Generating Syntactically Controlled Paraphrases without Using Annotated Parallel Pairs

Kuan-Hao Huang and Kai-Wei Chang
University of California, Los Angeles

EACL 2021
Paraphrase Generation Tasks

• Build a paraphrase model
  • Input sentence $\rightarrow$ output paraphrase
• Supervised approaches
  • Need many annotated paraphrase pairs for training
• Unsupervised approaches
  • Generated paraphrases are not diverse in syntax
• Our goal
  • Generate syntactically diverse paraphrases
  • Train a paraphrase model without using annotated paraphrase pairs
Syntactically Controlled Paraphrase Generation

- Control the syntax of output paraphrases [Iyyer+ 2018]
- Give model target parse templates as control signals
- Challenges
  - No ground truths
  - Syntactic control

```
We are going to have a picnic if it is a sunny day tomorrow.

If tomorrow is a sunny day, we will have a picnic.

We will have a picnic if it is a sunny day tomorrow.

Let's have a picnic if tomorrow is a sunny day.
```
Syntactically Controlled Paraphrase Generator (SynPG)

- Disentangle a sentence into syntactic and semantic embedding
  - **Syntactic embedding:** encoded from constituency parse
  - **Semantic embedding:** encoded from bag of words
Syntactically Controlled Paraphrase Generator (SynPG)

• Learn a decoder to reconstruct the input sentence
  • Reconstruction: no need for annotated paraphrase pairs
  • Disentanglement: ability to control syntax
Syntactically Controlled Paraphrase Generator (SynPG)

- Generate a paraphrase by replacing the syntactic embedding with the target parse information
Word Dropout

- Add word dropout to encourage the model to generate words not appearing in the input sentence

![Diagram](Image)
Evaluation on Syntactic Control Ability

- Consider paraphrase pair \((s_1, s_2)\)
  - Get the parse \((p_1, p_2)\)
  - \((\text{input sentence, parse template}) = (s_1, p_2)\)
  - Ground truth is \(s_2\)
- Evaluation metrics
  - **BLEU score**: similarity between the prediction and \(s_2\)
  - **Template matching accuracy (TMA)**: how accurately the prediction follows \(p_2\) based on exact match on top-2 layers of parse tree
- Test on four paraphrase datasets (ParaNMT, Quora, PAN, and MRPC)
Comparison with Unsupervised Models

• Compared models
  • Back Translation (BackTrans), Vanilla VAE (VAE), Syntax-Infused VAE (SIVAE) [Zhang+ 2019]
  • Better syntactic control ability
Comparison with Supervised Models

- Compared models
  - Seq2seq Model (Seq2seq), Syntactically Controlled Paraphrase Network (SCPN) [Iyyer+ 2018]
- Competitive performance to supervised models
- No need for annotated data → boost performance by considering larger corpus

![Comparison with Supervised Models Diagram](image)
## Paraphrase Examples

| Template | Generated Paraphrase |
|----------|----------------------|
| Original | can you adjust the cameras?  
(S (NP) (VP) (.) )  
(SBARQ (ADV) ( ) (S) ( , ) (SQ) ( . ) )  
(S (PP) ( , ) (NP) (VP) ( . ) ) | you can adjust the cameras.  
well, adjust the cameras, can you?  
on the cameras, you can adjust them? |
| Original | she doesn’t keep pictures from her childhood.  
(SBARQ (WH) (ADV) (VP) (SQ) ( . ) )  
(S (“”) (NP) (VP) (””) (NP) (VP) ( . ) )  
(S (ADV) (NP) (VP) ( . ) ) | why doesn’t she keep her pictures from childhood.  
“ she doesn’t keep pictures from her childhood ” she said.  
perhaps she doesn’t keep pictures from her childhood. |
Focus on text classification tasks
Generate syntactically paraphrases for data augmentation
More robust against syntactically adversarial attacks [Iyyer+ 2018]
We present SynPG to generate syntactically controlled paraphrases without using annotated paraphrase pairs. SynPG uses a novel architecture to disentangle a sentence into semantics and syntax. Extensive experimental results demonstrate the superiority of SynPG. We show that SynPG can help to improve the model robustness.

Code and pre-trained models are available at https://github.com/uclanlp/synpg

Thank You!