Graph Enhanced Contrastive Learning for Radiology Findings Summarization

Jinpeng Hu♠♥* , Zhuo Li♠♥* , Zhihong Chen♠♥, Zhen Li♠♥
Xiang Wan♥♦†, Tsung-Hui Chang♠♥†
♠The Chinese University of Hong Kong (Shenzhen)
♥Shenzhen Research Institute of Big Data
♦Pazhou Lab, Guangzhou, 510330, China

ACL2022

Code:https://github.com/jinpeng01/AIG_CL

2022. 6. 15 • ChongQing

Leibniz-Institut für Sozialwissenschaften

Reported by Yang Peng

Artificial

1.Introduction

2.Method

3.Experiments

gesis
Leibniz-Institut
für Sozialwissenschaften

Introduction

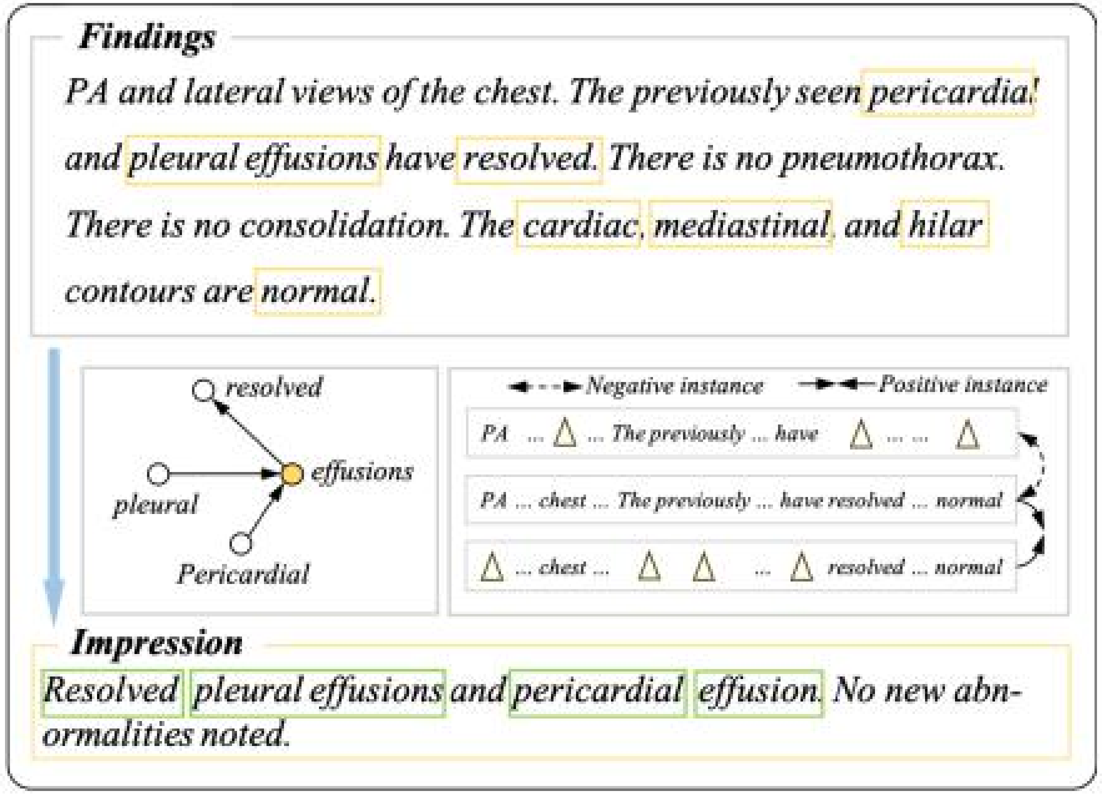


Figure 1: An example of the findings and corresponding impression, where the relation information, as well as positive and negative examples, are also shown in the figure. Note that △ represents the removed word.

