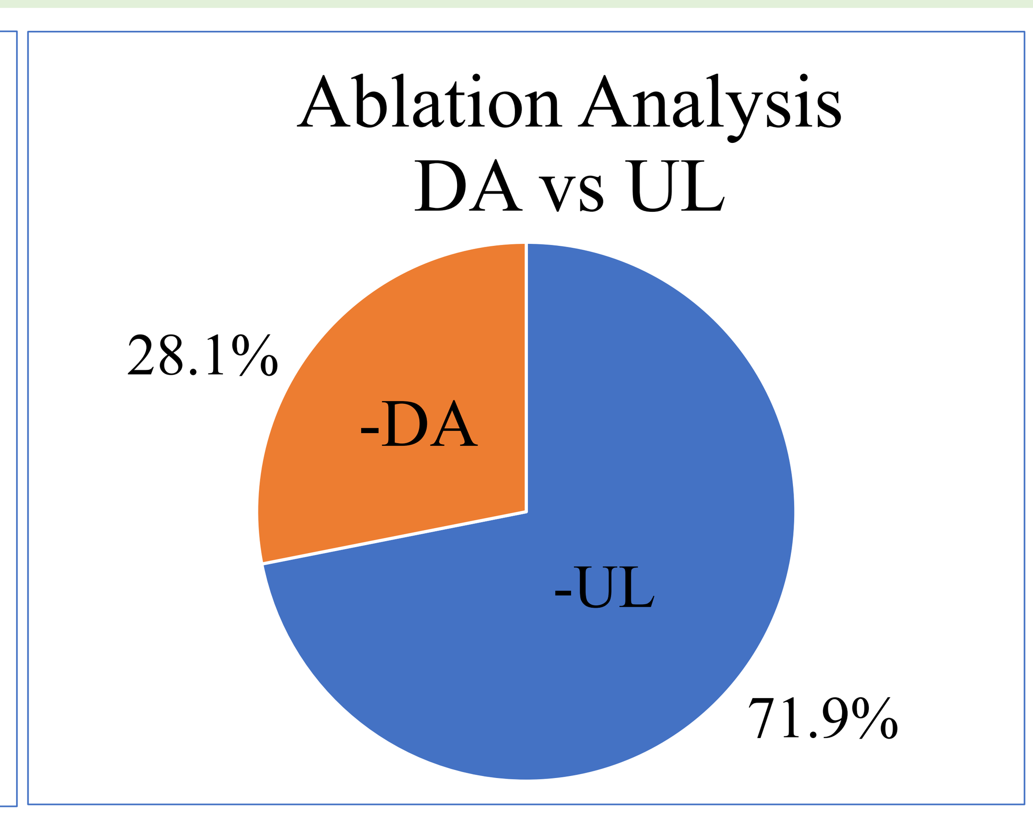
Balance generation performance and promote imbalance learning across different frequency tokens

https://github.com/woqingdoua/TIMER

Experiments on IU X-RAY[1]







- TIMER obtains significant improvements on infrequent token generations while maintaining performance on frequent tokens by multiple evaluation metrics.
- Unlikelihood loss (UL) have a greater contribution than dynamic adaptation (DA) in promoting model performance.

(1) Demner-Fushman, Dina, et al. "Preparing a collection of radiology examinations for distribution and retrieval."

(2) Jing, Baoyu, Pengtao Xie, and Eric Xing. "On the automatic generation of medical imaging reports." *arXiv preprint arXiv:1711.08195* (2017).

(3) Chen, Zhihong, et al. "Cross-modal memory networks for radiology report generation." *arXiv preprint arXiv:2204.13258* (2022).

(4) Smit, Akshay, et al. "CheXbert: combining automatic labelers and expert annotations for accurate radiology report labeling using BERT." *arXiv preprint arXiv:2004.09167* (2020).

Acknowledgements

This work was supported by a gift from West Cancer Foundation, Ralph E. Powe Junior Faculty Enhancement Award, and the National Science Foundation with award number IIS-2245920.