Evolved Transformer for Conditional Text Generation

Artificial Intelligence and Machine Learning 2021

Yuri Sala
Giovanni Tangredi
Luca Viscanti

Politecnico di Torino

1859

Text generation

Mainly adopted in conversational agent.

- Unconditional text generation
- Conditional text generation

Conditional text generation

Goal: aims to learn and predict the next words to obtain meaningful and coherent sentences based on a seed.

Take consideration of other aspect for the generation

- Context
- Topic
- Emotion

Our Goal

- Analyze CTRL model
- Adapt the network to the new dataset (COCO)
- Improve the generation
- Evaluate the results with BLEU metrics

Conditional Transformer Language Model (CTRL)

The generation of next words is always conditioned on a control code, that allow to predict which words are more suitable for context.

- Goal-oriented conversation
- Chatbots
- Query & Answering system

CTRL architecture

- Stack of 48 Transformer encoder blocks
- Sinusoidal positional encoding
- Tied-embedding softmax (input - output)

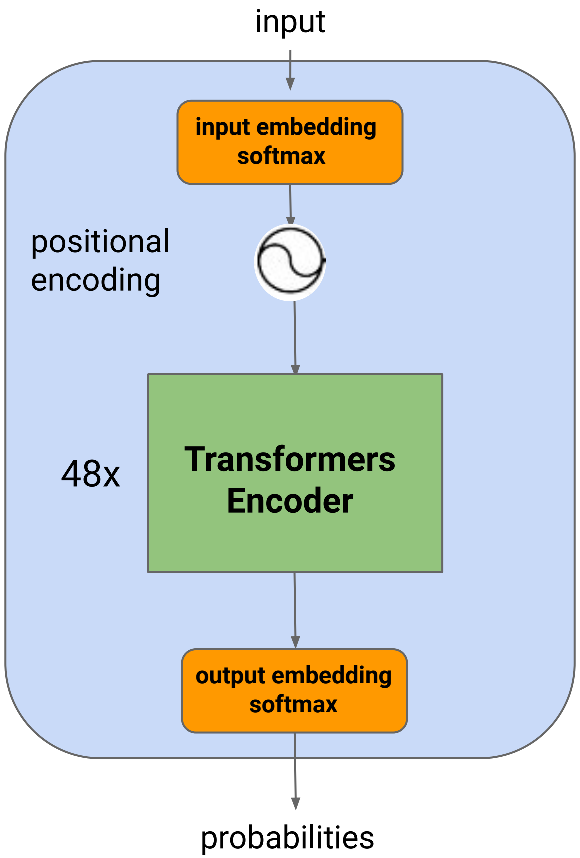


Figure 1: CTRL architecture

“CTRL: A CONDITIONAL TRANSFORMER LANGUAGE MODEL FOR CONTROLLABLE GENERATION”, Salesforce Research., arXiv:1909.05858v2, 20 Sep 2019

Common Objects in Context (COCO)

Large-scale object detection, segmentation and captioning dataset.
Divided in categories grouped in super-categories:

- Super-categories as Control Codes.
- Annotation from images as text, split in train and validation set.
- Short sentences.

New Control Codes: ***Person, Vehicle, Indoor, Outdoor, Appliance, Kitchen, Electronic, Furniture, Food, Accessory, Animal, Sports***.

Fine-tuning over the COCO dataset

- Starting point: CTRL pre-trained model checkpoint.
- Tokenizer for input sentences pre-trained from Huggingface library
- Trained over a million of sentences with different Control Codes split in 12 different files.
- Just one epoch

Fine-tuning

- The model generate good and coherent sentences.
- The new control codes permit to generate phrases that remain in topic.

Sports A woman and child flying kites on a sandy beach.
Sports A person riding skis across a snow covered field.
Kitchen A person making a stack of pancakes in their hands.
Food A person holding a bowl of fruits with water.