Method

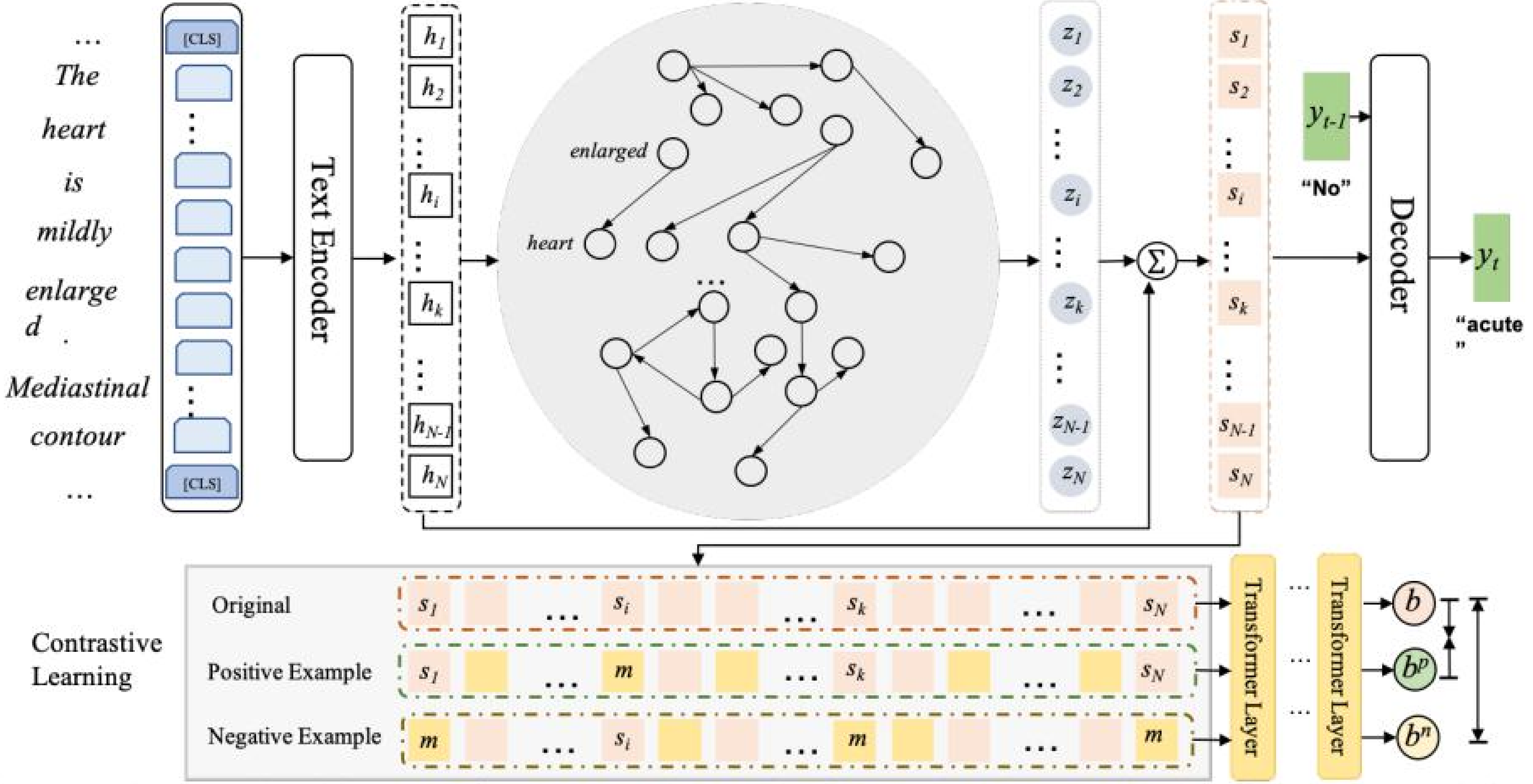
