Longformer: The Long-Document Transformer

Modifications from Huggingface's Implementation

All models require a `global_attention_size` specified in the config, setting a global attention for all first `global_attention_size` tokens in any sentence. Individual different global attention sizes for sentences are not supported. This setting allows running on TPUs where tensor sizes have to be determined.

`_get_global_attn_indices` in `longformer_attention.py` contains how the new global attention indices are specified. Changed all `tf.cond` to if confiditions, since global attention is specified in the start now.

To load weights from a pre-trained huggingface longformer, run `utils/convert_pretrained_pytorch_checkpoint_to_tf.py` to create a checkpoint.
There is also a `utils/longformer_tokenizer_to_tfrecord.py` that transformers pytorch longformer tokenized data to tf_records.

Steps to Fine-tune on MNLI

Prepare the pre-trained checkpoint

Option 1. Use our saved checkpoint of `allenai/longformer-base-4096` stored in cloud storage

```
gsutil cp -r gs://model-garden-ucsd-zihan/longformer-4096 .
```

Option 2. Create it directly

```
python3 utils/convert_pretrained_pytorch_checkpoint_to_tf.py
```

[Optional] Prepare the input file

```
python3 longformer_tokenizer_to_tfrecord.py
```

Training

Here, we use the training data of MNLI that were uploaded to the cloud storage, you can replace it with the input files you generated.

```
TRAIN_DATA=task.train_data.input_path=gs://model-garden-ucsd-
zihan/longformer_allenai_mnli_train.tf_record,task.validation_data.input_path=gs://mode
garden-ucsd-zihan/longformer_allenai_mnli_eval.tf_record
INIT_CHECKPOINT=longformer-4096/longformer
PYTHONPATH=/path/to/model/garden \
    python3 train.py \
    --experiment=longformer/glue \
    --config_file=experiments/glue_mnli_allenai.yaml \
    --
params_override="${TRAIN_DATA},runtime.distribution_strategy=tpu,task.init_checkpoint=
 \
    --tpu=local \
    --model_dir=/path/to/outputdir \
    --mode=train_and_eval
```

This should take ~ 3 hours to run, and give a performance of ~86.