CoNT: Contrastive Neural Text Generation

Aditya Jain (aj780)
Abhinay Reddy Vongur (av730)
Parth Hasmukh Jain (pj269)

Related Work and Problem

- Why contrastive learning ?
  - Contrastive learning is used to combat exposure bias.
  - It helps in generalizing the model for unseen inputs.
  - Introduces sequence level loss - crucial because generation models are evaluated on sequence level scores (e.g. BLEU)
- Naive Contrastive Learning (NaiveCL)
  - Only uses from-batch examples as negatives samples
  - Loss function treats all negative samples the same.
- Previous applications of contrastive learning in Text Generation
  - Generated hard negatives by perturbing the ground truth

CoNT

- Generate hard negatives from the model predictions.
- Rank hard negatives based on their similarity with ground truth.
- N-pairs contrastive loss based on cosine similarity and the rank.

$$\mathcal{L}_{\text{N-Pairs}} = \sum_{(\boldsymbol{y}^+, \boldsymbol{y}^-) \in \mathcal{P}} \mathcal{L}(\boldsymbol{y}^+, \boldsymbol{y}^-) = \sum_{(\boldsymbol{y}^+, \boldsymbol{y}^-) \in \mathcal{P}} \max\{0, \cos(\mathbf{z}_{\boldsymbol{x}}, \mathbf{z}_{\boldsymbol{y}^-}) - \cos(\mathbf{z}_{\boldsymbol{x}}, \mathbf{z}_{\boldsymbol{y}^+}) + \xi\}.$$

- Use the learned similarity function in the inference stage.

$$\boldsymbol{y}^* = \arg\max_{\hat{\boldsymbol{y}}} \{\alpha \cdot \cos(\mathbf{z}_{\boldsymbol{x}}, \mathbf{z}_{\hat{\boldsymbol{y}}}) + (1 - \alpha) \prod_{t=0}^{n} p(\hat{y}_t | \boldsymbol{x}, \hat{\boldsymbol{y}}_{<t})\},$$

Results

Model	BLEU-3/4		ROUGE-L	METEOR	CIDEr
T5-base	28.76	18.54	34.56	23.94	9.4
T5-large	43.01	31.96	42.75	31.12	15.13
T5-base-CONT	42.6	31.42	43.15	32.05	15.96
Reproduced results	29.3	20.6	49.8	28.9	12.67

Results on Common Sense Generation task

```
pot butter add egg crack: Then, the chef adds the butter, then the egg and the cheese, stirring the pot.

design piece burn wood tool: A person shows how to design a piece of wood with a burning tool.
```

Example Output from the model for given concepts