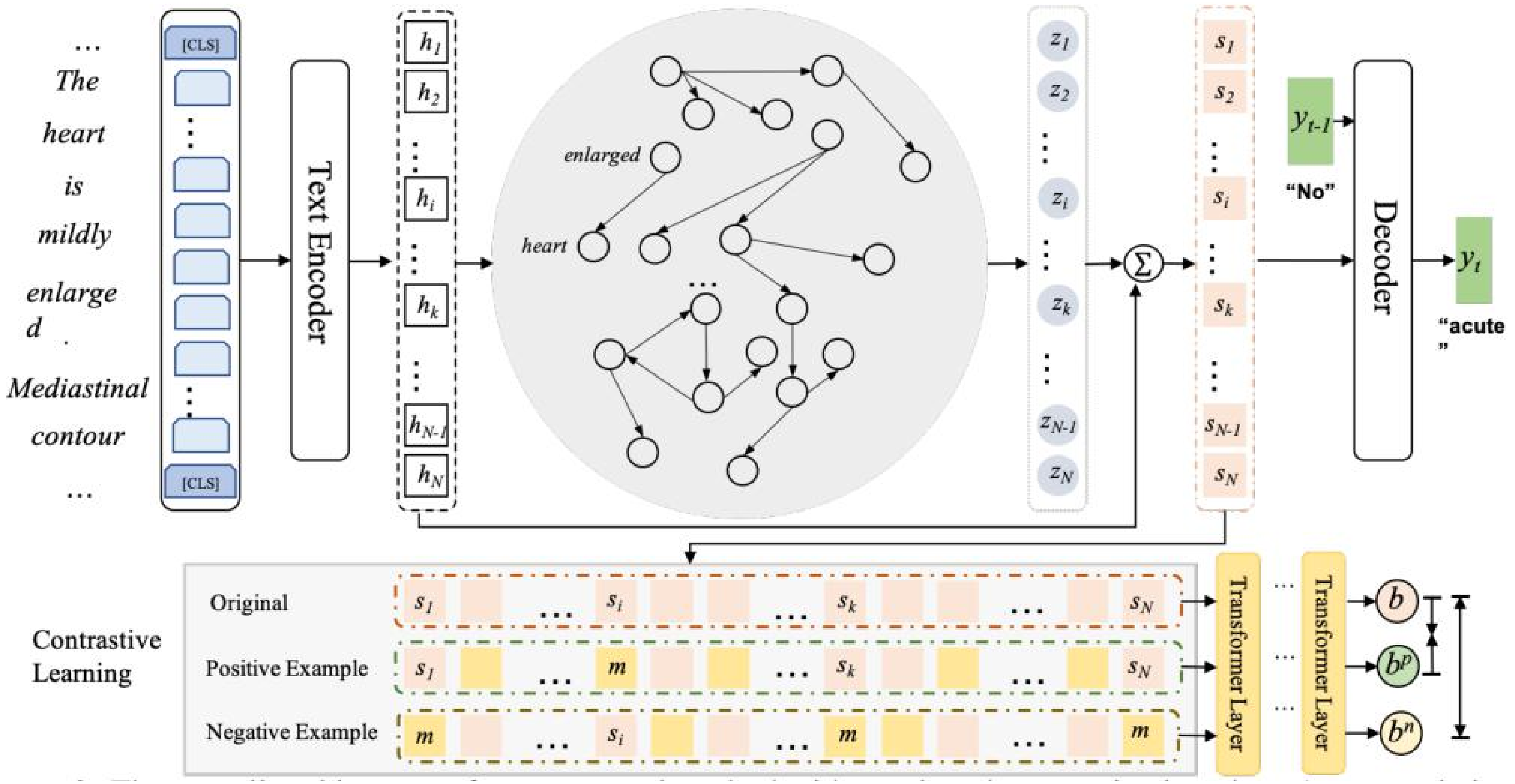


