Contrastive Representation Learning for Exemplar-Guided Paraphrase Generation

Project By:
1. Harshit Gupta (2020114017)
2. Pratyaksh Gautam (2020114002)

Introducing The Problem Statement

The goal of Exemplar-Guided Paraphrase Generation (EGPG) is to produce a target sentence that matches the style of the provided exemplar while preserving the source sentence's content information.

New Idea and Approach: Designing two contrastive losses with regard to the content and style.

The target sentence has two characteristics: first, it has the same substance as the source sentence, and second, it has the same style as the exemplar sentence.

source (X)	*what is the easiest way to get followers on quora ?*
exemplar (Z)	*how do i avoid plagiarism in my article ?*
target (Y)	*how do i get more followers for my quora ?*
retrieved (Y')	*what are the better ways to ask questions on quora ?*

Dataset Characteristics

We train and evaluate our model on 2 datasets. They are:

1. **QQPos:** Formal sentence pairs.

Link: https://www.kaggle.com/competitions/quora-question-pairs/data

2. **ParaNMT:** Informal sentence pairs trained on a back translation task for English language.

Link: https://metatext.io/datasets/paranmt-50m

We take around 93k sentences for training, 3k for validation and 3k for testing purposes.

NOTE: These datasets contain source and target sentences ONLY!

How do we get Exemplar Sentences??

To obtain Exemplar data, we actually create them from scratch using the above dataset and the algorithm outlined below. *Link:* https://arxiv.org/pdf/2005.08417.pdf

Algorithm 1 Searching Exemplar Sentences

Require: dataset $\mathbb{D} = (\mathbb{D}_X, \mathbb{D}_Y)$

1: **for** Y in $\mathbb{D}_Y$ **do**
2: find the sentence set $\mathbb{C}_1 \subseteq \mathbb{D}_X$ that each $C \in \mathbb{C}_1$ satisfies $|len(C) - len(Y)| \leq 2$
3: find the sentence set $\mathbb{C}_2 \subseteq \mathbb{C}_1$ that for each $C \in \mathbb{C}_2$, the number of shared words between C and Y, denoted by c, satisfies $c + 2 \leq len(Y)$
4: find the exemplar $Z \in \mathbb{C}_2$ which has the smallest POS tag sequence editdistance with Y
5: **end for**

The Main Model Architecture

The Main Model Architecture can be summarized below as:

The Style Encoder is a BERT-base architecture while the Content Encoder is a Bi-Directional GRU. The decoder is a single layered GRU.

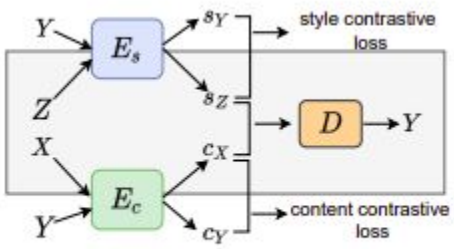

Figure 1: An Overview of Our Model

Some Important Parameters

1. We use 300-d GloVe Embeddings.
2. Hidden dimension size of the Content Encoder is 512 while for Style Encoder it is 768.
3. The learning rate used is 0.0001 and Epochs are set to around 15.
4. Maximum sentence length taken is 15.
5. Teacher forcing ratio is 1.0
6. Regarding the triple loss function, the balancing parameters are set to 0.1 each while the temperature parameter is set at 0.5

Understanding The Loss Function

The Loss Function used in this approach can be seen below as:

$$L = \sum_{i=1}^{n} L_i^{nll} + \lambda_1 L^{ccl} + \lambda_2 L^{scl}$$

It is the sum of:

1. Cross Entropy Loss (NLL+Internal Softmax)
2. Content Contrastive Loss: Between style represented **target** and **exemplar** sentences.
3. Style contrastive loss: Between content represented **source** and **target** sentences.

Content Contrastive Loss

$$\mathcal{L}_{X_i}^{ccl} = -log\frac{exp(c_{X_i} \cdot c_{Y_i}/\tau)}{exp(\frac{c_{X_i} \cdot c_{Y_i}}{\tau}) + \sum\limits_{\substack{j \neq i \\ T \in \{X,Y\}}} exp(\frac{c_{X_i} \cdot c_{T_j}}{\tau})} \tag{6}$$

$$\mathcal{L}_{Y_i}^{ccl} = -log\frac{exp(c_{Y_i} \cdot c_{X_i}/\tau)}{exp(\frac{c_{Y_i} \cdot c_{X_i}}{\tau}) + \sum\limits_{\substack{j \neq i \\ T \in \{X,Y\}}} exp(\frac{c_{Y_i} \cdot c_{T_j}}{\tau})} \tag{7}$$

$$\mathcal{L}^{ccl} = \sum_{i=1}^{n}(\mathcal{L}_{X_i}^{ccl} + \mathcal{L}_{Y_i}^{ccl}) \tag{8}$$

Style Contrastive Loss

$$\mathcal{L}_{Y_i}^{scl} = -log\frac{exp(s_{Y_i} \cdot s_{Z_i}/\tau)}{exp(\frac{s_{Y_i} \cdot s_{Z_i}}{\tau}) + \sum\limits_{\substack{j \neq i \\ T \in \{Z,Y\}}} exp(\frac{s_{Y_i} \cdot s_{T_j}}{\tau})} \tag{9}$$

$$\mathcal{L}_{Z_i}^{scl} = -log\frac{exp(s_{Z_i} \cdot s_{Y_i}/\tau)}{exp(\frac{s_{Z_i} \cdot s_{Y_i}}{\tau}) + \sum\limits_{\substack{j \neq i \\ T \in \{Z,Y\}}} exp(\frac{s_{Z_i} \cdot s_{T_j}}{\tau})} \tag{10}$$

Why The Above 2 Losses ?

