Conditional text generation: leveraging slangs of the web

Luca Bajardi, Ludovico Bessi, Francesco Saracco

Department of Mathematical Sciences
Politecnico di Torino

February 22^{nd}, 2021

Overview

Introduction to the problem

There are mainly two types of Text Generation:

- Unconditional TG: solely depends on the content of training data $\rightarrow$ little diversification
- Conditional TG: also influenced by external factors (context, mood, etc.) $\rightarrow$ more control over generated text

Introduction to the problem

There are mainly two types of Text Generation:

- Unconditional TG: solely depends on the content of training data → little diversification
- Conditional TG: also influenced by external factors (context, mood, etc.) → more control over generated text

Introduction to the problem

There are mainly two types of Text Generation:

- Unconditional TG: solely depends on the content of training data $\rightarrow$ little diversification
- Conditional TG: also influenced by external factors (context, mood, etc.) $\rightarrow$ more control over generated text

Model

CTRL is a 1.63-billion parameter conditional language model based on Transformer architecture, released by Salesforce. It exploits control codes to trigger specific style, content and task-specific behaviour, learned from training data.

Model

Given a sequence $x = (x_1, \ldots, x_n)$, where each x_i is taken from a fixed set of symbols, the aim of language modeling is to learn the probability distribution $\mathbb{P}(x)$ that produces the sequences.

$$\mathbb{P}(x) = \prod_{i=1}^{n} \mathbb{P}(x_i | x_{<i})$$

Thus, using the chain rule, language modeling is easily framed in the next-token prediction task.

Model

Given a sequence $x = (x_1, \ldots, x_n)$, where each x_i is taken from a fixed set of symbols, the aim of language modeling is to learn the probability distribution $\mathbb{P}(x)$ that produces the sequences.

$$\mathbb{P}(x) = \prod_{i=1}^{n} \mathbb{P}(x_i | x_{<i})$$

Thus, using the chain rule, language modeling is easily framed in the next-token prediction task.

Model

CTRL is a conditional language model, meaning that the training phase is always conditioned on a control code c and the distribution learned is $\mathbb{P}(x|c)$.

$$\mathbb{P}(x|c) = \prod_{i=1}^{n} \mathbb{P}(x_i|x_{<i}, c)$$

This variation provides a form of control over the generation of new text, which will be produced accordingly to the code provided.

Model

CTRL is a conditional language model, meaning that the training phase is always conditioned on a control code c and the distribution learned is $\mathbb{P}(x|c)$.

$$\mathbb{P}(x|c) = \prod_{i=1}^{n} \mathbb{P}(x_i|x_{<i}, c)$$

This variation provides a form of control over the generation of new text, which will be produced accordingly to the code provided.

Model: examples

Wikipedia Anarchism is

Model: examples

Wikipedia Anarchism is a political philosophy that advocates the abolition of all forms of hierarchy and domination, including capitalism, patriarchy, racism, sexism, heterosexism and other oppressive social structures. The term...

Model: examples

Wikipedia Anarchism is a political philosophy that advocates the abolition of all forms of hierarchy and domination, including capitalism, patriarchy, racism, sexism, heterosexism and other oppressive social structures. The term...

Books Anarchism is

Model: examples

Wikipedia Anarchism is a political philosophy that advocates the abolition of all forms of hierarchy and domination, including capitalism, patriarchy, racism, sexism, heterosexism and other oppressive social structures. The term...

Books Anarchism is the only true and practical form of Socialism. It has been said that Socialism will never be established until it can be proved to work in practice. The same may be said, with equal truth, about Anarchism. But if we are not able to prove its practicability at present...

Microsoft COCO Dataset

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

A large bus sitting next to a very tall building.

Microsoft COCO Dataset

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

A large bus sitting next to a very tall building.

Main task

Original: Several signs written in Arabic under a cloudy sky.

Main task

Original: Several signs written in Arabic under a cloudy sky.

caption Several signs written in Arabic

Main task

Original: Several signs written in Arabic under a cloudy sky.

caption Several signs written in Arabic and English.

Fine tuning the baseline model

Objective Fine tune CTRL model and make it learn to complete a caption in a style similar to COCO captions.

Tools PyTorch, Tensorflow, Microsoft Azure virtual machine (6 vCPUs, 56 GB RAM, Tesla K80 GPU)

Limitations CTRL model too big to fit K80's (and many other GPU's) main memory $\rightarrow$ CPU training, small training set

Procedure Ran 250+250+500 iterations and selected the best model w.r.t. metrics explained below

Fine tuning the baseline model: some example output

Original	CTRL
Four elephants with people riding on top of them.	Four elephants with people riding them in a field.
A white truck parked next to a double decker bus.	A white truck parked next to a building.
A group of people in a park some with kites	A group of people in a field flying kites.
Several signs written in Arabic under a cloudy sky.	Several signs written in Arabic and English.
A picture of an oven with food baking inside.	A picture of an oven with a pizza on it.

Influence style through multiple inputs

Objective Learn to complete captions in a COCO fashion, but with a different style

Why Could be used to adapt a general learned task to a specific context

How Exploit extraneous text (in our case from Wikipedia and Reddit)

We tried two different approaches.

Influence style through multiple inputs

Objective Learn to complete captions in a COCO fashion, but with a different style

Why Could be used to adapt a general learned task to a specific context

How Exploit extraneous text (in our case from Wikipedia and Reddit)

We tried two different approaches.

First approach: mixture input

Recalling the idea of mixture probability distribution, we tried to make the model learn

$$p_{c,e}(x) = \pi \cdot p_c(x) + (1 - \pi) \cdot p_e(x)$$

where $p_c(x)$ is the distribution generating COCO captions text and $p_e(x)$ is the distribution generating the extraneous text. We built two training sets containing a mix of captions and extraneous text with 3:1 proportion (so, π set to 0.75 in the previous eq.) and used them to train two new control codes: "formal" and "informal".