Figure 2: The overall architecture of our proposed method with graph and contrastive learning. An example input and output at $t-1$ and t step are shown in the figure, where the top is the backbone sequence-to-sequence paradigm with a graph to store relation information between critical words and the bottom is the contrastive learning module with specific positive and negative examples. m refer to a mask vector.

Method



input sequence $\mathcal{X} = x_1, x_2, \cdots, x_N$,

find a sequence $\mathbf{Y} = \{y_1, \ldots y_i, \ldots, y_L\}$

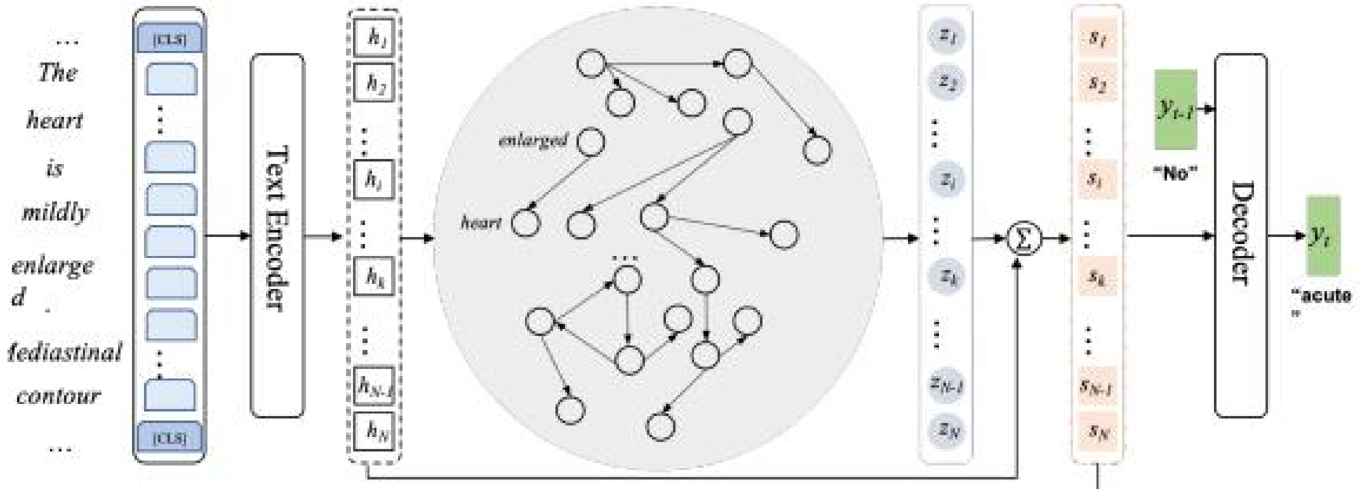
$$p(\mathbf{Y} \mid \mathcal{X}) = \prod_{t=1}^{L} p\left(y_t \mid y_1, \ldots, y_{t-1}, \mathcal{X}\right) \quad (1)$$

The model is then trained to maximize the negative conditional log-likelihood of $\mathcal{Y}$ given the $\mathcal{X}$:

$$\theta^* = \arg\max_{\theta} \sum_{t=1}^{L} \log p\left(y_t \mid y_1, \ldots, y_{t-1}, \mathcal{X}, A; \theta\right) \quad (2)$$

Method

Graph Enhanced Encoder



each token x_i

hidden state $\mathbf{h}_i$

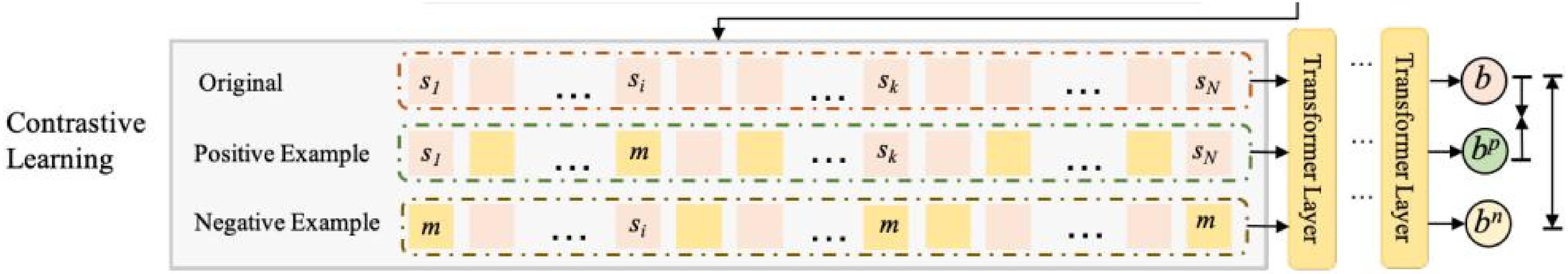
$$[\mathbf{h}_1, \mathbf{h}_2, \cdots, \mathbf{h}_n] = f_{te}(x_1, x_2, \cdots, x_n) \quad (3)$$