Some probable benefits of using the above 2 loss functions can help are:

1. Gradient vanishing, a serious problem when training the encoder-decoder model, can be resolved by adding extra losses to the encoders' output.
2. Since the contrastive losses limit the model's freedom of parameter adjustment by requiring the encoder and decoder to simultaneously concentrate on their respective tasks of feature extraction and sentence generation, the overfitting problem may be avoided.

Adding Attention

The paper doesn't go on to specifically state how and where to pay attention, but we accomplish the same thing by using the following strategy. Scaled Dot Product Attention is what we use. The Content Encoder and Decoder are regarded as the typical encoder-decoder model. In order for the model to learn and pay attention to style-based parameters as well, we concatenated style representations with the content representations when we passed them.

Results Generated

We generated ROUGE-L scores between the generated sentences and the target sentences (ground truth)

Metric	QQPos	ParaNMT
ROUGE1: Precision	0.4004	0.5416
ROUGE1: Recall	0.3754	0.5179
ROUGE1: F-Measure	0.3861	0.5271
ROUGE2: Precision	0.1711	0.3116
ROUGE2: Recall	0.1640	0.3014
ROUGE2: F-Measure	0.1670	0.3052
ROUGEL: Precision	0.3442	0.4890
ROUGEL: Recall	0.3246	0.4696
ROUGEL: F-Measure	0.3329	0.4770
BLEU	0.0995	0.1196
METEOR	0.3234	0.4813

Some Generated Results

We generated sentences using the testing data and have presented a few them here

On the Para-NMT Dataset

```
Source:  chemical and physical compatibility has been demonstrated with the following <UNK> <PAD> <PAD> <P
AD> <PAD>
Target:  chemical and physical stability has been demonstrated with the following solvents <PAD> <PAD> <PA
D> <PAD>
Exemplar:  [CLS] chemical and physical compatibility has been demonstrated with the following diluents [SE
P]
Generated:  <PAD> chemical and quality has been demonstrated with <EOS> <EOS> <PAD> <PAD> <PAD> <PAD> <PAD
>
```

```
Source:  gentlemen . . . my client is n ' t saying another word . <PAD>
Target:  my client wo n ' t tell you anything . <PAD> <PAD> <PAD> <PAD> <PAD>
Exemplar:  [CLS] i can't abandon my patients. [SEP] [PAD] [PAD] [PAD] [PAD]
Generated:  <PAD> i client is n ' t saying another word . <EOS> <PAD> <PAD> <PAD>
----------------------------------
```

```
Source:  well , i hope you know that when it comes to these sorts of situations
Target:  i hope you know if something like that happens , you can trust me .
Exemplar:  [CLS] i want my men off this boat. i am countermanding [SEP]
Generated:  <PAD> i , i hope you know that comes it comes to these kinds <EOS>
----------------------------------
```

```
Source:  the doctor ' s optimism about the pace of his recovery turned out to be
Target:  the doctor was not mistaken in his optimism about the pace of recovery . <PAD>
Exemplar:  [CLS] these segments are determined in relation to the axis of reference. [SEP] [PAD]
Generated:  <PAD> the doctor ' s expressed of hopes about of the ' a to <EOS>
----------------------------------
```

On the QQPos Dataset

```
Source:  which are the best books to learn html , css and javascript ? <PAD> <PAD>
Target:  what are the best books for learning css , javascript and php ? <PAD> <PAD>
Exemplar:  [CLS] which are the best books to learn html, css and java [SEP]
Generated:  <PAD> which are the best books to to learn c++ and c++ ? <EOS> <PAD>
----------------------------------
```

```
Source:  what are some good badminton rackets ? <PAD> <PAD> <PAD> <PAD> <PAD> <PAD> <PAD> <PAD>
Target:  which is the best badminton racket ? <PAD> <PAD> <PAD> <PAD> <PAD> <PAD> <PAD> <PAD>
Exemplar:  [CLS] what are the best philosophy podcasts? [SEP] [PAD] [PAD] [PAD] [PAD] [PAD]
Generated:  <PAD> what are some best badminton ? <EOS> <PAD> <PAD> <PAD> <PAD> <PAD> <PAD> <PAD>
----------------------------------
```

```
Source:  why did the indian government ban rs . 500 and rs . 1000 currency ?
Target:  did the indian government ban the 500 rs & 1000 rupees notes ? <PAD> <PAD>
Exemplar:  [CLS] why has the modi government banned the 500 and 1000 rupee [SEP]
Generated:  <PAD> why the indian indian ban 500 and 1000 rupee and and 500 and <EOS>
----------------------------------
```

```
Source:  what are the applications of linear algebra in biology ? <PAD> <PAD> <PAD> <PAD> <PAD>
Target:  what is the application of linear algebra to economics ? <PAD> <PAD> <PAD> <PAD> <PAD>
Exemplar:  [CLS] what is the process of converting joules to watts? [SEP] [PAD] [PAD]
Generated:  <PAD> what is the applications of linear algebra in biology ? ? <EOS> <PAD> <PAD>
----------------------------------
```

Thank You! :)
Any suggestions?