First approach: mixture input

Recalling the idea of mixture probability distribution, we tried to make the model learn

$$p_{c,e}(x) = \pi \cdot p_c(x) + (1 - \pi) \cdot p_e(x)$$

where $p_c(x)$ is the distribution generating COCO captions text and $p_e(x)$ is the distribution generating the extraneous text. We built two training sets containing a mix of captions and extraneous text with 3:1 proportion (so, π set to 0.75 in the previous eq.) and used them to train two new control codes: "formal" and "informal".

Second approach: incremental learning

Issue The previous approach does not exploit the knowledge obtain during the fine tuning phase.

As a second try, we ran 250 additional iterations with the previously trained control code "caption" on the extraneous text. This produced two "new" models: one generating formal written captions and the other generating informal written captions, both with the "caption" control code.

Second approach: incremental learning

Issue The previous approach does not exploit the knowledge obtain during the fine tuning phase.

As a second try, we ran 250 additional iterations with the previously trained control code "caption" on the extraneous text. This produced two "new" models: one generating formal written captions and the other generating informal written captions, both with the "caption" control code.

Peculiarities of our proposal

- Input-oriented rather than architecture-oriented

Peculiarities of our proposal

- Input-oriented rather than architecture-oriented
- Easy to test different settings (only parameter π has to be tuned)

Peculiarities of our proposal

- Input-oriented rather than architecture-oriented
- Easy to test different settings (only parameter π has to be tuned)
- Flexibility (could be extended to more than 2 sources of text at once)

Metrics

- BLEU: modified form of precision to compare a candidate sentence against multiple references.

Metrics

- BLEU: modified form of precision to compare a candidate sentence against multiple references.
- POS-BLEU: calculation of the classic BLEU score on the POS tags of the words instead of the words themselves.

Metrics

- BLEU: modified form of precision to compare a candidate sentence against multiple references.
- POS-BLEU: calculation of the classic BLEU score on the POS tags of the words instead of the words themselves.
- SELF-BLEU: compare candidate sentence against all other outputs.

Metrics

- BLEU: modified form of precision to compare a candidate sentence against multiple references.
- POS-BLEU: calculation of the classic BLEU score on the POS tags of the words instead of the words themselves.
- SELF-BLEU: compare candidate sentence against all other outputs.
- Formality level with BART

Baseline model results

Metric	250 iters	500 iters	1000 iters
BLEU2	0.534	0.541	0.539
BLEU3	0.499	0.507	0.505
BLEU4	0.462	0.470	0.467
BLEU5	0.409	0.419	0.416

Metric	250 iters	500 iters	1000 iters
P-BLEU2	0.684	0.687	0.685
P-BLEU3	0.624	0.629	0.627
P-BLEU4	0.571	0.578	0.574
P-BLEU5	0.512	0.520	0.516

Metric	250 iters	500 iters	1000 iters
S-BLEU2	0.773	0.766	0.763
S-BLEU3	0.604	0.592	0.587
S-BLEU4	0.421	0.425	0.419
S-BLEU5	0.324	0.306	0.309

Figure: Results for baseline models

Mixture input results

Metric	baseline	wikipedia	reddit
BLEU-4	0.470	0.464	0.467
AFL	0.258	0.248	0.249
AIL	0.751	0.752	0.751

Figure: BLEU and formality level

Mixture input results

Original: A living room and dining area with a large fireplace.

Mixture input results

Original: A living room and dining area with a large fireplace.

Infomal caption: A living room and dining area with

Mixture input results

Original: A living room and dining area with a large fireplace.

Infomal caption: A living room and dining area with a large screen tv on the wall.

Mixture input results

Original: A man in a white shirt and black shorts is holding his tennis racket.

Mixture input results

Original: A man in a white shirt and black shorts is holding his tennis racket.

Formal caption: A man in a white shirt

Mixture input results

Original: A man in a white shirt and black shorts is holding his tennis racket.

Formal caption: A man in a white shirt and hat is holding an umbrella.

Incremental learning results

Metric	baseline	wikipedia	reddit
BLEU-4	0.470	0.422	0.436
AFL	0.249	0.250	0.326
AIL	0.751	0.750	0.674

Figure: BLEU and formality level

Incremental learning results

Original: A group of people at a table with plates and glasses.

Incremental learning results

Original: A group of people at a table with plates and glasses.

Informal caption: A group of people at a table

Incremental learning results

Original: A group of people at a table with plates and glasses.

Informal caption: A group of people at a table having fun and eating pizza.

Incremental learning results

Original: A baby boy wearing a hat and holding an umbrella

Incremental learning results

Original: A baby boy wearing a hat and holding an umbrella

Formal caption: A baby boy wearing

Incremental learning results

Original: A baby boy wearing a hat and holding an umbrella

Formal caption: A baby boy wearing a striped shirt and bow tie.

Noteworthy examples

Formal caption:
A group of people are gathered together to listen and learn about the latest developments in science or technology.

Noteworthy examples

Formal caption:

A group of people are gathered together to listen and learn about the latest developments in science or technology.

Infomal caption:

A group of people are standing around a table having fun posing for the camera.

Noteworthy examples

Formal caption:

An old, black and white photo of a man riding on top of an elephant.

Noteworthy examples

Formal caption:

An old, black and white photo of a man riding on top of an elephant.

Infomal caption:

An old, black and white photo of a man that looks like Moe from The Simpsons.

Future directions

- Finetune BART

Future directions

- Finetune BART
- Experimentation with different extraneous texts

Future directions

- Finetune BART
- Experimentation with different extraneous texts
- Different weights in Mixture