$$\mathbf{z} = f_{ge}(\mathbf{h}, A), \quad (4)$$

$$\mathbf{s} = \mathrm{MLP}([\mathbf{h}_1 \oplus \mathbf{z}_1, \mathbf{h}_2 \oplus \mathbf{z}_2, \cdots, \mathbf{h}_n \oplus \mathbf{z}_n]), \quad (5)$$

Method

Contrastive Learning



$$sim(\mathbf{b}^1, \mathbf{b}^2) = \frac{\mathbf{b}_1^\top \mathbf{b}_2}{\|\mathbf{b}_1\| \cdot \|\mathbf{b}_2\|}$$

$$\mathbf{b} = f_{ce}(\mathbf{s}), \tag{6}$$

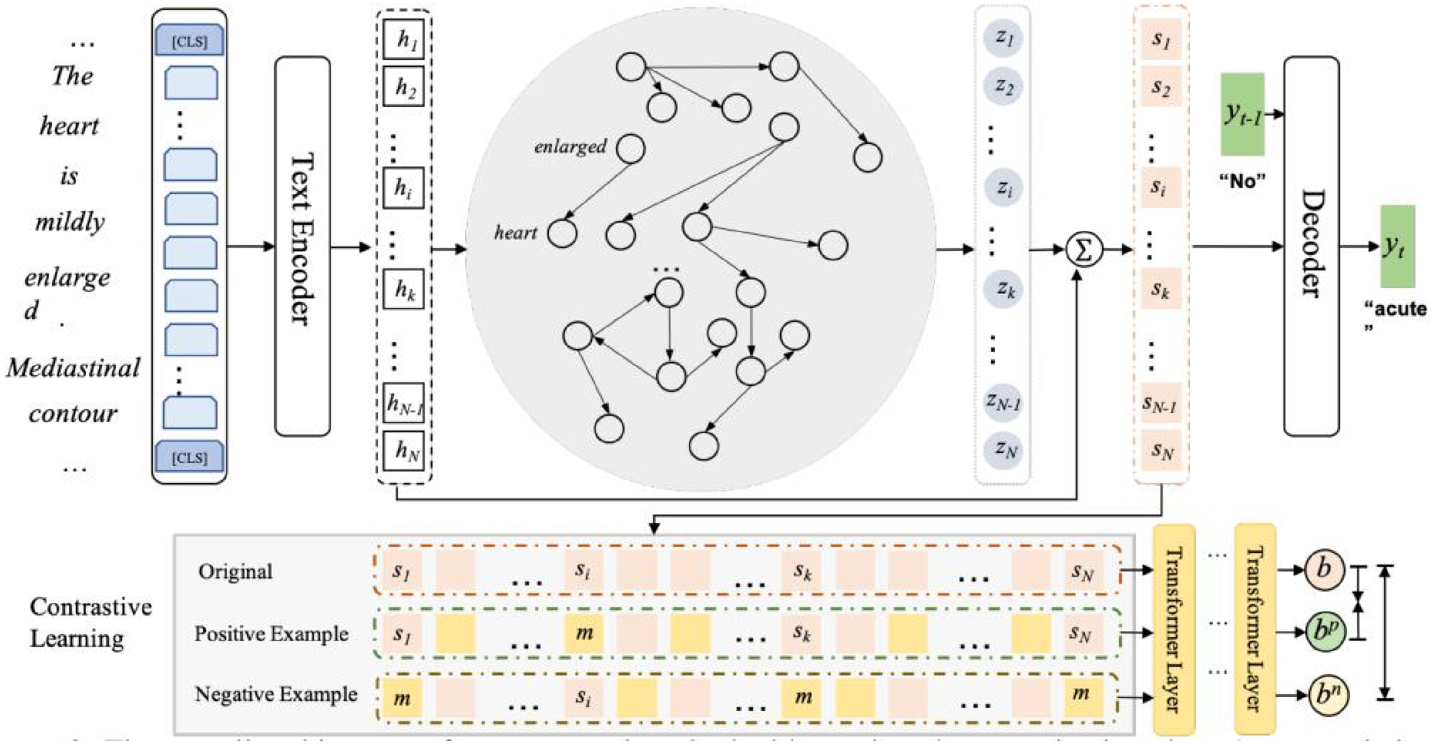
$$\mathbf{b}^p = f_{ce}(\mathbf{s}^p), \tag{7}$$

