what is GPT2

Generative Pre-trained Transformer 2(GPT-2) is a language model. It can generate content based on the input. The GPT-2 is based on the Transformer and it has 4 model sizes, which are GPT-2 small, GPT-2 medium, GPT-2 large and GPT-2 extra large. It is pre-trained with a large scale of content dataset and can be used to predict the next word or generate articles based on the assigned input.

As GPT2 is a pre-trained model, the workflow could download the parameters and finetune it with your own workload. A quick example can be found at https://huggingface.co/distilgpt2

Install

For GPT2, there are two recommended ways to use it.

1. API from huggingface
   GPT2 API from huggingface is commonly used. In huggingface, GPT2 is included in the transformers package. You can use pip to install transformers.
   - pip install transformers
   - As the transformers package is compatible with both pytorch(https://pytorch.org/get-started/locally/) and tensorflow(https://www.tensorflow.org/install/pip), you can install either one based on your preference.
2. wrapper from gpt-2-simple(https://github.com/minimaxir/gpt-2-simple)
   This is a python package that wraps the GPT2 with easy-to-use interface. You can use pip to install it
   - pip install gpt-2-simple
   - You also need to install the tensorflow>=2.5.1.

Usage

1. API from huggingface
   - from transformers import pipeline, set_seed
   - generator = pipeline('text-generation', model='gpt2') #pipeline is used as the inference function in huggingface API. This step is to declare the function and the model for inference.
   - set_seed(42) # set the random seed
   - sequences=generator("Hello, I'm a language model,", max_length=30, num_return_sequences=5) # start generation content
   - # for training or finetune, please reference to https://huggingface.co/transformers/v2.0.0/examples.html
2. gpt-2-simple
   - import gpt-2-simple
   - gpt-2-simple.download_gpt2(model_name="124M") #download the pre-trained parameters

- sess = gpt-2-simple.start_tf_sess()
- gpt-2-simple.finetune() # finetune with your workload. For the details of parameters, please refer to the github documentation or the colab example attached later.
- gpt-2-simple.generate(sess, run_name='run1') # start generate content

Useful links

- quick cde example using google colab https://colab.research.google.com/drive/1VLG8e7YSEwypxU-noRNhsv5dW4NfTGce#scrollTo=4RNY6RBI9LmL
- get start with huggingface https://huggingface.co/docs/transformers/installation
- huggingface gpt2 doc https://huggingface.co/docs/transformers/model_doc/gpt2