Our variation

- Starting from Transformer
- Updates with Evolved Transformer

Transfomer

Natural language processing (NLP) relying on attention layer instead of recurrence

Encoder-decoder structure with:

- positional encoding (based on sinusoids)
- 6x encoder
- 6x decoder

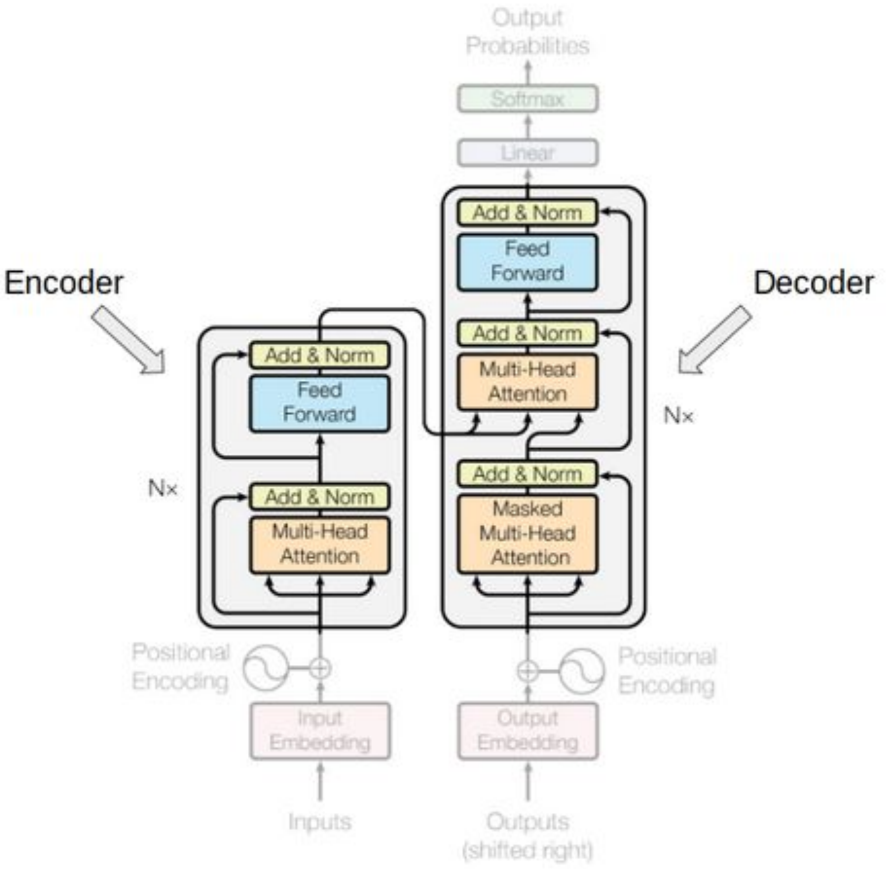


Figure 2: Transformer architecture

“Attention Is All You Need”, Vaswani et al., arXiv:1706.03762v5, 6 Dec 2017

Evolved Transformer

Novel architecture found out with Neural Architecture Search (NAS)

Main differences:

- grouping two blocks in a single one
- separable and wide convolutions
- Gated linear unit (GLU)