$$\mathbf{b}^n = f_{ce}(\mathbf{s}^n), \tag{8}$$

$$l_{con} = -\log \frac{e^{\mathrm{sim}(\mathbf{b}_i, \mathbf{b}^p)/\tau}}{\sum_{j=1}^{N} \left(e^{\mathrm{sim}(\mathbf{b}_i, \mathbf{b}^p)/\tau} + e^{\mathrm{sim}(\mathbf{b}_i, \mathbf{b}^n)/\tau}\right)} \tag{9}$$

Method

Decoder



$$y_t = f_e(\mathbf{s}_1, \mathbf{s}_2, \cdots, \mathbf{s}_n, y_1, \cdots, y_{t-1}), \tag{10}$$

$$L = l_{ge} + \lambda l_{con}, \tag{11}$$

where l_{ge} is the basic sequence-to-sequence loss, and λ is the weight to control the contrastive loss.

Experiments

DATA	TYPE	TRAIN	DEV	TEST
OPENI	REPORT #	2400	292	576
	AVG. WF	37.89	37.77	37.98
	AVG. SF	5.75	5.68	5.77
	AVG. WI	10.43	11.22	10.61
	AVG. SI	2.86	2.94	2.82
MIMIC -CXR	REPORT #	122,014	957	1,606
	AVG. WF	55.78	56.57	70.67
	AVG. SF	6.50	6.51	7.28
	AVG. WI	16.98	17.18	21.71
	AVG. SI	3.02	3.04	3.49

Table 1: The statistics of the two benchmark datasets with random split for OPENI and official split for MIMIC-CXR, including the numbers of report, the averaged sentence-based length (AVG. SF, AVG. SI), the averaged word-based length (AVG. WF, AVG. WI) of both IMPRESSION and FINDINGS.

Experiments

DATA	MODEL	ROUGE R-1	ROUGE R-2	ROUGE R-L	FC P	FC R	FC F-1
OPENI	BASE	62.74	53.32	62.86	-	-	-
	BASE+CL	63.53	54.58	63.13	-	-	-
	BASE+GRAPH	63.29	54.12	63.03	-	-	-
	BASE+GRAPH+CL	**64.97**	**55.59**	**64.45**	-	-	-
MIMIC-CXR	BASE	47.92	32.43	45.83	58.05	50.90	53.01
	BASE+CL	48.15	33.25	46.24	58.34	51.58	53.70
	BASE+GRAPH	48.29	33.30	46.36	57.80	51.70	53.50
	BASE+GRAPH+CL	**49.13**	**33.76**	**47.12**	**58.85**	**52.33**	**54.52**

Table 2: Comparisons of baselines and our method on OPENI and MIMIC-CXR datasets. R-1, R-2 and R-L refer to ROUGE-1, ROUGE-2 and ROUGE-L, respectively. P, R and F-1 represent precision, recall, and F1 score.

