



Politecnico
di Torino

Conditional Text Generation using CTRL

Lukas Canciani Graziani

s277237@studenti.polito.it

Matteo Gally

s277935@studenti.polito.it

Carlo Iurato

s277979@studenti.polito.it

Introduction

Text generation

- Unconditional

Text generation based only on the text data provided for training.

- Conditional

Text generation based also on external factors.

CTRL

Salesforce's model for conditional text generation based on control codes.

Task description

Conditional text generation using:

- COCO dataset
- CTRL model
- Evaluation metrics (BLEU, POS-BLEU, SELF-BLEU)



Dataset

COCO Dataset

Coco is a dataset of images and captions, used for several tasks.



A horse carrying a large load of hay and two people sitting on it.



The man at bat readies to swing at the pitch while the umpire looks on.

Dataset

- Caption extracted from two json files
- Train dataset creation (40.000 captions)

‘A panoramic photo of a kitchen and dining room.’
‘The two people are walking down the beach.’

- Evaluation dataset creation (200 captions)

‘A bicycle is parked by a bench at night.’

- Generation dataset creation (200 captions)

‘A bicycle is parked by ~~a bench at night~~.’



Experiments

- Make TF-Records

Conversion of raw text file into a new compressed, and compatible with tensorflow platform.

- Fine-tuned model

Training parameters:

- Learning rate (0.1 , 0.01 , 0.001)
- Iterations (1 , 5 , 10)



Experiments

- Caption generation

Generation parameters:

- Penalty (1 , 1.2 , 2)
- Top-k (1 , 5 , 10)
- Temperature (0.1 , 0.2 , 0.5)



Results

- Generated text evaluation

Evaluation metrics:

- BLEU
- SELF-BLEU
- POS-BLEU

Results

- Results obtained

Mean BLEU	Mean SELF-BLEU	Mean POS-BLEU
0.538	0.554	0.567

Fig.1 - Mean values for BLEU, POS-BLEU, SELF-BLEU.

- Generate captions with best parameters

Parameters used, lr=0.1, iterations=10, temperature=0.5, top-k=1, penalty=1.2

BLEU	SELF-BLEU	POS-BLEU
0.597	0.497	0.589

Fig.2 - Values for BLEU, POS-BLEU, SELF-BLEU using best parameters.

'A giraffe is resting by some rocks, and a small dog is sitting next to it.'

Conclusion & Future works

- Conclusion



- Future works

- Use the entire COCO dataset
- Use different control codes
- Use 1000 captions for generation

Thanks for the attention

You can follow our work on GitHub:

[https://github.com/Carlo13gen/ConditionalTextGeneration.](https://github.com/Carlo13gen/ConditionalTextGeneration)



Politecnico
di Torino