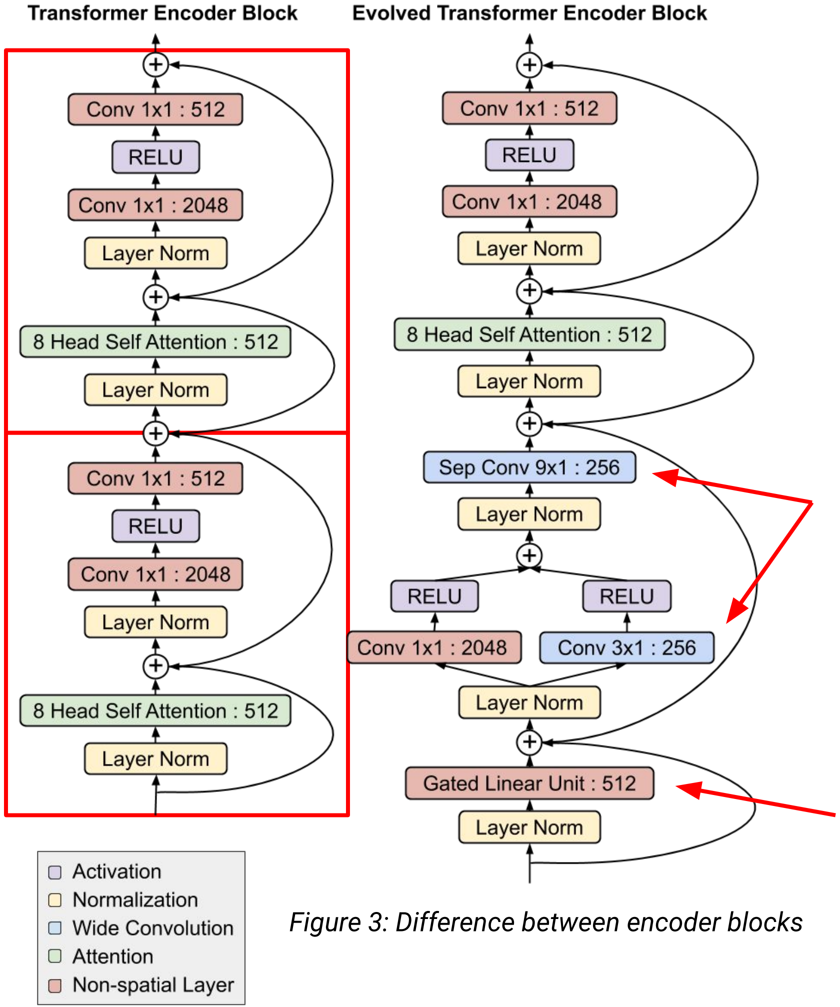


Figure 3: Difference between encoder blocks

“The Evolved Transforme”, So et al., arXiv:1901.11117v4, 17 May 2019

Our proposal

Two main variations:

- Lower number of layers
- Different encoder block

CTRL network:

Our network:

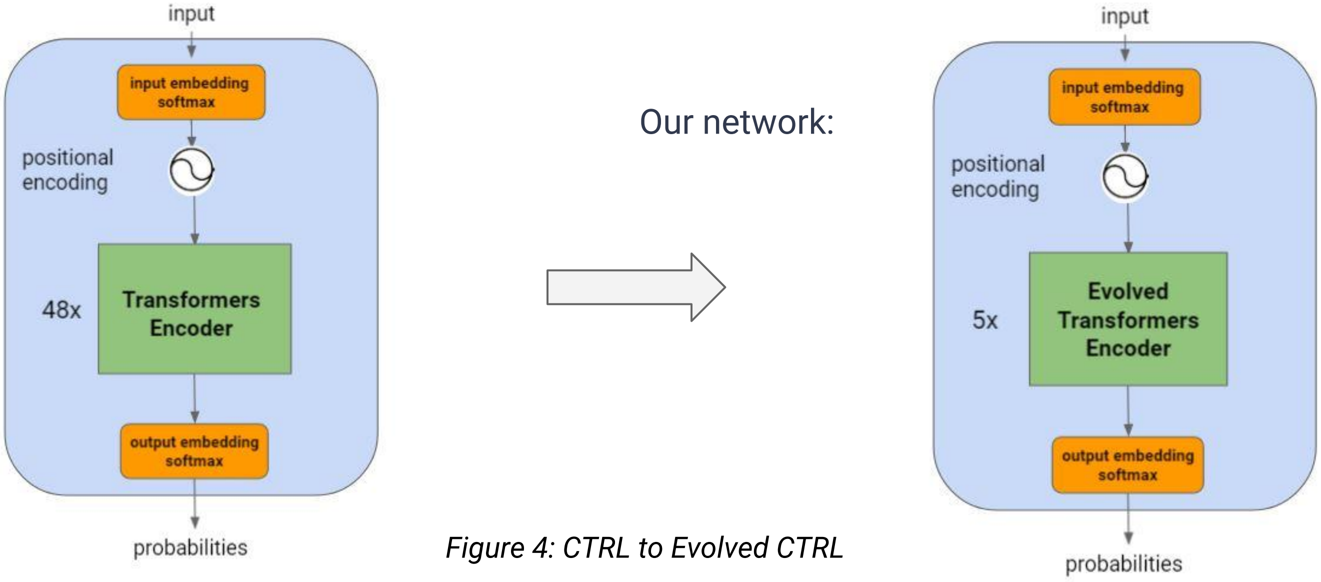


Figure 4: CTRL to Evolved CTRL

Other ideas

While performing our studies we also thought about other kind of changes that we could have applied to the network.
Here we point out some examples:

- Variation of the positional encoding
- Variation of the loss function
- Possibility to add some *decoder* block

Metrics - BLEU

- Evaluate the generated sentences with reference sentences taken from real world data. Simple to calculate and widely adopted.
- Try to match n-grams from the candidate sentences with n-grams from the references.
- An individual N-gram score is the evaluation of just matching grams of a specific order, such as single words (1-gram) or word pairs (2-gram or bigram) or more.
- Cumulative scores refer to the calculation of individual n-gram scores at all orders from 1 to n and weighting them by calculating the weighted geometric mean, usually the weights are equally distributed .
- We used Cumulative BLEU scores from 1 to 4, as notation we used BLEU-N.
- For calculate the score for more than 1 sentence , we calculate the score for each for each one and then we calculated the arithmetic mean

Metrics - SELF-BLEU

- Used to evaluate the similarity of the generated text from the Machine
- The lower the value the more different the generated sentences are.
- Very similar to BLEU but instead the references used for comparison are the rest of the generated sentences.

Metrics - POS-BLEU

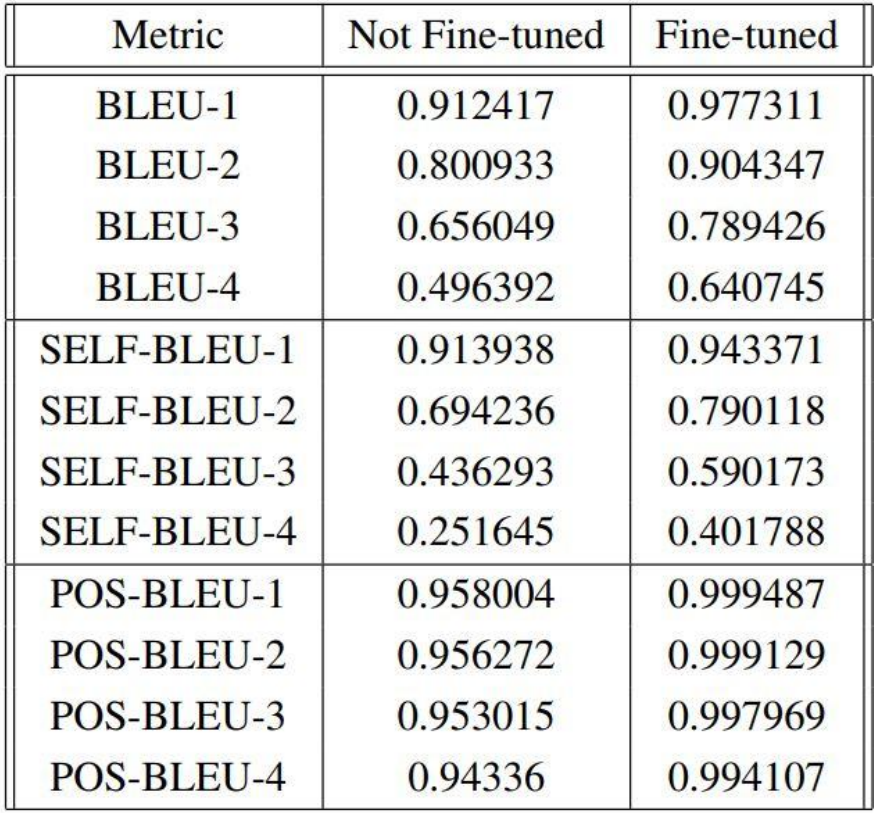
- Implementation of BLUE using the part of speech tagging.
- The candidates and the references are converted using POS Tags and the the BLUE-N scores are calculated
- Used to evaluate if the candidates are structured very similarly with the references