Experiments

MODEL	OPENI RANDOM SPLIT R-1	OPENI RANDOM SPLIT R-2	OPENI RANDOM SPLIT R-L	MIMIC-CXR OFFICIAL SPLIT R-1	MIMIC-CXR OFFICIAL SPLIT R-2	MIMIC-CXR OFFICIAL SPLIT R-L	MIMIC-CXR RANDOM SPLIT R-1	MIMIC-CXR RANDOM SPLIT R-2	MIMIC-CXR RANDOM SPLIT R-L
LEXRANK (Erkan and Radev, 2004)	14.63	4.42	14.06	18.11	7.47	16.87	-	-	-
TRANSEXT (Liu and Lapata, 2019)	15.58	5.28	14.42	31.00	16.55	27.49	-	-	-
PGN (LSTM) (See et al., 2017)	63.71	54.23	63.38	46.41	32.33	44.76	-	-	-
TRANSABS (Liu and Lapata, 2019)	59.66	49.41	59.18	47.16	32.31	45.47	-	-	-
ONTOLOGYABS† (Gharebagh et al., 2020)	-	-	-	-	-	-	53.57	40.78	51.81
WGSUM (LSTM)† (Hu et al., 2021)	64.32	55.48	63.97	47.48	33.03	45.43	54.97	43.64	53.81
WGSUM (TRANS)† (Hu et al., 2021)	61.63	50.98	61.73	48.37	33.34	46.68	56.38	44.75	55.32
OURS	**64.97**	**55.59**	**64.45**	**49.13**	**33.76**	**47.12**	**57.38**	**45.52**	**56.13**

Table 3: Comparisons of our proposed models with previous study on the OPENI and MIMIC-CXR with respect to ROUGE metric. † refers to that the results is directly cited from the original paper.

Experiments

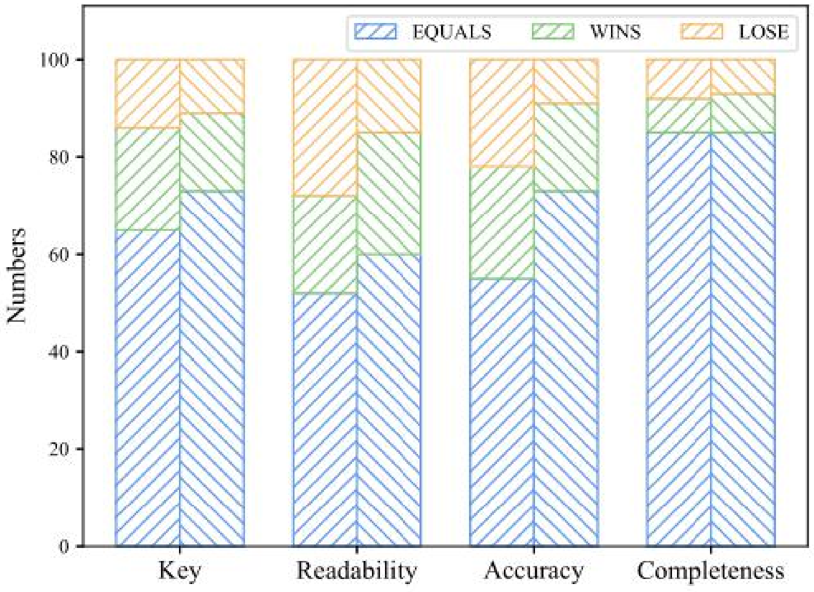


Figure 3: The results of human evaluation, where forward and backslash represent that BASE+GRAPH+CL versus the reference and BASE, respectively. Yellow, green and blue represent that our model loses, equal to competitors and wins.