Final results – CTRL

Fine-tuned vs not fine-tuned

In both results only one epoch of training was run

The more high the n-grams the more we can see the fine-tuned model does better of the not fine-tuned one



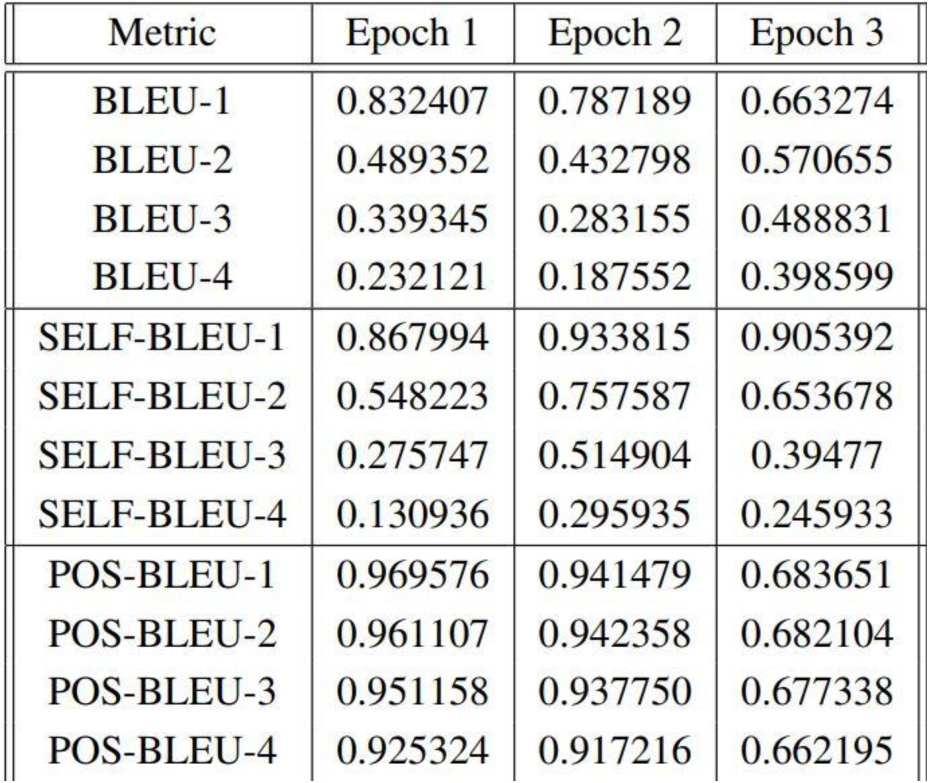
Metric	Not Fine-tuned	Fine-tuned
BLEU-1	0.912417	0.977311
BLEU-2	0.800933	0.904347
BLEU-3	0.656049	0.789426
BLEU-4	0.496392	0.640745
SELF-BLEU-1	0.913938	0.943371
SELF-BLEU-2	0.694236	0.790118
SELF-BLEU-3	0.436293	0.590173
SELF-BLEU-4	0.251645	0.401788
POS-BLEU-1	0.958004	0.999487
POS-BLEU-2	0.956272	0.999129
POS-BLEU-3	0.953015	0.997969
POS-BLEU-4	0.94336	0.994107

Figure 5: BLEU Metrics evaluation

Final results – Evolved CTRL

Metrics for the Evolved transformer for 3 epochs

Still even the best results are worse than not fine-tuned CTRL



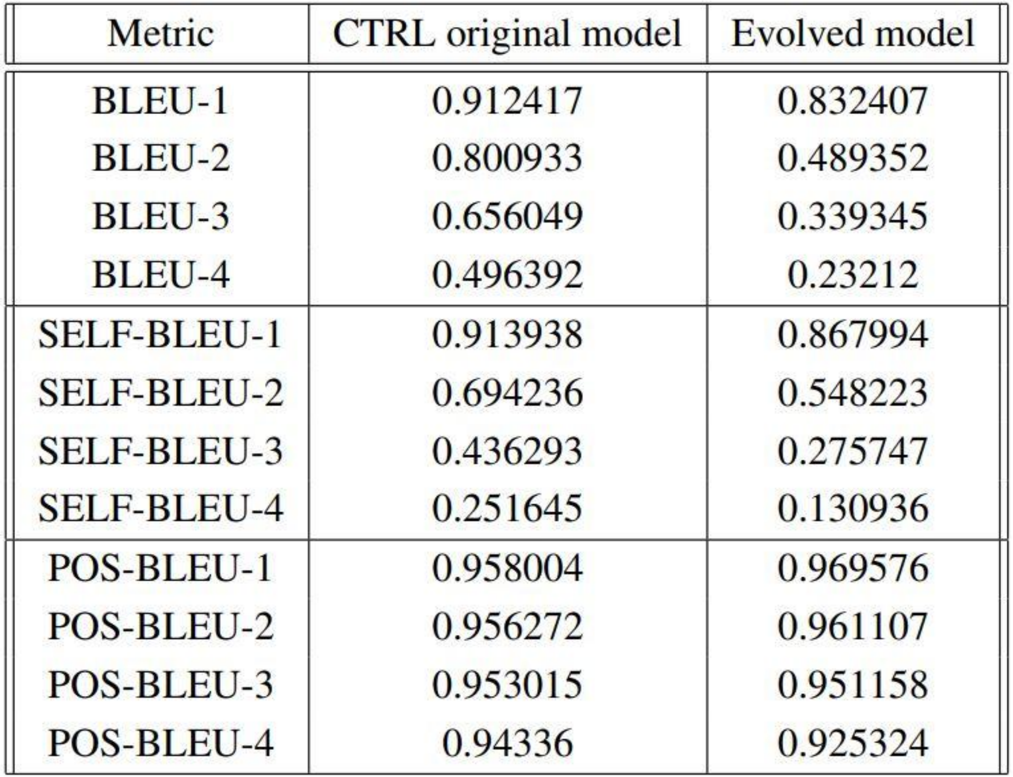
Metric	Epoch 1	Epoch 2	Epoch 3
BLEU-1	0.832407	0.787189	0.663274
BLEU-2	0.489352	0.432798	0.570655
BLEU-3	0.339345	0.283155	0.488831
BLEU-4	0.232121	0.187552	0.398599
SELF-BLEU-1	0.867994	0.933815	0.905392
SELF-BLEU-2	0.548223	0.757587	0.653678
SELF-BLEU-3	0.275747	0.514904	0.39477
SELF-BLEU-4	0.130936	0.295935	0.245933
POS-BLEU-1	0.969576	0.941479	0.683651
POS-BLEU-2	0.961107	0.942358	0.682104
POS-BLEU-3	0.951158	0.937750	0.677338
POS-BLEU-4	0.925324	0.917216	0.662195

Figure 6: BLEU Metrics for different epochs

Final results – CTRL vs Evolved CTRL

Comparison between Not fine-tuned CTRL and Evolved CTRL

Score of the evolved CTRL are worse in almost all cases



Metric	CTRL original model	Evolved model
BLEU-1	0.912417	0.832407
BLEU-2	0.800933	0.489352
BLEU-3	0.656049	0.339345
BLEU-4	0.496392	0.23212
SELF-BLEU-1	0.913938	0.867994
SELF-BLEU-2	0.694236	0.548223
SELF-BLEU-3	0.436293	0.275747
SELF-BLEU-4	0.251645	0.130936
POS-BLEU-1	0.958004	0.969576
POS-BLEU-2	0.956272	0.961107
POS-BLEU-3	0.953015	0.951158
POS-BLEU-4	0.94336	0.925324

Figure 7: BLEU Metrics comparisons

Final results – Generated text

Generated text from fine-tuned CTRL

- Kitchen A knife and fork cut into four bowls of food.
- Vehicle A car is parked near a kite in the grass.
- Wikipedia A car is parked on a street with all luggage.

Generated text from Evolved CTRL

- Appliance A man making his which three cut their park sauce their straight one py middle on.
- Indoor A living room with wine
- Animal A group of zebra eating
- Sports Two dogs playing in car

Thanks for your attention!