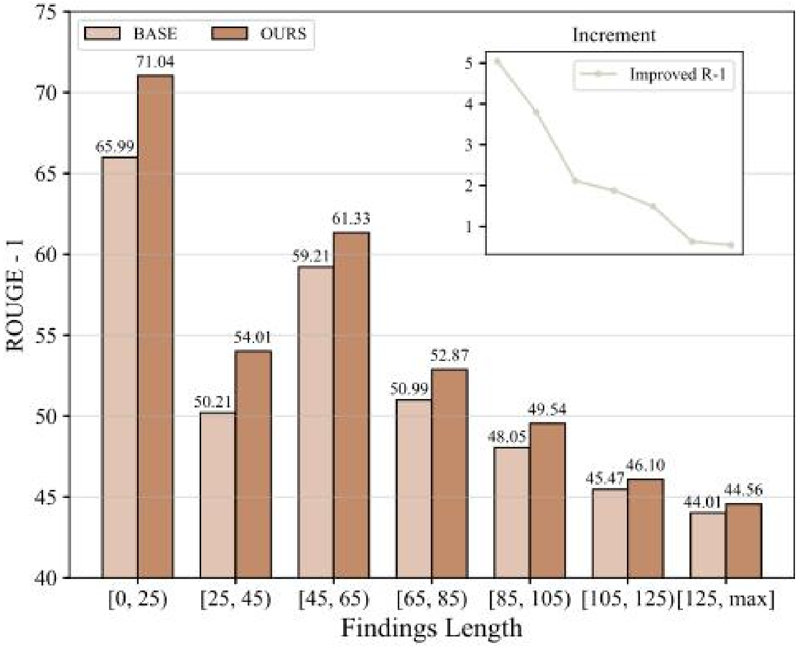


Figure 4: R-1 score of generated impressions from BASE and our model on the MIMIC-CXR test set, where OURS represent the BASE+GRAPH+CL.

Experiments

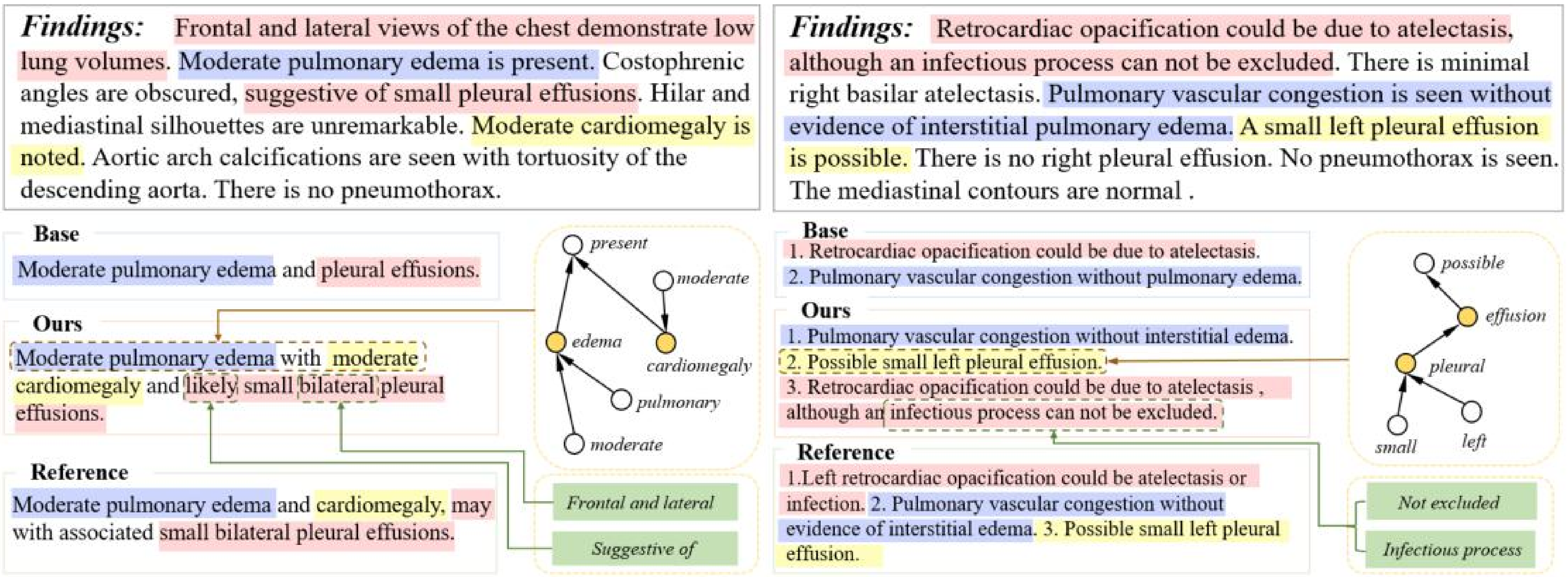


Figure 5: Examples of the generated impressions from BASE and BASE+GRAPH+CL as well as reference impressions. The yellow nodes in the graph indicate that these words are contained in entities.

Thank you!